

Pohjoismainen rekisteritutkimuspäivä — Register-based research in the Nordic Countries — possibilities and practices — Helsingissä 5.4.2011

Kaikissa Pohjoismaissa on rikkaat terveys- ja sosiaalirekisteriaineistot sekä tutkimusmyönteinen tietolainsäädäntö. Välillä maat kilpailevat rekisteritutkimuksen kärkipaikasta, mutta yhä useammin viisi Pohjolan maata yhdistää voimansa yhteisten tutkimusten läpiviemiseen. Rekisteritutkimuksen tukikeskus (ReTki) järjesti 5.4.2011 THL:n tiloissa koko päivän kestäneen Pohjoismaisen rekisteritutkimusseminaarin.

ReTkin johtaja Irma-Leena Notkola avasi seminaarin ja kertoi, että tilaisuus on järjestetty erityisesti Suomen tarpeita varten. Siksipä aamupäivän aikana keskityttiin kartoittamaan kotimaan tilannetta. Kimmo Koivunen Tieteen tietotekniikan keskukselta (CSC) kertoi julkisen tiedon hyödyntämistä edistävistä hankkeista. Yleinen tavoite on, että julkisen sektorin tiedot pitäisi saada nykyistä tehokkaampaan käyttöön. Tästä on strateginen periaatepäätös EU-tasolla, joka on toiminut lähtökohtana lähinnä kuluneen vuoden aikana tehdyille työille Suomessakin.

Vanhasen ja Kiviniemen hallitusten ohjelmissa on todettu, että aineistot pitäisi harmonisoida ja myöhemmin selvittää, miten niihin pääsisi paremmin käsiksi. Eduskunnan tulevaisuuskomitean selvityksen mukaan Suomessa on hyviä ai-

neistoja, mutta korkeat kustannukset estävät niiden hyödyntämistä. Avoin kysymys on myös se, kuinka järjestelmät saadaan toimimaan yhteen. Liikenne- ja viestintäministeriö on myös perustanut työryhmän aineiston saatavuuden ja uudelleenkäytön parantamiseksi. Valtiovarainministeriössä kaksi työryhmää työskentelee julkisen tiedon hyödyntämisen parissa. Ensimmäinen keskittyy selvittämään aineistojen uudelleenkäytön tarpeita ja vaikutuksia julkishallinnon kannalta, kustannuskysymyksiä ja sitä, kenen kaikkien pitäisi päästä aineistoihin käsiksi. Työryhmän tänä vuonna valmistuvassa työssä on tarkoitus myös tehdä toimenpide-ehdotuksia siitä, miten toivotut vaikutukset saadaan aikaan romuttamatta koko järjestelmää ja minkälaisia muutoksia lainsäädäntöön tarvitaan. Toinen työryhmä määrittelee teknisiä rajapintoja aineistojen hyödyntämiseen.

Koivunen kertoi myös opetus- ja kulttuuriministeriön asettaman Tutkimuksen tietoaaineistot -työryhmän selvityksestä, jonka tuloksena on alkuvuodesta julkaistu raportti otsikolla ”Tieto käyttöön – Tiekartta tutkimusten sähköisten tietoaaineistojen hyödyntämiseksi”. Selvityksen mukaan on ainakin kolmentyyppisiä aineistoja: 1) julkisen sektorin hallinnoimia tietoaaineistoja, joiden käyttöä tietosuoja-asiat tms. eivät rajoita; 2) julkisen sektorin ja tutkimusjärjestelmän tuottamia tai hallinnoimia tietoaaineistoja, joiden käyttöön liittyy erityisrajoituksia tai vaateita liittyen esim. tietosuojaan ja tekijänoikeuksiin; 3) tutkimusjärjestelmän tuottamat tiedot, jotka pitäisi saada laajempaan käyttöön. Selvityksen mukaan pitäisi luoda selvä kansallinen aineistopolitiikka, joka mahdollistaa aineis-

tojen helpon saatavuuden ja käyttämisen. Lisäksi tulisi luoda kannusteita ja palkitsemisjärjestelmiä, jotka varmistaisivat hyvälaatuisten aineistojen tuottamisen jatkossakin. Rahoituksen olisi oltava pitkäaikaisista, aineistot pitäisi inventoida ja kuvata riittävällä tarkkuudella ja nämä tiedot pitäisi saattaa helposti löydettäväksi.

