

# Tietovarantojen elinkaarisuunnittelu on tutkimuksessa vielä uutta

ANTTI TUOMI-NIKULA & PÄIVI HÄMÄLÄINEN

## Johdanto

Viime aikoina on esitetty toistuvasti vaatimuksia julkisen datan avaamisesta yhteiseen käyttöön (esim. keskustelu YP:n numerossa 1/2014). Sektoritutkimuslaitosten hallussa onkin valtava määrä dataa: rekistereitä, tilastoja ja muuta systemaattisesti kerättyä tietoa, jota on viime aikoina avattu yhä enemmän. Näiden lisäksi on kuitenkin vielä suuri määrä tutkimushankkeissa syntyneitä tietoaineistoja, jotka ovat hajallaan eri tutkimusryhmien hallussa. Ne ovat hankkeiden ulkopuolelle pitkälti näkymättömiä ja vailla jatko-työdyntämisen mahdollisuuksia. Sosiaali- ja terveysalan asiantuntijalaitosten yhteenliittymän SOTERKON (2014; VNK 2011a) tietovaranto-ohjelma koki tärkeäksi saada yleiskuvan siitä, millaisista tietovarannoista SOTERKON puitteissa on kysymys ja millainen nykytila on suhteessa lakien, asetusten ja ohjelmien antamiin tavoitteisiin hyvästä tietovarantojen tiedonhallinnasta. Tämän vuoksi selvitettiin, millaisia tietovarantoja SOTERKO-tutkimusohjelmissa käytetään ja miten niiden elinkaarta hallitaan.

SOTERKO-tietovarantotyön käynnistyesä yhteiskunnallinen keskustelu julkisen sektorin tietovarantojen merkityksestä ja niiden avaamisesta oli noussut keskiöön, samoin keskustelu tietojen turvaamisesta ja yhteiskäyttöisyyden lisäämisestä. Valtioneuvoston periaatepäätös julkisen sektorin digitaalisten aineistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä annettiin maaliskuussa 2011 (VNK 2011b). Julkisen sektorin tietoaineistoihin oli runsaasti viittauksia pääministeri Jyrki Kataisen hallitusohjelmassa saman vuoden kesäkuulta (VNK 2011c). Laki julkisen hallinnon tietohallinnon ohjauksesta (634/2011) annettiin samoihin aikoihin. Valtioneuvoston asetus tietoturvallisuudesta valtion-

hallinnossa (VNK 2010/681) oli annettu jo aiemmin. Valtiovarainministeriön avoimen tiedon ohjelma käynnistyi vuonna 2013 (VM 2013).

Kaikkien ohjelmien kiteytetty pääviesti on tietovarantojen hallinnoinnin kehittäminen siten, että tiedot ovat yhtäältä turvassa, myös yksityisen suojan piirissä, ja toisaalta hyödynnettävissä monipuolisesti muihinkin tarkoituksiin kuin siihen, mitä varten ne on ensimmäiseksi kerätty. Myös kansainvälisellä tasolla tietovarantotyöhön on alettu kiinnittää yhä enemmän huomiota. Muun muassa Euroopan unioni (2014) ja OECD (2013) pyrkivät eritoten tietosuojakysymysten selkiyttämiseen ja tiedon yhteiskäyttöisyyden parantamiseen.

## Selvityksen toteutus

Tietovarantoselvitys tehtiin kaksiosaisena. Ensimmäisessä osassa kerättiin perustiedot viiden SOTERKON piiriin kuuluvan laitoksen (THL, TTL, STUK, Fimea ja Kela) tietovarannoista. Toisessa osassa kerättiin tarkemmat tiedot SOTERKON viiden varsinaisen tutkimus- ja kehittämishankkeen (kuvio 1) käyttämistä ja koostamista tietovarannoista keskittyen elinkaaren hallintaan ja sen eri vaiheisiin. Selvityksen ensimmäinen osa tuotti perustiedot yhteensä 60 tietovarannosta, joista 29:stä saatiin toisessa osassa yksityiskohtaiset tiedot. Selvitys suunniteltiin SOTERKON tietovaranto-ohjelmassa, jonka jäsenet ovat osallistuneet myös tämän artikkelin kommentointiin.

