

# **Kansalliset terveysrekisterit: tiedon hyödynnettävyys tutkimuskäytössä**

Tietojärjestelmätieteiden  
pro gradu -tutkielma

Laatija:  
Aleksanteri Haka

Ohjaaja:  
Prof. Reima Suomi

29.8.2022  
Turku

Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

Pro gradu -tutkielma

**Oppiaine:** Tietojärjestelmätiede

**Tekijä:** Aleksanteri Haka

**Otsikko:** Kansalliset terveystietorekisterit: tiedon hyödynnettävyys tutkimuskäytössä

**Ohjaaja:** Prof. Reima Suomi

**Sivumäärä:** 69 sivua + liite

**Päivämäärä:** 29.8.2022

Tutkielmassa tarkastellaan Suomen kansallisia terveystietorekistereitä, tiedon toisiokäytön arkkitehtuuria ja tiedon hyödynnettävyyttä tutkimuskäytössä. Tutkielmassa erityisen tarkastelun alla ovat THL:n terveystietorekisterit ja niiden sisältämä tieto. Tavoitteena on muodostaa kuva siitä, kuinka hyödynnettävää terveystietorekistereiden tieto on toisiokäyttötarkoituksiin ja luoda pohja mahdolliselle jatkotutkimukselle tutkielmassa koostetun viitekehityksen avulla.

Tutkielman teoriaosuudessa käsitellään kirjallisuuskatsauksen avulla Suomen terveystiedon toisiokäytön arkkitehtuuria, terveystiedon hyödynnettävyyttä ja tiedon toimituksessa vaikuttavia eri viranomaisia ja osapuolia. Terveystiedon toisiokäyttöön on viimeisimpänä vaikuttanut erityisesti vuonna 2019 voimaan astunut laki sosiaali- ja terveystietojen toissijaisesta käytöstä (552/2019). Tiedon toisiokäyttölain myötä pääasiallisesti tietolupaviranomaiseksi asetettiin Findata. Teoriaosuudessa käsitellään myös THL:n terveystietorekistereitä ja niiden käyttötarkoituksia.

Teoriaosuuden lopuksi käydään läpi aiempaa tutkimusta terveystietorekisterien ja avoimien tietolähteiden sisältämän tiedon arvioinnista ja hyödynnettävyydestä. Aiemman tutkimuksen avulla muodostetaan tutkimuksen teoreettinen viitekehys terveystietorekistereiden tiedon hyödynnettävyyden arvioinnille. Tiedon hyödynnettävyyttä arvioitiin kuuden eri tiedon attribuutin avulla: laatu, tarkkuus, täydellisyys, ajantasaisuus, johdonmukaisuus ja saatavuus.

Tutkielman tutkimuskysymyksiin etsitään vastauksia laadullisin menetelmin ja aineiston keruu suoritettiin puolistrukturoituina teemahaastatteluina. Haastattelut analysoitiin teemoittelun avulla. Haastatteluita oli yhteensä kahdeksan kappaletta. Haastateltavat koostuivat lääke- ja terveydenhuollon alan markkinatutkimusyriyksessä työskentelevistä tutkijoista sekä THL:n rekisterivirkahenkilöistä.

Tutkielman tulosten mukaan terveystietorekistereiden tiedon hyödynnettävyyteen toisiokäytössä tällä hetkellä eniten vaikuttavat tietojen saatavuus ja ajantasaisuus. Tietojen saatavuuteen ja ajantasaisuuteen vaikuttavat erityisesti THL:n aineistokäsittelyyn allokoitujen resurssien, tietopyyntöjen käsittelyn prosessit ja metatietoaineistojen kehitys.

**Avainsanat:** terveystietorekisteri, potilastietorekisteri, toisiokäyttö, terveystieto, reaali maailman tieto

## SISÄLLYS

<b>1</b>	<b>Johdanto</b>	<b>7</b>
	1.1 Tausta	7
	1.2 Tutkimuskysymykset	9
	1.3 Tutkimuksen rajaus	10
	1.4 Tutkimuksen rakenne	10
<b>2</b>	<b>Terveystiedon hyödyntäminen</b>	<b>12</b>
	2.1 Potilastietorekisterit	12
	2.2 Reaalimaailman tieto	12
	2.3 Terveystiedon tiedon käyttö	13
	2.3.1 Tiedon käyttö	13
	2.3.2 Tiedon ensisijainen käyttö	14
	2.3.3 Tiedon toissijainen käyttö	15
	2.4 Asiakas- ja potilastietojen toissijaisen käytön arkkitehtuuri	17
	2.5 Laki sosiaali- ja terveystietojen toissijaisesta käytöstä ja Findata	19
	2.6 Terveystiedon ja hyvinvoinnin laitoksen terveystietorekisterit	21
	2.6.1 THL:n rekisterit	21
	2.6.2 Hoitoilmoitusrekisterit (HILMO)	24
	2.6.3 Terveystiedon kansalliset laaturekisterit	24
<b>3</b>	<b>Teoreettinen viitekehys</b>	<b>26</b>
	3.1 Viitekehäksen muodostaminen	26
	3.2 Aiempi kirjallisuus ja viitekehäykset	29
<b>4</b>	<b>Tutkimusasetelma</b>	<b>34</b>
	4.1 Tutkimuksen lähestymistapa	34
	4.2 Tutkimusmenetelmät	34
	4.3 Tiedonkeruumenetelmät	35
	4.3.1 Menetelmien valinta	35
	4.3.2 Haastateltavat	37
	4.3.3 Tutkimusyhteyden esittely	38
	4.3.4 Haastatteluiden sisältö	39
	4.4 Tutkimusaineiston analyysi	40

	5
4.5 Tavoite ja ennakoidut tulokset	42
<b>5 Tutkimustulokset</b>	<b>44</b>
5.1 Laatu	44
5.2 Tarkkuus	46
5.3 Täydellisyys	47
5.4 Johdonmukaisuus	49
5.5 Saatavuus	50
5.6 Ajantasaisuus	56
<b>6 Johtopäätökset ja keskustelu</b>	<b>59</b>
6.1 Aiempi kirjallisuus	59
6.2 Keskeiset johtopäätökset	59
6.3 Kontribuutio, tutkimuksen rajoitukset ja jatkotutkimus	63
<b>Lähteet</b>	<b>65</b>

**KUVIOT**

Kuva 1 Terveysthuollon tiedon käyttö	14
Kuva 2 Tietojen luovutusperusteet	20
Kuva 3 Tiedonkäsittelyn keskeiset toimijat	21

**TAULUKOT**

Taulukko 1 Tiedon hyödynnettävyyden attribuutit	28
Taulukko 2 Tutkimuksen haastattelut	37
Taulukko 3 Teemoitteluanalyysin vaiheet	40

# 1 Johdanto

## 1.1 Tausta

Kansallisten sosiaali- ja terveydenhuollon rekistereiden tiedolla on olennainen rooli terveydenhuollon ja kansan hyvinvoinnin ajankohtaisessa tutkimuksessa ja tilastoinnissa. Terveydenhuollossa asiakastietoa kerätään sen ensisijaiseen käyttötarkoitukseen, eli esimerkiksi potilaan hoitamis- ja tutkimistarkoituksiin. Kyseisen tiedon toisiokäyttö, eli muun muassa tieteellinen tutkimus, tilastointi, kehittämis- ja innovaatiotoiminta, viranomaisvalvonta, opetus ja tietojohdaminen ovat nopeasti yleistyneet ja tehostuneet ensiökäytön (ensisijaisen käytön) rinnalle järjestelmien ja palveluiden kehityksen ansiosta (Lähtenmäki ym. 2021, Huovila & Lähesmaa 2019). Suomessa terveystieto kulkeutuu terveydenhuollon asiakaskohtaisista asiakas- ja potilastietojärjestelmien tietoaltaisiin ja Kanta-arkistoon. Tietoaltaissa olevan tiedon luvittamisesta toisiokäyttöön on vastuussa vuonna 2019 voimaan tulleen toisiolain myötä perustettu lupaviranomainen Findata. Findatan kanssa yhteistyötä tekevä sosiaali- ja terveysministeriön alaisuudessa toimiva Terveyden ja hyvinvoinnin laitos (THL) ylläpitää useaa eri terveystietorekisteriä, jotka ovat toisiokäyttötarkoituksiin luotuja tietoaltaita. Tässä tutkimuksessa tarkastellaan THL:n terveystietorekistereiden tämän hetken tilaa ja niiden hyödynnettävyyttä toissijaisena tietolähteenä tutkimuksessa.

Potilastieto- ja terveystietorekisterit ovat digitalisoituneet Pohjoismaissa viime vuosikymmenten aikana. Vuodesta 2009 asti Pohjoismaissa on valtiollisesti implementoitu sähköisiä kansallisia potilastietoaltaita, metatietojärjestelmiä, biopankkeja ja niiden välillä toimivia järjestelmiä, joiden tavoitteena on alun perin ollut terveydenhuollon järjestelmien tehokkuuden ja tuottavuuden parantaminen (Aarnio ym. 2020, Gilstad ym. 2016, Nøhr ym. 2018). Järjestelmien- ja tietoaltaiden kehityksen tavoitteet ovat muuttuneet vuosien varrella sidosryhmälähtöisemmiksi (Nøhr ym. 2018). Järjestelmät mahdollistavat sen, että terveydenhuollon ammattilaiset pystyvät toimenpidettä tehtäessä kirjata tietoa, joka tallentuu potilastietorekisteriin. Hoitoalasta riippuen toimenpiteen tiedot on kirjattava yhteen tai useampaan tietojärjestelmään. Tietokannat sisältävät reaaliaikaista tietoa potilaiden käyttämistä lääkkeistä, heille tehdyistä hoitotoimenpiteistä ja terveydentilasta. Tämän kaltaista kerättyä tiedoksi kutsutaan reaali maailman tiedoksi (real-world data, RWD). (Rannanheimo & Jauhonen 2018.)