Erlaisia haasteita näyttäisi olevan suunnitelmien tiellä. Raportissa on esitetty selkeitä suosituksia ja toimenpide-ehdotuksia moneen keskeiseen asiaan. Koivunen esitteli myös visiota aineistoinfrastruktuurista, jossa olisi semanttisia eri alojen klustereita ja antoi esimerkkejä, kuinka asiat voisivat toimia yhteen ja minkälaisia palveluita infrastruktuurissa voisi olla. Yksi osa tätä kansallista infrastruktuuria olisi MIDRAS-hankkeessa hahmoteltu malli rekisteritietojen etäkäyttöön. Tämä on Suomessa vasta lastenken-gissään, kun taas useassa muussa Pohjoismaassa etäjärjestelmät ovat olleet todellisuutta jo pitkään. Eri selvitysten yhteisenä keskeisenä sanomana on, että lainsäädäntöä muuttamalla, tietorakenteita harmonisoimalla ja kustannuksia mataltamalla eri sektorien tiedot on saatava nykyistä laajempaan käyttöön. Myöskään tietosuojalainsäädännön liian tiukkoja tulkintoja ei pitäisi käyttää tekosyynä aineistojen avaamisen ja tietojen hyödyntämisen vaikeuttamiselle.

Yleisökysymyksessä tiedusteltiin kuinka Tanskassa vaikeiksi todetut aineistojen keräämiseen ja tuottamiseen liittyvät vastuuongelmat on ratkaistu Suomessa. Kinnusen mukaan Suomessa on varsin hyvät aineistontuottamisprosessit, eikä asiaa ole aineistojen uusikäyttöä tarkastelevissa selvityksissä juurikaan nostettu esille. Mielestämme

tämä olisi kuitenkin erittäin tärkeä näkökulma huomioonotettavaksi aineistoinfrastruktuuria rakennettaessa. Olisiko THL:n hoitoilmoitusrekisterin eli Hilmon laatu parempi, jos tietoja käytettäisiin järjestelmällisemmin sairaaloiden laadunhallinnassa? Kuinka kauan perusterveydenhuollon toimintaa rekisteröivä AvoHilmo olisi ollut kattava, jos kentän tietojärjestelmät olisivat olleet yhdenmukaiset tai kaikki järjestelmätoimittajat yhtä ripeitä tarvittavien muutosten toteuttamisessa? Jäävätkö yksityisten palveluntuottajien avokäyntitiedot erityisesti perusterveydenhuollon (työterveyshuollon) osalta rekisterien kannalta harmaaksi alueeksi tulevaisuudessakin ilman riittävää keppiä tai porkkanaa?

Marianne Johnson Rekisteritutkimuksen tukikeskuksesta kertoi kaksivuotisesta MIDRAS-hankkeesta (Micro Data Remote Access System), joka päättyi huhtikuussa 2011. Hankkeen tarkoituksena oli – esteiden kartoittamisen lisäksi – pilotoida rekistereiden etäkäyttöjärjestelmää, jossa aineistoja ei kopioitaisi ja luovutettaisi rekisterinpitäjiltä tutkijoille, vaan tiedot voisi analysoida omalta koneelta käsin etäkäyttäen tietoturvallisten yhteyksien avulla rekisterinpitäjien tarjoamaa tutkimuspalvelintä. Tutkijat toivoivat, että etäkäyttöjärjestelmä olisi luotettava, joustava, käyttäjäystävällinen, tehokas, halpa ja vuorovaikutteinen. Lisäksi aineistojen pitäisi olla hyvin määriteltyjä ja dokumentoituja, riittävän monipuolisia ja lupaprosessien nopeita. Rekisterinpitäjät puolestaan toivoivat järjestelmän olevan turvallisen, muokattavissa olevan ja kustannustehokkaan. He myös toivoivat lisärahoitusta järjestelmän rakentamiseen ja ylläpitämiseen. Erityisenä ongelmana pidettiin myös sitä, etteivät tutkijat tunne tarpeeksi aineistoja eivätkä siten pysty tekemään hyviä hakemuksia.

Hankkeen loppuraportti toimenpide-ehdotuksineen ja pilotin kuvauksineen valmistuu keväällä 2011 ja sen kommentointikierros päättyy syksyllä 2011. Johnson mainitsi järjestelmän mahdollisiksi rahoittajiksi Sitran, Tekesin ja Suomen Akatemian sekä eri ministeriöt, kunhan työnjaosta sovitaan. Hän arveli myös, että käytännön toteuttaminen vaatii pitkäaikaista ja laajennettua hanketta, jossa järjestelmää kehitetään yhteistyössä hyödyntäen sekä MIDRAS-hankkeen että Tilastokeskuksen oman erillisen etäkäyttöjärjestelmä-hankkeen kokemuksia. Valitettavasti vuoden 2012 talousarvion osalta konkreettinen ehdotus tulee valmistumaan kuitenkin liian myöhään. Toivottavasti tämä tärkeä ja mielenkiintoinen hanke ei lykkäänny liian kauaksi tulevaisuuteen.

