

Avoimesti tieteestä!: Seppo Parkkila – Avoin data, globaali etu

Äänitteen kesto: 16 min

Litterointimerkinnät

sa-	sana jää kesken
(sana)	epävarmasti kuultu jakso puheessa tai epävarmasti tunnistettu puhuja
(-)	sana, josta ei ole saatu selvää
(--)	useampia sanoja, joista ei ole saatu selvää
, . ? :	kieliopin mukainen välimerkki tai alle 10 sekunnin tauko puheessa

[musiikkia]

Avoimesti tieteestä, keskustelua tieteen, tutkimuksen ja oppimisen avoimuudesta ja vastuullisuudesta.

Henriikka Mustajoki: Tervetuloa. Avoimesta tieteestä -podcastin pariin. Minä olen Henriikka Mustajoki, ja tänään vieraanamme on Tampereen yliopiston anatomian professori Seppo Parkkila, tervetuloa.

Seppo Parkkila: Kiitoksia.

Henriikka Mustajoki: Jos puhutaan anatomian tutkimuksesta, niin minkälaisen datan kanssa siinä pyöritään? Miltä se data näyttää, mistä se koostuu?

Seppo Parkkila: Meidän tutkimuksessa suuri osa sitä on tämmöstä geenidataa. Ja se mitä me hyvin paljon tänä päivänä tehdään on se, et me analysoidaan sitä julkisesti, avoimesti saatavilla olevaa genomidataa. Toinen osa on tällästä proteiinidataa, proteiinien rakenteisiin liittyvää dataa, joka on myöskin avoimesti saatavilla. Ja jos puhutaan genomidatasta taikka proteiinirakenteisiin liittyvästä datasta, niin nämähän on olleet avoimesti saatavilla jo monia, monia vuosia. Se on ollu ikään kuin se standardi, mistä lähdetään liikkeelle. Totta kai mä tiedän, että varmasti monilla ryhmillä on sellaista dataa olemassa, jota ei avoimesti ole saatavilla. Mutta kyl me tänä päivänä pystytään tosi paljon käyttämään avoimesti saatavilla olevaa dataa meidän tutkimuksessa.

Henriikka Mustajoki: Kuka sen avoimen datan on alun perin tuottanu? Ja onko se kuitenkin, ku sanoit että on tehty jo pitkään, niin onko yhä vaik kymmenen vuotta sitten tuotettu data käyttökelposta?

Seppo Parkkila: No tuohon jälkimmäiseen kysymykseen voin sanoa, että kyllä se on käyttökelpoista. Että jos puhutaan tämmösestä genomidatasta, siellähän on virheitä, mutta tavallaan sitä dataa rakennetaan aikasemmin tuotetun datan päälle koko ajan. Eli se data monipuolistuu ja resoluutiotaso koko ajan paranee. On

tarkempia ja tarkempia proteiinimalleja, ja se että niitä on esimerkiksi jostaki tietystä proteiinista vaikkapa viisi mallia taikka kymmenen mallia, niin se aina parantaa sitä. Alun perin tietysti joku ollu se ensimmäinen. Ja sillä tavalla tää koko ajan kehittyy ja paranee tää laatu. No ketkä tätä on sitten tuottaneet, niin kyllä ne on nimenomaan tutkimusryhmät eri puolilta maailmaa. Nämä tutkimusryhmät on tuottaneet sen datan, ehkä julkaisseet siitä jotakin, ja sitten se on avoimesti saatavilla muiden tutkijoiden käytössä.

Henriikka Mustajoki: Usein kuulee semmosta huolta, et jos mä avaan usein aika vaivalla kerätyn datan, niin enemmänkin se on semmonen kokemus et mä menetän jotain. Millä tavalla sun tutkimusalalla syntyy se motivaatio jakaa sitä dataa? Saako siitä meritoitumispisteitä, viitataan siihen, vai mihin se sellanen jakamisen kulttuuri sun mielestä perustuu?

Seppo Parkkila: Tuo on erittäin hyvä kysymys. Siihen on itse asiassa aika vaikea vastata yhdellä ja oikealla tavalla, voi olla monenlaisia vastauksia. Kyllä varmaan trendi on aika lailla sellanen tänä päivänä, että jos tutkimusryhmässä tuotetaan joku data-aineisto, niin sitten julkastaan se ensimmäinen työ. Joka usein on sitten ehkä se merkittävä työ ja julkastaan mahdollisimman korkeatasoisella foorumilla. Ja usein on myös niin, että se kyseinen julkasusarja vaatii sen datan avaamista. Eli se on ikään kuin lähtökohta se, että kun sinä julkaiset siitä omasta aineistostasi dataa, niin sen jälkeen sen on oltava avoimena. Se on ikään ku vähän tällanen että pakotetaan avaamaan sitä dataansa.