Potilastietokannoista tieto on mahdollista siirtää toisiokäyttöä varten luotuihin terveysrekistereihin. Suomessa tiedot välittyvät muun muassa potilastietoaltaista manuaalisesti ja automaattisesti järjestelmien avulla THL:n eri terveysrekistereihin (Martti & Viitanen 2016). Tietoa hyödyntävät valtiolliset toimijat, lääketieteelliset sidosryhmät ja muut tutkijat. Potilaskohtaisen reaali maailman tiedon saatavuus mahdollistaa tiedon analysoinnin ja siten tarjoaa mahdollisuuden uudenlaisten hoitojen, lääkkeiden ja innovaatioiden kehittämiseen. Lääketeollisuudelle tarkka ja luotettava tutkimustieto on tärkeää lääkekehityksen kasvavien kulujen ja regulaatioympäristön takia. Tietoa tarvitaan myös terveydenhuollon tehokkuuden, menetelmien ja kustannusten tarkkailuun. (Rannanheimo & Jauhonen 2018.)

Pohjoismaat ovat hyvin edistyneitä terveydenhuollon järjestelmien ja potilastietorekistereiden kehittäjinä. Järjestelmät keräävät potilaskohtaisesti ja hoitoalakohtaisesti yksityiskohtaista tietoa terveydenhuollon asiakkaiden hoidoista. Tutkittuja tietoja ovat muun muassa reseptitiedot, lääkeannostukset, vastaanottokäynnit ja potilaiden diagnoosit. (Nøhr ym. 2020.)

Potilas- ja terveystieto on kertynyt omissa muodoissaan eri potilasjärjestelmien tietoaletsiin (Gilstad ym. 2016). Suomessa eri terveydenhuoltoalan toimijat ovat ajan saatossa voineet kehittää järjestelmänsä eri palveluntoimittajien kanssa, minkä vuoksi käytössä on erilaisia järjestelmiä. Tämän lisäksi eri hoitoalat vaativat omanlaistaan tiedonkeruuta, minkä takia eri aloille on syntynyt omia järjestelmiä ja tiedonkeruumenetelmiä. Tietolähteiden moninaisuuden lisäksi on tunnistettu eri tekijöitä, jotka luovat haasteita tiedon käsittelyyn. Näitä ovat tietotarpeet, tiedon kirjaaminen, metatieto ja tiedonkäsittelyn luvat. (Rannanheimo & Jauhonen 2018.)

Terveydatan reaali maailman hyödynnettävyys voi osoittautua haasteelliseksi myös tiedon laadun takia. Järjestelmien ja tietokantojen perustamisen alkuperäiset tarkoitukset ovat jättäneet toissijaiseksi tiedon laadun ja sen laatukriteerien määrittelemisen tietokantojen kehittämisessä (Nøhr ym. 2018). Koska potilas- ja asiakastietojärjestelmien ensisijainen käyttötarkoitus ei ole tiedon laadun seuranta, voi data-analyysin kannalta tärkeitä tietoja olla puutteellisesti kirjattu tai tietoa ei välttämättä ole kirjattu ollenkaan terveydenhuollon asiakaskohtamisissa (Rannanheimo & Jauhonen 2018). Esimerkiksi Timosen ja kollegoiden (2018) tutkimuksen mukaan Suomen e-reseptidatasta 85,9 % sisälsi poikkeavuuksia, joista 7,2 % oli selittämättömiä poikkeavuuksia. Lisäksi 28,4 % tutkituista e-resepteistä ei sisältänyt tarvittavaa olennaista dataa reseptissä määrätystä lääkkeestä (Timonen ym. 2018). Puutteellisen datan lisäksi tietojen metatiedot eivät



välttämättä ole riittävällä tasolla saatavilla, eivätkä tietojärjestelmät yleensä tue riittävän yhdenmukaista tietojen kirjaamista kansallisella tasolla (Rannanheimo & Jauhonen 2018). Metatietojen huono saatavuus hankaloittaa tietojen tilaamisprosessia, koska ilman metatietoja ei voida tietää minkälaista tietoa terveysrekisteristä on mahdollista tilata. Näiden ongelmakohtien lisäksi tiedon saatavuus on varsin hidasta lupa- ja aineistoprosessien takia, minkä takia tieto on usein myöhään hyödynnettävissä (Rannanheimo & Jauhonen 2018).

Koska THL:n terveysrekistereistä saatavaa reaali maailman dataa on mahdollista käyttää ajankohtaisiin lääketieteellisiin innovaatioihin ja terveyttä edistäviin tutkimustarkoituksiin, on tärkeää tutkia, millä tasolla terveysrekisterit ovat ja kuinka hyödynnettävää niistä saatava tieto on tutkimustarkoituksiin. Asiatuntijahaastatteluiden ja kirjallisuuskatsauksen avulla tämä tutkimus pyrkii laajentamaan ymmärrystä terveysrekistereiden nykytilasta ja niiden tiedon hyödynnettävyydestä. Ymmärrys reaali maailman tiedon hyödynnettävyydestä auttaa hahmottamaan, kuinka terveysrekistereitä, sen tiedon rakenteita ja niiden ympärillä toimivia järjestelmiä olisi kehitettävä.

## 1.2 Tutkimuskysymykset

Tutkimuksen tavoitteena on saada yleinen kuva THL:n terveysrekistereiden tilasta ja rekistereiden datan hyödynnettävyydestä toissijaisena tietolähteenä tutkimuskäyttöön.

Tutkimuskysymyksenä on:

- Kuinka hyödynnettävää THL:n terveysrekistereiden reaali maailman tieto on kansallisesti tutkimuskäyttöön?

Vastaus etsitään seuraavien apukysymyksien avulla:

- Mitkä THL:n terveysrekistereiden piirteet vaikuttavat tiedon hyödynnettävyyteen?
- Mitkä ovat terveysrekistereiden kehityksen haasteita?

### 1.3 Tutkimuksen rajaus

Tutkimuksen tarkastelun kohteeksi on rajattu THL:n terveystietorekisterit, sillä niillä on olennainen rooli Suomen kansalaisten terveyden ja terveydenhuollon toisiokäyttöisessä tutkimuksessa. Terveystietorekistereitä tarkastellaan lääketieteen tietorekistereiden piirteiden arviointiin luotuihin viitekehyksiin perustuvan viitekehysten käsitteiden avulla.

Tutkimuksessa haastateltiin viittä rekistereiden eri osa-alueiden asiantuntijaa. Heidän lisäksi haastateltiin myös kolmea tutkija-asiantuntijaa, jotka työksensä tilaavat rekistereistä tietoa ja luovat niistä analyyseja. Rekisteriasiantuntijat valittiin eri osa-alueilta, koska tutkimuksen kannalta on tärkeää saada mahdollisimman laaja ymmärrys rekistereistä ja niiden sisältämistä tiedoista. Haastattelut toteutettiin suomeksi ja etäyhteydellä. Haastatteluiden teemoina olivat teoreettisen viitekehysten perustuvat kysymykset rekistereiden piirteistä ja niiden kehityksen vaiheista. Haastatteluja ei rajattu tiettyihin aiheisiin, vaan haastateltavat saivat vapaasti kertoa heidän tärkeiksi kokemiaan aiheita ja näihin teemoihin saatettiin siirtyä, mikäli ne koettiin tärkeiksi tutkimuksen kannalta.

### 1.4 Tutkimuksen rakenne

Tutkimuksen ensimmäisessä luvussa esitellään tutkimuksen aihe ja siihen liittyvä tausta. Luvusta on mahdollista saada kuva tästä tutkimuksesta ja sen tarkoitusperistä.

Tutkimuksen toisessa luvussa esitellään käytännön nykytilanne tutkimukselle esittelemällä THL:n toiminta terveystietorekistereiden parissa sekä tutkielmassa käsiteltävät rekisterit aiemman kirjallisuuden avulla. Esittelyyn sisältyy tiivistetty tämänhetkinen kuva terveystietorekistereistä, niiden rakenteista ja toiminnoista.

Kolmannessa luvussa muodostetaan teoreettinen viitekehys potilastietorekistereiden tiedon hyödynnettävyyden arvioinnille. Tutkielmassa oleva viitekehys on hybridi aiemman kirjallisuuden muodostamista viitekehysistä tietorekistereiden ja niiden piirteiden arvioimisesta tutkimuskäyttöön. Viitekehyksessä on koostettu terveystietorekistereiden arviointiin soveltuvia attribuutteja, joiden avulla on mahdollista arvioida kuinka hyödynnettävissä rekisterit ja niistä saatava tieto on.

