

Hallitseva ja houkuttava *big data* – tieteen ja teknologiantutkimuksen aihioita

■ Minna Ruckenstein

Lokakuaisessa tieteen- ja teknologian tutkimuksen konferenssissa San Diegossa ei jäänyt epäselväksi, että *big data* on muotia. Kymmenien tutkijoiden puheenvuorot kertoivat siitä, että on tärkeää olla mukana etulinjassa hahmottamassa ja hallitsemasa nousevaa tieteenalaa, *big data sciencea*. Toistaiseksi *big data*lle ei ole suomenkielistä vastinetta, mutta yleensä siitä puhuttaessa tarkoitetaan niin suurten aineistojen analyysiä, ettei niitä voi käsitellä perinteisin keinoin.

Olemme seuranneet viimeisen kahden vuoden ajan Kuluttajatutkimuskeskuksen henkilökohtaista mittaamista käsittelevässä Salwe-ohjelmaan liittyvässä tutkimusprojektissamme (yhdessä Mika Pantzarin, Veera Mustosen ja Krista Laguksen kanssa) *big data*n nousua tutkimusmaailman ilmiöksi. Huomiota herättävää on ollut tapa, jolla se leikkaa koko tiedeyhteisön läpi. Niin ekonometrian, tietojenkäsittelytieteen kuin sosiologiankin aloilla puhutaan digitaalisten aineistojen kertymisestä, aineistonkäytön haasteista ja aineistojen avaamista mahdollisuuksista. Algoritmien laskennallinen voima mietityttää sekä aineisto-osaajia että niitä, jotka näkevät algoritmien vaikuttavan yhä salakavallammalla tavalla siihen, miten meitä ohjailaan.

Isoja datamääriä kertyy nykyään vaivihkaa. Lukuisilla organisaatioilla on käytössä tietoa tuottavia aineistonkeruumenetelmiä. Esimerkiksi kirjastoilla, kaupoilla ja sairaaloilla on tällaista jatkuvasti karttuvaa aineistoa. Joissakin tapauksissa yritykset, esimerkiksi operaattorit, keräävät tietoa myös lain velvoittamina.

Aineistojen kertyminen on herättänyt toivota niiden hyödyntämisestä, ja konsultit lupaavat tuottavuuden hyppäyksiä, kun aineistoja opitaan hyödyntämään. Tiedonlouhinnalla viitataan

menelmään, jolla suurten aineistojen kohinasta pyritään löytämään oleellista tietoa. Louhinnan ajatellaan tuottavan läpinäkyvyyttä, tehostavan toimintaa ja rakentavan reaaliaikaisia takaisin-kytkentöjä, joiden avulla tekemisiä voi jatkuvas- ti arvioida.

Ensisilmäyksellä aiheeseen liittyvässä keskustelussa on paljon tuttua. San Diegon konferenssissa tietomassoja koskevan keskustelun taustalle sijoitettiin historiallisia jatkuvuuksia: tilastoja, ihmisten seuranta ja tayloristisia unelmia yhteiskunnan ja yksilöiden suorituskyvyn optimoinnista. Tietovarantojen käyttöä ja analysointia kuvattiin myös yrityksenä ohjailta kuluttajia heistä kerrytyllä aineistolla, jolloin aineistoanalyysi muistuttaa lähinnä aiempaa hienoviritteisempää ja reaaliaikaisempaa asiakassegmentointia.

Michel Foucault'n näkemykset sisäistetyttä kontrollista tulivat esille useissa puheenvuoroissa, kuten myös Max Weberin rationalismin rautahäkki. Lisäksi keskusteluissa vilahti James Scottin teoksessa *Seeing like a state* esittämä näkemys valtion tavasta kategorisoida ja hallita asioita. Tästä näkökulmasta kiinnostus suurta tietoa-aineistoihin on uudenlainen tapa kohdentaa huomio mitattavaksi taipuviin asioihin ja ilmiöihin ja kaventaa näin näkökulmaa ihmiseen ja yhteiskuntaan. Suurten aineistojen analyysi kasvattaa entisestään tehokkuuteen ja suoritus- tusten mittaamiseen keskittyvää byrokratiaa, jota ihmiset toteuttavat vastatessaan heille asetettuihin tavoitteisiin.

Hallintaan liittyvien kehityskulkujen toteutumisen todennäköisyyttä korostavat toiveet tiedonlouhinnan auktorisoivasta ja legitimoivasta voimasta. Suurten aineistojen analyysi voi toimia kaikkialle yhteiskuntaan levittäytyvä-

nä kirjanpitojärjestelmänä, joka kertoo kenestä on hyötyä tai keneen kannattaa luottaa. Googlen toiminnassa nähdään piirteitä tällaisesta yhä syvemmälle ihmisten arkeen levittäytyvästä kontrollikoneistosta, jota suuri osa ihmisistä on tietämättään mukana kannattelemassa.

Huomiota herättävää San Diegon puheenvuoroissa oli niiden journalistinen ote. Tapa, jolla digitaaliset aineistot energisoivat keskustelua hallinnan muodoista ja innosta ohjailla ihmisiä, perustui muiden aiheiden parissa tehtyyn tutkimukseen. Termin big data käsittehistoriaa ei avattu keskusteluissa: kukaan ei esimerkiksi pohtinut vakavasti sitä, miten se on löytänyt tiensä tiedemaailman muoti-ilmiöksi. Tiedonlouhinnan konkreettisia sovelluksia tai käytäntöjä ei myöskään juuri käsitelty.

