



**UNIVERSITY
OF TURKU**

This is a self-archived – parallel-published version of an original article. This version may differ from the original in pagination and typographic details. When using please cite the original.

AUTHOR	MATTI LINDBERG
TITLE	Rekisteriaineistojen tutkimuskäytön nykytilasta ja tulevista suuntalinjoista
YEAR	2025
VERSION	Publisher's PDF
CITATION	Lindberg, Matti Rekisteriaineistojen tutkimuskäytön nykytilasta ja tulevista suuntalinjoista YHTEISKUNTAPOLITIIKKA 90 (2025):1, p. 91-96
LICENSE	CC BY NC ND

Rekisteriaineistojen tutkimuskäytön nykytilasta ja tulevista suuntalinjoista

MATTI LINDBERG

Tällä hetkellä ei ole kootusti saatavilla tilastotietoja siitä, miten suuri osa suomalaisten yliopistojen ja tutkimuslaitosten tuottamasta tutkimuksesta perustuu rekisteriaineistoihin. Joka tapauksessa lienee selvää, että rekisteritietojen tutkimuskäyttö on kasvanut voimakkaasti viimeisen noin kymmenen vuoden aikana. Lähitulevaisuudessa rekisteriaineistojen tutkimuskäytön voidaan olettaa kasvavan entisestään tiedon tuottamisessa tapahtuneen murroksen myötä (Ruotsalainen ym. 2021; Ruotsalainen 2017). Suomalaisissa yliopistoissa ja tutkimuslaitoksissa toimivien tutkimusryhmien kansainvälisen kilpailukykyyn edistämiseksi – ja mahdollisuuksille tuottaa politiikan tekemisen kannalta relevanttia tutkimusta – keskeistä ei kuitenkaan ole pelkästään tilastoviranomaisten ja -laitosten koostamat laadukkaat rekisteriaineistot. Rekisteriaineistoja käytetään päänsääntöisesti lupaviranomaisten ylläpitämässä etäkäyttöympäristössä eli Tilastokeskuksen FIONAssa tai Findatan Kapselissa. Aineistojen kehittämisen ohella keskeisiä kehittämiskohteita ovat siten etäkäyttöympäristöt sekä aineistojen käyttöä ohjaavat käyttöluvaprosessit. Lupaviromaiset, eli käytännössä Tilastokeskus ja Findata, eivät ehkä koe olevansa tiedepoliittisia toimijoita. Kuitenkin lupaviranomaisen käyttöluvaprosessit sekä palvelujen ja aineistojen hinnoittelu vaikuttavat suoraan siihen, millä aikataululla ja missä laajuudessa, sekä millä kustannuksilla, rekisteriaineistoja hyödyntäviä tutkimuksia on mahdollista toteuttaa Suomessa.

Käyn tässä avauksessani läpi rekisteriaineistojen tutkimuskäytön lähtökohtia ja tuon esiin joitakin käyttöluvaprosessin ongelmia. Keskityn käyttöluvaprosessiin, joka liittyy rekisteriaineistojen tutkimuskäyttöön etäkäyttöympäristössä

tietosuojalain määrittelemien yleisten periaatteiden mukaisesti. Jätän siis tarkasteluni ulkopuolelle toisio lain puitteissa tapahtuvan lääketieteellisen rekisteritutkimuksen erityisongelmat.¹ Näkökulmani liittyy käyttöluvaprosessiin sisäänrakennettujen tulkinnanvaraisuuksien aiheuttamiin haasteisiin rekisteriaineistoja hyödyntävien tutkimushankkeiden suunnittelulle ja hallinnoimiselle. Keskustelen lopuksi kahdesta käyttöluvaprosessin kehittämisen tulevaisuuden suuntalinjasta.