Tietovarannon käsite selvityksessä on laaja. Tietovaranto voi olla mikä tahansa informaation kokonaisuus, jota tutkimushankkeissa käytetään, koostetaan ja käsitellään. Se voi olla erityyppistä dataa, kuten rekisteritietoja, tilastoja, haastattelu-

ja kyselyaineistoja tai mittautuloksia. Data voi olla mikrotasoisista tai aggregoitua. Numeerisen datan lisäksi tietovarannot voivat koostua myös fyysisistä näytteistä tai kvalitatiivisista aineistoista. Yhteistä tietovarannoille on se, että ne ovat suoraan hankkeissa tehtyihin tutkimuksiin liittyviä. Näin ollen esimerkiksi hallinnollisia tietovarantoja ei ole otettu mukaan selvitykseen. Tietovarantojen tyyppit on eritelty kuviossa 1.<sup>1</sup>

Vastauksissa korostuivat ennen kaikkea rekisteriaineistot sekä kysely- ja haastatteluaineistot, joita molempia oli kahta vaille kaikissa hankkeissa. Yksi tietovaranto saattaa koostua useammankin tyyppisestä aineistosta, mistä syystä aineistojen kokonaissumma (83) on suurempi kuin hankkeiden tietovarantojen (60; kuvio 1, liitetaulukko 1).

Kuvion 1 vasemmalla puolella on lueteltu SOTERKO-tutkimusohjelmat ja niiden koordinaatiosta vastuussa olevat laitokset, ja niiden alapuolella ovat erikseen vielä Fimea ja Kela. Oikealla taas on lueteltu hankkeiden tietovarantojen aineistotyyppit sen mukaan, mitä kunkin hankkeen tietovarantoihin sisältyy. Myös aineistotyyppien osuudet kaikista aineistoista on laskettu. Luokka ”muut” sisältää vaikeasti määriteltäviä ja rajattavia tietovarantoja, joissa voi olla osia useammasta eri aineistotyyppistä tai joita kyselyn vastaajat eivät olleet tarkemmin määritelleet.

Tietovarantojen aineistotyyppit kuvaavat koordinaattorilaitosten tutkimuksellisia painotuksia. TTL:n hankkeissa, jotka keskittyvät työelämän tutkimukseen, korostuvat ennen kaikkea kyselyt ja haastattelut. THL on huomattava rekisterinpitäjä, ja sen terveyttä ja hyvinvointia koskevissa hankkeissa onkin käytetty kyselyjen ja haastattelujen lisäksi runsaasti rekisteriaineistoja. STUKin altistustutkimuksissa korostuvat rekisteriaineistojen ohella mittaukset. Fimean ja Kelan tietovarannot muodostavat selvityksessä selvästi erillisen kokonaisuuden, koska niihin oli listattu kyselyssä suppeampien tutkimushankkeiden tietovarantojen lisäksi myös suuria perusrekistereitä sekä laajoja kyselytutkimuksia.

Kaikkiaan 60 tietovarannosta oli ilmoitettu raportointivalmiiksi 34. Yhdeksässä tapauksessa aineiston keruu tai saanti oli kesken, ja neljässä tapauksessa sitä vasta suunniteltiin. Kymmenen

tietovarantoa oli luokiteltu aineiston valmistilaltaan luokkaan ”muu” – näistä valtaosa oli jatkuvasti päivittyviä aineistoja.

Aineistojen keruun ja saamisen käytännön haasteita selvitettiin. Oli kyseessä sitten tutkimusryhmän itsensä keräämä tai muilta aineistohaltijoilta vastaanotettu valmis aineisto, selvästi suurimmat haasteet olivat resurssien puute, lupaprosessien hitaus sekä aineiston saannin tai keruun hitaus yleisesti. Lisäksi muutamassa tapauksessa haasteina mainittiin muun muassa aineiston laatu, aineistosta maksettava hinta sekä tiedon hankala löydettävyys.

