

Aluetutkimus big data -aikakaudella

Daria Gritsenko

Aluetutkimus tieteenalana saattaa ensi näkemältä vaikuttaa yhteensopimattomalta big data -teknologian kanssa, sillä aluetutkimuksen ytimessä on alueellisesti määriteltyjen paikkojen syvä kontekstuaalinen ymmärrys historian, kielen ja instituutioiden tarkastelun avulla. Big data -tutkimus perustuu sen sijaan alueellisesti rajaamattomaan analyysiin algoritmeja käyttäen. Siinä pyritään löytämään lainalaisuuksia ja trendejä aineiston, ei teorian lähtökohdista. Onko näillä kahdella tieteenalalla siis mitään annettavaa toisilleen? Esseen tarkoituksena on tarkastella big data -teknologian tuomia mahdollisuuksia ja rajoitteita aluetutkimuksen kentällä.

Aluetutkimus monitieteisenä tieteenalana

1900-luvun puolivälissä yhdysvaltalaisen ja länsieurooppalaisten yliopistojen ”kaukaisten kulttuurien” laitokset avasivat ovensa kieli- ja kirjallisuustutkijoiden lisäksi sosiologeille, taloustieteilijöille ja politiikan tutkijoille, joiden tarkastelukohteeksi muodostuivat tutkimien- sa alueiden senhetkiset yhteiskunnat (Colonomos 2016, 65). Näin syntyi aluetutkimuksen tieteenala, yhteiskuntatieteellisen tutkimuksen uusi haara, joka pyrki luomaan yleiset puitteet maapallon eri alueiden kuvailulle ja selittämiseksi: oli kyseessä sitten Neuvostoliitto, Lähi-Itä, Afrikka, Latinalainen Amerikka tai Kiina. Kielitaidolla tuli olla tutkimuksessa tärkeä rooli, mutta tieteenalojen välisiä rajoja tulisi ylittää (Rafael 1994). Tästä onkin seurannut joidenkin mieltämä ristiriita aluetutkimuksen ja tieteenalojen välillä. ”Historian lopun” jälkeisellä aikakaudella (Fukuyama 1992) kylmän sodan projektina pidetyn aluetutkimuksen arvoa on kyseenalaistettu entistä useammin (Rafael 1994). Silloin jonkin tietyn alueen tieteen- alakohtaisen tarkastelun merkityksen korostamisesta on muodostunut tärkeä tapa oikeuttaa aluetutkimuksen olemassaoloa (Bates 1997).

On selvää, että yli puolen vuosisadan pituisen historiansa aikana aluetutkimus on tuottanut ja popularisoinut monia tärkeitä metodologisia saavutuksia. Ensinnäkin maantieteilijöiden ja antropologien jo kauan sitten tunnustama ajatus siitä, että *kontekstilla on väliä* on tuotu syvällisen kenttätutkimuksen avulla eksplisiittisesti taloustieteen, politiikan tutkimuksen ja yhteiskuntatieteiden aloille. Aluetutkimus on myös rutiininomaisesti haastanut tieteenalojen

Yhdysvalta- ja Eurooppa-keskisiä oletuksia. Szantonin (2004) mukaan valtavirtatieteenalat ovat oikeastaan aluetutkimuksen erityistapauksia, Yhdysvalta- ja Eurooppa-tutkimusta. Lisäksi aluetutkimus on ottanut muun muassa ankkuroidun teorian (*grounded theory*), narratiivintutkimuksen ja etnografian välineitä osaksi monimutkaista tutkimusstrategiaansa ja sen kautta luonut useita metodologisia innovaatioita (Bates 1997). Tämän seurauksena tutkijat ovat irrottautuneet tulkinnoista, jotka pohjautuvat varhaisen aluetutkimuksen yhdelle keskeisimmistä teemoista *universaalin* ja *erityisen* väliselle kahtiajaolle (Colonos 2016, 66), ja kehittäneet uusia menetelmiä paikkakohtaiselle tutkimukselle, jonka tuloksena on syntynyt sittemmin myös eri tieteenaloille levinneitä kontekstuaalisesti ja kulttuurisesti juurtuneita selityksiä.