Pohjoismaisilta vierailta oli pyydetty lyhyitä esityksiä oman maansa tilanteesta. Ivan Thaulow Tanskan Tilastokeskuksesta kertoi aluksi hiukan heidän mikrotason tietoja sisältävistä aineistoistaan ja niiden perusyksiköistä sekä linkkaushierarkioista. Esimerkiksi verotustiedot kerätään sekä yksilöille että yrityksille ja haluttaessa voidaan yksikkötunnusteiden ja niiden hierarkioiden avulla muodostaa aineisto, jonka avulla tarkastellaan vaikkapa samalla työpaikalla työskentelevien henkilöiden terveyttä. Aineistojen yhdistämismahdollisuuksia on loputtomasti, ja myös Tanskan Tilastokeskuksen ulkopuolisia aineistoja voidaan käyttää tutkimuksissa ja linkkauksissa, kunhan aineistot toimitetaan heille. Thaulowin mukaan he tekevät aineistoja yli 300 tutkimukseen vuosittain. Koko Tanskan väestön kattavat rekisteriaineistot ovat saatavilla vuodesta 1980 lähtien, jopa pitemmältäkin ajanjaksolta. Etäkäyttöjärjestelmät ovat olleet käytössä jo yli 10 vuotta. Lupa annetaan sellaiseen aineistoon, joka

tarvitaan tutkimuksen läpiviemiseksi.

Kaikki luvat kulkevat tutkimuspalveluyksikön kautta, joka koordinoi kaikkea tutkimusta Tanskan Tilastokeskuksen aineistoilla. Se käsittelee ja hyväksyy lupahakemukset sekä valmistelee aineistot. Tietoja ei koskaan luovuteta fyysisesti tutkijalle, vaan tunnisteeton aineisto valmistellaan lopulliseen muotoon, ja se on etäkäyttöyhteyksien, salasanojen ja turvakoodien takana. Sama yksikkö neuvoo tutkijoita sekä uusissa että vanhoissa tutkimusprojekteissa. Periaatteessa tutkimuskäyttöä ei ole rajoitettu Tanskaan, vaan myös muiden maiden tutkijat voi saada aineistot käyttöönsä, kunhan heillä on kiinteä yhteys johonkin tanskalaiseen tutkimuslaitokseen, jonka Tanskan Tilastokeskus on arvioinut vakavasti otettavaksi tieteelliseksi toimijaksi.

Kaikki tutkimuskoneet ovat kokonaan erillään tilastotuotannosta. Tarvittaessa tutkijat voivat maksaa myös uusista tietokoneista, jotka valjastetaan pyörittämään vain heidän aineistojansa. Samalla vähennetään etäkäyttöjärjestelmää pyörittävien palvelimien liiallista kuormitusta. Etäkäyttöjärjestelmässä tutkijoiden käytettävissä on lukuisia yleisiä tilasto-ohjelmistoja, mutta tutkija ei koskaan saa tulostaa aineistoa yksilötasolla. Tilastotason tulokset lähetetään tutkijalle sähköposteilla, jotka taltioidaan ja niiden sisältöä tarkastetaan pistokokein. Lupaprosessi ja aineiston muodostaminen on nopeaa. Eettisen toimikunnan käsittelyä ei tarvita ja lupaan tarvittava Tanskan Tilastokeskuksen johtajan allekirjoitus saadaan parissa päivässä ja tutkimusaineisto on pääsääntöisesti käytettävissä noin kuukauden kuluessa hakemuksesta. Aika valtava ero Suomen tilanteeseen, jossa jopa vuosien odotusajat sekä luvan ja aineiston saamiseen ovat valitettavaa todellisuutta.

Vigdis Kvalheim Norjan sosiaalitieteiden datapalveluista kertoi heidän kansallisista käytännöistään. Eri laitoksilla on erilaiset säännöt ja käytännöt aineistojen tutkimuskäyttöön, mikä on rajoittanut aineistojen tutkimuskäyttöä. Lainsäädännöllä on pyritty samanaikaisesti parantamaan tietojen tutkimuskäytön mahdollisuuksia ja tietosuojaa. Rekisterinpitäjille on myös annettu määrääjat aineistojen muodostamiseen luvan saannin jälkeen: 30 päivää yhden rekisterin käytössä ja 60 päivää useamman aineiston yhdistämisessä. Norjassa etäkäyttöjärjestelmä on nimeltään RAIRD, mutta sille vasta haetaan rahoitusta.