”Tieteellinen tutkimus” ja minimointiperiaate käyttöluvan arviointikriteereinä

Rekisteriaineistojen tutkimuskäyttöä sääntelevät ennen kaikkea tietosuojalaki (1050/2018) sekä siihen pohjautuva tietosuovaltuutetun toimiston ohjeistus (2025; 2010). Tietosuojalaki tai tietosuovaltuutetun toimiston ohjeistus eivät yritä määrittellä millään tapaa tarkasti, mitä ”tutkimuksella” tarkoitetaan tai milloin tutkimukselliset kriteerit saada rekisteritiedot käyttöön täyttyvät. Vielä vähemmän ne tarjoavat vastauksia moniin keskeisiin tutkimushankkeen suunnittelun käytännön kysymyksiin, kuten erityisesti siihen, että miten laaja tutkimus voi olla niin tutkimustehtävien, tutkijoiden lukumäärän kuin pyydettyjen aineistojen osalta. Selvää on, että tutkimusaiheiden ja menetelmien moninaisuus huomioon otettuna tarkasti määriteltyjen kriteerien antaminen edellisiin kysymyksiin olisi lakitekstissä mahdotonta. Tietosuovaltuutetun toimiston ohjeistus korostaakin lupaviranomaisen oman harkin-

¹ Käyttöluvaprosessin ja aineistokustannusten kriittistä toisio lain alaisessa lääketieteellisessä tutkimuksessa ks. Reito ym. 2022.

nan keskeistä merkitystä käyttöluvan myöntämisessä (Tietosuojavaltuutetun toimisto 2010).

Tietosuojalaki edellyttää, että henkilörekisterin (so. rekisteriaineiston) tutkimuskäyttö perustuu tutkimussuunnitelmaan ja että tutkimuksella on vastuullinen johtaja tai siitä vastaava ryhmä (Tietosuojalaki 31 §). Lain mukaan tutkimussuunnitelman pitää olla ”asianmukainen” (Henkilötietolaki 14 §), mutta tämän tarkempia määrittelyjä ei laissa anneta. Tietosuojavaltuutetun toimiston ohjeiden (2010) mukaan lupaviranomaisen pitää voida varmistua siitä, että tutkimus täyttää ”yleisesti hyväksytyt” tieteellisen tutkimuksen tunnusmerkit ennen kuin lupaviranomainen voi myöntää tutkimukselle käyttöoikeuden hallinnoimissa rekisteritietoihin. Tällaisia ”yleisesti hyväksytyjä” tunnusmerkkejä ovat tutkimusongelman ja tutkimuksen tavoitteiden selkeä kuvaus, ja se että tutkimus perustuu tieteenalakohtaisesti relevanttiin tutkimusmenetelmään (Tietosuojavaltuutetun toimisto 2010). Lisäksi pyydettyjen tietojen käyttötarkoitus tulee määritellä ”riittävän” yksilöidysti (Tietosuojavaltuutetun toimisto 2010). Tietosuojavaltuutetun toimiston ohjeet (2010) eivät kuitenkaan anna tarkkoja määritelmiä siitä, mitä tässä yhteydessä voidaan pitää ”riittävänä”.

Tietojen käyttötärpeen perustelun ohella muita keskeisiä käyttöluvan arviointikriteerejä ovat tietojen minimointiperiaate sekä tietosuojan vaikutusten arviointi. Tietojen minimointiperiaatteen mukaan rekisterinpitäjän tulee kerätä ylläpitämänsä rekisteriin vain niitä tietoja, jotka ovat olennaisia rekisterin määritellyn käyttötarkoituksen kannalta (Tietosuojavaltuutetun toimisto 2024; Euroopan parlamentin ja neuvoston asetus 2016/679). Koska rekisteriaineistoa käyttävä tutkimus on kyseisen tutkimusaineiston rekisterinpitäjä, on rekisteritietojen tutkimuskäyttö alisteinen tietojen minimointiperiaatteelle. Tietosuojaan liittyvien riskien tunnistaminen ja näihin vastaaminen ovat osa käyttöluvaprosessia hakijan tekemän tietosuojan vaikutusten arvioinnin muodossa. Lupaviranomaiset edellyttävät hakijalta tietosuojan vaikutusten arviointia erityisesti siinä tapauksessa, kun käyttöluvaa haetaan erityisen arkaluonteisiin tietoihin. Näitä ovat esimerkiksi terveyttä ja sosiaali-etuuksia, rikoksia, poliittista tai uskonnollista vakaumusta tai seksuaalista suuntautumista koskevat tiedot (Tietosuojavaltuutetun toimisto 2025b).

