

## **Käytännön näkökulma tutkimusaineistojen avoimuuteen**

tiedeasiantuntija *Jyrki Hakapää*, Suomen Akatemia

Avoin-seminaari, 21.5.2013, Helsinki

Hyvät kuulijat,

Tutkimusaineistojen ja -tulosten avoimuus ovat modernin tieteen kulmakiviä. Viimeiset puoli vuosituhatta tiede on periaatteellisesti tukeutunut kaikille kiinnostuneille avattuun työskentelyyn, keskusteluun ja tulkintaan, eli siitä lähtien kun kirjapainotaito mahdollisti laajan tutkimusaineistojen ja -tulosten levittämisen sekä niiden vertailun toisiinsa.

Kirjastojen ja arkistojen avulla on vuosisatojen kuluessa voitu hahmotella mahdollisuutta tai jopa pyrkimystä kohti tällaista kaikenkattavuutta. Nyt elämme keskellä uutta vaihetta, keskellä digitaalisten kokoelmien, verkostojen ja pilvien käyttöönottoa. Ne ovat uusia kirjahyllyjä, joita rakentaen pyrimme nopeammin, kattavammin ja pysyvämmiin pitämään jatkuvasti ulottuvillamme kaiken maailman tiedot – niin raakadatat kuin niiden tulkinnatkin – ja luomaan entistä nopeammin uutta tietoa ja uusia käsityksiä itsestämme ja ympäristöstämme.

Se idealismista, sillä minulta on pyydetty puheenvuoroa käytännön näkökulmasta. Se ei aina ole kovin kiitollinen lähtökohta, varsinkaan nykyisessä tilanteessa, jossa uudet toimintaperiaatteet saati -tavat eivät suinkaan ole vielä vakiintuneet. Mitä tulee tieteellisten aineistojen ja tulosten julkaisemiseen, takanamme on suhteellisen yhdenmukainen vuosisata vakiintuneine ja käytännössä kaikkien tieteen tekijöiden hyväksymine tieteellisine konferensseineen ja julkaisusarjoineen. Nyt tieteen käytäntöihin on lisätty uusia yhteydenpito- ja julkaisutapoja, eivätkä ne vielä ole kaikkien tutkijoiden tai tieteenalojen ulottuvilla tasa-arvoisesti. Kenen näkökulmasta tällöin olisi hyväksi puhua?

Yritän asettua tutkimusaineistojensa avaamisen mahdollisuuksia pohtivien tutkijoiden saappaisiin. Mielestäni tämä näkökulmien moninaisuus on otettava lähtökohdaksemme pohtiessa tutkimusaineistojen avaamisen mahdollisuuksia ja tukemisen tarpeita. Tutkimusaineistoja on monenlaisia ja niiden kunkin avaaminen sähköisessä muodossa internetistä saataviksi vaatii erilaisia

tieteenala- tai jopa tutkimuskohtaisia ratkaisuja. Ratkaisujen etsiminen ja löytäminen kuitenkin kannattaa, koska avoimuus on tehokas ja tuloksekas tapa edistää tutkimusta.

Työskentelen Akatemiassa tiedeasiantuntijana kulttuurin ja yhteiskunnan tutkimuksen yksikössä, minkä takia pohdin asiaa tässä esityksessä tämän yksikön toimialalle kuuluvien tieteenalojen kautta. Yksikön, tai pikemminkin sen tukeman toimikunnan, alalle sattuu osumaan myös kaksi Akatemian kannalta tärkeää tietoaarkistoa: *Yhteiskunnallinen tietoaarkisto* ja kielitieteellistä tutkimusaineistoa säilövä *FIN-CLARIN*.<sup>1</sup> Nämä kaksi arkistoa, ja vain ne, esiintyvät vuosittain hakuilmoituksessamme kohdassa, jossa opastetaan tutkijoita siitä, miten hakemuksessa on kerrottava tutkimusaineiston käyttämisestä, säilyttämisestä, suojaamisesta sekä myöhemmän käytön mahdollistamisesta:

”Akademia suosittelee, että hankkeet luovuttavat kokoamansa yhteiskuntatieteellisen aineiston Yhteiskuntatieteellisen tietoaarkiston käyttöön. Samoin Akademia suosittelee, että hankkeissa luodut kieliaineistot saatetaan muiden tutkijoiden käyttöön FIN-CLARIN -järjestelmän kautta.”