Toisin kuin Tanskassa, Norjan Tilastokeskus voi luovuttaa tiedot tutkijoiden käyttöön. Norjan sosiaalitieteiden datapalvelu pyrkii omalta osaltaan poistamaan aineistojen käytön esteitä. Peruseriaatteena on se, että aineisto on keskeistä tieteen tekemiselle ja kun parannetaan pääsyä aineistoihin, niin samalla tuetaan tutkimusta. Sosiaalitieteiden puolella aineistojen infrastruktuuria on edistetty ja rahoitettu jo 50 vuotta. Vasta hiljattain konteksti on kuitenkin alkanut muuttua ja on ymmärretty, että kalliisti tuotettuja aineistoja on käytetty aivan liian vähän ja mekanismit niiden saamiseksi tutkimuskäyttöön ovat olleet liian kömpelöitä tutkimuksen kannalta. Nykyään aineistojen helpompaan saatavuuteen tähtäävässä yhteistyössä ovat mukana Norjan tutkimusneuvosto, tietosuojaviranomaiset, Tilastokeskus ja tutkijayhteisö. Toisin kuin Suomessa, Norjan datapalvelu arkistoi mielellään myös rekisteriaineistoja, esimerkiksi vuodesta 1992 alkaen vuosittain poimittavaa 10 prosentin väestötosta työikäisistä norjalaisista, joita seurataan mm. työvoima-, terveys- ja sosiaalirekisterien tietojen osalta. Tällainen väylä Norjan Tilastokeskuksen aineistoihin osoittautui todella suosituksi

tutkijoiden piirissä, mutta se oli käytännössä kallis ja johti pyyntöjen lisääntyessä kapasiteettiongelmiin ja pitkiin jopa puolen vuoden jonoihin aineiston saamiseksi. Vuoteen 2007 asti aineistot olivat ilmaisessa käytössä, mutta sen jälkeen niiden kalleus on vähentänyt aineiston käyttöä. Ongelmat on tunnistettu ja nyt uutta aineistoa tehdään 20 prosentin väestötöksellä. Myös aineiston saatavuutta aiotaan parantaa.

Ruotsin tiedenevoston johtaja Magnus Stenbeck esitteli aluksi Ruotsin rekisterijärjestelmää, myös sen laajennettuja osia, kuten terveydenhuollon laaturekistereitä ja biopankkiaineistoja. Suurin ongelma on aineistojen sijaitseminen eri paikoissa. Valtakunnallisia toimijoita ovat Tilastokeskus ja Sosiaalivaltio, mutta 71 kansallisesta laaturekisteristä vastaa 20 maakäräjää ja biopankkiaineistot ovat vielä enemmän hajallaan.

Eri aineistojen tutkimuskäytön parantamiseksi ruotsalaiset kävivät huolellisesti läpi kaikki varteenotettavat lainsäädännöt JUDI-hankkeessaan. Suurten tutkimustietokantojen perustamista hankaloittaa monimutkainen ja osittain ristiriitainen tietosuojaa ja tutkimuseettistä arviointia koskeva lainsäädäntö. Näin rekisteriaineistoja keräävä kansallista tietoaarkistoa ei voida perustaa ainakaan toistaiseksi, vaikka tällaiselle toiminnalle olisikin tarvetta: tällaiselle yleisaineistolle ei voida myöntää eettistä lupaa, joka on edellytys aineiston käyttöluvan saamiselle. Lupahakemuksiin ja tietopyyntöihin toivotaan myös Ruotsissa selkeämpiä sääntöjä. Ehkä Suomen ja Tanskan kaltaiselle Rekisteritutkimuksen tukikeskukselle olisi tarvetta myös Ruotsissa? Yksi tapa ratkaista lainsäädöllinen umpisolmu olisi nostaa (eri rekisteritietoja yhdistävä) tutkimus yhtä tärkeäksi rekisterin perustamisen syyksi kuin muutkin mahdolliset syyt.

Tämä kuitenkin vaatisi helppoa pääsyä lähtörekistereihin, jota taas rajoitetaan eri säädöksillä, joita voi olla mahdotonta purkaa.