Toiseksi aluetieteellisiin hankkeisiin sisäänrakennettu moni- ja poikkitieteisyys on johtanut laajamittaiseen käsitteiden lainaamiseen, risteyttämiseen ja vertailevien lähestymistapojen painottamiseen (Katzenstein 2001). Aluetutkimuslaitosten tutkijoiden taustat ovat usein eri tieteenaloissa, minkä seurauksena yhteiset seminaarit ja muu yhteistyö altistavat väistämättä tutkijan eri tieteenalojen vaikutukselle. Nämä käytännön olosuhteet pakottavat aluetutkijan astumaan oman mukavuusalueensa ulkopuolelle ja antavat ideoita ja tukevat luovaa käsitteenmuodostusta. Szanton (2004) näkeekin aluetutkimuksen sukupuoli-, etnisyy-, ja maaseutututkimuksen sekä muiden 1970-luvulla ilmestyneiden poikkitieteellisten alojen edeltäjänä. Lisäksi aluetutkimuksen ytimessä alusta alkaen ollut *kieli* mullisti yhteiskuntatieteitä perustavanlaatuisesti, kun tutkijoiden huomio kiinnittyi ”kieleen produktiivisena ja vaikuttavana voimana itsessään” (Gibson-Graham, 2004). Kielen performatiivisuuden ja diskurssin vallan korostaminen kriittisen dekonstruoinnin välineenä toi aluetutkimuksen yhteiskuntatieteiden niin sanotun tulkinnallisen käänteen etujoukkoihin. Nämä metodologiset saavutukset, kuten seuraavaksi perustelen, ovat oleellisia myös big data -aikakaudella.

Digitaaliset ihmistieteet ja big data

Laajimmassa merkityksessään data tarkoittaa kaikkea materiaalia, joka on tallennettu tai tallennettavissa. Tämän määritelmän valossa *big data* tarkoittaa yksinkertaisesti suurta tietomäärää. Datan koko, jota tavallisesti mitataan giga- ja terabitteinä, on kuitenkin vain yksi big datan monista puolista. (McAfee et al. 2012.) Digitalisaatio eli fyysisten kohteiden ja prosessien integroiminen osaksi digitaalista ulottuvuutta johtaa datan määrän moninkertaiseen kasvuun. Digitaalisten teknologioiden omaksuminen kaikilla ihmiselämän osa-alueilla johtaa siihen, että olemassaolevan datan määrä kasvaa suunnattomalla nopeudella. Tiedon määrän ja kasvunopeuden lisäksi big datan ominaisuuksiin kuuluu myös monimuotoisuus. Satelliiteista, sosiaalisesta mediasta ja seurantavälineistä saatavan aineiston monimutkaisuus ja monipuolisuus onkin haastanut perinteiset toimintamallit paitsi arkielämässä, myös tutkimuksessa (Mayer-Schoenberger ja Cukier 2013). Big datan määrä, nopea kasvu ja monimuotoisuus (englanniksi kolme v:tä eli *volume*, *velocity* ja *variety*) erottavat sen sekä määrällisestä että laadullisesta ”vanhasta” datasta, sillä sen työstäminen vaatii enemmän laskennallista tehoa, uusia laskennallisia paradigmoja ja uusia metodologisia lähestymistapoja.

Mitä lisäarvoa big data tuo yhteiskunta- ja ihmistieteisiin? Big data voi auttaa tutkijoita päätöksentekoon, mieltymyksiin, identiteetteihin ja käyttäytymiseen liittyvien monimutkaisten kysymysten tarkastelussa (Gonzalez-Bailon 2013). Big datan äännekkäimmät puolestapuhujat ennustavat aineistolähtöisen ja empiirisen tutkimusotteen voittoa teoriapainotteesta ja tulkinnallisesta, jonka myötä ihmistieteet nousisivat kuvailevasta ja selittävästä tieteenalasta

”oikean”, ennustamiseen kykenevän tieteen tasolle (Anderson 2008). Ennakoiva analytiikka – yksi big datan keskeisistä lupauksista – olettaa, että big datan avulla yhteiskunnallisia, taloudellisia ja kulttuurisia ilmiöitä on mahdollista ennustaa luotettavasti, vaikkamme täysin ymmärtäisikään niiden taustalla vaikuttavia syy-seuraussuhteita. Ennakoivan analytiikan menetelmään kriittisesti suhtautuvat painottavat kuitenkin, että tällaisesta tieteellisen käytännön muuttumisesta tekniseksi käytännöksi seuraa vakavia metodologisia ongelmia, sillä big data ei kykene luomaan kuvaa, joka olisi ”vivahteikas ja ihmisten tutkimiselle aidosti sensitiivinen” (Ekbia et al. 2016, xi).

