

# Kolme periaatetta rekisteritutkimuksen eettisen kestävyuden arviointiin

ANTTI VEILAHTI

## Johdanto

Viime aikoina viranomais- ja muiden henkilörekisterien tietosuojaan liittyvät kysymykset ovat puhuttaneet muun muassa Vastaamo-tietovuodon yhteydessä. EU:n tietosuoja-asetuksen myötä rekisteriaineistojen hallintaa on kehitetty ja rekisterien siirtämisestä, säilyttämisestä ja yhdistämisestä on tarkat säännökset, joilla pyritään suojaamaan yksilön oikeuksien ja tietosuojan toteutumista. Kysymys rekisteritutkimuksen – ja yleensä määrällisen tutkimuksen – eettisistä puolista liitetäänkin usein vahvasti tietosuojaan ja kysymykseen yksilöstä oikeudesta itseään koskeviin tietoihin. Muilta osin määrällisen tutkimuksen eettiset kysymykset palautuvat usein keskusteluun tieteellisten normien ja käytäntöjen toteutumisesta, kuten reliabiliteetista ja validiteetista, joilla pyritään parantamaan turvaamaan tieteen luotettavuus, oikeutus julkisessa keskustelussa, taloudelliset resurssit, jatkuvuus sekä tieteellinen yhteistyö.

Yhteiskuntatieteellinen rekisteritutkimus kohdistuu yleensä kuitenkin ihmisiin, ja siksi rekisteritutkimukseen liittyy erityisesti sosiaalisten ja moraalisten arvojen toteutumiseen liittyviä eettisiä kysymyksiä (Abbas ym. 2012). Näitä eettisiä kysymyksiä voidaan arvioida esimerkiksi tutkimuseettisen neuvottelukunnan laatimien eettisen ennakoarvioinnin periaatteiden näkökulmasta (Kohonen ym. 2019). Lainsäädännöllisesti rekisteritutkimus ei eroa muista tutkimuslajeista muilta osin kuin liittyen tutkittavien suostumukseen. Erityisesti myös rekisteritutkimuksen yhteydessä tulisi arvioida ”tutkittaville henkilöille, heidän läheisilleen ja mahdollisesti tutkijalle itselleen koituvat riskit [ja] haitat”.

Miten näitä riskejä ja haittoja sitten tulisi ymmärtää? Määrällisessä tutkimuksessa haitat ja riskit on perinteisesti liitetty juuri tietosuojaan ja on ajateltu, että jos henkilö ei ole tunnistettavissa eivätkä tiedot joudu väärin käsiin, hänelle ei voi tutkimuksesta syntyä haittoja. Tällaista näkökulmaa voidaan pitää liberalistisena ja yksilöllistävänä: riskejä ja haittoja arvioidaan juridisluontoisesti siitä näkökulmasta, mitä välittömiä yksilöllisiä seurauksia tutkimuksesta on yksilöitävissä.

Tilastollisesti tarkasteltuna yksilölle voi kuitenkin koitua riskejä ja haittoja esimerkiksi tutkimuksen tuottamien ryhmäkonstruktioiden kautta, vaikka vaikutuksia ei yksittäisen yksilön tasolla kyettäisi tarkasti määrittämään. Yksi yhteiskuntatieteellisen rekisteritutkimuksen tärkeimmistä tavoitteista on erilaisten ryhmitysten kuvaaminen ja tuottaminen. Olemme hyvin monenlaisten erojen summa ja se, mitä erotteluja korostetaan ja merkityksellistetään erilaisissa tutkimusasetelmissä vaikuttaa yksilön sosiaalisiin suhteisiin, häneen kohdistuviin stereotyyppioihin sekä yksilön asemaan yhteiskunnassa. Pahimmillaan tällaiset erottelut voivat johtaa ryhmien essentialisoimiseen ja jopa kuvattujen ilmiöiden performatiiviseen tuottamiseen. Tässä avauksessa ehdotan määrällisen tutkimuksen eettisen kestävyuden arviointiin kansainvälisessä keskustelussa nousutta niin kutsuttua tarkoituksilähtöistä (*built-for-purpose*) tutkimusotetta, jonka avulla rekisteritutkimukseen liittyviä riskejä ja haittoja voidaan pyrkiä vähentämään.