Mainitut kaksi arkistoa tunnetaan tutkimusaineistojen säilyttämisen ja tutkijoiden käyttöön avaamisen asiantuntemuksen keskuksina. Ne toteuttavatkin tehtävänsä ansiokkaasti: esimerkiksi *Yhteiskuntatieteellisen tietoaarkiston* vuosittain vastaanottamien arkistokokoelmien määrä on nyt noussut yli sadan, ja jatkokäyttöpyyntöjä on viime vuosina tullut lähes säännöllisesti yli viisi sataa per vuosi. Tietoaarkisto on myös järjestänyt tiedotustilaisuuksia ja koulutusta tutkijoille, jotta he tietäisivät lähettää aineistojaan arkistoon ja että he yhdessä arkiston henkilökunnan kanssa muokkaisivat aineistonsa avoimuuden kannalta käytettävään kuntoon. Tietoaarkiston menestys näkyy myös Akatemian vastaanottamissa rahoitushakemuksissa: yhteiskuntatieteiden tutkijat kertovat säännöllisesti antavansa tutkimusaineistonsa arkistoon säilytettäväksi projektinsa jälkeen, erityisesti silloin kun kyse on sarjallisesta tai tilastollisesta materiaalista.

Tässä ilmeneekin merkittävä asia: tutkimusaineiston luonne rajoittaa suuresti sen avaamista muiden tutkijoiden käyttöön. Numerollisesti esitettävää tilastollista tai kvantitatiivista aineistoa on helppo käsitellä sekä sovittaa lakipykälien hyväksymään muotoon jopa silloin, kun kyse on ihmistieteistä. Kvalitatiivista tai narratiivista lähdeaineistoa pystytään avaamaan muiden tutkijoiden käyttöön paljon harvemmin: esimerkiksi *Yhteiskuntatieteellisen tietoaarkiston* kokoelmissa tällaisella materiaalilla on hyvin pieni merkitys. Humanistisessa ja yhteiskunnallisessa tutkimuksessa tällä lähdemateriaalilla, esimerkiksi haastatteluilla ja havainnointiraporteilla, on kuitenkin erittäin

---

<sup>1</sup> <http://www.fsd.uta.fi/fi/> ja <https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/KielipankkiEtusivu>

merkittävä osuus. Nähdäkseni meillä ei vielä ole keinoja, jotka mahdollistaisivat tämäntyyppisen aineiston avaamisen. Tutkijoiden argumentit ja keskustelu kvalitatiivisesta lähdeaineistosta käsittelee mieluumminkin sitä, miten tarkkoja henkilötietoja sisältävä ja kuvaileva aineisto on pidettävä turvassa tai korkeintaan luotettavan ja yhteisen tutkimusprojektin parissa työskentelevän ryhmän piirissä, eikä sitä millä lailla se saataisiin muiden tutkijoiden käyttöön. Tämä on harmittavaa sikälkin, että lähes kaikkien tahojen esittämien tutkimusten suurten haasteiden osalta muistetaan nykyään aina huomauttaa, että niiden tarkasteluun tai jopa ratkaisemiseen tarvitaan merkittävää ihmistieteiden osaajia ja tuloksia. Mutta näiden alojen tutkimusaineistojen käyttöönotto kokonaisuudessaan on hankalaa.

Samanlainen kohtalo on historiallisia aineistoja käsittelevien tutkijoiden – ei pelkästään historian, mutta kaikkien tätä aineistoa käyttävien humanististen ja yhteiskuntatieteellisten tutkijoiden – osana. Naapurissamme *Kansallisarkiston* saleissa istuneet tälläkin hetkellä joukko tutkijoita joko puhtaaksikirjoittaen lähdeaineistoa muistiinpanoihinsa tai ottaen niistä valokuvia myöhempää lukemista varten. Näitä tutkimusaineistoja on kuitenkin vaikea saada avatuksi, sillä on epäselvää, kenen omaisuutta ne oikein ovat ja kenen lupaa niiden avaamiseen oikein tarvitaan. Tutkijat ottavat säännöllisesti kiittäen vastaan kirjastojen ja arkistojen uudet digitoidut kokoelmat, mutta samalla tutkijoiden itsensä tekemää arkistokokoelmien tulkitsemista ja dokumentointia ei saada laajempaan käyttöön. Outi Hupaniityn viime vuonna julkaiseman selvityksen *Tutkijan ääni ja sähköiset aineistot* päähuomio oli tutkijoiden tahto päästä vahvemmin vaikuttamaan digitoitavien aineistojen valintaan.<sup>2</sup> Historiallisten aineistojen avaaminen vaatisi kokoelmien ylläpitäjien ja tutkijoiden tavoitteiden yhteensovittamista.