Tutkimusaineistojen etäkäyttöjärjestelmä MONA on jo toiminnassa Ruotsin Tilastokeskuksessa. Se on niin sanottu keskitetty tietovarasto, jossa voidaan yhdistellä Ruotsin Tilastokeskuksen aineistoja ja muita sinne erikseen lähetettyjä aineistoja. Jotta muiden rekisterinpitäjien aineistot saataisiin helpommin käyttöön, on erillisessä CO-DIR-aloitteessa tavoitteena linkittää eri toimijoiden rekisteriaineistot siten, että eri toimijoiden aineistot ovat yhteiskäytössä saumattomasti. Käytännössä tämä linkitys tarkoittaisi, ettei erillistä kansallista tietoaarkistoa tarvittaisi, vaan yhteys rekistereihin otettaisiin vain silloin, kun muodostetaan tarvittavat eettiset luvat saaneen tutkimussuunnitelman mukaista aineistoa. Tutkimusaineisto tallentuisi sitten niin sanotulle ”luottamuspalvelimelle” (federation server), jonka ylläpidosta vastaisi luotettava toimija, jolla on riittävä tietojenkäsittelyllinen kapasiteetti. Tällöin kenenkään toimijan ei koskaan tarvitsisi nähdä aineistoa, jota ei tarvita. Vastaava, vaikkakin tietomäärältään pienempi järjestelmä, on jo käytössä Australiassa, joten suunnitelma ei ole pelkkää science fictionia. Tällainen ratkaisu voisi ratkaista monia tunnistettuja ongelmia sekä mahdollistaa aineistojen saannin ilman lainsäädännön uudistamista.

Magnus Stenbeck kertoi, että myös pohjoismaisella tasolla on käynnissä hanke (Nordic Task Force for Open Access to National Data Repositories), jossa pyritään edistämään pohjoismaisten tietokantojen vapaata käyttöä. Jo hiukan parempi koordinaatio tehostaisi yhteistyötoimintaa huomattavasti, kun kaikkien eri yhteistyöhankkeiden ei tarvitsisi selvittää samoja asioita uudestaan ja uudestaan.

Iltapäiväsessiossa eri maiden edustajat esittelivät rekisteritutkimuksen esimerkkejä. Elisabeth Lynge Kööpenhaminan yliopistosta kertoi kohortti- ja pitkittäistutkimuksen helppoudesta Pohjoismaissa. Esimerkiksi pohjoismaisella syöpärekisteriaineistolla on muodostettu ammattialistuksia mittaava pitkittäinen tutkimusaineisto, jossa on 385 miljoonaa henkilövuotta ja 2,8 miljoonaa syöpää. Mutta rekisteritutkimuksella voidaan myös saada ristiriitaisia tuloksia. Esimerkiksi vuonna 2005 arvioitiin mammografian vähentävän Tanskan rintasyöpäkuolleisuutta neljänneksellä. Viittä vuotta myöhemmin samassa lehdessä julkaistu tutkimus julisti, ettei kuolleisuus olekaan vähentynyt syöpäseulonnan myötä. Ensimmäinen tutkimus perustui rekistereihin, kun taas jälkimmäinen ainoastaan tilastotietoihin. Tarkemmassa analyysissä kuolleisuuden vähenemisen tosin osoitettiin olevan noin 13 prosenttia eli puolet aiemman tutkimuksen tuloksesta. Rekisteriin perustuvalla tutkimuksella voitiin kuitenkin tarkastella eri muuttajien yhteisvaikutuksia, mitä tilastotietojen avulla ei voida tehdä.

Stein Emil Vollset Norjan syntymärekisteristä kertoi kokemuksia maailman vanhimman kansallisen syntymärekisterin tutkimushyödyntämisestä. Norja aloitti syntymärekisterinsä jo vuonna 1967, kaksi vuosikymmentä ennen Suomea. Syntyneitä lapsia on jo 2,5 miljoonaa ja kun heidän vanhempansa lasketaan mukaan, on rekisterissä jo neljä miljoonaa ihmistä – monessa polvessa. Tutkimuksissa on voitu yhdistää esimerkiksi isoäitien, äitien ja heidän lastensa tiedot yhteen ja tarkastella raskaudenaikaisten tautien (erityisesti raskausmyrkytyksen) periytyvyyttä ilman geenitietoja. Toisessa tutkimuksessa puolestaan oli yhdistetty yhdeksän eri kansallista rekisteriä, joiden avulla

seurattiin kaikkia ennenaikaisena syntyneitä lapsia ja tutkittiin heidän sosiaalista selviytymistään. Työttömyyttä ja rikollisuutta lukuun ottamatta ennenaikaisilla oli suhteellisesti muita enemmän sosiaalisia ja terveydellisiä ongelmia, ja sitä selvemmin mitä varhaisemmin raskaus päättyi, vaikka suurin osa lapsista menestyikin ongelmitta. Syntymärekisterin aineistolla oli tutkittu myös sisaruksia, joista toinen oli syntynyt hedelmöityshoidoista ja toinen spontaanista raskaudesta. Rekisteristä pystyttiin etsimään hedelmöityshoidoista syntyneille optimaaliset verrokkit eli samojen vanhempien spontaanisti alkunsa saaneet lapset. Kaikki esitellyt tutkimukset oli julkaistu huippulehdissä, kuten British Medical Journal, New England Journal of Medicine ja Lancet. Siinä on hyvä tavoite suomalaisillekin rekisteritutkijoille, sillä miksi me haluaisimme tyytyä vähempään? Norjalaiset ovat anomassa suursausta rekistereihin – kymmeniä miljoonia öljykruunuja. Tähän verrattuna on suomalaisten panostus rekistereihin hävettävä: Norjan syntymärekisterissä on töissä parikymmentä henkeä, Suomessa alle 1,5 henkilötyövuotta.