Neljännessä luvussa esitellään tutkimuksen empiirinen osuus. Tutkielman tutkimuksen tiedonkeruumenetelmänä on teemahaastattelu. Haastatteluryhmänä toimivat THL:n terveystietorekistereiden ja niiden tietojen eri asiantuntijat. Heidän lisäksi tutkimuksessa haastateltiin tutkijoita, jotka hyödyntävät THL:ltä saatavaa tietoa

tutkimuksissa. Tutkimuksen haastattelukysymykset rajautuvat tutkielman pääteeman, terveysrekistereiden piirteiden, tietovarantojen hyödynnettävyyden ja tutkimuksen teoreettisen viitekehyksen ympärille. Luvussa tarkastellaan myös haastatteluiden analyysia, sen menetelmiä ja tutkimuksen yleistä tavoitetta.

Tutkielman lopuksi käsitellään tutkimuksen tuottamia johtopäätöksiä ja luodaan yleinen kuva terveysrekistereiden sisältämän reaali maailman tiedon hyödynnettävyydestä tutkimuskäyttöön kansallisella tasolla tutkielman teoreettisen viitekehyksen avulla. Tämän lisäksi pohdinnan alla ovat mahdolliset jatkotutkimuksen aiheet.

## 2 Terveystiedon hyödyntäminen

Tässä luvussa käsitellään terveystiedon toisiokäytön periaatteita ja sen arkkitehtuuria Suomessa. Näiden lisäksi luvussa esitellään THL:n terveystietorekisterit. Terveystiedon toisiokäyttöä varten Suomessa on olemassa laaja-alainen ja aktiivisesti kehityksen alla oleva tietojärjestelmien ja tietoaaltaiden verkko, joka toimittaa eri valtakunnallisille toimijoille tietoa aktiivisesti. Terveystietoa on mahdollista tilata lakimääritteiseltä lupaviranomaiselta lupamenettelyn ja tietopyynnön avulla. Vuonna 2019 voimaan astunut SOTE-tiedon toisiokäyttölaki on virtaviivaistanut tiedon tilaamista ja lupaprosessien käsittelyä (Laki sosiaali- ja terveystietojen toissijaisesta käytöstä 552/2019). Kyseinen laki määritteli lupaviranomaisen aseman tiedon toisiokäytön käsittelyssä, ja lupaviranomainen nykyään toimii yhtenä linkkinä tiedon hakijan ja valtakunnallisten rekisteriviranomaisten välillä. (Lähtenmäki ym. 2021.)

### 2.1 Potilastietorekisterit

Potilastietorekisterit ovat terveydenhuollon tiedon säilytys-, hyödyntämis- ja yhdistelyjärjestelmiä. Niiden olemassaolon tarkoituksena on parantaa potilaiden terveyttä kehittämällä potilaiden hoidon laatua ajankohtaisen informaation avulla ja pääasiallisena tavoitteena on potilaiden hoidon arviointi (De Groot ym. 2017). Potilastietorekisterin määritelmänä voidaan pitää *tietorekisteriä*, joka kerää yhdenmukaisia tietoja tietyn sairauden, tilan tai altistumisen määrittämisen populaation tiettyjen tulosten arvioimiseksi ja joka palvelee yhtä tai useampaa ennalta määritettyä tieteellistä, kliinistä tai poliittista tarkoitusta” (Gliklich ym. 2010). Rekistereiden sisältämät tiedot ovat yksilöllisesti tai aggregoidusti maan lainsäädännön mukaan jaettavissa toissijaisten hyödyntäjien käyttöön. Valtiolliset ja viralliset toimijat ja virastot voivat vaatia rekistereiden tietoja erilaisiin tarkoituksiin (De Groot ym. 2017). Toisiokäyttöön tarkoitettuihin terveystietorekistereihin tieto pääosin päätyy potilastietojärjestelmien rajapintoihin luotujen automaattisten tiedonvälitysjärjestelmien kautta.

### 2.2 Reaalimaailman tieto

Reaalimaailman tieto on pääasiassa lääketieteellisissä merkityksissä käytetty termi. Reaalimaailman tiedolla tarkoitetaan potilaiden terveydenhuollon asiakaskohtaamisissa kerättyä tai potilaiden itsensä tallentamaa ajankohtaista tosielämän tietoa, jota kertyy eri terveydenhuollon tai siihen liittyvien palveluiden rakentamiin järjestelmiin ja

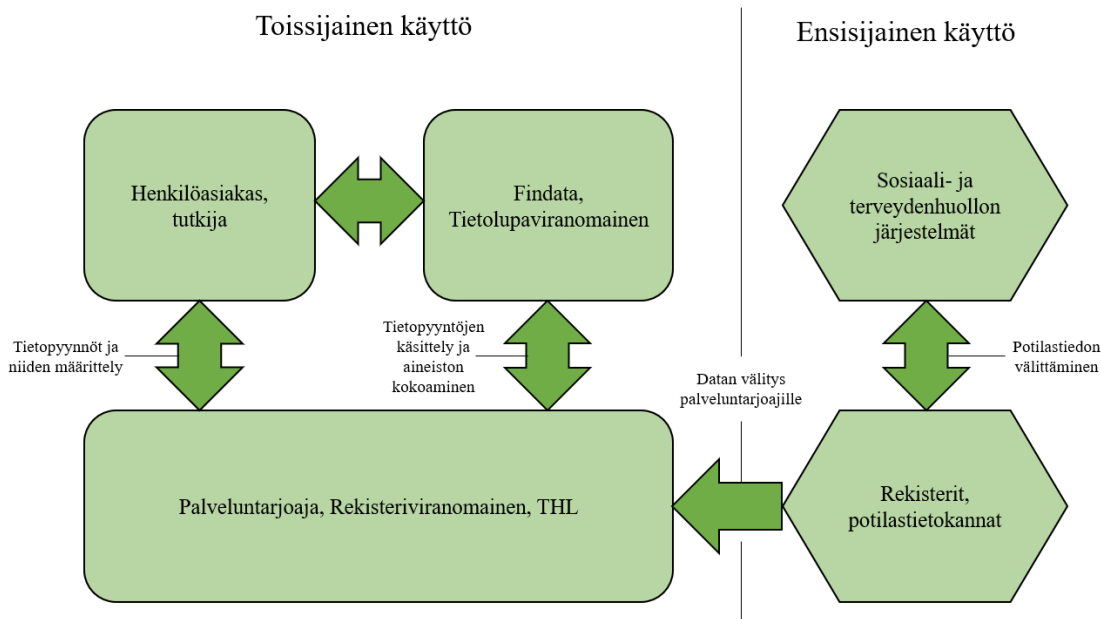
rekistereihin (Garrison ym. 2007, Rannanheimo & Jauhonen 2018). Reaalimaailman tieto on ihmisten terveyteen ja hyvinvointiin liittyvää tietoa, jota hyödynnetään lääketieteellisessä päätöksenteossa ja jota ei siis ole kerätty perinteisten kontrolloitujen kokeiden ja tutkimusten avulla (Garrison ym. 2007). Kliinisistä tutkimuksista ja kokeista saatuun tietoon verrattuna reaalimaailman tiedosta saadut tulokset ovat yleistettävämpiä sekä saatavilla halvemmin ja nopeammin (Garrison ym. 2007, Sørensen ym. 1996).

Reaalimaailman tiedon käyttö lääketieteellisessä tutkimuksessa on yleistynyt maailmanlaajuisesti sen ominaisuuksien takia (De Groot ym. 2017). Suomessa reaalimaailman tietoon perustuvaa tietoa tarvitsevat muun muassa Fimea ja Valvira lääkkeiden myyntilupiin, lääkinnällisten laitteiden valvontaan ja lääkkeiden ja laitteiden turvallisuuden liittyvässä regulaatiossa (Rannanheimo & Jauhonen 2018). Terveysrekistereihin, biopankkeihin ja terveydenhuollon tietoaltaisiin kerättyä terveydenhuollon tietoa voidaan siis kutsua reaalimaailman dataksi. Tietoaltaalla tarkoitetaan tietyn tarkoituksen toiminnan tarpeisiin rakennettua tietoaaineistojen kokoelmaa, joka voi koostua monen eri järjestelmän tuottamista tiedoista (Rannanheimo & Jauhonen 2018). Kehityksen myötä reaalimaailman tiedon rekistereiden määrä on kasvanut eksponentiaalisesti. Yhtenä kasvun syynä on ollut ymmärryksen kasvaminen reaalimaailman tiedon arvokkuudesta lääketieteellisessä tutkimuksessa. (Sørensen ym. 1996.)

## **2.3 Terveydenhuollon tiedon käyttö**

### **2.3.1 Tiedon käyttö**

Sosiaali- ja terveydenhuollon palveluissa syntyvää tietoa käytetään useisiin eri tarkoituksiin. Käyttötarkoitukset ovat jaettu ensisijaiseen ja toissijaiseen käyttöön. Ensisijaisella käytöllä tarkoitetaan tiedon hyödyntämistä sosiaali- ja terveydenhuollon palvelutilanteissa. Toissijaisella käytöllä tarkoitetaan muita, toisiokäyttölaissa määriteltyjä käyttötarkoituksia kuin ensisijaisia käyttötarkoituksia. Nämä käyttötarkoitukset ovat esitelty tarkemmin seuraavissa alaluvuissa.