Minimointiperiaatteen tulkinna-varaisuus ja hankkeen koko

Rekisteriaineistoihin käyttöluvaa hakevan tutkimusryhmän haasteena on sovittaa yhteen useita toisistaan riippuvia epävarmoja prosesseja, kuten rahoituksen hakeminen, tutkimushankkeen suunnittelu ja tutkijoiden rekrytointi sekä käyttöluvien hakeminen aineistoihin. Rahoituksen ja käyttöluvan hakemisen yhteensovittamista helpottaa olennaisesti se, että käyttöluvan saaminen rekisteriaineistoon ei edellytä rahoituksen osoittamista lupahakemuksessa kuvatulle tutkimukselle. Käyttöluvaa rekisteriaineistoihin ei siis tarvitse hakea jollekin tietyllä Suomen Akatemian tai jonkin muun tahon rahoittamalle tieteellisen arvioinnin läpikäyneelle ”tutkimushankkeelle”, vaan sitä voi hakea käyttöluvahakemusta varten vasten muodostetulle ”käyttöluvahankkeelle”.

Käyttöluvahakemuksessa nimettyjen rekisteriaineistojen tutkimuskäyttö on siten mahdollista aloittaa jonkin aiemman rahoituksen turvin. Itse asiassa mitä laaja-alaisempi käyttöluvan taustalla oleva ”käyttöluvahanke” on tutkimuskysymysten ja aineistojen käyttötarkoitusten määrittelyissä, sitä enemmän vastaavalla tutkimusryhmällä on liikkumavaraa organisoida tutkimuksen tekeminen omasta näkökulmastaan parhaalla mahdollisella tavalla. Tietyin käyttöluvahankkeen mukaisia aineistoa voivat hyödyntää useat, teemoiltaan käyttöluvahankkeen teemojen kanssa riittävän samankaltaiset, Suomen Akatemian ja muiden vastaavien tahojen rahoittamat varsinaiset tutkimushankkeet. Voidaan väittää, että laaja-alaisiin ja tätä kautta erittäin kalliisiin aineistokokonaisuuksiin pystyvät käytännössä hakemaan käyttöluvia vain kaikkein menestyneimmät ja verkottuneimmat tutkimusryhmät, jotka voivat olla paitsi varmoja rahoituksen jatkuvuudesta myös pystyvät jakamaan kustannuksia verkostojensa sisällä.

Rekisteritutkimuksen yksi keskeinen etu on se, että aineistojen suuret koot sekä laaja-alaiset tietosisällöt mahdollistavat sellaisten tieteellisesti kunnianhimoisten tutkimusten tekemisen, joiden toteuttaminen ei muun tyyppisillä aineistoilla olisi mahdollista (esim. Smith Jervelund & De Montgometry 2020; Hovde Lyngstad & Skardhamar 2011). Tietojen pyytämiseen mahdollisimman laaja-alaisesti ohjaa voimakkaasti myös se, että tutkimusta aloitettaessa ei yksinkertaisesti voida etukäteen tietää, mitkä tekijät osoittautuvat tärkeiksi tai millaisia jatkokysymyksiä alkuperäisistä tutki-

muskysymyksistä polveutuu. Analyysit ovat usein niin kompleksisia, että tietojen tarvetta ja kohde-ryhmiä on usein mahdotonta määritellä tarkasti etukäteen – ne elävät tutkimuksen edetessä. Toki tutkimukselle voidaan pyytää uusilla lupahakemuksilla uusia tietoja tutkimuksen edetessä, kuten usein tehdäänkin. Jatkuvien aineistopyyntö- hakemusten lähettäminen kuitenkin kuormittaa tutkimusryhmiä ja lisää kustannuksia sekä yleisesti ottaen hankaloittaa hankkeiden suunnittelua.