Vollset kertasi myös Norjan rekisteri-infrastruktuurin ja merkittävät lainsäädännölliset uudistukset. Verrattaessa muiden Pohjoismaiden vastaaviin rekistereihin vain sairaaloiden poistoilmoitusrekisteri oli auttamattoman ”uusi” Norjassa; linkkauksen mahdollistavaa tietoa on saatavilla vasta vuodesta 2008 lähtien. Viidentoista kansallisen rekisterin lisäksi Norjassa on myös 19 laaturekisteriä – toki vähemmän kuin Ruotsissa, mutta missä ihmeessä ovat suomalaiset laaturekisterit? Lainsäädännön uudistuksille olisi tosiaan korkea aika myös Suomessa!

Keskustelussa tuli ilmi, että Tanskassa erillisten rekistereiden määrä on viime aikoina vähentynyt,

kun samaa ”tietoputkea” pitkin kootaan useamman rekisterin tietoja. Norjassa asia ei mene sillä tavalla, vaan kaikki rekisterit ovat erillisiä omine tiedonkeruineen. Suomessa AvoHilmon yhteydessä kerätään myös rokotusrekisterin tiedot ja samanlaisen alle vuorokauden viiveellä tiedot rekisteriin tuottavan tiedonkeruun toivoisi yleistyvän myös muihin terveysrekistereihin. Suomessa olisi erillisten laaturekistereiden sijaan syytä integroida tarvittavat rakenteelliset tiedonkeruut potilastietojärjestelmiin, jossa ne palvelisivat suoraan kliinistä työtä ja josta ne voitaisiin automaattisesti siirtää myös hoitoilmoitusrekisteriin ”lisälehtinä”. Näin tehdään jo vaativan sydänpotilaan ja psykiatrian osalta. Jos tietojärjestelmien uudistus on liian kankeaa, niin tietojen keruu vaikkapa nettipohjaisella lomakkeella THL:n palvelimella ajaisi pitkälti saman asian. Keskeistä olisi joka tapauksessa saada yhdessä päätettyä, minkälainen yhteinen rakenteellinen tietosisältö kullekin lisälehdelle tarvitaan.

Seuraavaksi Martin Hällsten Tukholman yliopistosta kertoi kokemuksistaan ruotsalaisten sosiaalirekistereiden hyödyntämisestä. Hänen mukaansa sosiaalitieteellisessä tutkimuksessa on käytetty kauemmin rekisteritietoja kuin erillisiä kyselylomakkeita ja yhdeksi varhaisimmista rekisterilinkkauksista sisältäväksi ruotsalaiseksi tutkimukseksi hän mainitsi Gunnar Boaltin vuonna 1947 julkaiseman tutkimuksen, jossa koulutietoja oli yhdistetty muihin rekisteritietoihin. Ruotsissa on jo kauan kerätty ja säilytetty otostietoja kunkin kuun 5., 15. ja 25. päivänä syntyneistä ja 15. päivänä syntyneistä on muodostettu oma rekisterinsä vuoden 1950 väestölaskennasta lähtien.

Kuuluuista on vuonna 1964 aloitettu Metropolitan-tutkimus, joka keräsi vuonna 1953 syntyneiden tukholmalaisien tietoja laajasti.

Hankkeen tiedonkeruu päättyi vuonna 1986 suureen skandaaliin, sillä tietosuojaviranomaiset pitivät hankkeen saamia tutkimuslupia riittämättöminä seurantatietojen jatkuvaan keräämiseen. Aineistosta oli poistettava henkilötunnukset ja muut yksittäisen henkilön tunnistamisen mahdollistavat tiedot. Metropolitan-myrsky riehui kauan Ruotsissa, ja se vaikuttaa edelleenkin tietosuojalainsäädäntöön.