## Luokitukset ja politiikka

Tilastotieteen ja tilastollisen tutkimuksen historia liittyy valtioiden kehittymiseen eikä ole sattumaa, että englanninkielinen sana ”statistics” viittaa juuri kantasanaan ”state” eli valtio. Teoksessaan *The Politics of Large Numbers: A History of Statistical Reasoning* ranskalainen sosiologi Alain Desrosières (1998) kuvaa sitä, miten valtioiden synnylle ja olemassaololle keskeinen ”väestön” käsite tuotettiin tilastollisilla tekniikoilla. Ilman tilastollisia tekniikoita – eräänlaista ilmiötä numeerisesti merkityksellistävä semiotiikkaa – valtio nykyisessä hallinnallisessa, väestön käyttäytymistä ohjaavassa mielessä ei olisi mahdollinen. Olisi hankala puhua esimerkiksi suomalaisista, jos meillä ei olisi yhteistä kuvaa suomessa asuvien käyttämästä ”tyypillisestä” kielestä, tulotasosta tai elintavoista.

Toisaalta väestön tuottaminen ei ole vain poliittinen kysymys, vaan samalla kun väestö tuotetaan, tuotetaan myös politiikka käsitteenä. Toinen keskeinen väestöjen syntyminen ja tuottamiseen liittyvä ajattelija, niin ikään ranskalainen Michel Foucault (2010), on teoksessaan *Turvallisuus, Alue ja Väestö: hallinnallisuuden historia* kuvannut modernille ajalle tyypillistä erilaisten ihmisten ohjaamiseen liittyvien tekniikoiden kehitystä niin kutsutulla ”police”-verbillä. Sillä ei alun perin viitattu rikollisuuden torjuntaan, vaan keskiajan jälkeen kehittyneeseen yleiseen pyrkimykseen ohjata yksilöä vallitsevien normien mukaisesti. Tällaisten vallankäytön muotojen systematisointi mahdollisti kansallisvaltioiden ja väestöjen kaltaisten käsitteiden synnyn sekä tilastoihin tai ”väestötieteseen” perustuvan yksilöiden ohjaamisen.

Nykyään sekä julkiset että yksityiset toimijat tuottavat tilastoilla hyvin monenlaisia ryhmiä ja luokituksia. Sosiaalisen konstruktionismin perinteen mukaan tilastot eivät kuvaa jotain erillistä, objektiivista tai ulkoista todellisuutta, vaan erilaiset ryhmitykset ja luokitukset osallistuvat kuvaamaan todellisuuden tuottamiseen korostamalla tiettyjä erotteluja toisten kustannuksella (Berger & Luckmann 1966). ”Luokittelu on inhimillistä”, kuten tieteellisen tiedon sosiologian klassikko *Sorting Things Out* (Bowker & Star, 1999) asian ilmaisee. Myös määrällisiä menetelmiä käyttävät tieteenalat toimivat ”performatiivisesti” eli sosiaalisia verkostoja ja hegemonisia identiteettejä tuottavina aktiivisina toimijoina, ei vähiten ihmistieteissä (vrt. Butler 1990; Law & Singleton 2000).

## Syntyperän tuottaminen rekisteritutkimuksissa

Identiteettien ja toimijuuden tuottamisen näkökulmasta yksi eettisesti hankalimpia kysymyksiä on syntyperään liittyvät luokitukset. On helpompaa ottaa politiikan kohteeksi vaihtuvat elämäntilanteet kuin tekijät, joihin yksilö ei voi edes periaatteessa vaikuttaa. Yhtenä esimerkkinä tästä on sukupuoleen liittyvä tutkimus, joka on toisaalta tärkeää yhdenvertaisuuden arvioimiseksi, mutta voi samalla tuottaa sukupuoleen liittyviä stereotyyppioita. Monista kyselytutkimuksista löytyy jo kohta ”muu sukupuoli”, joka antaa yksilölle lisää valtaa sen suhteen, miten oma, henkilökohtainen sukupuolikäsitys heijastuu tutkimustuloksissa. Viranomaisrekisterit ovat valitettavasti olleet kehityksessään hitaampia.

Toinen tällä hetkellä ajankohtainen aihe liittyy ”maahanmuutosta” käytävään keskusteluun. Esimerkiksi Helsingin sanomat uutisoi 3.12.2020 otsikolla ”Kelan tuore tutkimus: Maahanmuuttajat saivat viidesosan toimeentulotuista ja lähes 30 prosenttia Kelan työttömyysturvasta” ja julkaisi 5.12.2020 jatkoartikkelin, jossa todettiin, että ”Kelan tuore selvitys osoittaa, että nimenomaan pakolaisina ja turvapaikanhakijoina Suomeen tulleilla maahanmuuttajilla on suuria vaikeuksia työllistyä”.