Itse kerättyjen aineistojen keruun kesto raportoitiin 12 tietovarannosta. Yhdessä tapauksessa keruuseen oli kulunut aikaa 1–3 kuukautta ja neljässä 4–12 kuukautta. Seitsemässä tapauksessa aineistojen keruu oli kestänyt yli vuoden. Muilta aineistotoimittajilta hankittujen, valmiiden aineistojen saannin kesto käyttöluvan pyytämistä aineiston toimittamiseen ilmoitettiin tarkemmin 10 tietovarannosta. Yksi oli saatu alle viikossa ja neljä alle vuodessa. Viiden aineiston saaminen oli kestänyt yli vuoden. Valmiista aineistoista maksettava hinta oli ilmoitettu 7 tietovarannosta. Yhtä vaille kaikki olivat maksaneet 1 500–10 000 euroa. Yksi aineisto oli ollut tätä kalliimpi.

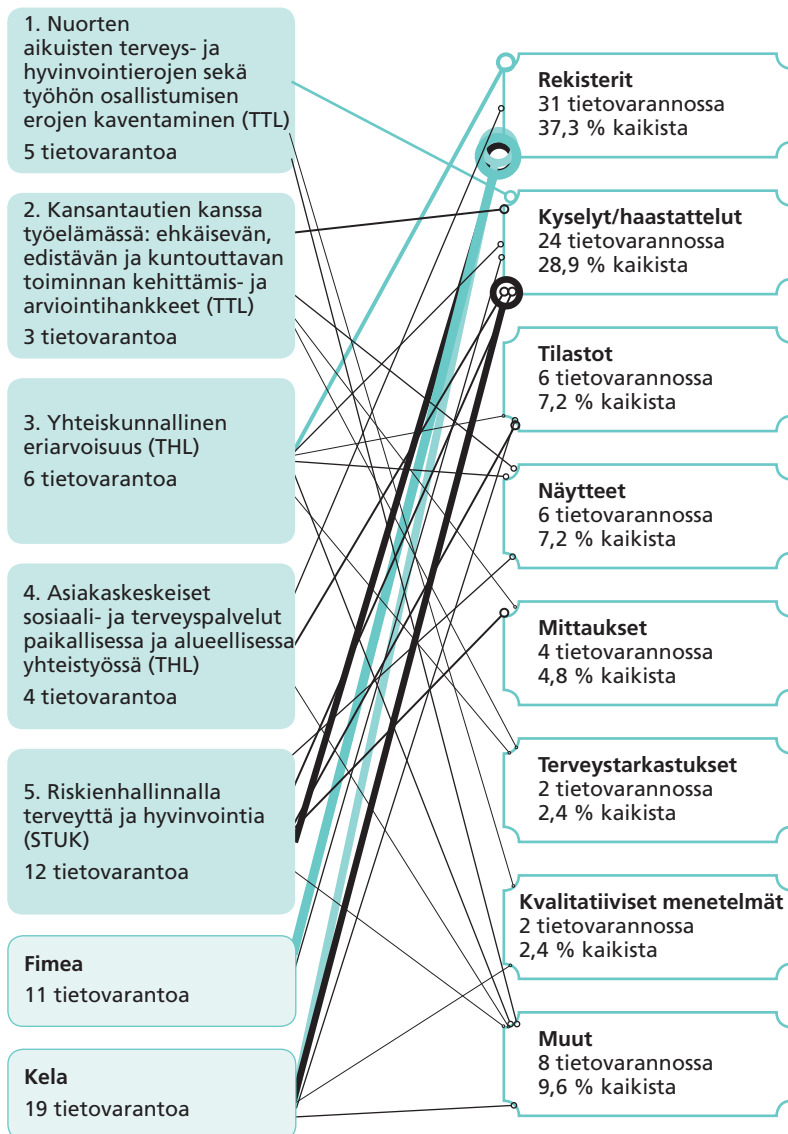
## Tietovarantojen elinkaarisuunnittelu

Kyselyn ydinosan muodostivat tietovarantojen elinkaareen liittyvät kysymykset. Kyselyllä pyrittiin selvittämään, miten aineistoja säilytetään, miten niiden tietoturva huolehditaan sekä miten ja missä aineisto on kuvattu. Tämä osa kyselystä on kohdistettu ainoastaan varsinaisten SOTERKO-tutkimusohjelmien tietovarantoihin, ja se tuotti vastaukset yhteensä 29 tietovarannosta.

Kysyttäessä, kenellä on pääsy aineistoihin, suurin osa vastaajista ilmoitti aineiston ylläpitäjän, käyttöluvan saaneet sekä tutkimusryhmän jäsenet. Muut vaihtoehdot saivat vain muutaman hajamaininnan. Tästä päätellen käyttölupa-asioihin on yleisesti ottaen kiinnitetty melko hyvin huomiota.