Yhteenvetona voidaan sanoa että, tutkijoilla on useita hyvin perusteltuja motiiveja pyytää tietoja mahdollisimman kattaviin tietosisältöihin ja muodostaa mahdollisimman suuria hankkeita niin aineistojen kuin käyttäjien määrän osalta – ainakin rahoituksen niin sallissa. Kuitenkin tietojen käyttötarkpeiden perustelujen arvioimisesta ja minimointiperiaatteen toteutumisesta huolehtimisesta tulee lupaviranomaisille potentiaalisesti sitä vaikeampaa, mitä enemmän hankkeella on käytössä tietoja ja mitä suurempi on tietojen käyttävien tutkijoiden joukko. Lupaviranomaisilla on siten ymmärrettävä motivaatio pyrkiä kontrolloimaan rekisteritutkimusta tekevien hankkeiden kokoa. Tätä kirjoitettaessa lupaviranomaisista erityisesti Tilastokeskus on kommunikoinut aikeistaan ottaa vuonna 2025 käyttöön uusia käytäntöjä liittyen suurien hankkeiden jakamiseen pienempiin osakokonaisuuksiin tietojen minimointiperiaatteen mukaisesti sekä yleisesti tarkentaa käyttölupien myöntämisen perusteita (Tilastokeskus 2024a). Minimointiperiaatteen tulkinnanvaraisuus tekee sen soveltamisesta hankkeen koon rajaamiseen kuitenkin ongelmallista.

Tietosuojavaltuutetun toimiston ohjeistus (2025; 2024; 2010) ei selvennä sitä, miten tietojen minimointiperiaate ja tietojen käyttötarkoituksen perusteleva suhteutus toinen toisiinsa. Jos luvanhakijan pitää pystyä perustelevaan, miksi pyydetty tiedot ovat tutkimuksen toteuttamisen kannalta tarpeellisia, eikä tämä samalla perustelee myös sen, että pyydettyjen tietojen joukossa ei ole tarpeettomia tietoja. Ylipäänsä on epäselvää, mihin tiedon tarpeiden perusteluissa voidaan vetää raja hankkeen tiedontarpeille kokonaisuutena verrattuna hankkeessa toimivien yksittäisten tutkijoiden tiedontarpeisiin. Hieman kärjistäen voidaan väittää, että jos minimointiperiaatteen toteutumista tulkittaisiin tiukasti siten, että kaikilla hankkeen tutkijoilla tulee olla täsmälleen sama perusteltu syy käyttää täsmälleen kaikkia samoja tietoja, ei käyttölupia myönnettäisi yksittäis-

tä tutkimuspaperia suuremmille kokonaisuuksille. Epäselvää on myös se, voiko tietojen käyttötarkoitusta perustella yksinkertaisesti sillä, että pyydettyjen tietojen saaminen parantaisi edellytyksiä tuottaa tieteellisesti merkittävää tutkimusta? Tarkeemmin sanoen, voiko tietojen tarvetta perustella käyttölupahakemuksessa sillä, että pääsy yksityiskohtaisiin ja laaja-alaisiin tietoihin parantaa edellytyksiä tuottaa korkeatasoista tutkimusta, vaikka tietojen minimointiperiaatteen mukaisesti tutkimus olisikin mahdollista suorittaa suppeammilla tiedoilla kuin mitä hakemuksessa pyydetään – tosin tässä tapauksessa myös tutkimuksen laatu olisi ”suppeampi”.