Muuttoliikkeeseen liittyvä julkinen keskustelu on hyvä esimerkki siitä, että tieteellisen tiedon keskeinen yleisö on julkinen ja demokraattinen keskustelu ja että tutkimustuloksia julkaisevat tutkijat eivät välttämättä jaa samoja arvoja tutkimuksia lukevan yleisön kanssa (vrt. Wynne 1995; Irwin & Michael 2003). Juuri tämä voi johtaa niin sanottuihin *performatiivisiin* vaikutuksiin, koska se, miten muuttoliikkeeseen liittyviä eroja ja toisaalta niiden suhdetta ”valtaväestöön” kehystetään, ohjaa identiteettien ja stratifikaatiota tuottavien rakenteiden muodostumista, mikä puolestaan voi vaikuttaa esimerkiksi työnantajien rekrytointipäätöksiin – tutkimuksella saatetaan siis tuottaa niitä eroja, jotka on alun perin otettu tutkimuksen kohteeksi.

Kelan aihetta koskevassa tutkimusraportissa pyrittiin käyttämään mahdollisimman neutraaleja, yleisessä käytössä olevia käsitteitä ja erittelemään ilmiöiden taustalla olevia elämäntilanteita (Jauhinainen & Raivonen 2020). Julkisessa keskustelussa tutkimukseen liitetty retoriikka oli kuitenkin tutkimusraportissa käytettyä voimakkaampaa. On tär-

keää huomata, että tällaiset mahdollisesti formatiiviset vaikutukset eivät sinänsä ole tutkimusraporttien ominaisuuksia, vaan ne liittyvät siihen laajempaan sosiaaliseen verkostoon, jossa tutkimustulosten analysointi, raportointi ja lukeminen tapahtuu (vrt. Irwin & Michael 2003; Latour 2005).

Kyse ei kuitenkaan ole vain siitä, voiko tutkija edes periaatteessa vaikuttaa tekstin lukijassa aiheuttamiin mielleyhtymiin, vaan myös tutkimusaineiston ja -asetelman rakenteeseen liittyvistä – ja siten eettisen ennakoarvioinnin kannalta tärkeistä – kysymyksistä.

## Tarkoituslähtöisyys eettisesti kestäväen rekisteritutkimuksen perustana

Vaikka tilastollista tutkimusta koskevista eettisistä kysymyksistä on niukasti kansainvälistä keskustelua verrattuna esimerkiksi laadullisesta tutkimuksesta käytyyn keskusteluun, Michael Zyphur ja Dean Pierides (2017) esittävät, että monia eettisiä sudenkuoppia voitaisiin välttää, jos tilastollinen tutkimus räätälöitäisiin tarkoituksellisesti (*built-for-purpose*). Se tarkoittaa sitä, että tutkimuksen tulisi olla suunniteltu eettisesti vastaamaan *spesifeihin ongelmiin ja tilanteisiin kontekstin kannalta relevantein käsittein* – ei niinkään yleistilastoiksi, joissa vain kuriositeetin vuoksi kuvataan vaikkapa tiettyä vähemmistöryhmää ilman tarkemmin määriteltyjä tavoitteita.

Esimerkiksi palveluiden kehittämisen näkökulmasta voisi olla perusteltua tutkia muualta Suomeen muuttaneille tarjottavia erityispalveluita vaikkapa vertailemalla palvelujen kehittymistä ja vaikuttavuutta eri vuosina mutta saman ryhmän sisällä. Mutta tarvitaanko sellaista differentiaalisia identiteettejä tuottavaa tutkimusta, jossa yleisellä tasolla verrataan ”valtaväestöä” ja ”muualta tulleita”? Eroavathan kaikki Suomessa asuvat lähtökohdiltaan. On alkoholistien lapsia, oppimisvaikeuksista kärsiviä, lapsena huostaanotettuja tai väkivaltaa kohdanneita.

Jokaisen vertailun yhteydessä tulisikin kysyä, *mihin tarkkarajaiseen ongelmaan* vertailulla halutaan selvyyttä, ja voisiko kysymyksiin vastata jotenkin relevanttimmin kuin kokonaisia väestöryhmiä erottelemalla. Laskeutumalla ongelmien ja aineistojen kanssa satelliittiperspektiivistä lähemmäs ruohonjuuritasoa voidaan paremmin eritellä niitä tekijöitä, jotka tosiasiallisesti aiheuttavat haasteita yksilöiden erilaisissa elämän-

tilanteissa ja suhteessa palvelujärjestelmään. Samalla tulee kysyä, onko valittu tutkimusasetelma kaikkein *tarkoituksenumukaisin* näiden haasteiden kannalta, jos kyseessä on esimerkiksi työnantajan stereotyyppioihin liittyvä työllistymisongelma eikä esimerkiksi erityisryhmän sosiaalietuuksien käyttö sinänsä.