Kuva 1 Terveydenhuollon tiedon käyttö (mukaillen Huovila & Lähesmaa 2019)

### 2.3.2 Tiedon ensisijainen käyttö

Terveydenhuollon tiedon ensisijaisella käytöllä tai ensiökäytöllä tarkoitetaan sen hyödyntämistä ja käyttämistä siihen tarkoitukseen kuin tieto on alun perin kerätty, kuten terveydenhuollon asiakkaiden terveyden edistämiseen tai potilaan akuutin vaivan tutkimiseen palvelutilanteessa. Ensiökäytössä käytettävät asiakas- ja potilastiedot kerätään tapahtumatasoisesti samoissa terveydenhuollon potilaiden asiakaskohtaamisissa, näytteenotossa ja muissa palvelutilanteissa, joissa niitä hyödynnetään potilaan hoitoon. (Lähtenmäki ym. 2021, Lehto & Neittaanmäki 2018.)

Terveydenhuollon palvelutilanteessa järjestelmien avulla kerätty potilaskohtainen tieto yhdistyy potilaan omiin henkilökohtaisiin historiatietoihin eri tietovarantoihin, joiden sisältämää dataa on kerätty useiden eri järjestelmien avulla. Koska eri terveydenhuollon organisaatioilla on useita eri potilastietojärjestelmä käytössään ja niiden käyttö on muun muassa kunta- ja organisaatiokohtaista, on tietovarantojen ja niiden välillä olevien integraatioiden tehtävänä yhdistää eri järjestelmien kautta kerääntynyttä tietoa omiksi kokonaisuuksikseen. (Lehto & Neittaanmäki 2018, Martti & Viitanen 2016.)

Sosiaali- ja terveydenhuollon organisaatiot, jotka keräävät ensiökäyttöön potilastietoa, ovat Suomessa asiakastietolain mukaan velvoitettu välittämään tietoa Suomessa toimiville eri organisaatioille, jotka ylläpitävät ensiö- ja toisiokäyttötarkoituksiin luotuja terveystietorekistereitä ja asiakastietoarkistoja. Näistä

toimijoista keskeisimpiä ovat sosiaali- ja terveysministeriö (STM), THL, Kela, Väestörekisterikeskus ja Valvira (Lehto & Neittaanmäki 2018). Tiedot välittyvät organisaatioiden tietovarantoihin pääasiassa automaattisesti kirjauksien valmistuttua eri tiedonvälityspalveluiden avulla (Martti & Viitanen 2016).

### 2.3.3 Tiedon toissijainen käyttö

Toissijaisella tiedolla tarkoitetaan alkuperäiseen tarkoitukseensa jo valmiiksi kerättyä tietoa, jota on mahdollista selata ilman erillistä tiedonkeruuta (Hearst ym. 1988). Toissijaisten tietolähteiden käyttö lääketieteellisessä tutkimuksessa on kannattavaa, koska tietolähteissä olevan tiedon myötä tiedonkeruuta ei tarvitse itse suorittaa. Se, että data on jo olemassa, mahdollisesti siis tehostaa tutkimusprosessia ja nopeuttaa tutkimustulosten löytämistä (Sørensen ym. 1996).

Reaalimaailman datalla on olennainen rooli toissijaisena tiedonlähteenä lääketieteellisessä tutkimuksessa. Terveysrekistereiden ja tietoaaltaiden sisältäessä kattavasti tietoa potilaista olisi vastaavanlaisen yleistettävissä olevan datan kerääminen hyvin kallista ja hidasta yksittäistä tutkimusta varten. Terveystietojen tietovarantojen sisältämän datan monipuolisuus mahdollistaa sen käyttämisen moniin eri terveydenhuollon menetelmiä, laatua ja regulaatiota edistäviin tarkoituksiin. (Rannanheimo & Jauhonen 2018.)

Suomessa terveystiedon toissijaisessa käytössä terveydenhuollossa kerääntyneitä potilastietoja on mahdollista hyödyntää useisiin eri tarkoituksiin, joiden edellytykset ovat määritelty 1.5.2019 voimaan astuneessa sosiaali- ja terveystietojen toisiokäyttölaissa (Laki sosiaali- ja terveystietojen toissijaisesta käytöstä 552/2019). Toisiokäytön hyödyntämisen tarkoituksia terveystiedoille on määritelty tarkemmin toisiokäyttölaissa (hyödyntämisen tarkoitusten selitykset ovat tiivistelmiä toisiokäyttölain lakiteksteistä):

- Kehittämisen- ja innovaatiotoiminta (37 §)
  - Kehittämisen- ja innovaatiotoiminnalla terveystiedon toisiokäytössä tarkoitetaan liiketoiminta- ja muun tiedon hyödyntämistä terveystietojen kanssa, kun tarkoituksena on luoda ja kehittää uusia aiempia parempia sosiaali- ja terveydenhuollon tuotteita ja palveluita, jotka edistävät kansanterveyttä tai sosiaaliturvaa, tai suojella yksilöiden terveyttä ja hyvinvointia.

- Tieteellinen tutkimus
- Tilastointi (38 §)
  - Tieteellistä tutkimusta ja tilastointia varten on mahdollista saada lupaviranomaiselta tietolupa yksilöllisten henkilötietojen hyödyntämiseen tutkimuksessa.
- Opetus (39 §)
  - Terveydenhuollon asiakastietoja on mahdollista käyttää niitä käsittelevän henkilöstön ja terveydenhuollon alaa opiskelevien opetukseen, mikäli se on opetuksen kannalta välttämätöntä. Tunnisteellista tietoa voi käyttää vain harvinaistilanteissa. Opetustarkoituksiin edellytyksenä on tietojen käsittelylle myönnetty tietolupa.
- Viranomaisen suunnittelu- ja selvitystehtävä (40 §)
  - Viranomaisen suunnittelu- ja selvitystehtävää varten voidaan käyttää asiakastietoja ja organisaatioiden henkilötietoja, mikäli suunnittelua ja selvitystä ei voida suorittaa ilman kyseisten tietojen hyödyntämistä.
- Tietojohdaminen (41 §)
  - Terveydenhuollon palveluntuottajilla on oikeus käsitellä itse tuottamaansa tunnisteellista tietoa organisaationsa toiminnassa, mikäli se on välttämätöntä palveluntuottajan toiminnan tuottamisessa. Myös kunnilla ja kuntayhtymillä on oikeus käsitellä tunnisteellisesti kuntansa asiakastietoja tietojohdamista tarkoituksissa.
- Viranomaisohjaus ja -valvonta (42 §)
  - Terveydenhuollon tietojen hyödyntämistä viranomaisten ohjauksessa ja valvonnassa tarkoitetaan sitä että, tietoa on mahdollista käyttää terveydenhuollon ammattilaisten ja yksiköiden toiminnan valvontaa varten.
- Tilastoviranomaisen tiedon käsittely (36 §)
  - Aiempien määritelmien lisäksi tilastoviranomainen voi käsitellä rekistereissä olevia yksilöllisiä ja tunnisteellisia henkilötietoja ilman erillistä tutkimuslupa tai tietopyyntökäsittelyä. Tilastoviranomaisella on oikeus saada maksutta, salassapitosäännösten ja muiden rajoitusten rajoittamatta tietoa tiedon välittämistehtäviä varten.



Terveydenhuollon organisaatiot voivat siis itse hyödyntää keräämiään asiakas- ja potilastietojaan toiminnan kehittämiseen sekä päätöksenteon tukena (Martti & Viitanen 2016). Sen lisäksi terveydenhuollon eri rekistereihin kertynyttä tietoa on mahdollista yhdistellä toisiokäyttölain mukaan toisiokäyttöön. Tietoa voidaan hyödyntää muun muassa tilastotietoina, vertailutietokantoina, avoimena datana ja aineistopimintoina tutkimuksiin. Ulkopuolisten toimijoiden on mahdollista hyödyntää tietoa tutkimuslupamenettelyn myötä. Tutkimusluvan tietojen luovuttamiseen voi antaa toisiokäyttölaissa määritelty lupaviranomainen eli Findata, joka tarkistaa toissijaisen tiedon hyödyntämisen tarkoitukset ja edellytykset (Findata 2022a).

Tietolupaa tai tietopyyntöä voi tutkimustarkoituksiin hakea yksityinen tai julkinen toimija, kuten esimerkiksi yritys tai lopputyötä tekevä opiskelija, mutta luvan saaminen edellyttää hakemuksen tekemistä ja tutkimussuunnitelman luovuttamista lupaviranomaiselle hakemusprosessissa (tietoluvan ja tietopyynnön eroista tarkemmin luvussa 2.4). Tutkimustarkoituksesta, otoksen koosta sekä tietolupaa hakevasta toimijasta riippuen lupa ja rekisterivirkahenkilön tietopoimintaan kuluttama työaika maksavat hakijalle lupaviranomaisen määrittelemän summan verran. Terveysrekisteritietoja ei saa hyödyntää tutkimukseen ilman aineiston käyttö lupaa. Toissijaista tietoa on mahdollista hyödyntää kertaluonteisesti (yksittäiset tutkimukset) tai jatkuvaluonteisesti (tilastot, joihin tiedon tulee päivittyä reaaliajassa). (Findata 2022a.)