Asia erikseen on se, että tietoturvaan liittyvät riskit voivat kasvaa hankkeen koon kasvaessa, vaikkakaan tämä ei ole mitenkään itsestään selvää. Hankkeen koon, mitä tulee niin aineistojen kuin käyttäjien määrään, ja tietoturvariskien kesken ei ole mitään selkeää yhteyttä tai ainakaan lupaviranomaiset eivät ole antaneet mitään selkeitä tietoturvaan perustuvia kriteerejä hankkeiden maksimikoolle. Ymmärrettävästi tällaisia kriteerejä olisikin vaikeaa määritellä. Erityisen sensitiivisiä tietoja pyydetessä käyttöluvan hakijan tulee suorittaa tietosuoja-arviointi (ns. DPIA -lomakkeen täyttäminen). Tietosuoja-arvioinnin tekeminen epäilemättä selvittää käyttäjän vastuita pyydettyjen aineistojen tietoturvalisistä käytöstä huolehtimisessa. Nykymuotoiset tietosuojan vaikutusten arviointi -lomakkeet eivät kuitenkaan sovellu erityisen hyvin sen arviointiin, milloin jokin hanke on kasvanut tietoturvan näkökulmasta ”liian suureksi”.

Rekisterisaineistojen etäkäytön kehittämisen suuntalinjoja

Keskustelen seuraavassa kahdesta mahdollisesta tulevaisuuden suuntalinjasta kehittää rekisteriaineistojen etäkäyttöä ja käyttölupaprosessia tietoturvan ja tietojen minimointiperiaatteen toteutumisen edistämiseksi.

Ensimmäisessä suuntalinjassa lupaviranomaiset pyrkivät hallinnollisteknisillä uudelleen järjestelyillä tarkentamaan lupaehtoja ja aineistojen käytön valvontaa. Hallinnollisteknisillä toimenpiteillä tarkoitan esimerkiksi käyttölupahakemusten aiempaa ”tiukempaa” arviointia ja aiempaa tarkempien perustelujen pyytämistä tietojen käyttötarkoituksesta; lupaviranomaisten muodos-

tamia otosaineistoja sen sijaan että käyttöluhia annettaisiin kokonaisaineistoihin; (aiempaa tarkempaa) tutkimustulosten ennakkotarkastusta; sekä suurien hankkeiden pilkkomista. Tilastokeskuksen kommunikoimat suunnitelmat vuoden 2025 aikana tapahtuvista uudistuksista (Tilastokeskus 2024a; 2024b), kuten suurien hankkeiden jakaminen pää- ja alihankkeisiin sekä etäkäyttöön luovutettavien aineistojen aiempaa tarkempi rajaaminen tutkimuksen kannalta välttämättömiin tietoihin, asettuvat selvästi tämän suuntalinjan alle.

Hallinnollisteknisiin toimenpiteisiin liittyvä potentiaalinen ongelma on kustannusten nousu. Selvää on, että lupaviranomaisten lisääntynyt työtaakka lisää kustannuksia, jotka lupaviranomaiset siirtävät joko kokonaan tai osittain tutkimusryhmiltä veloitettaviin käsittely- ja palvelumaksuihin. Myös tutkimusryhmien epäsuorat kustannukset tulevat kasvamaan, kun käyttöilupaprosessi monitukaistuu. Tällaisia epäsuoria kustannuksia ovat muun muassa lupahakemusten kirjoittamiseen ja aineistojen käytön hallinnointiin sekä käyttöluapäätösten odottamiseen kuluva aika. Sikäli kuin suuria hankkeita tullaan tulevaisuudessa jakamaan pienempiin osakokonaisuuksiin, kuten ainakin Tilastokeskuksen suunnitelmissa on (Tilastokeskus 2024a), potentiaalinen ongelma tutkimusryhmille tulee olemaan aineistojen käyttökustannusten huomattava kasvu. Vaikka edellä kuvatun kaltaiset hallinnollistekniset toimenpiteet onnistuisivatkin edistämään minimointiperiaatteen toteutumista, voidaan kuitenkin epäillä, missä ne määrin onnistuvat edistämään tietoturvan toteutumista. Tämän tyyppiset uudistukset eivät millään tapaa muuta sitä etäkäyttöympäristöjen perustoimintoa, että niissä on käytössä (pseudonymisoituja) henkilötietoja suuresta joukosta suomalaisia. Käyttöilupaprosessin mahdollisesta tiukentumisesta huolimatta etäkäyttöympäristöissä olisi edelleen teoriasa mahdollista yrittää tunnistaa henkilöitä yhdistelemällä tietoja useista eri tietolähteistä. Hallinnollisteknisten toimenpiteiden potentiaalinen ongelma tai rajoitus on siis tarkemmin määriteltynä niiden eräänlainen kustannustehottomuus: niihin voi liittyä huomattava kustannusten nousu käyttäjille, mutta niillä saavutetaan vain marginaalisia parannuksia tietoturvaan.