Big data voidaan käsittää paitsi tutkimusmenetelmänä, myös yhteiskunnallisena ilmiönä. Kriittinen datatutkimus (*critical data studies*) on poikkitieteellinen tieteenala, jonka juuret ovat moraalifilosofiassa, yhteiskunnallisessa oikeustieteessä, tiede- ja teknologiatutkimuksessa ja antropologiassa, joka tarkastelee big datan oikeudellisia ja eettisiä haasteita, muun muassa immateriaalioikeuksiin, yksityisyydensuojaan ja tietoiseen suostumukseen liittyviä kysymyksiä. Yksilöitä koskevan suuren tietomäärän (Morozov 2013) ja kerätyn materiaalin analysointiin käytettyjen algoritmien läpinäkyvyys (Burrell 2016) sekä valtion ja kansalaisten, pienten ja suurten yritysten ja tiedon tuottajien ja omistajien välisen ”datakuilun” kasvaminen (Boyd ja Crawford 2012) kuuluvat tutkijoiden tunnistamiin ja problematisoiimiin aiheisiin, joiden joukko kasvaneet mitä todennäköisemmin tulevaisuudessa.

Big data ja aluetutkimus

Big datan integroiminen osaksi aluetutkimusta vaatii sen ymmärtämistä sekä teknologiana että yhteiskunnallisena ilmiönä. Ensinnäkin metodologinen kritiikki ja etenkin väite, jonka mukaan big data -teknologia syrjäyttäisi perinteisen epistemologian ja loisi malleja, jotka eivät pystyisi selittämään yhteiskunnallisia ilmiöitä, voidaan torjua tukeutumalla myös vastaisuudessa aluetutkimuksen tiedon kontekstuaalisuuden merkitystä vaalivaan vahvaan perinteeseen. Verrattuna empiiriseen yhteiskuntatieteeseen, joka on vertailevan ulottuvuuden joko läsnäollessa tai sen puuttuessa aina jollain tavalla paikkaan ja alueeseen sidottua, aluetutkimuksen kentällä toimivat tutkijat osaavat erinomaisesti tulkita ehdollisia tutkimustuloksia laajemmassa teoreettisessa kontekstissa tunnustaen kontekstuaalisuuden oleellisena osana selitysten rakentamista. Teoriankehitys tapana analysoida yhteiskunnallista todellisuutta voidaan hedelmällisesti yhdistää suureen tietomäärään ja uusimpiin laskumenetelmiin. Sen lisäksi aluetutkimuksen kiinnostus kielentutkimukseen tekee laskennallisen kielitieteen menetelmien käytön suurten tekstuaalisten aineistojen, etenkin (digitalisoitujen) painettujen, digitaalisen ja sosiaalisen median aineistojen, analysoimiseksi laajalti käytetyn laadullisen tekstinanalyysin loogiseksi jatkuvuudeksi (ks. esim. Koltsova ja Shcherbak 2015). Näin ollen aluetutkimuksella on pitkäaikaisen moni- ja poikkitieteellisen käytäntöjensä ansiosta mahdollisuus tuoda kontekstuaalista sensitiivisyyttä uuteen big data -tekniikan avulla tehtävään tutkimukseen.

Toiseksi on arvioitava big datan ominaisuuksia, kuten sitä erityiskontekstia, jossa venäläistä big dataa tuotetaan. Valtion pyrkimys säädellä tietovirtoja, rajoittaa datan säilytystä ulkomaisilla palvelimilla sekä asettaa Internet-sivuille ja -palveluille määräaikaista ja pysyviä toimintakieltoja, vaikuttavat paitsi aineistonkeruun käytännön mahdollisuuksiin, myös aineiston analyysiin ja sen tulkintaan. Vaikka venäläisiin institutionaaliin ja sääntelyyn liittyviin käytäntöihin asetetut rajat ovat tutkijoille jokseenkin tuttuja (Nocetti 2015), käyttäjien luoman sisällön potentiaalisista vääristymistä tiedetään edelleen vähän (Nisbet et al. 2017). Sensuuria ja sensuurikäytäntöjen ja aineiston vääristymien suhteita käsittelevä tutkimus on varhaisessa

vaiheessa (Vendil Pallin 2017). On lisäksi vaikeaa arvioida digitaalisista lähteistä kerätyn materiaalin, kuten blogitekstien, keskustelupalstaviestien tai sosiaalisen median aineiston, edustavuutta koko väestöön suhteutettuna. Venäläiseen big dataan liittyviä haavoittuvuuksia, epävarmuuksia, ja oikeudellisia ja eettisiä ristiriitoja on mahdollista ratkaista vain osittain aluetutkimuksen työkalujen avulla. Selvitäksemme niistä haasteista, joita venäläinen big data luo yhteiskunnallisena ilmiönä, meidän tuleekin ottaa käyttöön kriittisen datatutkimuksen kehittämisiä menetelmiä kuten big datan kansantaloustiedettä (Boyd & Crawford 2012; Mann 2018) ja kriittisten ihmistieteiden *Kulturkritik*-lähestymistapaa (Presner 2015).