Tutkimuksen eettisyys, tarkoituksenumukaisuus ja relevanssi tutkittavien ilmiöiden ja haasteiden kannalta on kysymys, johon ei ole oikeaa ja väärää vastausta. Eettisiä riskejä ei voida koskaan kokonaan välttää, mutta aina voidaan pyrkiä aiempaa osuvampiin, tarkoituksenumukaisempiin ja eettisiin riskejä vähentäviin lähestymistapoihin. Seuraavaksi esitän kolme holistista periaatetta, joiden avulla rekisteritutkimuksen eettisiä riskejä voidaan pyrkiä vähentämään: yksilön tarpeiden asettamisella etusijalle; enemmistö-konstruktioiden välttämällä ryhmien erityistarpeita kunnioittaen; sekä aineiston ymmärtämisellä subjektiviteettien ja toimijuuden tuottajana.

### Periaate 1: aineiston yksiköt subjektiviteetin ja toimijuuden tuottajina

Tutkimuseettisesti tärkeät kysymykset alkavat viimeistään silloin, kun mietitään tutkimukseen soveltuvaa aineistoa. Aineisto toimii ongelmien kehystämisen ja subjektivaation pohjana. Kun muutoliikettä koskevaa tutkimusta päätetään tehdä esimerkiksi Kelan asiakkaista koostuvan rekisteriaineiston kautta vertaamalla tiettyjä asiakkaita toisiin, kehystetään muutoliikkeeseen liittyviä ongelmia käytettyjen tilastoyksikköjen kautta.

Voidaan esimerkiksi kysyä, jos Kelan tutkimusraportin (Jauhiainen & Raivonen 2020) heijastamaa työntekijöiden ja työnantajien kohtaantongelmaa tarkasteltaisiin työnantajista koostuvalla rekisteriaineistolla, kehystettäisiinkö ongelmaa toisin. Olisiko *Helsingin Sanomat* voinut vaikkapa uutisoida, että ”työnantajista X prosenttia on haluttomia palkkaamaan työnhakijoita, jotka kokevat itsestään poikkeavina”?

Työnantajien syrjivä asenne voisi hyvin johtaa havaittuihin eroihin toimeentuloturvaetuuksien käytössä kuitenkin vastuullistamatta siitä työntekijöitä. Sen sijaan tutkimusraporttia seuranneessa keskustelussa keskityttiin puolueen välisiin eroihin siinä, ”kuinka suurena ongelmana ne pitävät heikkoa työllistymistä”, ikään kuin vallitsisi yksimielisyys itse ongelman kehystämisestä – siis työntekijöiden vastuuttamisesta työnantajien sijaan tilanteen kannalta keskeisimpinä toimijoina.

## **Periaate 2: yksilön tarpeet ja niiden kannalta keskeiset mittarit**

Jos tutkimuksen tarkoituksellisuus ja relevanssi tarkasteltavan aiheen kannalta otetaan eettisesti kestävä tutkimuksen lähtökohdaksi, on pohdittava sitä, minkälaiset mittarit ovat tutkimustavoitteiden kannalta kaikkein tarkoituksellisia. Mitä ne kuvaavat ja miten käytetyt mittarit toimivat vastuuttamisen ja subjektivaation välineinä? Voidaanko mittareiden antamia tuloksia käyttää tutkittavien kannalta vähemmän tutkimuksen mukaisiin kysymyksenasetteluihin?

Esimerkiksi jos tutkimuksen tavoitteena on yhdenvertaisuuden toteutumisen arviointi, millaiset mittarit ovat yhdenvertaisuuden toteutumisen kannalta kaikkein keskeisimpiä. Ja mitä edes tarkoitetaan yhdenvertaisuudella? Tarkoitetaan-ko sillä samankaltaisuutta ja symmetristä oikeudenmukaisuutta John Rawlsin esittämällä tavalla; huomioidaanko kulttuuriset siteet, joita Will Kymlicka on peräänkuuluttanut; ja mikä arvo annetaan yksilön erityisille lähtökohdille, autenttisuudelle ja dialogiselle identiteetille, kuten Charles Taylor on kommunitaristisessa mallissaan kysynyt?