## **2.4 Asiakas- ja potilastietojen toissijaisen käytön arkkitehtuuri**

Asiakas- ja potilastiedon hyödyntämisessä toisiokäyttötarkoituksiin kertaluonteisesti on neljä prosessivaihetta: 1. tietotarpeen määrittely, 2. tietolupaprosessi, 3. tietoaineiston toimittaminen ja 4. tiedon hyödyntäminen (Huovila & Lähesmaa 2019). Prosessissa toimitettava tieto kulkeutuu eri järjestelmien, toimijoiden ja käsittelijöiden kautta. Prosessiin osallistuu yleensä itse tiedon hyödyntäjä, sote-palvelunantaja, rekisterinpitäjä, tietolupaviranomainen ja tiedon luovuttaja (palveluoperaattori). Prosessilla ja tiedon luovutukseen kehitetyllä arkkitehtuurilla varmistetaan tiedon tietoturvallinen välitys laillisin perustein luvan saaneelle tiedon hyödyntäjälle. Sosiaali- ja terveystietoa on mahdollista myös hyödyntää jatkuvaluonteisesti toissijaisiin tarkoituksiin. Tällöin tiedon pääkäyttötarkoitukset ovat organisaatioiden ja muiden johtamiseen, ohjaukseen ja valvontaan liittyviä (Martti & Viitanen 2016).

Terveydenhuollon potilas- ja asiakastietoa kerätään palvelutilanteissa ja muissa toiminnoissa sosiaali- ja terveydenhuollon organisaatioiden omien potilas- ja asiakastietojärjestelmien sekä potilastiedon erillisjärjestelmien avulla alueellisiin tietoaaltaisiin. Näitä tietoja ovat muun muassa asiakkuustiedot, lääkitystiedot, kuolemansyytiedot ja potilaskohtaiset kliiniset tutkimusaineistot (Huovila & Lähesmaa 2019). Sote-organisaatioilla on omia organisaatio- ja kuntakohtaisia järjestelmiä ja niiden kehitys ei ole maatasolla yhtenäistä. Sote-organisaatioiden järjestelmien, ulkoisten rekistereiden ja tietoaaltaiden välille on luotu rajapintoja ja alueellisia integraatoratkaisuja, joiden avulla potilastietoa välittyy keskeisimpiin valtakunnallisiin toisiotarkoituksiin luotuihin terveysrekistereihin. Näitä tietojärjestelmäratkaisuja- ja kokonaisuuksia kutsutaan yleisesti käsitteellä tietoaalusta (Huovila & Lähesmaa 2019). Potilastietoa toimitetaan automaattisesti ja manuaalisesti integraatioiden ja rajapintojen avulla tilastotietokantoihin, THL:n terveysrekistereihin, Kelan Kanta-palvelujen Potilastiedon arkistoon, Reseptikeskukseen ja Sosiaalihuollon asiakastiedon arkistoon (Martti & Viitanen 2016) Kyseisille rekistereille ja arkistoille on luotu tietoaaineistojen kuvausjärjestelmiä, joiden avulla tietoa hakevan on mahdollista selvittää, minkälaista ja minkä muotoista tietoa on rekistereissä (Huovila & Lähesmaa 2019). Aineistojen kuvausjärjestelmät ovat kehitetty rekisterikohtaisesti.

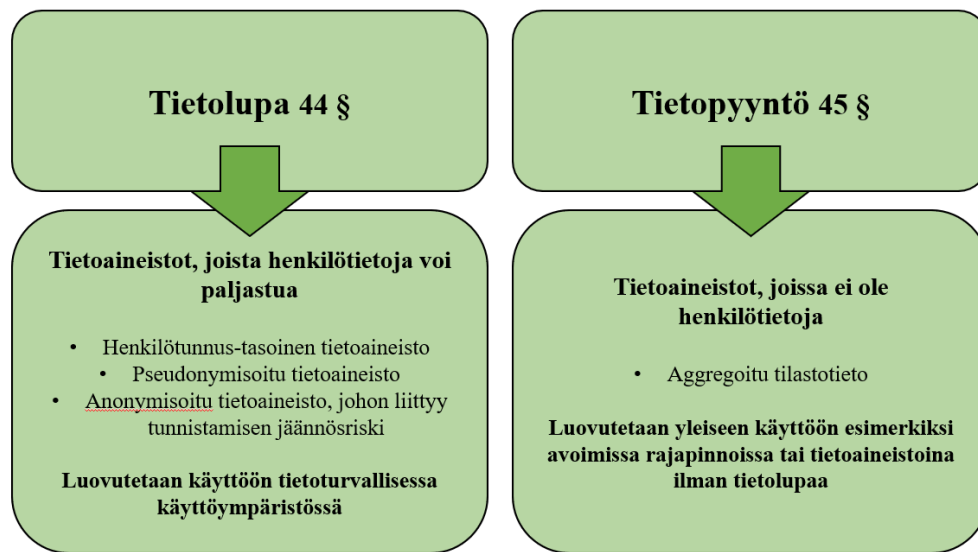
Toisiokäyttöön terveys-, potilas- ja muuta tietoa on mahdollista hakea tietopyynnön tai tietolupahakemuksen avulla lupaviranomaisen kautta. Tietolupahakemus on tehtävä silloin, kun haetaan yksilötasoisista aineistoa. Tällöin tietoluvan mukaista aineistoa on mahdollista käsitellä ja analysoida vain tietoturvallisessa etäkäyttöympäristössä. Tietopyyntö tehdään silloin kun on tarve tilastotasoiselle aineistolle. Tällöin aineisto toimitetaan hakijalle itselleen. Lupaviranomainen käsittelee tietopyynnön tai tietolupahakemuksen ja tekee päätöksen tietojen luovuttamisesta toisiokäyttölain mukaisten ohjeistuksien ja valvontaviranomaisen periaatteiden mukaan. Tiedon luovutusprosessi riippuu hakemuksen tekijän tiedon käyttötarkoituksesta. Päätökset tietolupapäätöksen saaneista hankkeista ovat julkisia ja Findata julkaisee niiden tiedot verkkosivuillaan. (Findata 2022b.)

Toisiolakia sovelletaan pääasiassa rekisteritutkimuksiin. Rekisteritutkimuksella tarkoitetaan tutkimusta, jossa hyödynnetään toiseen tarkoitukseen kerättyjä rekisteritietoja. Toisiolakia ei siis sovelleta Fimealle ilmoitettaviin kliinisiin lääketutkimuksiin. Findata ei myöskään voi myöntää lupia kaikkien toisiolain piirissä olevien rekisterinpitäjien aineistoihin. Tietoluvan hakeminen tai tietopyynnön tekeminen

on mahdollista vain henkilöille, joilla on Suomen väestötietojärjestelmään kirjattu henkilötunnus. (Findata 2022b.)

## **2.5 Laki sosiaali- ja terveystietojen toissijaisesta käytöstä ja Findata**

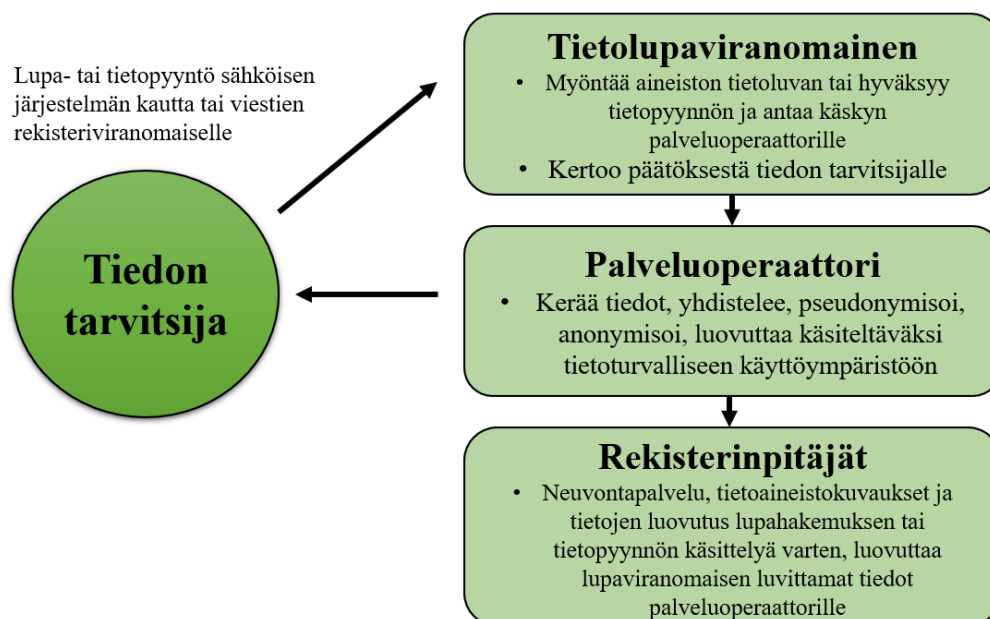
Vuonna 2019 toukokuun 1. päivänä voimaan astunut laki sosiaali- ja terveystietojen toissijaisesta käytöstä luotiin tarpeesta yhtenäistää sosiaali- ja terveydenhuollon potilas- ja muiden henkilötietojen toisiokäyttöä ohjaavaa lainsäädäntöä ja sujuvoittaa tiedon luovuttamiseen liittyvää aiemmin rinnakkain toiminutta lupamenettelyä. Laissa on huomioitu kiihtynyt palveluiden digitalisoituminen, niiden integraatiot ja tietojen käsittelyn tietoturva-vaatimukset. Aiemmin historiassa tietoluvat eri rekistereiden tietoihin on luovutettu pääasiassa rekisteripitäjien toimesta. Toisiokäyttölain myötä luvat henkilötiedollisiin tietoihin myöntää keskitetysti pääasiallisesti tietolupaviranomainen Findata, jonka lupakäsittelyä varten on lain myötä luotu oma tietopyyntöjen hallintajärjestelmä. Aineiston käyttö lupapäätöksen voi tehdä myös rekisterinpitäjä, mutta datan anonymisointiprosessin jälkeen tiedot luovuttaa tiedon pyytäjälle tietolupaviranomainen. Tietolupaviranomainen luovuttaa tiedot lupaa vastaan yhteistyössä rekisterinpitäjän kanssa tietoturvallisessa käyttöympäristössä riippuen siitä, sisältääkö pyydetty aineisto yksilöitävissä olevaa henkilötietoa. Mikäli tieto ei sisällä henkilötietoja, eli on aggregoitua, riittää tiedon saamiseksi normaali tietopyyntö ja pyydetty tieto on vapaasti julkaistavissa yleiseen käyttöön. (Huovila & Lähesmaa 2019, Laki sosiaali- ja terveystietojen toissijaisesta käytöstä 2019.)