Toinen tulevaisuuden suuntalinja liittyy etäkäyttöjärjestelmien toimintalogiikan perusteelliseen uudistamiseen. Tältä osin voitaisiin tarkastella sitä, missä määrin uudentyypisten teknologisten ratkaisujen avulla olisi mahdollista moni-

puolistaa tai organisoida täysin uudella tapaa se, millaisia rajapintoja etäkäyttöympäristöissä on erityyppisten tietojen, tai ihmiskäyttäjien ja tekoälypohjaisten automatisoitujen käyttäjien, kesken.

Havainnollistan argumenttiani tulevaisuuden etäkäyttöympäristön toimintaa kuvailevan esimerkin avulla. Haluan tutkia rekisteriaineistojen avulla, miten lapsuuden- ja nuoruudenaikaiset kasvuympäristöt – kuvattuna sosioekonomisten ja demografisten rekisteritietojen avulla – vaikuttivat vuonna x syntyneiden lasten kohortissa siihen, miten henkilö menestyi aikuisena työmarkkinoilla ja koulutuksessa. Muotoilen aineistopyynnön etäkäyttöympäristön pääkäyttäjälle. Kustannusten ja käsittelyaikojen minimoimiseksi pääkäyttäjä ei ole ihminen vaan jonkinlainen tekoälypohjainen automatisoitu järjestelmä. Määrittelen aineistopyynnössäni, mitä tietoja tarvitsen miltäkin ajankohdalta niin lapsista kuin näiden vanhemmista. Pääkäyttäjä muodostaa määrittelemäni aineiston yhdistelemällä rekisteritietoja eri lähteistä. Pääkäyttäjä ei lähetä minulle (pseudonymisoitua) henkilötietoaineistoa, vaan sen pohjalta luodun synteettisen aineiston (synteettisistä aineistoista ks. esim. Rieman 2022; Hradec ym. 2022). Analysoin synteettistä aineistoa tavalliseen tapaan: koodaan muuttujia, estimoin malleja, ja niin edelleen. Huomaan tutkimukseni edetessä, että tarvitsen lisätietoja, esimerkiksi tiedot lapsen syntymäpainosta ja mahdollisista diagnooseista, jotka voivat haitata koulunkäyntiä. Teen pääkäyttäjälle täydentävän aineistopyynnön, jonka pohjalta pääkäyttäjä toimittaa minulle uuden synteettisen aineiston. Kun olen tyytyväinen muuttujiini ja malleihini, lähetän niiden muodostamisessa käytetyt koodit pääkäyttäjälle. Pääkäyttäjä ajaa lähettämäni koodit alkuperäisessä henkilötietoaineistossa ja lähettää tulokset minulle. Olen saanut tieteellistä julkaisuani varten tulokset, jotka perustuvat henkilötietoja sisältävän ”oikean” aineiston analyysin ilman, että minulla on ollut missään vaiheessa pääsyä muuhun kuin täysin anonyymiin synteettiseen aineistoon.