Kaksoiskansalaisuudesta kolmikansalaisuuteen?

Aluetutkijat, mukaan lukien Venäjä-tutkijat, pitävät big data -metodologiaa uutena ”hienona” kenttänä ja haluavat selvittää, mitä merkitys big datalla on sen perinteisen vahvuuden, syvällisen kontekstuaalisen ymmärryksen kannalta. Robert Hall on teoksessaan *Area Studies: With Special Reference to Their Implications for Research in the Social Science* kirjoittanut aluetutkijoiden olevan sekä oman tieteenalansa että alueellisen tutkimuskohteensa asian tuntijoina ”kaksoiskansalaisia” (Rafael 1994). Entä minkälainen on aluetutkija big data -aikakaudella? Syvästi ahdistunut akateemisen identiteettikriisin uhri (ei mitään aihealuetta kattavasti hallitseva), jokapaikanhöylä vai tietojenkäsittelytieteilijöiden kanssa saumattomasti toimiva tiimipelaaja? Uskon, että vastaus kysymykseen määrittyy aluetutkijoiden oman asenteen perusteella. Seuraavaksi esittelenkin kolme mahdollista lähestymistapaa big datasta kiinnostuneille aluetutkijoille.

Ensimmäinen näistä kolmesta on pragmaattinen lähestymistapa. Big data on uudenlainen aineistotyyppi, johon liittyy erilaisia tiedonlouhinnan ja koneoppimisen tekniikoita. Ottaen huomioon, että jokaisella aineistotyyppillä on omat rajoituksensa, on tärkeää ymmärtää ja tunnustaa big dataan liittyvät haasteet. Samalla on lähes pakollista sisällyttää big data osaksi talouden, yhteiskunnan ja kulttuurin digitaalisen muutoksen tutkimusta. Pragmaattinen lähestymistapa sopii hyvin tutkijoille, jotka nauttivat poikkitieteellisestä yhteistyöstä ja ovat valmiita ottamaan osaa yhteisiin hankkeisiin tietojenkäsittelytieteilijöiden kanssa.

Toinen lähestymistapa on integroiva. Big data on osa digitaalisten ihmistieteiden ja laskennallisten yhteiskuntatieteiden laajempaa trendiä, jonka tavoitteena on valaista yhteiskunnallisten, taloudellisten ja kulttuurien ilmiöiden monimutkaisuutta. Tämä trendi tuo jälleen esiin, että tieteenalojen rajat ylittävät teemat, kysymykset ja ongelmat ovat ensiarvoisen tärkeitä. Integroivan lähestymistavan lähtökohtana ovat tosielämän ongelmat, ei johonkin tieteenalaan juurtunut teoreettinen viitekehys. Se pyrkii kuvailemaan ja tarkastelemaan niitä erilaisin menetelmien ja mallien avulla päämääränään maksimaalinen kattavuus. Konseptuaalinen ja metodologinen lainailu on integroivan lähestymistavan ytimessä.

Kolmas lähestymistapa on kriittinen. Big data on ennustamiseen, optimointiin ja ongelmanratkaisuun pakkomielleisesti suhtautuvan yhteiskuntamme ajanhenki, *Zeitgeist*. Aluetutkimus antaa meille keinoja kritisoida big data -menetelmää ja nostaa jälleen esiin sekä tulkinnan (*Verstehen*) että kritiikin metodologisen merkityksen korostaen, että sillä kontekstilla, jossa big dataa tuotetaan ja mobilisoidaan, on merkitystä. Big data -liike on paremman tieteellisen käytännön utopia siitä huolimatta, että epäarvoisuus ja oikeudelliset ja eettiset haasteet aiheuttavat sille huomattavia ongelmia. Kriittisen lähestymistavan tavoitteena on paljastaa big dataan ja sen konteksteihin liittyviä haavoittuvuuksia, epävarmuuksia ja ristiriitoja.