Kuva 2 Tietojen luovutusperusteet (mukaillen Huovila & Lähesmaa 2019)

Tietolupahakemus ja tietopyynnöt on mahdollista tehdä osoitteessa <https://lupa.findata.csc.fi/> ja Findatan vaatimat tiedot ja ohjeet löytyvät tietolupaviranomaisen sivustoilta <https://findata.fi/palvelut/tietolupahakemukset>. Yksilöllisten tietojen luovuttamiseen vaaditaan tietoluvan lisäksi tiedonhyödyntämissuunnitelma, millä tarkoitetaan tutkimussuunnitelmaa tai vastaavaa, jossa kerrotaan tietojen käyttötarkoitus, käsittelijät, tietojen säilytyspaikka, säilytyksestä vastuussa olevat toimijat sekä tietojen käsittelyn oikeudellinen peruste (Findata 2022b).

Toisilaki (Laki sosiaali- ja terveystietojen toissijaisesta käytöstä 552/2019) määrittelee tietolupaviranomaisen, Findatan, roolin tietolupapäätöksiä parissa. Findatalla on valtuudet tehdä tietolupapäätöksiä sellaisten organisaatioiden rekistereiden aineistoista, jotka ylläpitävät terveys- ja potilastietorekistereitä. Tietolupaviranomainen vastaa eri aineistojen asianmukaisesta yhdistämisestä ja aineistojen välittämisestä tiedon hakijalle ja luvan saaneelle. Tietolupaviranomaisen alaisuudessa toimii palveluoperaattori, joka käytännössä kerää rekisterinpitäjiltä tiedot ja yhdistelee ne hakijan toiveiden mukaisesti. Palveluoperaattori myös neuvoo tietojen hyödyntäjää tietotarpeiden määrittelyssä ja lopullisesti luo tietoturvallisesta käyttöympäristön tietojen käsittelyyn. (Findata 2022b.)



Kuva 3 Tiedonkäsittelyn keskeiset toimijat (mukaillen Huovila & Lähesmaa 2019)

Tietojen yhdistelyssä tiedot luovuttavan rekisterivirkahenkilön on toisiokäyttölain mukaan käytettävä minimointiperiaatetta: henkilötietoja tulisi käsitellä mahdollisimman vähän ja yksittäisten henkilöiden mahdollinen tunnistaminen tiedoista tulisi estää. Nämä saavuttaakseen käytetään tiedon aggregointia, anonymisointia ja pseudonimysointia. Aggregoidulla tiedolla tarkoitetaan anonymia summatason tietoa, anonymisoinnilla yksittäisen henkilön tunnistamisen mahdollisuuden poistamista ja pseudonimysoinnilla yksittäisen henkilön yksilöivien tietojen muuntamista ns. pseudotunnisteiksi (Lehto & Neittaanmäki 2018).

Toisiokäyttölain mukaan tietolupaviranomaisella on oikeus periä korvausta tarjoamistaan tietojen luovutus-, yhdistely- ja toimituspalveluista. Mikäli tiedot pyydetään rekisteristä, joka ei ole tietolupaviranomaisen hallinnan alla, välittää tietolupaviranomainen tälle ulkoiselle rekisterin ylläpitäjälle vastaavan korvauksen tehdyistä palveluista. (Huovila & Lähesmaa 2019.)

## 2.6 Terveyden ja hyvinvoinnin laitoksen terveysrekisterit

### 2.6.1 THL:n rekisterit

Terveyden ja hyvinvoinnin laitos (THL) on Suomen sosiaali- ja terveysministeriön alaisuudessa toimiva valtiollinen toimija. THL:n tehtävät ja toimiala on määritelty THL:n toimintaa määrittelevässä laissa (Laki Terveyden ja hyvinvoinnin laitoksesta,

31.10.2008/668). Laitoksen tehtävänä on väestön hyvinvointia ja terveyttä seuraamalla kehittää toimenpiteitä hyvinvoinnin edistämiseksi, tutkia sosiaali- ja terveydenhuollon toimintaa, antaa asiantuntijakonsultointia muille alan viranomaisille, harjoittaa tutkimustoimintaa väestön terveyden ja hyvinvoinnin edistämiseksi, ylläpitää sosiaali- ja terveystieteen koodistoja ja luokituksia, osallistua sosiaali- ja terveystieteen kansainväliseen toimintaan ja toimia tilastolain (Tilastolaki 23.4.2008/280) määrittelemänä tilastoviranomaisena. Tilastoviranomaisena THL:llä on oikeus kerätä Suomen kansalaisista tietoa ja ylläpitää tietorekistereitä tilastolain ja toisiokäyttölain määrittelyjen mukaan tilastointi- ja toisiokäyttötarkoituksiin. (Laki Terveyden ja hyvinvoinnin laitoksesta 31.10.2008/668.) Laitoksen toiminta ja järjestelmien kehitys ovat riippuvaisia valtion talousarviorahoituksesta ja hankkeiden myötä saadusta ulkopuolisesta rahoituksesta (THL 2021).

THL tarjoaa erilaisia palveluita valtion, kuntien ja maakuntien päättäjille, sote-alan toimijoille, järjestöille, tutkijoille ja kansalaisille. Palveluiden tarkoituksena on auttaa terveys- ja sosiaalihuollon kehitystä. Toisiokäytön pääasiallisia palveluita ovat koodistopalvelu, aineistokatalogi, raportointi ja tilastointi ja terveystietorekistereiden tiedon toisiokäyttö. (THL 2022)

- Koodistopalvelu
  - Koodistopalvelu on luotu terveydenhuollossa käytettävien termien, tietosisältörakenteiden, lomakerakenteiden, testirakenteiden, luokitusten, rekisteritietojen ja nimikkeistöjen valtiollisen yhteneväisyyden ja laadun ylläpitämiseksi. Järjestelmästä on mahdollista hakea kyseisiä tietorakenteita maksuttomasti. (THL 2020a)
- Aineistokatalogi
  - Aineistokatalogi on THL:n rekistereilleen luoma aineistojen kuvausjärjestelmä, jonka avulla on mahdollista etsiä kuvaukset ja metatiedot THL:n keskeisimmistä aineistoista ja niiden muuttujista. THL:n aineistokatalogista haettaessa tietoa, on niitä mahdollista selata aakkosjärjestyksessä. Aineistoja on tällä hetkellä 459. (THL 2020b)

- Raportointi ja tilastointi
  - THL julkaisee ja tuottaa kaikille avoimia ja maksuttomia tilastotietoja ja raportteja sosiaali- ja terveydenhuollon eri aihealueilta. Tilastotietoa on mahdollista hakea useiden eri verkkoasiointipalveluiden kautta, kuten THL:n omilta sivustoilta tai SOTKANet palvelusta. (sotkanet.fi)
- Terveys- ja sosiaalihuollon rekistereiden tiedon toisiokäyttö
  - THL:n ylläpitämien terveys- ja sosiaalihuollon rekistereiden tietoa on mahdollista saada avoimena datana raportointipalveluiden, ohjelmistorajapintojen (SOTKANet Rest API) ja verkkopalvelujen avulla tai räätälöitynä tietopyynnön avulla Findatan kautta. Osa tilatuista aineistoista toimitetaan edelleen suoraan THL:n rekisterivirkahenkilöiden toimesta tietopyynnön tekijöille. Rekistereistä on mahdollista tilata aggregoitua avointa tietoa tietopyynnöllä tai tarkempaa yksilöllistä tietoa tietoluvan ja tietopyynnön avulla.

THL ylläpitää 15 valtakunnallista terveyden- ja sosiaalihuollon henkilökisteriä (Laki sosiaali- ja terveysalan tutkimus- ja kehittämiskeskuksen tilastotoimesta 409/2001, Laki terveydenhuollon valtakunnallisista henkilökistereistä 566/1989). Rekisterit ovat tietosisällöltään suppeampia kuin ensiökäytön kliniset henkilökisterit, mutta niiden etuna ovat tietojen pitkät aikasarjat. Näihin aineistotietokantoihin on kerätty tietoa suomalaisista vuodesta 1969 alkaen. Tietokannat on luotu datan toisiokäyttöä varten auttamaan suomalaisten terveyden, hyvinvoinnin ja terveydenhuollon palvelujärjestelmien tutkimuksessa. Tietokantojen aineistot ovat vuosien saatossa kertyneet monella tapaa muun muassa kyselyiden, näytteiden ja tutkimusten myötä syntyneellä tiedolla.