Aineistojen sijaintia ja tietoturvaa koskevat kysymykset tuottivat useampia vastauksia, mikä kuvanee aineistojen erilaisia tietoturva-asteita (osa sisältää henkilötunnisteista dataa, osa ei) sekä myös aineistojen olomuotoa (enimmäkseen dataa, mutta myös fyysisiä näytteitä). Valtaosa vastaajista ilmoitti, että aineistoja säilytetään joko ylläpito-

1 Kaikki tietovarannot esitetään liitetaulukossa 1, joka on julkaistu osoitteessa [www.julkari.fi](http://www.julkari.fi). Taulukossa esitellään vastauksissa luetellut tietovarannot tärkeimpine perustietoineen ja aineistotyyppineen. Taulukkoon on merkitty myös tietovarantojen valmistila.



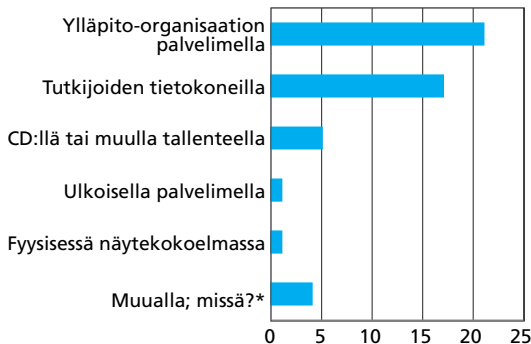
Kuvio 1. SOTERKO-tutkimusohjelmien sekä Fimean ja Kelan tietovarannot ryhmiteltyinä aineistotyyppin mukaan.

organisaation palvelimella tai tutkijoiden tietokoneilla, tai molemmilla samanaikaisesti. (Kuvio 2.)

Myös aineiston tietoturva koskeva kysymys tuotti paljon erilaisia vastauksia. Lähes kaikki vastaajat ilmoittivat useampia tietoturvan varmistamisen keinoja, joista useimmin oli mainittu ”pääsy aineistoon vaatii kirjautumista/salasanaa”, ”varmuuskopiointi määräjoiin” ja ”säilytys lukitussa huoneessa”. Erikseen oli mainittu mm. ”arkistointi”, kulunvalvonta” ja ”alkuperäinen data ja käyttödata tallennettu erikseen”. Tietoturva ja -suoja oli siis käsitetty varsin laajasti.

Aineiston kuvailua selviteti useammalla ky-

symyksellä. Noin kahdesta kolmasosasta tietovarantoja oli tehty rekisteriseloste, ja suurimmasa osassa vastauksia oli mukana selosteen teksti tai linkki siihen. Osa selosteista oli organisaation standardien mukaisia strukturoituja ja osa vapaista kuvauksia. Tämän takia kyselyssä oli toinen kysymys, jossa pyydettiin määrittelemään aineiston kuvailua tarkemmin. Yli puolet ilmoitti, että aineisto oli kuvailtu ”vapaasti”. Lähes puolet taas ilmoitti, että se oli kuvattu organisaation ohjeiden mukaisesti. Huomionarvoista on, että vain yksi vastaaja ilmoitti, että aineiston kuvailu on tehty jonkin tietyn metatietomallin mukaan.



\*Etäkäyttö Tilastokeskuksesta (n=2),  
Lukitussa tietokoneessa (n=1),  
arkistonpitäjän lukitussa kaapissa (n=1)

Kuvio 2. Aineistojen säilyttäminen.

Erikseen kysyttiin vielä, missä aineisto on kuvattu. Kaksi kolmasosaa aineistoista oli kuvattu tutkimussuunnitelmassa. Lisäksi lähes puolet oli kuvattu organisaation keskitetyssä aineistokatalogissa ja noin neljäsosa tutkimuksen tuottamassa artikkelissa. Tämänkin kysymyksen vastauksien perusteella voidaan tulkita, että erilaiset aineistotyytit ovat vaikuttaneet siihen, miten ja missä ne on kuvattu.