Ruotsin Tilastokeskus on muodostanut valmiiksi suuria tietokantoja (Linda, Louise/Lisa, Multigenerationsregister), joihin on kerätty valmiiksi valtavasti rekisteritietoja, joita voidaan analysoida itsenäisesti tai yhdistämällä muita rekisteritietoja. Tällaiset eivät ole olleet mahdollisia Suomessa, sillä pysyvien rekisterien perustaminen vaatii yleensä lainmuutoksen tai tietosuojalautakunnan luvan. Tätä ei yleensä edes yritetä, koska luvan saamista pidetään epätodennäköisenä.

Hällsten kertoi myös varsin yksityiskohtaisesti monista ruotsalaisista rekistereistä ja valotti ongelmia, joita liittyy esimerkiksi kotitalouksien ja koulutustason rekisteriperusteiseen määrittelyyn Ruotsissa. Yhdeksi erityisen kiinnostavaksi ja ainutlaatuisiksi rekisteriksi hän mainitsi vuodesta 1968 kerätyn kaikkia miehiä koskevan kutsunta-aineiston, jossa on tietoja esimerkiksi pituudesta, painosta, älykkyystesteistä ja myös psykologin arvioimista ei-kognitiivisista kyvyistä, kuten sosiaalisesta kypsyydestä, psykologisesta energiasta ja tunne-elämän vakaudesta. Suomen armeijatiedot ovat olleet varsin vähäisessä käytössä.

Lopuksi Mika Gissler THL:stä kertoi pohjoismaisen yhteistyön mahdollisuuksista. Samankaltaisen demokraattisen yhteiskuntarakenteen, sosiaaliturvan ja yleisen terveydenhuoltopalveluiden saatavuuden lisäksi Pohjoismaissa on kerätty aineistoja jo pitkään. Ensimmäiset

rekisterit ovat 1940- ja 1950-luvulta ja helpot linkkaukset mahdollistavat yksilötunnisteet ovat käytettävissä 1960-luvulta lähtien. Lisäksi rekistereitä saa käyttää tutkimukseen, eikä siihen tarvita henkilön antamaa pelkkiä rekisteritietoa käytettäessä. Rekisteriaineistojen laatu on myös pääsääntöisesti todettu hyväksi, vaikka onkin vaikea saada tutkimusrahoitusta aineistojen validiteettitarkasteluihin. Monet suomalaiset validiteettitutkimukset ovat 1980- ja 1990-luvulta! Yhteistyötä on tehty monien eri rekisterien tiimoilta ja on selvää, että Pohjoismaiden 25,6 miljoonan asukkaan väestöpohja alkaa riittää jo harvinaisempienkin päätetapahtumien tarkasteluun, vaikka minkään yksittäisen maan osalta asia ei olisi kaan niin.

Gissler kertoi kokemuksia pohjoismaisesta yhteistyöstä eri hankkeissa. Ensimmäisessä oli tarkasteltu koulutustason vaikutusta ennen-aikaisiin syntymiin ja hanke oli osoittautunut todella menestykselliseksi ja tuottanut jo täyden tusinan julkaisuja. Toinen esimerkki koski mielenterveyspotilaiden elinajanodotteita, jotka olivat selvästi matalampia kuin muun väestön vaikka erot olivatkin ajan myötä hiukan liudentuneet. Päihdeongelmien osalta ero ei kuitenkaan ollut kehittynyt odotetulla tavalla, ja tarkoituksena onkin vielä pyrkiä vertailemaan eri maiden päihdepolitiikkaa tämän löydöksen valossa. Kolmannessa hankkeessa seurataan erilaisten hedelmöityshoitojen terveysvaikutuksia ja sadantuhannen lapsen seurannalla päästäänensimmäistä kertaa kiinni myös harvinaisiin päätetapahtumiin. Neljännen hankkeen osalta Gisslerillä oli kuitenkin kerrottavanaan varoituksen sana. Aineistoa kerättiin kuudesta maasta, joista kaikki eivät kylläkään olleet pohjoismaita. Hanke oli edennyt hitaasti, kun tutkimusideat ja asetelmat olivat muuttuneet jat-

kuvasti ja lupien saaminenkin oli venynyt, eikä usean vuoden jälkeen hankkeesta ollut vielä ainuttakaan julkaisua. Lisäksi oli tullut kiistaa siitä, kuka omistaa aineistot, missä niitä säilytetään ja kenellä on oikeudet niiden käyttöön. Tällaiset ongelmat lienevät ratkaistavissa, mutta vievät aikaa. Ehkäpä tiukempi projektinhallinta ja heti alussa sovitut vastuut ja pelisäännöt mahdollisten ongelmien varalta auttaisivat varautumaan tällaisiin yllätyksiin.