Tietojenkäsittelytiedettä ja kulttuurintutkimusta yhdistävä professori Phoebe Sengers Cornellin yliopistosta muistutti puheenvuorossaan, että big data on käsitteenä epämääräinen ja harhaanjohtava sen tuottaessa näennäistä yhtenäisyyttä. Hän korosti meidän kaikkien tavalla tai toisella linkittyvän osaksi tätä kehityskulkua. Siinä mielessä kyse on yhtä laaja-alaisesta ilmiöstä kuin internet. Kehityksen alkuvaiheissa myös internet herätti keskusteluja uhkista ja mahdollisuuksista, mutta nyt sen ominaislaatu jo ymmärretään paremmin. Internetiä on hyödynnetty niin vakiintuneiden yrityslogiikkojen haastamisessa kuin monopoleja lähentelevien kaupallisten palveluiden rakentamisessa. Se on synnyttänyt täysin uusia tapoja kerätä ja jakaa tietoa tai järjestäytyä poliittisesti.

Oman tutkimuksemme perusteella suurten tietovarantojen analyysissa on laadullisesti uutta se, että yksilökohtaista aineistoa voidaan yhdistää lukuisista eri lähteistä ja toisaalta aineistot saattavat olla varsin epäyhtenäisiä. Ne voivat koostua eri paikkoihin jätetyistä digitaalisista jäljistä tai olla kirjoitettuja viestejä, terveystietoja tai arkistoituja kuvia. Tällainen aineistojen moninaisuus haastaa käsitystä tiedon ja tutkimuksen luonteesta ja tulee epäilemättä tuottamaan uudenlaisia tiedon ja politiikan yhteenkietoutumisia.

Olennaista on, ettei aineistoanalyysi tuota pelkkää informaatiota ongelmanratkaisutilanteisiin. Tieto syntyy ja tulee merkitykselliseksi käytännöissä, joiden avulla tiedollisia aukkoja täytetään ja ongelmia ratkotaan. Suurten aineistojen analyysityö tekee tämän erityisen näkyväksi, koska se vaatii usein prosessimaita työskentelyä. Tietoa louhitaan yhteistyössä muiden kanssa ja aineistolle esitetään yhä uusia kysymyksiä. Kyse on jatkumosta, jossa kuljetaan pikkuhiljaa kohti mielenkiintoisia vastauksia tai aineiston esittämisen tapoja. Antropologit ovat verranneet tätä etnografisen tiedon tuottamiseen, jossa lähtökohtana on tiedon tuotannon avoimuus. Vasta kun aineisto on kerätty ja saatu kesytettyä, tiedetään millaisiin kysymyksiin se voi tuottaa vastauksia.

Hedelmällisempää onkin ajatella aineistoanalyysia erilaisten yhteisöjen ja yhteistoiminnan muotojen tuottajana. Ne taas vuorostaan käyttävät digitaalisia aineistoja omiin tarkoituksiperiinsä. Vahvat toimijat, kuten Google, Amazon, Walmart tai NSA, hyödyntävät digitaalista aineistoa, jotta ne oppisivat tunnistamaan ihmisten aikeita, tarpeita ja liikkeitä. Yksittäisille ihmisille hyöty on huomattavasti rajallisempi kuin silloin, jos aineistoanalyysi aidosti tukisi ihmisiä heidän hyvinvointipyrkimyksissään tai arvoprojekteissaan. Tästä näkökulmasta toiveissa on, että aineistojen analyysimenetelmät kehittyisivät tavoilla, joilla ne hyödyntäisivät digitaalisia jälkiä jättäviä ihmisiä ja yhteiskuntaa laajemmassakin mielessä.

San Diegossa Googlen datakeskuksia tutkineelta ShinJoung Yeolta kysyttiin, uskoiko hän aineistoanalyysin ratkaisevan eriarvoisuuden liittyviä ongelmia. Yeo vastasi, ettei eriarvoisuuksien lievittämisessä ole kyse aineiston vähyydestä vaan sosiaalisesta ja poliittisesta tahdosta. Aineistoanalyysia voi kuitenkin hyödyntää myös tuttujen epäkohtien esittämiseen uudenlaisilla tavoilla. Näin tiedonlouhintaan liittyvät kehityskulut haastavat tutkijoita sekä käyttämään mielikuvitustaan että tekemään eettisesti kestäviä valintoja tutkimustyössään.

Kuluttajatutkimuskeskuksessa olemme sekä opetelleet aineistoanalyysin käytäntöä (yhteis-

äino

Löydät www.aino.net sivuiltamme uuden
AINO DESIGN MAGAZINEN videoineen.
Tilaa omaan s-postiisi.

FREDA 33, HELSINKI
MA-PE 10.30 - 18.00
LA 10.30 - 15.00
PUH. 611 611
WWW.AINO.NET



työssä Studio Mindin Pasi Karhun kanssa) että hahmotelleet digitaalisia aineistoja uudenlaisten yhteiskunnallisten kenttien, toiminnan alueiden ja kyvykkyyksien näkökulmasta. Näin aineistojen hyödyntäminen linkittyy myös tapoihin, joilla digitaaliset jäljet löytävät paikkansa ihmisten arjessa ja tekemisissä. Kiinnostavaa ei ole digitaalinen aineisto sinänsä vaan kaikki se, mitä aineistojen ympärillä tapahtuu. Uudeksi tutkimusalueeksemme on hahmottunut datan sosiaalinen elämä. Viime kädessä on kysymys siitä, millaisiin tarkoituksiin digitaalisia aineistoja ja analyysimenetelmiä halutaan taivutella. Pelkällä

puheella ja hehkutuksella big datasta on jo tullut yhteiskunnallinen toimija. Tärkeää on nyt ymmärtää syvällisemmin tämän uuden toimijan luonnetta, vaikutuksia ja vaikuttavuutta.

Kirjoittaja on filosofian tohtori ja Kuluttajatutkimuskeskuksen erikoistutkija.