Aineistojen elinkaarta koskevan osion viimeinen kysymys koski aineistojen pitkäaikaisäilytystä. Enemmistö vastaajista ilmoitti, että aineisto tullaan säilyttämään, ja useimmiten sähköisenä. Alle kolmasosa ilmoitti, että aineisto joko arkistoidaan, anonymisoidaan tai tuhotaan tutkimuksen päätyttyä. Tämä on huomattavan alhainen lukema, kun otetaan huomioon, että henkilötietolain (523/1999) mukaan ”tutkimusaineisto, jonka henkilötiedot eivät enää ole tarpeen tutkimuksen suorittamiseksi tai sen tulosten asianmukaisuuden varmistamiseksi tulee, joko a) hävittää, b) siirtää arkistoitavaksi tai c) sen tiedot tulee muuttaa sellaiseen muotoon, ettei tiedon kohde ole niistä tunnistettavissa”. Voi tosin olla, että jotkut vastaajista eivät olleet vastauksissaan tarkoittaneet varsinaista tutkimusdataa vaan tutkimuksen tuottamia, julkaistavia aineistoja.

## Tietovarantojen luovuttaminen ja jatkokäyttö

Kyselyn loppuosa keskittyi aineistojen jakamiseen ja jatkokäyttöön tutkimusryhmän ulkopuolella. Osion ensimmäisessä kysymyksessä kysyttiin, mis-

sä tutkimuksen vaiheessa aineisto jaetaan tai jaettaisiin muille. Selvästi suurin osa vastasi ”ei koskaan” ja perusteli useimmiten syyksi, että aineiston käyttö on rajattu vain kyseiseen tutkimukseen. Myös mm. aineistojen sisältämiin henkilötietoihin vedottiin. Vain kolmen tietovarannon aineisto jaettiin eteenpäin heti, kun se on kerätty. (Kuvio 3.)

Kyselyllä haluttiin myös erikseen selvittää, millä tavoin hankkeissa on pyritty tai aiotaan pyrkiä edistämään aineistojen jatkokäyttöä. Aineistojen sisältö ja prosessit oli pääosin dokumentoitu, ja lisäksi aineiston tiedot olivat usein saatavilla verkossa. Vain kahden aineiston tiedot oli syötetty organisaation tai muun tahon aineistokatalogiin. Tämä siitä huolimatta, että aikaisemmassa kysymyksessä ”Missä aineisto on kuvattu?” ilmoitettiin, että peräti 13 aineistoa oli kuvattu organisaation aineistokatalogissa. Tämä viittaa siihen, että moni vastaaja on käsittänyt jatkokäytön mahdollistamisen konkreettisemmaksi asiaksi kuin vain aineiston kuvailuksi.

Teknisluonteisten seikkojen lisäksi mainittiin usein myös sopimus- ja käyttöluupa-asiat. Yksittäisinä aineistojen jatkokäytön edistämisen keinoina mainittiin mm. jatkohyödyntämisen aktiivinen markkinointi, datan saatavuus tutkimusraportin liitteenä sekä aineiston tallennus Yhteiskuntatieteelliseen tietoarkeeseen (FSD).

Lisäksi selvitettiin syitä, jotka estävät aineistojen jatkokäyttöä. Eniten mainintoja saivat rajoitukset käyttöluvuissa ja sopimuksissa. Monen aineiston osalta jatkokäytön esti erityisesti se, et-



\*Tutkimusraportin jälkeen (n=1), ei tietoa (n=3)

\*\*Perustarkastuksen ja raportoinnin jälkeen (n=5), ei tietoa (n=1)

\*\*\*Käyttö rajattu vain tähän tutkimukseen (n=7), Käyttöön liittyy saatekirjeen rajauksen vuoksi rajoituksia tutkimusryhmän ulkopuolella

Kuvio 3. Aineistojen jakaminen: ”Missä tutkimuksen vaiheessa jaatte tai jakaisitte aineistonne muille?”

tä tutkittavilta ei ollut saatu suostumusta jatkokäyttöön. Lisäksi kyselyssä haluttiin vielä selvittää, millä ehdoin jatkokäyttö olisi mahdollista. Yli puolessa aineistoja edellytyksenä oli aineiston muokkaus, kuten anonymisointi ja taulukointi, ennen luovutusta. Lähes puolessa tapauksista mainittiin lisäksi käyttöluupa.