Esimerkiksi palvelujärjestelmän piiriin pääsemisen nopeus työttömyyden kohdatessa voisi olla sellainen mittari, joka ei korosta erityisryhmän luonnetta vaan palvelujärjestelmän vastuuta ja toimijuutta. Tällöin ei tarvitse verrata muita ryhmiä vaan ainoastaan henkilöitä, jotka ovat tosiasiallisesti kohdanneet työttömyyden.

Sen sijaan yleisten työttömyys- tai työllisyyslukujen vertailu väestöryhmien välillä ei ole välttämättä perusteltua, koska niissä käsitellään yhtä aikaa palveluiden toimivuuden lisäksi muun muassa valikoitumiseen liittyviä tekijöitä. Vaikka esimerkiksi psyykkiset ongelmat olisivat yleisempiä jossain tietyssä ryhmässä, niihin liittyvien yksilöllisten haasteiden kannalta ei väestöryhmällä ole välttämättä mitään merkitystä, tai ainakin tällaisten merkitysten määrittämiseen tulisi käyttää intersektionaalisesta tutkimusperinteestä ammentavia, yksilöiden välisten erojen moninaisuutta korostavia tutkimusmenetelmiä. Onkin usein parempi keskittyä tutkimusasetelmassa spesifien psykiatristen tekijöiden vaikutuksiin tai esimerkiksi psykiatristen palveluiden toimivuuteen – ei niinkään demografisiin eroihin psykiatristen ongelmien ilmaantuvuudessa.

Erityisesti väestöryhmiä koskevassa tutkimuksessa tulisikin välttää mittareita, jotka eivät kyke-

ne erottelemaan niitä yksilöllisiä tarpeita, joita erilaisten palveluiden tai etuuksien käyttö heijastaa, koska tällöin korostetaan ryhmien välisiä erottelua ja differentiaalisia identiteettejä yksilöiden tarpeiden sijaan. Ryhmien erotteluun pyrkivät mittarit ovat ongelmallisia myös siksi, että emme koe kaikkien erottelujen esille tuomista samalla tavalla tarpeelliseksi. Emme esimerkiksi tuota tilasto- ja pääkaupunkiseudulla syntyneiden ja muualta Suomesta muuttaneiden välisestä eroista vaikkapa kaupungin puistopalveluiden käytössä. Jos tutkimuksen tarpeellisuutta ei arvioida yksilöiden tarpeiden näkökulmasta vaan yleisenä kuriositeettina, on vaara, että tutkimus palvelee tai ruokkii erilaisia identiteettipoliittisia tavoitteita.

## **Periaate 3: ryhmien erityistarpeet ja enemmistö-konstruktioiden välttäminen**

Edustuksellisuuden näkökulmasta rekisteriaineistot eivät edusta vain yksilöitä vaan myös ryhmiä. Siten myös ryhmien tarpeiden ymmärtäminen on keskeinen periaate määrällisen tutkimuksen tarkoituksellisuutta arvioitaessa. Tilastollisessa tutkimuksessa luodaan luokituksia, joihin liitetyt yksilöt voivat itse kokea edustavansa ryhmittelyjä vastaavia sosiaalisia luokitteluja tai jotka usein muiden näkökulmasta liitetään tällaisiin ryhmiin. Ryhmien tarpeita tulisi arvioida juuri tämän edustuksellisuuden kautta.

Yksi peruste makrotason luokitusten käyttämiseen on yhdenvertaisuuden toteutumisen tarkastelu, jolloin ainakin implisiittisesti vedotaan Rawlsin ”symmetriseen” yhdenvertaisuuskäsitteeseen kulttuuristen siteiden ja dialogisten identiteettien sijaan, jolloin haasteena on se, että yhdenvertaisuutta tarkastellaan nimenomaan ”valtaväestön” tarpeiden näkökulmasta. Näennäisesti neutraaleista luokituksista huolimatta ihmisryhmiin kohdistuva tutkimus saattaa siten heijastaa erilaisia vallankäytön muotoja ja stereotyyppioita sekä kulttuurisia ja biologisia erotteluita tuottavia kytköksiä.

Esimerkiksi muuttoliikkeeseen liitettyssä rekisteritutkimuksessa on ollut tapana verrata tiettyä ryhmittelyä – esimerkiksi ”Suomen ulkopuolelta muuttaneita” – toiseen ryhmittelyyn, jota on luonnehdittu ”kantäväestöksi” (Tervola & Verho 2014) tai ”valtaväestöksi” (Jauhiainen & Raivonen 2020). Tällöin voidaan kysyä, missä määrin eri ryhmiä vertaileva tarkastelu edistää sellaisia tarpeita, joita muualta Suomeen muuttaneet kokevat omien ryhmiensä edustavan. Vai vastaako tutkimus ennemminkin sel-

laisiin tarpeisiin, joiden koetaan edustavan ”enemmistöä” tai ”valtaväestöä” ja jotka voivat olla ristiriidassa erityisryhmän edustamiin tarpeisiin nähden?