Mut sitten kun mainitsit tuon meritoitumiskysymyksen, niin kyllähän se on ihan relevantti lähtökohta myöskin tänä päivänä ajatella niin. Että kun se sinun vaivalla tuottama data on sitten saatettu avoimeksi, niin kyllä sä sitä kautta saat näkyvyyttä omalle tutkimuksellesi paljon ja siihen sinun dataan viitataan. Täähän on tärkeä aina, että kun muut julkaisevat datasta, jonka joku toinen on tuottanut, niin siihen sitten tulee asianmukainen viittaus. Ja se on osa sitä meritoitumista. Sitä kauttahan sä niitä siteerauksia saat, kun sä julkaiset hyviä tutkimuksia ja julkaset hyvää dataa myöskin.

Henriikka Mustajoki: Ehkä monella alalla on vähän niin, että siitä datasta ei vielä saa niitä viittauksia. Mutta ilmeisesti sun tutkimusalalla se dataviittaus on jo sillä tavalla vakiintunut käytäntö, että se toimii ja siihen voidaan luottaa?

Seppo Parkkila: Näin ajattelen. Ainakin oma kokemus on ollu, että ku me julkasuja kirjoitetaan, joissa on käytetty avointa dataa, niin kyllä me erittäinkin tarkasti se sitten siteerataan et mistä se data on hankittu. Ja sehän on reilua, koska jos näin ei tehtäisi, niin siinä saattais sitten tulla sellanen käsitys, että tämä ryhmä on itse tuottanut sen datan. Ja siitähän ei oo kysymys, vaan me ainoastaan hyödynnämme sitä dataa, jonka joku toinen on tuottanu.

Henriikka Mustajoki: Semmonen toisen hyödyntäminen siihen viittaamatta, niin sit ollaan jo varmaan aika lähellä plagioinnin kenttää. Ja sinnehän ei nyt kukaan tutkija vapaaehtoisesti halua lähteä huitelemaan.

Seppo Parkkila: Juuri näin. Se on aivan vähintäänkin yhtä tärkeää viitata asianmukaisesti siihen datan lähteeseen oikein kuin missä tahansa muussa tekstissä jossa viittauksia käytetään.

Henriikka Mustajoki: Sä mainitsit aikasemmin, et tavallaan julkaisufoorumit pakottavat siihen datan avaamisen ja jossain määrin jotkut rahoituslähtetkin. Miten sä ajattelet, mimmonen rooli esimerkiksi nyt julkaisukanavilla on ollu siihen että se data on avointa? Vai miten tää on lähtenyt liikkeelle että siitä on nyt tullu enemmänki vakiokäytäntö ku mitään muuta tällä tutkimusalalla? Et mikä sen taustalla on, mikä on ollu se voima joka on siirtäny sen siitä että tämä on minun siihen että tämä on yhteisesti kaikkien käytettävissä?

Seppo Parkkila: Siinäki on varmaan monta tekijää. Itse ensinnäki ajattelen sillä tavalla, että suhtautumisessa dataan on tapahtumassa koko ajan sellainen sukupolven muutos. Mä oon huomannu sen, että ihan omassaki tutkimuksessa ja omassa tutkimusryhmässä, ni nuoret tutkijat jotenkin ajattelevat lähtökohtaisesti hyvin edistyneesti tästä asiasta. Ja he näkevät sen hyödyn, mikä sitä avoimesta datasta on. En tarkota sitä, että kaikki vanhemmat tutkijat sitten eivät ymmärtäisi, mut minusta siinä on tapahtunu sellasta positiivista kehitystä vuosikymmenten aikana.