Kysyttäessä, miten ja keneltä haetaan käyttöluupa aineistoon, ylivoimaisessa enemmistössä vastauksista mainittiin aineiston vastuuhenkilö sekä tutkimus- tai johtoryhmä. Lisäksi muutamassa tapauksessa mainittiin ylimalkaisemmin aineiston omistava organisaatio. Vain kolmen tietovarannon jatkokäyttöön liittyi jokin tietty lisenssi. Niin ikään kolmen tietovarannon jatkokäyttöön liittyi maksuja. Kahdeksan tietovarannon jatkokäytön ilmoitettiin olevan maksutonta.

## Johtopäätökset

Selvityksen tärkein tulos on se, että tietovarantoihin ja niiden elinkaarisuunnitteluun liittyvät käsitteet, periaatteet ja säädökset ovat tutkijoille vielä pääosin epäselviä ja jäsentymättömiä. Erityisesti tämä korostuu elinkaaren loppupäässä eli aineistojen pitkäaikaissäilytyksessä, jakamisessa ja jatkokäytön mahdollistamisessa tutkimuksen jälkeen. Näihin asioihin oli kiinnitetty sangen vähän huomiota. Tutkijoiden huomio on selvästi keskittynyt omaan tutkimukseen. Sen jälkeinen aika on tietovarantonäkökulmasta vielä uutta ja vähemmän tunnettua aluetta.

Erityistä huomiota selvityksessä kiinnitti se, että suurimmassa osassa kyselyn vastauksia tietovarantojen jatkokäyttö oli tehty mahdolltomaksi jo tutkimusta suunniteltaessa, kun aineiston käyttö oli rajattu vain kyseiseen tutkimukseen. Moni sopimustekninen asia voisi olla helposti ratkaistavissa. Esimerkiksi suunnittelemalla suostumuslomake huolellisesti voitaisiin ratkoa jatkokäytön pulmia jo ennalta, kun tutkittavilta pyydetäisiin suostumus tutkimuksen aineiston jatkokäytölle. Todennäköisesti myös tietoturva- ja suojaasiat tulisivat näin suunniteltua pidemmällä tähtäimellä jo ennen tietovarannon muodostumista.

Vaikka tietovaranto olisikin tutkimuksen päättymisen jälkeen vapaasti muiden käytettävissä, sen käyttökelpoisuus jää kyseenalaiseksi ilman aineiston systemaattista dokumentointia ja kuvailua. Aineistojen kuvailua koskevien vastausten kirjo ja osittainen ylimalkaisuus antavat vii-

teitä siitä, että metatietotyö ei ole vielä lyönyt itseään läpi tutkijakunnassa. Edessä lienee vielä pitkä tie avoimesti saatavilla oleviin yhtenäisesti kuvailtuihin metatietoihin. Vaikka metatietotyötä tehdään kansallisen tason hankkeissa, kattavia ratkaisuja ei ole vielä syntynyt. Asia tuskin alkaa todenteolla edistyä, elleivät organisaatiot itse ota sitä osaksi tutkimustoiminnan rutiineja. Yhtenä suunnannäyttäjänä on THL:n aineisto- ja järjestelmärekisteri, jota ollaan avaamassa osittain siten, että THL:n ulkopuolisillakin tutkijoilla on mahdollisuus selata THL:n kaikkien tutkimusaineistojen ja tietojärjestelmien metatietoja.

Kyselyn vastausten perusteella aineistot lienevät pääosin turvassa, mutta tietosuoja ja -turvaratkaisut ovat monenkirjavia ja osin riskialttiita. Joitakin aineistoja säilytettiin vain fyysisellä tallennusvälineellä, kuten muistitikulla, mutta suurin osa oli kuitenkin tallessa palvelimilla. Etäkäyttöratkaisut ovat vasta tekemässä tuloaan.

Aineistojen elinkaaren loppu (eli hävittäminen tai pitkäaikaissäilytys) on vaikea ongelma, johon tutkijat eivät ole löytäneet ratkaisua. Tämä on ymmärrettävää, sillä yhteisiä periaatteita ja ratkaisuja ei yksittäisten tutkimusaineistojen osalta vielä ole. Joitakin pitkäaikaissäilytykseen verrattavia palveluja on jo olemassa – esimerkiksi Yhteiskuntatieteellinen tietoarkisto – mutta kansalliset, tietealojen yli ulottuvat ratkaisut puuttuvat vielä.