Tietoa välitetään niin fyysisesti paperidokumentteina kuin tiedostoina ja sähköisinä lomakkeina. Tiedon välityksessä perusterveydenhuollon tiedon laaduntarkastusta on mahdollista suorittaa tarkastamalla tiedot perusterveydenhuollon järjestelmistä, vertaamalla Koodistopalvelun tietoja CodeMaster-tietojärjestelmäpalvelun tietoihin, THL:n tiedon laadun seurannassa tai mikäli tiedon hyödyntäjä löytää tiedon laadussa huomautettavaa toisiökäytön yhteydessä (Martti & Viitanen 2016). THL:n terveysrekisterit sisältävät henkilötietoja muun muassa potilastiedon ja terveydenhuollon

asiakkaiden palvelukohtaamisissa kerääntyneen tiedon muodossa. Nämä tiedot välittyvät pääosin THL:n rekistereihin reaaliaikaisesti niiden kirjaamisesta, lääkärin tai muun palvelun toimittajan toimesta, automaattisten tiedonvälitysjärjestelmien avulla. AvoHilmoon tiedot perusterveydenhuollosta välittyvät päivittäin poimintaohjelmilla suoraan potilastietojärjestelmistä (Martti & Viitanen 2016) Koska terveysrekistereihin kertyvä reaali maailman data on arvokasta tutkimukselle, toimii se hyvin usein toissijaisena tiedonlähteenä suomalaisten terveydenhuollon ja lääkemarkkinoiden tutkimuksessa. Alla esitellään toisiokäyttöön yleisimmin käytetyt THL:n terveysrekisterit, joita käsitellään tämän tutkielman tutkimusosiossa.

### 2.6.2 Hoitoilmoitusrekisterit (HILMO)

Hoitoilmoitusjärjestelmä on valtakunnallinen sosiaali- ja terveydenhuollon tiedonkeruu- ja raportointijärjestelmä. Järjestelmä koostuu sosiaalihuollon hoitoilmoitusrekisteristä, terveydenhuollon hoitoilmoitusrekisteristä ja perusterveydenhuollon avohoidon hoitoilmoitusrekisteristä. HILMO-rekistereistä toissijaiseen tutkimuskäyttöön saatava avoin data on aggregoitua. Rekisterin sisältämät yksilöidyt datat ovat salassa pidettäviä ja niiden tutkimuskäyttö edellyttää tietolupaa. Tietolupaa on mahdollista hakea sosiaali- ja terveysalan tietolupaviranomaiselta Findatalta. Datan luovuttamisesta ja aggregoinnista vastaavat Terveyden ja hyvinvoinnin laitoksella toimivat rekisterivirkahenkilöt. Tiedon toimittaminen maksaa rekisterivirkahenkilön tekemän tuntityön verran. (THL 2021.)

### 2.6.3 Terveydenhuollon kansalliset laaturekisterit

Laaturekistereillä tarkoitetaan THL:n tietokokonaisuuksia, joihin kerätään tarkkaa tietoa suomalaisten potilaiden saamista hoidoista ja niiden tuloksista. Laaturekistereitä on useita ja ne on luotu yksittäin, jotta tietoja voidaan tallentaa potilaskohtaisesti eri hoidoista ja hoitomenetelmistä. THL:n laaturekisteriprojektiin kuuluu yhdeksän eri laaturekisteriä. Näitä ovat muun muassa diabetesrekisteri, hiv-rekisteri, munuaistautirekisteri ja Sydänrekisteri. Tietoihin sisältyy potilaiden välttämättömät henkilötiedot, hoitotiedot, palvelun suorittaneen yksikön ja terveydenhuollon ammattilaisen tiedot. Näitä rekistereitä pääasiassa käytetään tiettyjen sairauksien ja niiden hoitomenetelmien arviointiin ja seuraamiseen kansallisella tasolla. Laaturekisteritoiminnan tavoitteena on seurata terveydenhuollon palveluiden laatua kansallisesti ja mahdollistaa tutkimus näihin



aiheisiin liittyen. Laaturekisterit ovat osaltaan vielä kehityksen alla, ja niiden kehitys on vuoteen 2022 asti ollut hanke- ja projektikohtaista. Vuodesta 2023 eteenpäin laaturekisterikehitys siirtyy projektimuotoisesta kehityksestä pysyvään toimintaan. (Keinänen & Salonen, 2022.)

### 3 Teorettinen viitekehys

#### 3.1 Viitekehysten muodostaminen

Tiedon laadun arvioinnissa tarkkaillaan, kuinka hyvin tietokannan sisältämä tieto vastaa kysymyksiin toiminnosta, jota varten se on luotu. Laadun arviointia varten on luotu useita ehdotuksia tiedon ominaisuuksista, joilla sen laatua voidaan arvioida. Tiedon laadun arvioinnissa puhutaankin usein tiedon attribuuteista. Tiedon laadun arvioinnin attribuutteja on luotu useita ja tietoa arvioidessa on valittava kyseisen tiedon arviointiin sopivat attribuutit. (Jiang ym. 2008.)

Artikkeleissa, joissa arvioidaan terveystietorekistereitä ja niiden tiedon toissijaista hyödynnettävyyttä, otetaan usein huomioon terveystietorekistereitä ympäröivä arkkitehtuuri ja sen eri piirteiden vaikutus tiedon toissijaiseen käyttöön. Terveystietorekistereiden arkkitehtuurin arvioinnissa voidaan huomioida esimerkiksi järjestelmät, jotka välittävät tietoa terveystietorekistereihin, tietojen koodistojen määrittely ja yhtenäisyys, eri alan toimijat ja tiedon saatavuuden nopeuteen vaikuttavat tekijät (Martti & Viitanen 2016). Kun keskitytään arvioimaan terveystietorekistereiden tiedon hyödynnettävyyttä, on hahmoteltava terveystietorekistereiden tiedolle oleelliset ominaisuuspiirteet, eli attribuutit, joiden avulla terveystiedon hyödynnettävyyden taso on mahdollista ymmärtää ja arvioida. Terveystietorekistereitä ympäröivä arkkitehtuuri ja sen järjestelmät ovat tällöin osatekijöitä, jotka vaikuttavat tiedon hyödynnettävyyden eri arvioitaviin attribuutteihin. Esimerkiksi, mikäli tiedon hyödynnettävyyden kokonaisvaltaisessa arvioinnissa yksi attribuuteista on tiedon saatavuus eli kuinka helppoa esimerkiksi tutkijan on saada tietoa terveystietorekisteristä toissijaiseen käyttöön, voi tietojärjestelmien tiedon välityksen tapa ja nopeus vaikuttaa siihen, kuinka saatavilla tieto on. (Jiang ym. 2008.)

Tässä tutkielmassa terveystietorekistereiden sisältämän tiedon hyödynnettävyyden arvioimisen attribuutit on koostettu kirjallisuuskatsauksen avulla. Attribuutit on koottu terveystietorekistereiden tiedon laadun arviointia käsittelevistä artikkeleista. Viitekehystä on vahvistettu rekisterivirkahenkilöiden haastatteluilla. THL:n terveystietorekistereiden tiedon hyödynnettävyys toisiokäytössä on puolestaan arvioitu tutkimuksessa tehtyjen haastatteluiden ja niiden teemoitteluanalyysin avulla. Analyysiä käsitellään yksityiskohtaisemmin luvussa 4.4.

Kirjallisuuskatsaus kokoaa artikkeleita, joissa on luotu viitekehys lääketieteellisen tiedon arvioimiseen. Kyseisissä artikkeleissa ja niiden tiedon arvioinnin viitekehyksissä

on luotu muuttujia terveystietorekistereiden attribuuteille, jotka ovat oleellisia nimenomaan terveystietorekistereiden tiedon arvioinnissa. Yleistä lääketieteellisen tiedon arvioinnin viitekehyksissä on konkreettinen tiedon ja rekistereiden välttämättömien attribuuttien luetteleminen ja niiden mittaamisen periaatteet esiteltynä tarkemmalla tasolla muuttujien avulla. Terveystietorekistereiden tiedon yhteydessä välttämättömillä attribuuteilla tarkoitetaan terveystietorekistereiden ominaisuuksia, joita ilman rekistereiden sisältämä terveystieto ei olisi hyödynnettävissä tutkimuskäyttöön. Selvittämällä rekisterikohtaisesti jokaisen attribuutin taso niiden arvioinnin avulla on mahdollista hahmottaa, kuinka hyödynnettävää rekisterin sisältämä tieto on toissijaisessa käytössä.

Terveystietorekistereiden attribuutteja koottiin rekistereiden ja niiden tietoa käsittelevistä artikkeleista (Bloland & MacNeil 2019; Máchová ym. 2018; Máchová & Lněnička 2017; Attard ym. 2015; Kučera ym. 2013; Janssen ym. 2012; Sørensen ym. 1996). Attribuuttien arviointiperiaatteet on yhdistelty siten, että niiden avulla voidaan arvioida THL:n terveystietorekistereiden tiedon hyödynnettävyyttä mahdollisimman kattavasti.