Mutta sitten kun puhutaan näistä julkaisufoorumeista, niin kyl mä ajattelen, että niillä on ollu myöskin ja on koko ajan erittäin suuri rooli. Mä tässä tänä aamuna ihan mielenkiinnosta hieman katsoin esimerkiksi lehteä, joka on hyvin tunnettu lääketieteen alalla, British Medical Journal. Siellä on jo linjattu vuonna 2013, että British Medical Journal ei julkase enää mitään trialia, joka liittyy lääkekehitykseen taikka laitteisiin, joissa kirjoittajat eivät vakuuta että kaikki data on julkista, avoimesti saatavilla. Ja jos esimerkiksi The British Medical Journal tekee tällaisen vaatimuksen, niin kyllähän se ohjaa todella voimakkaasti kirjoittajia siihen avoimuuteen, se on lähtökohta.

Henriikka Mustajoki: Nyt pieni mielikuvitusleikki, katotaan mihin tämä meidät johtaa. Miltä tutkimus näyttäisi jos tää kehitys ois lähteny toiseen suuntaan, elikkä että ei jaettaisi? Voisitsä kuvitella et mitä sä voisit tutkia ja miltä se tutkimustyö näyttäis ja tutkimuksen tulevaisuus näyttäis jos data säilyis aina vaan jokaisen ryhmän omilla koneilla ja sitä ei jaettaisi?

Seppo Parkkila: Kyllä se meidän alalla näyttäisi tänä päivänä hyvin toisenlaiselta. Genomi- taikka proteomitason tutkimus olis hyvin, hyvin toisennäköstä ja huomattavasti vaikeampaa. Ja se mihin se olis johtanu, niin epäilisinpä hyvin vahvasti, että siitä datasta olisi tullu maksullista. Sillon se tarkoittais käytännössä sitä, että meidän tutkimusrahoituksesta erittäin merkittävä osa pitäisi suunnata niihin maksuihin mitkä sen datan hankkimisesta johtuis. Itse asiassa on tänä päivänä aika vaikea edes kuvitella, että mitä se vois olla. Mutta kyllä mä sanosin, että tämä nykyinen malli ja se että enemmän ja enemmän on avointa dataa saatavilla, niin se on globaalisti koko ihmiskunnan kannalta suuri etu. Ja nythän me on nähty esimerkiksi tämän COVID-pandemian yhteydessä se, kuinka valtavasti

tutkimusaineistojen avaaminen on simuloinut sitä tutkimusta. COVID-pandemia on tietysti kohdellu ihmiskuntaa tosi kovasti ja ollu tietysti sellanen asia, että ei ois kukaa toivonu että se ois tapahtunu. Mutta se on samalla myöski merkittävästi nyt näyttäny sen avoimen julkasemisen ja avoimen datan voiman.

Henriikka Mustajoki: Mä usein avoimen tieteen edistämistyössä kuulen semmosen väitteen, että avoin tiede on kauheen kallista. Ja nyt ku mä kuuntelen sua, ni mä mietin, että mikä se olis sit se hinta jos me ei avattais. Ku sä puhuit tosta, että rahoituksesta menisi iso osa datan ostamiseen. Jos sä mietit sen avaamisen hintaa, koska sä varmaan osaat myös kertoa mitä työtä se vaatii kun itse tuotettua dataa avataan? Et sehän ei ole itsestään muiden käytettävissä.

Seppo Parkkila: Vois sanoa ensinnäki sen, että se hinta olis ihmishenkiä lääketieteessä. Sillon kun puhutaan ihmishengistä, niin se on hirveen vaikea arvioida, että mikä se hinta oikeasti rahassa olis. Eli kyllähän tässä täytyy koko ajan nähdä tää globaali näkökulma kaikessa tieteenteossa. Että me emme tee sitä tutkimusta itsellemme, emmekä me tee sitä omalle yliopistollemme, emmekä me tee sitä edes omalle kotimaallemme, vaan sitä tehdään kansainvälisessä tiedeyhteisössä koko ihmiskunnan hyväksi. Ja se, että meidän tutkimusdata, meidän tiede on avointa, niin se tavattomasti auttaa muun muassa kaikkea sitä työtä mitä tehdä vaikkapa kehityksmaissa. Ja tällästä hintalappua on todella vaikea arvioida, mikä se hinta olisi, jos data ei olis avointa, jos tiede ei olis avointa. Tänä päivänä se sitä pitkälti kuitenkin todella on, ja hyvä niin.