Lopulta on tutkijaryhmä, joka käytännössä ei tuota tutkimusaineistoa. Jos tutkimuksen tehtävänä on tutkia ja tulkita merkkejä, tekstejä tai taideteoksia, minkälaista tutkimusaineistoa oikeastaan syntyy? Tarvitsemmeko filosofin muistiinpanoja hänen lukiessaan Aristotelesta, tai kirjailijan havaintoja hänen tarkastellessaan Proustin teoksia? Haluammeko tutkijoiden lukiessaan ja tulkitessaan tekemät muistiinpanot avoimiksi, vai kai jaksamme odottaa siihen asti, että he voivat työstää ne valmiiksi ajatuksiksi ja julkaisuiksi? Tämä kysymys on retorinen, mutta painottaessamme tutkimusaineistojen avaamista on muistettava, ettei sitä ole pidettävä itsestään selvästi kaikille tieteenaloille järkevänä toimintatapana. Siten sitä ei myöskään voi pitää tieteellisten toiminnan mittana.

---

<sup>2</sup> [http://www.sls.fi/media/pdf/Forskarnas\\_rost\\_slutrapport\\_28112011\\_FI.pdf](http://www.sls.fi/media/pdf/Forskarnas_rost_slutrapport_28112011_FI.pdf)

Suomalaisella tieteen kentällä on siis hyvin erilaisia tutkimusaineistoja. Vaikuttaa siltä, että olemme toistaiseksi vain yhden aineistotyyppin osalta pystyneet kehittämään toimintamallin, joka mahdollistaa aineistojen avaamisen tutkijoiden käyttöön.

\*\*\*

Moninaisuuden havainnoiminen merkitsee myös sitä, että tutkijoiden käytettävissä tulee olemaan yhä useampia ja useampia arkistoja sekä tietokantoja, joihin he voivat jättää aineistonsa säilöttäväksi ja avattavaksi. Kuten sanottu, Akatemia mainitsee nimeltä kaksi kotimaista arkistoa, mutta muitakin kotimaisia ja ulkomaisia vaihtoehtoja joko suunnitellaan tai on jo olemassa. Milloin suomalaiset yliopistot julkaisevat omat tutkimusaineistojen säilömiselle ja avaamiselle tarkoitettut tietokantansa? Keille tutkijoista sopii *Zenodo*,<sup>3</sup> Euroopan komission tukema ja CERN:in ja OpenAire:n vastikään avaama tietokanta, johon tavoitellaan erityisesti eurooppalaisella tutkimusrahoituksella tehdyn tutkimuksen aineistoja ja julkaisuja? Onko tutkijoiden verkostoitumisen edistämiseen kehitetyn sosiaalisen median *academia.edu:n*<sup>4</sup> juuri avaama avoimen tutkimusaineiston sivusto vakavan tutkimustyön vastakohta vai varteenotettava vaihtoehto?