Tutkimuksen tietoaineistot -hanke toteutti tutkimusdatan hallinnan tilannekartoituksen yliopistoissa (OKM 2014) samoihin aikoihin kun SOTERKO-tietovarantoselvitys tehtiin. Sen perusteella myös korkeakoulumaailmassa on tunnistettu samoja aineistonhallinnan ongelmia kuin tutkimuslaitoksissa. Aineistojen säilyttämiskäytännöt ovat kirjavia, ja tutkijat kaipaavat ohjeistusta metatietotyöhön. Myös tutkijoiden tietynlainen mustasukkaisuus omista tietovarannoistaan mainittiin yhtenä jatkokäyttöä mahdollisesti estävänä tekijänä. SOTERKO-tutkimusohjelmissa eräänä selittävänä tekijänä nousi esiin se, että useammasta aineistosta koostettu, monilähteen tietovaranto on huomattavan työläs hallita elinkaaren suunnittelussa erityisesti tietosuojan ja käyttöoikeuksien osalta. Voi olla, että monessa tapauksessa tietovarannon saamiseen jatkokäytettävään muotoon puuttuu osaamista ja tietoa mm. anonymisoinnista ja käyttöehtojen määrittämisestä.

Tutkimusmaailmassa on perinteisesti julkaistu tutkimusraportti ja kootusti tuloksia, mutta

itse tutkimusdatan julkaiseminen ja avoin jakaminen on vielä uutta. Tutkimusaineistojen osalta kysymys on erityisesti metatiedon, mutta soveltuvien osin myös anonymisoidun tutkimusda-

tan, jakamisesta. Avoin datan prosessoinnin ja avoimen julkaisemisen osaamista olisikin vahvistettava.

## KIRJALLISUUS

- Avoin tiede ja tutkimus -hanke (ATT): hankkeen verkkosivusto, 2014. [avointiede.fi](http://avointiede.fi) (luettu 15.7.2014).
- Euroopan unioni: EU:n tietosuojasetuksen uudistaminen. [ec.europa.eu/justice/data-protection/](http://ec.europa.eu/justice/data-protection/) (luettu 15.7.2014).
- Henkilötietolaki 22.4.1999/523.
- Laki viranomaisten toiminnan julkisuudesta 21.5.1999/621.
- Laki julkisen hallinnon tietohallinnon ohjauksesta 10.6.2011/634.
- OECD: Strengthening Health Information Infrastructure for Health Care Quality Governance: Good Practices, New Opportunities and Data Privacy Protection Challenges, OECD Health Policy Studies, OECD Publishing 2013. [dx.doi.org/10.1787/9789264193505-en](https://doi.org/10.1787/9789264193505-en) (luettu 15.7.2014).
- OKM: Tutkimusdatan hallinnan tilannekartoituksen raportti. Helsinki: Opetus- ja kulttuuriministeriö, 2014. [www.tdata.fi/documents/10180/43697/Tutkimusdatan+hallinnan+tilannekartoituksen+raportti/f1a5b1a1-9c71-45b0-9b20-a91f889054d8](http://www.tdata.fi/documents/10180/43697/Tutkimusdatan+hallinnan+tilannekartoituksen+raportti/f1a5b1a1-9c71-45b0-9b20-a91f889054d8) (luettu 15.7.2014).
- SOTERKO: Sosiaali- ja terveysalan asiantuntijalaitosten yhteenliittymä, hankkeen verkkosivusto, 2014. [www.soterko.fi](http://www.soterko.fi) (luettu 15.7.2014).