Yleisinä ongelmina pohjoismaiselle yhteistyölle Gissler piti lupaprosessien ja aineiston saamisen hitautta. Rekisteriaineistot ovat usein myös varsin kalliita. Myös tietosuojasäännökset voivat vaihdella, eivätkä tutkijat voi aina valita aineistojen yhdistämisen ja aggregaattitason analyysien välillä. Lisäksi etäkäyttäjärjestelmät ovat hyvin eri pisteissä eri maissa. Yhteistyö Tanskan kanssa on ollut helppoa, kunhan toimittaa aineiston sinne ja käyttää sitä etäkäyttäjärjestelmän kautta. Mutta miten toimitaan sitten, kun kaikissa maissa on omat etäkäyttäjärjestelmät? Tämäkin ongelma on ratkaistava seuraavien vuosien aikana.

Gissler totesi rahoitusmahdollisuuksien olevan valitettavan rajoitettua Pohjoismaissa. Muita vaihtoehtoja ovat EU ja ei-eurooppalaiset rahoituslähteet, ehkä myös lääkeyhtiöt ja muut yksityiset toimijat. Kaikki nämä voivat kuitenkin olla hankalia aineistojen ja tulosten omistussuhteiden suhteen, joten kannattaa tarkistaa, mitä onkaan lupautunut tekemään rahoitusta saadakseen. Yhtenä erikoisena ilmiöni Gissler piti niin sanottua ”Catch 22” -ongelmaa: viranomaiset vaativat lääkeyrityksen varmistavan lääkkeen turvallisuuden, mutta rekisterinpitäjät eivät saa luovuttaa aineistoa haluttuun tarkoitukseen, koska tutkimusta ei katsoa tieteelliseksi. Jos lääkeyhtiöt koetaan isoik-

si pahoiksi möröiksi, jotka haluavat aineistoa käyttöönsä vain omaa etuaan ajaakseen, niin riittäisikö kokonaisaineistojen luovuttamisen sijaan perusteltu satunnaisotos? Etäkäyttäjärjestelmä, jolla voidaan seurata mitä analyyseja aineistosta tehdään, auttaisi varmasti myös yhteisen sävelen löytämisessä.

Loppupäättelmänä Gissler totesi, että yhteistyötä tarvitaan isompien aineistojen saamiseksi. Hyviä esimerkkejä yhteistyötutkimuksesta löytyy niin syöpä-, sairaala- kuin syntymärekisterienkin kanssa. Rekisteritutkimus yhteistyönä on siis mahdollista ja sitä pitäisi edistää mahdollisimman paljon. Yhteisestä ja ainakin yhteensovivasta infrastruktuurista olisi varmasti etua. Kokemusta pitäisi saada kasattua niin, ettei jokainen yhteistyöprojekti joudu tekemään samoja virheitä.

Loppukeskustelu jäi pitkän päivän päätteeksi vähäiseksi. Keskustelua kuitenkin käytiin tutkimustulosten vaikutuksesta yhteiskunnalle ja todettiin, ettei julkaistu artikkeli useinkaan ole tarpeeksi, vaan rekisteritutkimuksilla pitäisi olla myös välitön vaikutus toimintatapoihin. Kokouksessa mietittiin myös sitä, miten rekisteritutkimuksen tekemistä voitaisiin opettaa ja miten rekistereitä voitaisiin käyttää tehokkaammin. Yleensä tutkijoiden sisäänajo rekisteritutkimuksen maailmaan on se ongelmallisinta asia, kun ei ole tietoa, mistä aineistoja löytyy ja mitä niillä voi tehdä – ja toisaalta, mitä niillä ei voi tehdä. Lupien hankkimiseen on helppo antaa ohjeita kussakin maassa erikseen, mutta joillekin ehkä hieman yllättävänä asiana rekisteriaineistojen esikäsittelyyn ja analysoimiseen on erittäin vaikea antaa kattavia keit-

tokirjaohjeita. Sekundaarisen aineiston kanssa toimittaessa joudutaan tyypillisesti käyttämään hyvinkin tapauskohtaisia menetelmällisiä kompromisseja. Kannattaa muistaa, ettei rekistereiden todellisia mahdollisuuksia saada hyödynnettyä muiden analyyseja apinoimalla.

REIJO SUND

*VTT, tutkimuspäällikkö,
soveltavan tilastotieteen dosentti
Terveysten ja hyvinvoinnin laitos
Palvelujärjestelmän
tutkimusyksikkö*

MIKA GISSLER

*Professori, Turun Yliopisto ja
Pohjoismainen
kansanterveystieteen
korkeakoulu, Göteborg.
Tutkimusprofessori, Terveysten
ja hyvinvoinnin laitos,
Hyvinvoinnin ja terveyden
tietovarannot -yksikkö*