Identiteettien ja sosiaalisten rakenteiden tuottamisen riskejä ei voidakaan välttää muuttoliikettä koskevassa tutkimuksessa, koska ”muuttaja” (*migrant*) käsitteenä ja sen kehystäminen on aina ollut politiikan ja valtion tuottamisen ytimessä – ei vähiten kansallisvaltioiden syntyessä niin kutsutulla ”massamuuton aikakaudella” 1840–1914 (Hatton & Williamson 1998). Historiallisesti valtio käsitteenä on kehittynyt nimenomaan negatiivisesti määrittäneessä suhteessa muuttoliikkeeseen ”valtaväestön”, ”kansan” tai vastaavan enemmistön tarpeiden edustajana.

Kuten kriittisessä muuttoliiketutkimuksessa on todettu, ”maahanmuutossa” ei olekaan kysymys muuttamisesta tai sen edellyttämistä palveluista sinänsä (vrt. esim. kaupungistuminen, kosmopoliittisuus) vaan identiteettien tuottamiseen pyrkivästä asemoitumisesta tiettyyn ”maahan”, ja yleisemmin osaksi tiettyä ”kansallisvaltiota”, joka ei viittaa liikkeeseen ja toimijuuteen vaan yksilön ”paikkaan” (vrt. Nail, 2018). Asemoituminen, jolla on vahvoja konnotaatioita syntyperän kanssa, on väistämättä latautunut käsite, vaikka se pyrittäisiin kuvaamaan kuinka neutraalisti hyvänsä.

Tilastot ovat keskeisessä asemassa, kun tuotetaan erilaisia tarinoita tai hegemonisia identiteettejä (vrt. Bulter 1990) tyypillisistä ”suomalaisista”, ”valtaväestöstä” tai vaikkapa ”heteroseksuaalisuudesta”. Samalla tuotetaan differentiaalisia identiteettejä ”vähemmistöistä”, esimerkiksi ”valtaväestöön” suhteutettua kertomusta ”Suomen ulkopuolella syntyneistä”, ikään kuin heitä voitaisiin tutkia relevanttina kokonaisuutena ja tutkimus voisi edustaa tällaisen tilastollisena konstruktiona syntyneen luokittelun tarpeita sen sijaan, että tarkasteltaisiin niitä tosiasiallisia ryhmiä ja tarpeita, joita tutkimukseen osallistuvat itse kokevat edustavansa.

”Enemmistö” ei ole ongelmaton myöskään abstraktina käsitteenä, vaikka määrällisessä tutkimuksessa perinteisesti se on viittannut tilastoyksiköiden määränemmistöön. Tosiasiassa tilastollinen tutkimus edellyttää tutkimusyksiköiden muodostaman ”väestön” tai ”perusjoukon” etukäteistä muodostamista. Raja väestöön kuuluvista ja kuumattomista on onto-eettinen ja samalla poliittinen päätös, joka jo sinänsä edellyttää demokraattista tahtoa edustavan ”enemmistön” olemassaoloa, joten enemmistön tilastollinen määritelmä on väistämättä tautologinen.

Toisaalta voidaan kysyä, minkä erottelujen perusteella määränemmistö konstruoidaan – poikkeammehan me kaikki jollain tapaa toisistamme. Eri tieteenalat vastaavat tähän kysymykseen omista lähtökohdistaan. Esimerkiksi psykiatriassa normaalisuutta konstruoidaan psykometriikan keinoin, lähtemällä yksilöllisyyttä korostavasta liberaalistisesta tavoitteesta edistää yksilön sopeutumista vallitseviin normeihin. Yhteiskuntatieteellisessä tutkimuksessa tulisi kuitenkin käydä kriittistä keskustelua näiden normien sisällöstä. Siihen saadaan harvoin vastauksia rekisteritutkimuksella, mikä on jo sinänsä syy välttää ”enemmistöjä” tuotavia konstruktioita.

## Kohti kriittistä tiedolla ohjaamista

Yksi peruste muuttoliikettä koskevan akateemisen tutkimuksen puolesta on ollut se, ettei tilastojen tuottamista jätettäisi yksityisten, usein poliittisesti värittyneiden ajatushautomoiden vastuulle. Eettisiä periaatteita ja hyvää tieteellistä käytäntöä noudattaville tutkimuksille on tarvetta.