En yritä tässä yhteydessä arvioida, millä oletamuksilla edellä kuvailemani analyysitapa olisi tuottanut täsmälleen samat tulokset kuin tällä hetkellä Tilastokeskuksen FIONA- tai Findatan Kapseli -etäkäyttöympäristöissä käytössä oleva tapa. Tällä hetkellä FIONAssa ja Kapselissa toimitaan siten, että tutkijalla on pääsy oikeita henkilötietoja sisältäviin rekisteriaineistomoduuleihin, joista on poistettu suorat henkilötunnisteet

ja joiden tietoja voi yhdistellä toisiinsa pseudo-nymsoidun henkilönumeron avulla. Joka tapauksessa voidaan esittää kysymys, minkälaisen käyttöluvan olisin tarvinnut esimerkin mukaisen tutkimuksen toteuttamista varten. Jos pääkäyttäjänä toimiva tekoälypohjainen järjestelmä olisi tarpeeksi kehittynyt, tai jos tutkimuskysymykseni olisi ollut riittävän yksinkertainen, olisin voinut pyytää pääkäyttäjää tuottamaan tulokset suoraan, ilman että olisin tarvinnut pääsyä synteettiseen aineistoon. Olisinko tarvinnut käyttöluvan, koska ohjaamallani tekoälyllä on pääsy henkilötietoihin ja se toimii ikään kuin tutkimusryhmän jäsenenä? Mutta miksi olisin tarvinnut käyttöluvan, jos minulla ei kerran ole pääsyä henkilötietoihin – pelkkä tulosten näkeminen ei luonnollisestikaan voi vaatia käyttöluvaa.

Tämän keskustelunavauksen tarkoituksena ei selvästikään ole esittää valmista mallia tulevaisuuden etäkäyttöympäristön ja siihen liittyvän käyttö-

lupaprosessin toiminnasta. Sen sijaan tarkoitukseksi on herättää yhtäläisesti niin tutkijoiden kuin lupaviranomaisten parissa keskustelua siitä, miten rekisteritietojen tutkimuskäyttö tulisi organisoida tulevaisuudessa ja miten tulevaisuuden etäkäyttöympäristön tulisi toimia. Käyttölupaprosessin kehittämisen keskeisiä tavoitteita tulisi olla aineistojen saavutettavuuden parantaminen vaarantamatta tietoturvaa sekä kustannusten madaltaminen. Etäkäyttöympäristön toimintaan liittyvillä uudentyypisillä teknologisilla ratkaisuilla voisi olla keskeinen merkitys edellä mainittujen tavoitteiden saavuttamisessa. Joka tapauksessa keskeinen haaste lupaviranomaisille tulee olemaan yrittää pysyä mukana tiedontuottamisen ja tiedon tutkimuskäytön murroksessa. Ei ole selvää, missä määrin tähän haasteeseen vastaaminen tulee onnistumaan nykymuotoisten käyttölupaprosessin ja etäkäyttöympäristöjen toimintaa tehostamalla.