Kyllä mä omasta näkökulmasta ajattelisin, että se että me saadaan käyttää, voidaan käyttää paljon avoimia dataja, niin se on oikeastaan meidän oman tutkimuksemme elinehto tänä päivänä. Jos näin ei olisi, niin kyllä meidän täytyis oikeasti sitten fokusoida oma tutkimuskin aivan eri suuntaan. Ja se ois ehkä paljolti sellaista mitä me tehtiin vaikkapa 20 tai 30 vuotta sitten. Sillon meidän tutkimushankkeet oli aika sellasia, voi sanoa tänä päivänä aika pieniä ja hyvin fokusoituja. Me hankittiin vaikkapa jotain potilasnäytteitä pieni määrä omasta yliopistosairaalasta ja analysoitiin sitä muutaman potilaan näytesarjaa. Tänä päivänä näytemateriaali voidaan hankkia tuolta julkisista avoimista aineistoista ja sitten hyödyntää sitä, ja puhutaan aivan erilaisista potilaslukumääristä esimerkiksi.

Henriikka Mustajoki: Jos sä mietit Seppo sitä dataa jota sun tutkimusryhmissä tuotetaan ja sen avaamista. Millasia kokemuksia sul on sit siitä, mitä te olette tuottaneet siihen yhteiseen kansainväliseen aineistoon?

Seppo Parkkila: Meidän oma tutkimuksemme on enemmän fokusoitunu sen datan hyödyntämiseen. Me ollaan tuotettu aika vähän suuria data-aineistoja, mutta jonkun verran toki on. Ja sen avaaminen on ollu sillä tavalla itsestäänselvää, että kun ensimmäinen tutkimushanke siitä raportoidaan, niin lähtökohta on se, että siinä vaiheessa se raakadata avataan. Sitähän tosiaan nämä lehdet, julkasijat jo tänä päivänä pitkälti edellyttävät, ja sitä me tehdään omassa tutkimuksessa. Mutta haluan korostaa sitä, että meidän omat aineistomme ovat kuitenkin verraten pieniä. Mutta olemme olleet kyllä kovasti tyytyväisiä muiden aineistojen käyttäjiä.

Henriikka Mustajoki: Sä oot nyt hirveen hyvin kuvannut sen avoimen tieteen ja erityisesti nyt täs tapauksessa avoimen datan merkitystä omassa tutkimuksessa ja terveydenhuollon ja lääketieteen tutkimuksessa. Jos olis joku joka epäilis omalla tutkimusalalla, joka ei ehkä olis vielä samassa vaiheessa, niin mikä olis sellasta mikä vois motivoida ihmisiä avaamaan omaa dataansa tai hyödyntämään muiden dataa?

Seppo Parkkila: Jos ajatellaa että on kaksi vaihtoehtoa, avata sitä dataa tai olla avaamatta. Niin kyllä mä koen tänä päivänä, että jos sä päädyt siihen että sä et avaa omaa dataa, niin se on sellanen kuin käpertyminen nurkkaan. Sanoisin näin, että peli on menetetty kansainvälisessä kilpailussa, jossa tutkijan pitää näkyä ja toimia aktiivisesti. Ja jos tutkija avaa sen oman datansa mitä on tuottanu avoimesti, niin kyllä sitä tavattomasti arvostetaan tutkijapiireissä. Ainakin itse ajattelen näin, että siinä tutkija saa aivan erilaisen aseman. Tälläsiä yksityisyrittäjiä, jotka käpertyvät sinne nurkkaansa ja pitävät kaiken tiedon salassa, niin niitähän on ollu kautta historian. Mutta ei ne oo kyllä välttämättä niitä maailman menestyneimpiä tutkijoita. Et kyllä tänä päivänä kannattaa avoimesti toimia, kannattaa avata ne omat tutkimusdatat sitä mukaa kun se on toiminnan kannalta järkevää. Mä itse ajattelen juuri näin, että kun sä hankit tutkimustietoa, dataa, niin sä vaan tuotat sen oman ensimmäisen julkaisusi perustuen siihen. Ja viimeistään siinä vaiheessa ku sä julkaiset sen ensimmäisen artikkelin, niin se raakadata on sitte sen jälkeen avointa. Se on hyvin yksinkertainen lähtökohta ja se toimii varsin hyvin.

Henriikka Mustajoki: Mä oon kauheen inspiroitunu siitä että meillä on tutkimusaloja joissa avoimuus on jo lähtökohta ja normaali toimintatapa. Et tää ei oo enää mikään uusi asia, jota tavoitellaan, vaan avoin tiede on jo täällä tänään. Kiitoksia, Seppo Parkkila, tästä keskustelusta.

Seppo Parkkila: Kiitos.