Yhtä tärkeää on se, ettei eettisiä kysymyksiä rajata reliabiliteetin ja validiteetin kaltaisiin käsitteisiin tai juridisesti yksilöitäviin yksilöllisiin vaikutuksiin kuten tietosuojaan. Eettisiä riskejä tulisi arvioida erityisesti suhteessa ryhmäidentiteettien kautta syntyviin, tilastollisesti havaittaviin mutta yksilön kokemuksiin ja edustamiin riskeihin. Minkälaisiin ryhmiin yksilö liitetään, miten tällaiset ryhmittelyt tuotetaan politiikan kohteena ja miten yksilöiden ja ryhmien edustamat tarpeet toteutuvat valittuja luokituksia käytettäessä?

Ehkä keskeisin kontribuutio, jonka akateeminen tutkimus voi ajatuspajoille tarjota, ei ole toistaa ajatuspajojen omia asetelmia vaan toimia julkisena esimerkkinä sille, miten noudattaa eettisesti kestäviä tutkimusperiaatteita. Julkisissa keskusteluissa tieteellisille tutkimuksille annetaan usein suuri painoarvo ja menetelmälliset sekä periaatteelliset valinnat voivat legitimoida erilaisiin ryhmiin ja ryhmityksiin liittyviä pyrkimyksiä.

Eettisesti kestävien periaatteiden peräänkuuluttaminen ja esimerkillinen noudattaminen tieteellisessä tutkimuksessa on tärkeää myös siksi, että myös julkisella sektorilla tilasto- ja rekisteriaineistoja soveltavat yhä enemmän myös muut kuin yhteisiin eettisiin periaatteisiin sitoutuneet tutkijat. Rekisteriaineistojen ja niin sanotun ”Ison Aineiston” (*Big Data*) analysoinnin tekniset mahdolli-

suudet kuten koneoppimismallit ja ”tekoäly” ovat kehittyneet räjähdysmäisesti – myös julkisella sektorilla, esimerkiksi Kelassa. Isojen aineistojen tutkimuskäyttöön liittyy kysymyksiä muun muassa aineistojen omistajuudesta, julkisesta saatavuudesta (Savage & Burrows 2007) ja tiedon hallintojen vastuullisuudesta, kuten esimerkiksi viime aikoina tapahtuneissa Twitter-tilien sulkemisissa on havaittu. Mutta keskustelua on käyty myös siitä, miten aineistoa kerätään ja mitä seurauksia on siitä, että ihmiset kokevat yhä useammin olevan tietoon perustuvan valvonnan, arvioinnin ja analysoinnin kohteena, kuten Shoshana Zuboff (2019) kuvaa teoksessaan *The Age of Surveillance Capitalism*. Viime vuosikymmenellä niin sanottu kriittinen datatiede (*critical data science*) nousikin laajemmin kansainväliseen keskusteluun (Boyd & Crawford 2012; Dalton ym. 2016; Helles & Ørmen 2020).

Kriittistä näkökulmaa tiedolla ohjaamiseen ei tulisi sivuuttaa myöskään – ja erityisesti – silloin, kun kyse on julkisista toimijoista ja lakisääteistä rekisteritiedosta. Pohjoismaissa asuvat ovat tottuneet olemaan rekisteritiedon tuotannon kohteena aivan eri tavalla kuin monien muiden maiden

asukkaat. Esimerkiksi liberaalisuutta korostavassa anglo-saksisessa maailmassa yleisten rekisterien ylläpito olisi poliittinen mahdottomuus, ja edes ajatus koronarokotteista muodostettavasta viranomaisrekisteristä on nostattanut salaliittoteoreettikkojen karvat pystyyn. Meillähän tällainen rekisteri on jo olemassa.

Rekisteritietoa hallinnoivilla viranomaisilla, kuten Kelalla, THL:llä ja Tilastokeskuksella, on merkittävää valtaa sen suhteen, miten ja mihin tutkimustarkoituksiin julkisesti omistamiamme, pohjoismaalaisittain ainutlaatuisia rekisteritietoja käytetään – ja minkälaisen esimerkin rekisteritietoja käyttävät tutkijat antavat. Olemme varmasti yhtä mieltä rekisteriaineistojen monista hedelmistä: ilman niitä pohjoismaainen hyvinvointivaltio ei olisi koskaan noussut kukoistukseensa. Rekisteritietoihin liittyvään valtaan tulee kuulua myös vastuu: se, miten hyvinvointivaltio tulevaisuudessa näytetään eri ryhmille, on osittain myös rekisteritutkijoiden käsissä. Siksi on ensiarvoisen tärkeää pohdita tutkimuksesta koituvia eettisiä riskejä myös sosiaalisesta eikä vain yksilön oikeuksia korostavasta liberalistisesta näkökulmasta.