Riippumatta siitä, mikä lähestymistapa valitaan, aluetutkimuksen ja digitaalisten ihmis-

tieteiden leikkuupisteessä kiehuu jotain kiinnostavaa. Tieteenalarajojen yli ulottuva aito huoli kaikkialla läsnä olevan digitalisaation seurauksista on aihe, joka mahdollistaa uusien tutkimuksellisten kuvioiden ilmestymisen. Olen vakuuttunut siitä, että samanaikaisesti kun big data -menetelmiä kohtaan osoitettua kritiikkiä hyödynnetään, sekä kriittinen että positivistinen big data -tutkimus voivat tuoda uusia näkökulmia aluetutkimuksen saralle.

Englannista kääntänyt Kristiina Silvan

Lähteet

- Anderson, C. (2008) The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine* (June 23), http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.
- Bates, Robert H. (1997), Area studies and the discipline: a useful controversy?. *PS: Political Science and Politics*, 30:2, 166-169.
- Boyd, Danah & Crawford, Kate (2012), Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15:5, 662-679.
- Burrell, Jenna (2016), How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3:1, 1-12.
- Colonomos, Ariel (2016), *Selling the Future: The Perils of Predicting Global Politics*. London: Hurst.
- Ekbia, Hamid R., Sugimoto Cassidy R. & Mattioli, Michael (2016), Introduction. Teoksessa Cassidy R. Sugimoto, Hamid R. Ekbia & ja Michael Mattioli (toim.) *Big Data is not a Monolith*. Cambridge: MIT Press. xi-xxi.
- Fukuyama, Francis (1989), The end of history? *The national interest*, 16, 3-18.
- Gibson-Graham, J. K. (2004), Area studies after poststructuralism. *Environment and Planning A*, 36:3, 405-419.
- González-Bailón, Sandra (2013), Social science in the era of big data. *Policy & Internet*, 5:2, 147-160.
- Graham, Mark, Hale, Scott ja Stephens, Monica (2012), Digital Divide: The Geography of Internet Access. *Environment and Planning A* 44:5, 1009-10.
- Katzenstein, Peter J. (2001), Area and regional studies in the United States. *PS: Political Science and Politics*, 34:4, 789-791.
- Koltsova, Olessia & Shcherbak, Andrey (2015), 'LiveJournal Libra!': The political blogosphere and voting preferences in Russia in 2011-2012. *New Media & Society*, 17:10, 1715-1732.
- Mann, Laura (2018), Left to other peoples' devices? A political economy perspective on the big data revolution in development. *Development and Change*, 49:1, 3-36.
- Mayer-Schoenberger, Viktor & Cukier, Kenneth (2013), *Big Data. A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.
- McAfee, Andrew, Brynjolfsson, Erik (2012), Big data: the management revolution. *Harvard business review*, 90:10, 60-68.
- Morozov, Evgeny (2013), *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York, NY: PublicAffairs.
- Nisbet, Erik C., Kamenchuk, Olga, & Dal, Aysenur (2017), A Psychological Firewall? Risk Perceptions and Public Support for Online Censorship in Russia. *Social Science Quarterly*, 98:3, 958-975.
- Nocetti, Julien (2015), Contest and conquest: Russia and global internet governance. *International Affairs*, 91:1, 111-130.
- Presner, Todd (2015), *Critical Theory and the Mangle of Digital Humanities*. Teoksessa *Between Humanities and the Digital*. toim. Patrick Svensson ja David Goldberg. Cambridge: MIT Press. 55-68.

- Rafael, Vicente L. (1994), The cultures of area studies in the United States. *Social Text*, 41, 91-111.
- Szanton, David L. (2004), Introduction: The Origin, Nature, and Challenges of Area Studies in the United States. Teoksessa *The politics of knowledge: Area studies and the disciplines*. Ed. David L. Szanton. Berkley: University of California Press.
- Vendil Pallin, Carolina (2017), Internet control through ownership: the case of Russia. *Post-Soviet Affairs*, 33:1, 16-33.

Kiehtovimmat kirjalahjat



2690

Taru Anttonen & Milla Karppinen (toim.)

**SANKARITARINOITA TYTÖILLE
(JA KAIKILLE MUILLE)**



2690

Matti Mörttinen & Lauri Nurmi

**SAULI NIINISTÖ
- MÄNTYNIEMEN HERRA**



2990

Ulla Koskinen

SUOMESSA SELVIITYMISEN HISTORIAA



2690

Risto Isomäki

VIIDEN MEREN KANSA

Löydät nämä kirjat hyvinvarustetuista kirjakaupoista tai INTOKUSTANNUS.FI

into