- VNK: Valtioneuvoston asetus (1.7.2010/681) tietoturvallisuudesta valtionhallinnossa. Valtioneuvoston kanslia, 2010. [www.finlex.fi/fi/laki/ajantasa/2010/20100681](http://www.finlex.fi/fi/laki/ajantasa/2010/20100681) (luettu 15.7.2014).
- VNK: Valtioneuvoston asetus (27.1.2011/70) sosiaali- ja terveysalan asiantuntijalaitosten yhteenliittymästä. Valtioneuvoston kanslia, 2011a. [valtioneuvosto.fi/ajankohtaista/tiedotteet/tiedote/fi.jsp?oid=319869](http://valtioneuvosto.fi/ajankohtaista/tiedotteet/tiedote/fi.jsp?oid=319869) (luettu 15.7.2014).
- VNK: Valtioneuvoston periaatepäätös julkisen sektorin digitaalisten tietoaisteistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä 3.3.2011. Valtioneuvoston kanslia, 2011b. [valtioneuvosto.fi/toiminta/periaatepaatokset/periaatepaatos/fi.jsp?oid=322887](http://valtioneuvosto.fi/toiminta/periaatepaatokset/periaatepaatos/fi.jsp?oid=322887) (luettu 27.8.2014).
- VNK: Pääministeri Jyrki Kataisen hallituksen ohjelma 22.6.2011. Valtioneuvoston kanslia, 2011c. <http://valtioneuvosto.fi/hallitus/hallitusohjelma/fi.jsp> (luettu 15.7.2014).
- VM: Avoimen tiedon ohjelma. Valtiovarainministeriö, 2013. [http://www.vm.fi/vm/fi/05\\_hankkeet/02381\\_avoin\\_tieto/index.jsp](http://www.vm.fi/vm/fi/05_hankkeet/02381_avoin_tieto/index.jsp) (luettu 15.7.2014).
- Yhteiskuntatieteellinen tietoaarkisto (FSD): verkkosivusto. <http://www.fsd.uta.fi/fi> (luettu 15.7.2014).

## TIIVISTELMÄ

*Antti Tuomi-Nikula & Päivi Hämäläinen: Tietovarantojen elinkaarisuunnittelu on tutkimuksessa vielä uutta*

Julkisten tietovarantojen avaamisesta on viime vuosina keskusteltu runsaasti, ja paljon tietoa onkin jo avattu vapaaseen käyttöön. Hiljattain on myös tutkimuksellisiin tietovarantoihin ja niiden jatkokäytön mahdollisuuksiin alettu kiinnittää enemmän huomiota. Sosiaali- ja terveysalan asiantuntijalaitosten yhteenliittymän SOTERKOn tietovaranto-ohjelma koki tärkeäksi saada yleiskuvan siitä, millaisista tietovarannoista SOTERKOn puitteissa on kysymys sekä millainen nykytila on suhteessa lakien, asetusten ja ohjelmien antamiin tavoitteisiin hyvästä tietovarantojen tiedonhallinnasta.

SOTERKOn tietovaranto-ohjelma toteutti selvityksen, joka kohdistettiin SOTERKOn piiriin kuuluvien tutkimushankkeisiin ja tietoon, jota niissä käytetään, käsitellään ja tuotetaan, kuten rekisteri-, haastattelukysely-, tilasto- ja mittautustietoon sekä näytteisiin ja kvalitatiivisiin aineistoihin. Kysymykset kattoivat tietovarantojen koko elinkaaren tiedonkeruun suunnitte-

lusta tiedon jatkokäyttöön.

Selvitys osoitti, että tutkimustiedon hallinta sen elinkaaren eri vaiheissa on monenkirjavaa: dataa säilytetään niin tutkijoiden tietokoneilla, kassakaapeissa kuin muistitikuillakin, tiedon dokumentointi on ymmärretty monin eri tavoin ja tiedon keräämisen tiellä on esteitä (esim. resurssipula, hankalat lupakäytännöt ja korkeat hinnat). Näyttää siltä, että tietovarantoihin ja niiden elinkaarisuunnitteluun liittyvät käsitteet, periaatteet ja säädökset ovat tutkijoille vielä pääosin epäselviä ja jäsentymättömiä. Erityisesti tämä korostuu elinkaaren loppupäässä, aineistojen pitkäaikais säilytyksessä ja jakamisessa sekä varsinkin jatkokäytön mahdollistamisessa tutkimuksen jälkeen. Myös aineistojen kattava dokumentointi ja metatietotyö ovat vielä pitkälti lapsenkengissään.

Tutkimusmaailmassa on perinteisesti julkaistu tutkimusraportti ja tuloksia kootusti, mutta itse tutkimusdatan julkaiseminen ja avoin jakaminen on vielä uutta. Avoimen datan prosessoinnin ja avoimen julkaisemisen osaamista olisikin vahvistettava.