## KIRJALLISUUS

- Abbas, Fathelrahman A. & Babikir, Mohamed, O. & Mirghani, Mohamed E. S. & Kabbashi, N. A. (2012) Why Ethics In Research Are Crucial? *Advances in Natural and Applied Sciences* 6 (5), 660–663.
- Berger, Peter L. & Luckmann, Thomas (1966) *The Social Construction of Reality*. New York: Doubleday.
- Boyd, Danah & Crawford, Keith (2012) Critical questions for big data. *Information, Communication & Society* 15, 662–679.
- Bowker, Geoffrey C & Star, Susan L. (1999) *Sorting Things Out. Classification and Its Consequences*. London: MIT Press.
- Butler, Judith (1990) *Gender trouble: Feminism and the subversion of identity*. New York: Routledge.
- Dalton, Craig M. & Taylor, Linnet & Thatcher, Jim (2016) Critical data studies: a dialog on data and space. *Big Data & Society* 3, 1–9.
- Desrosières, Alain (1998) *The politics of large numbers: A history of statistical reasoning*. Harvard University Press.
- Foucault, Michel (2010) *Turvallisuus, alue, väestö: hallinnallisuuden historia*. College de Francon luennot 1977–1978. Helsinki: Tutkijaliitto.
- Hatton, Timothy J. & Williamson, Jeffrey G. (1998) *The age of mass migration: Causes and economic impact*. Oxford University Press on Demand.
- Helles, Rasmus & Ørmen, Jacob (2020) Big data and explanation: Reflections on the uses of big data in media and communication research. *European Journal of Communication* 35 (3), 290–300.
- Irwin, Alan & Michael, Mike (2003) *Ethno-epistemic assemblages: heterogeneity and relationality in scientific citizenship*. Science, Social Theory & Public Knowledge. Maidenhead: Open University Press.
- Jauhiainen, Signe & Raivonen, Leo (2020) *Maahanmuuttajien Kelan etuuksien käyttö vuonna 2018*. Työpapereita 157. Helsinki: Kela.
- Kohonen, Iina & Kuula-Luumi, Arja & Spoof, Sanna-Kaisa (2019) *Ihmiseen kohdistuvan tutkimuksen eettiset periaatteet ja ihmistieteiden eettinen ennakko arviointi Suomessa*. Tutkimuseettisen neuvottelukunnan julkaisu 3. Helsinki: Tutkimuseettinen neuvottelukunta.
- Latour, Bruno (2005) *Reassembling the Social. An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

- Law, John & Singleton, Vicky (2000) *Performing Technology's Stories: On Social Constructivism, Performance, and Performativity*. *Technology and Culture* 41 (4), 765–775.
- Nail, Thomas (2018) *The Political Centrality of the Migrant. Critical Perspectives on Migration in the Twenty-First Century*. In Marianna Karakoulaki, Laura Southgate & Jakob Steiner (eds.) *Critical Perspectives on Migration in the Twenty-First Century*. Bristol: E-International Relations Publishing, 15–27.
- Savage, Mike & Burrows, Roger (2007) *The Coming Crisis of Empirical Sociology*. *Sociology* 41 (5), 885–899.
- Tervola, Jussi & Verho, Jouko (2014) *Maahanmuuttajien sosiaaliturvan käyttö vuonna 2011*. Työpöytä 64. Helsinki: Kela.
- Wynne, Brian (1995) *Public Understanding of Science*. In Sheila Jasanoff & Gerald E. Markle & James C. Peterson & Trevor Pinch (eds.) *Handbook of Science and Technology Studies*. Thousand Oaks, CA: Sage Publications, 361–388.
- Zuboff, Shoshana (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.
- Zyphur, Michael & Pierides, Dean (2017) *Is Quantitative Research Ethical? Tools for Ethically Practicing, Evaluating, and Using Quantitative Research*. *Journal of Business Ethics* 143 (1), 1–16.

#### KIRJOITTAJA

Veilahti, Antti, PhD, FL, PsM, postdoc-tutkija, Kööpenhaminan yliopisto / tutkimuspäällikkö, Kela (antti.veilahti[at]gmail.com)