KIRJALLISUUS

- Euroopan parlamentin ja neuvoston asetus 2016/679 (2016) Asetus luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/46/EY kumoamisesta (yleinen tietosuojasetus). <https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=celex%3A32016R0679> (luettu 31.1.0.2024)
- Hovde Lyngstad, Torkild & Skardhamar, Torbjørn (2011) Nordic Register Data and Their Untapped Potential for Criminological Knowledge. *Crime and Justice* 40 (1), 613–645. <https://doi.org/10.1086/658881>
- Hradec, J. & Craglia, M. & Di Leo, M. & De Nigris, S. & Ostlaender, N. & Nicholson, N. (2022) Multipurpose synthetic population for policy applications. Luxembourg: Publications Office of the European Union (JRC128595, EUR 31116 EN). doi:10.2760/50072
- Reito, Alekski & Sanmark, Enni & Tuovinen, Timo & Seppälä, Toni T. & Kuitunen, Ilari & Ponkilainen, Ville & Ekman, Elina & Kauppila, Joonas H. (2022) Toisiolaki – lääketieteellisen tutkimuksen mahdollistaja vai tukahduttaja? *Suomen Lääkäri-lehti* 77, e30589
- Riemann, Robert (2022) Synthetic data. In Massimo Attoresi & Stefano Leucci (eds.) *TechSonar 2022–2023 Report*. European Union, 14–15. https://www.edps.europa.eu/system/files/2022-11/22-11-10_techsonar_report_22_23_en.pdf (luettu 31.1.0.2024)
- Ruotsalainen, Kaija & Jokimäki, Hanna & Pitkälä, Jukka & Nieminen, Jari & Pasila, Aura (2021) Hallinnolliset rekisterit ovat tilastotuotantomme kivijalka – nyt niiden käyttö on murroksessa. *Tilastokeskuksen artikkelit* 26.3.2021. <https://stat.fi/tietotrendit/artikkelit/2021/hallinnolliset-rekisterit-ovat-tilastotuotantomme-kivijalka-nyt-niiden-kaytto-on-murroksessa> (luettu 31.1.0.2024)
- Ruotsalainen, Kaija (2017) Rekisteritiedoilla tilastotuotannon tuottavuusloikka. *Tilastokeskuksen artikkelit* 21.12.2017. <https://stat.fi/tietotrendit/artikkelit/2017/rekisteritiedoilla-tilastotuotannon-tuottavuusloikka> (luettu 31.10.2024)
- Smith Jervelund, Signe & De Montgomery, Christopher J. (2020) Nordic registry data: Value, validity and future. *Scandinavian Journal of Public Health* 48 (1), 1–4. <https://doi.org/10.1177/1403494819898573>
- Tietosuojalaki (2018) <https://www.finlex.fi/fi/laki/ajantasa/2018/20181050#L6P35>
- Tietosuojavaltuutetun toimisto (2025) Tieteellinen tutkimus ja tietosuoja. Tietosuojavaltuutetun toimiston kotisivut. <https://tietosuoja.fi/tieteellinen-tutkimus> (luettu 3.2.2025)
- Tietosuojavaltuutetun toimisto (2025b) Vaikutustendarviointi. Tietosuojavaltuutetun toimiston kotisivut. <https://tietosuoja.fi/vaiikutustendarviointi> (luettu 3.2.2025)
- Tietosuojavaltuutetun toimisto (2010) Rekisteritutkimuksen tietosuojaopas tutkijoille ja tietopyyntöjä käsitteleville viranomaisille. <https://tietosuoja.fi/documents/6927448/10594424/Rekisteritutkimuksen+tietosuojaopas/dd6cd081-1557-6f77-8794->

5edc8555c557/Rekisteritutkimuksen+tietosujo-
opas.pdf?t=1541599095000 (luettu 31.1.0.2024)
Tietosuojavaltuutetun toimiston (2024) Tietojen mi-
nimointi. Tietosuojavaltuutetun toimiston kotisi-
vut. <https://tietosuoja.fi/tietojen-minimointi> (luet-
tu 31.1.0.2024)
Tilastokeskus (2024a) Tilastokeskuksen tutkimusaineis-
tojen tietosuoja vahvistuu (26.9.2024). (Tilastokes-

kuksen tutkijapalveluiden asiakastiedote). [https://
stat.fi/media/uploads/tup/mikroaineistot/tutkijati-
laisuus_yhteenveto_\(4\).pdf](https://stat.fi/media/uploads/tup/mikroaineistot/tutkijati-
laisuus_yhteenveto_(4).pdf) (luettu 8.11.2024)
Tilastokeskus (2024b) Ajankohtaista tutkijapalveluista
– Lokakuu 2024 (Tilastokeskuksen tutkijapalvelui-
den asiakastiedote). Tilastokeskuksen tutkijapalve-
luiden kotisivut. [https://stat.fi/tup/tutkijapalvelut/
index.html](https://stat.fi/tup/tutkijapalvelut/
index.html) (luettu 8.11.2024)

KIRJOITTAJA

Lindberg, Matti, VTT, tutkimuspäällikkö, Turun yliopisto (matti.lindberg[at]utu.fi)