Ihmetellessämme lähes päivittäin uusien tietokantojen syntyä on muistettava, ettei tilanne oikeastaan ole ainutlaatuinen. Jos keskitymme vain valitsemaan ja vankistamaan muutamia keskeisiä ja kaikenkattavilta vaikuttavia rakenteita, niin mitä tapahtuu niille moninaisille sähköisille tutkimusaineistoille, joita suomalainen tutkimus on viimeisen vuosikymmenen pioneerityönä jo tuottanut? Sähköisinä tutkimusaineistoina löytyy jo muun muassa kirjallisuuden kriittisiä editioita, kuten Aleksis Kiven *Nummisuutarit*,<sup>5</sup> tieteen termipankkeja, kuten *Perinteentutkimuksen terminologian* sivusto,<sup>6</sup> tai tietokantoja, kuten tietokanta *Suomen sodissa 1939–1945 menehtyneet*,<sup>7</sup> tai linkkilistoja ja portaaleja, kuten *Kansallisbiografia*-sivuston kokoelma elämäkerrallisista aineistoista.<sup>8</sup> Tutkijat ovat keränneet, tulkinneet, selittäneet ja lopulta julkaisseet tällaisia tutkimusaineistoja monien kirjastojen, arkistojen ja tieteellisten seurojen kotisivuilla. Niinkin hyvät keskitetyt tietoarkistot kuin meillä on olemassa, niin niiden ulkopuolelle jää joka tapauksessa hajanainen mutta arvokas kokoelma erilaisia aineistoja, joiden ylläpidosta ja kehittämisestä olisi myös pidettävä huolta.

---

<sup>3</sup> <http://www.zenodo.org/>

<sup>4</sup> <http://blog.academia.edu/post/49814995816/academia-edu-releases-embedded-data-sets-and-code>

<sup>5</sup> <http://elias.finlit.fi/nummisuutarit/index.php?action=text#itemId=1>

<sup>6</sup> <http://www.helsinki.fi/folkloristiikka/opiskelu/terminologia.htm>

<sup>7</sup> <http://kronos.narc.fi/menehtyneet/>

<sup>8</sup> <http://www.kansallisbiografia.fi/?p=links>

Mihin tahansa päädytäänkään, näen moninaisuuden tutkijan kannalta hyvänä asiana: yksinkertaisesti, silloin on olemassa useampia toimintatapoja, joista tutkija voi löytää sopivimman vaihtoehdon kokoamansa tutkimusaineiston säilyttämisen ja avaamisen varalle.

Vaihtoehtojen on kuitenkin oltava sellaisia, että aineistot säilyvät turvassa ja niiden säilytys on pitkäikäistä – historian tutkimuksen alalta väitelleenä puhun tällöin mielellään vuosisadoista. Jonkun tahon on siis arvioitava kunkin arkistointimahdollisuuden laatu. Lopullisesti tämän valinnan tulee tekemään tiedeyhteisö itse, mutta tässä ratkaisujen etsimis- ja kokeiluvaiheessa jota nyt elämme, tieteen tukijoilla ja rahoittajille lienee korostettu osuus. Niinpä sellaiset tahot kuin Euroopan komissio ja yksittäiset valtiot ovatkin tukemassa ja luomassa erilaisia ratkaisuja. Ne eivät kuitenkaan saisi unohtaa tutkijoiden omaa aktiivisuutta ja kokeiluja.

Varsinkin näin murrosvaiheessa meidän olisi pidettävä huolta siitä, että tutkijoilla todella olisi realistinen mahdollisuus pyrkiä aineistojensa avaamiseen. Tämä tarkoittaa moninaisten lakipykäläisten setvimistä, oli kyse sitten tutkimuksessa tarkasteltujen ihmisten oikeuksista tai alkuperäisen lähdemateriaalin omistuksesta. Se tarkoittaa opastuksen ja koulutuksen järjestämistä: tutkijan koulutukseen on lisättävä opetusta aineistojen valmistelusta arkistoitavaksi tai vaihtoehtoisesti luotava lisää työpaikkoja tietoasiantuntijoille, joiden tehtävä on opastaa tutkijoita ja muokata aineistot julkaistavaan kuntoon. Tutkimusrahoituksessa olisi syytä ottaa huomioon näiden uusien tehtävien suorittamisen aiheuttamat kulut. Se tarkoittaa myös avoimuuden arviointia ja arvostamista: tutkimusaineistojen julkaisemisen parissa tehty työ olisi huomioitavat tutkijan saavutuksia palkittaessa, tai viranhauissa tai tutkimusrahoituksen jaossa.

Mutta erityisesti se siis tarkoittaa moninaisuuden hyväksymistä: on yritettävä löytää yhteisiä koko tieteellisen toiminnan kattavia ratkaisuja ja tapoja, mutta on myös ymmärrettävä aikaisempien yritysten ylläpidon ja kehittämisen tarve sekä uusien ja irtiottavien kokeilujen mahdollisuus luoda ainutlaatuisia läpimurtoja.