Taulukko 1 Tiedon hyödynnettävyyden attribuutit

<b>Tiedon attribuutit</b>	<b>Selitys</b>	<b>Arviointi</b>
Laatu	Tiedon laatua voidaan arvioida tarkkailemalla tiedon todenmukaisuutta ja toistuvuutta. Tiedon laatuun vaikuttavat tiedon lähteet ja se kuinka tieto on mitattu ja tallennettu.	Todenmukaisuutta voidaan mitata tarkastelemalla eroa tiedon ja sen totuudenmukaisen reaali maailman arvon välillä. Toistuvuutta voidaan mitata vertailemalla tietoa eri lähteistä.
Tarkkuus	Onko tieto keskittynyt aiheen ja tutkimuksen kannalta oleelliseen tietoon. Tarkka data ei sisällä ylimääräistä tietoa, joka raskauttaa itse aiheeseen liittyvän tutkimuksen tekoa. Tiedon tarkkuuteen vaikuttaa se, onko tietoon sovittu aiheelle oleelliset ja pakollisesti tallennettavat datapisteet.	Tiedon tarkkuutta on mahdollista mitata tarkastelemalla itse tietoa ja sen sisältöä.
Täydellisyys	Tiedon täydellisyys muuttujana heijastaa sitä, onko kaikki relevantti tieto tallennettu tietokantaan ja käytettävissä. Tiedon aggregointi vaikuttaa tiedon täydellisyyteen.	Täydellisyyttä voidaan mitata katsomalla datapisteitä, jotka ovat tarkoitus merkitä, ja tarkkailemalla niiden raportointia ja tallentamista. Konkreettisesti täydellisyyttä voidaan mitata vertaamalla kokonaispopulaatiota ja tiedon populaation määrää.
Ajantasaisuus	Kuinka ajantasaista tieto on ja kuinka ajantasaisesti se on saatavilla. Liittyy tiedon saatavuuteen ja ennustettavuuteen.	Ajantasaisuutta on mahdollista mitata järjestelmien välisen tiedonkulun nopeudella ja tiedon toimituksen kestolla sen tilausajankohdasta lähtien.
Johdonmukaisuus	Onko tieto ajan myötä säilynyt johdonmukaisena, vai onko sen historiassa tapahtunut selittämättömiä muutoksia sisällössä tai muodossa. Liittyy metatietoihin ja muutoslokeihin. Järjestelmien uudistukset, muutokset ja kehitys vaikuttavat tiedon johdonmukaisuuteen.	Mahdollista mitata tarkkailemalla tietoa eri aikavälien välillä.
Saatavuus	Kuinka helppoa tietoa on saada. Tiedon saamisen kustannus ja luvat. Saatavuuteen liittyvät lupakäsittelyprosessit, ajantasaisuus, tiedon metatiedot ja niiden laatu.	Tiedon saatavuutta on mahdollista mitata tarkastelemalla tiedon tilaamisen järjestelmiä ja sen sääntöjä ja toimivuutta.

### 3.2 Aiempi kirjallisuus ja viitekehukset

Tiedon laadun useista arviointitavoista huolimatta ei ole olemassa yhteisymmärrystä parhaasta tavasta arvioida ja mitata tiedon ja rekistereiden laatua (Bloland & MacNeil 2019). Blolandin ja MacNeilin (2019) tavoitteena ja tarkoituksena oli luoda yhtenäinen tapa tiedon laadun arvioimiseen ja vakiinnuttaa tiettyjä termejä arvioimiselle. Heidän viitekehyksensä keskittyy tarkentamaan ja selittämään jo olemassa olevia potilastietokantoihin ja niiden datan muuttujiin liittyviä käsitteitä lisäämällä datan laadulle, käytettävyydelle ja hyödynnettävyydelle alamuuttujia. Kyseinen viitekehys keskittyy rokotekattavuustietojen laadun arvioimiseen, mutta on sovellettavissa muiden lääketieteellisten rekistereiden arviointiin.

Blolandin ja MacNeilin (2019) viitekehyksessä lääketieteellisen tiedon arvioinnille on määritetty kolme pääpiirrettä: tiedon laatu, datan käytettävyys ja datan hyödynnettävyys. Näille pääpiirteille he ovat määritelleet eri attribuutteja, joilla tietoa on mahdollista arvioida. Tiedon laadun attribuutit ovat tietojen oikeellisuus ja yhtäpitävyys. Tiedon käytettävyyden attribuutteja ovat puolestaan tietojen relevanssi, tehokkuus, täydellisyys, ajantasaisuus, eheys ja johdonmukaisuus. Tiedon hyödynnettävyydelle viitekehys ei määrittele attribuutteja, mutta hyödynnettävyyttä on Blolandin ja MacNeilin (2019) mukaan mahdollista arvioida tarkkailemalla, käytetäänkö tietoa päätöksenteossa. Viitekehysten mukaan tiedon laadun, käytettävyyden ja hyödynnettävyyden arvioinnissa oleellista on aloittaa tutkimus sieltä, missä tieto on alun perin kerätty. Tiedon keräämis- ja merkitsemisvaiheen puutteet johtavat ongelmiin tiedon myöhemmissä käsittely- ja tutkimisvaiheissa, mikä voi johtaa siihen, että tiedon arviointi näiden mittarien avulla on mahdotonta. Päätöksenteossa käytettävän tiedon kohdalla on oleellista, että tieto olisi mahdollisimman arvioitavissa näiden attribuuttien avulla. (Bloland & MacNeil 2019.)

Máchová ym. (2018) ja Attard ym. (2015) mukaan tiedon hyödynnettävyys on tiedon laadun arvioinnissa geneerinen ylätasoinen attribuutti. Heidän mukaansa tiedon hyödynnettävyyteen vaikuttavat se kuinka saatavilla, johdonmukaista ja täydellistä tieto on. Máchová ym. (2018) artikkeli keskittyy avoimien valtiollisten tietoportaalien hyödynnettävyyteen. Se kokoaa kirjallisuuskatsauksen avulla viitekehysten avoimien tietoportaalien tiedon hyödynnettävyydelle ja niiden arvioimiselle. Tiedon hyödyntämisen arviointia varten määrittelevät he pääasiallisina attribuutteina löydettävyyden, saatavuuden ja uudelleenkäytettävyyden. Löydettävyyteen ja

saatavuuteen vaikuttaa oleellisesti metatietojen olemassaolo ja laatu. (Máchová ym. 2018.)

Sørensen ym. (1996) esittelivät viitekehyksen toisiokäytön tietolähteiden arvioimiselle epidemiologisessa tutkimuksessa. Viitekehyksessä attribuutit, joiden nähtiin vaikuttavan toissijaiseen tietoon tutkimuksessa, olivat tiedon kattavuus, tarkkuus ja täydellisyys, tietolähteen koko, aikaväli jolta tietoa on rekisteröity, tiedon saatavuus ja kustannukset, tiedon formaatti ja viimeisenä tiedon yhdistettävyyys muihin tietolähteisiin. Kattavuudella Sørensen ym. (1996) viittaavat kohdepopulaation tietolähteeseen rekisteröityjen edustajien osuuteen. Tutkijoiden mukaan on tärkeää tietää, onko käsiteltävä tietolähde populaatioon pohjautuva vai onko sen muodostamiseen käytetty ennalta määritettyjä valintaprosesseja (Sørensen ym. 1996).

Sørensen ym. (1996) mukaan tarkkuuden varmistamiseksi on tärkeää välttää satunnaisia virheitä sekä systemaattisia virheitä tiedon rekisteröinnissä. Tutkijoiden mukaan tietolähteen laadun arvioinnissa on tärkeää huomioida myös puuttuvan tiedon määrä, koska puuttuvat tiedot vaikuttavat heikentää tiedon laatua. Epätarkat tai puuttuvat tiedot saattavat aiheuttaa vinoumia nollahypoteesia kohtaan. (Sørensen ym. 1996.)

Tietolähteen koon suhteen on Sørensen ym. (1996) mukaan välttämätöntä tietää kuinka monen henkilön ja muuttujan tiedot on rekisteröity lähteeseen. Lisäksi voi olla oleellista tietää eri muuttujien jakauma, sillä se saattaa vaikuttaa tutkimusasetelman muodostamiseen. Esimerkiksi jos tietolähde on hyvin suuri, jopa pieni yhteys tiedoissa voi antaa tilastollisesti merkittäviä tuloksia. Tästä syystä on tärkeää suhteuttaa tiedonlähteen koko kliiniseen relevanssiin pelkän p-arvon arvioimisen sijaan. (Sørensen ym. 1996.)

Rekisteröidyn tiedon aikaväli on merkittävä tekijä esimerkiksi sellaisissa pitkittäistutkimuksissa, joissa tutkitaan altistumisen sekä sairauden kehittymisen yhteyttä. Vaikka tietoa olisikin saatavilla pitkältä aikaväliltä, rekistereiden muotoa ja tietojen koodeja saatetaan päivittää, jolloin eri aikoina kerättyjen tietojen vertailtavuus voi kärsiä. Sørensen ym. (1996) mukaan usein ei ole selkeää kuka omistaa tietolähteen tai kenellä on oikeus käyttää sen sisältöä. Siksi on tärkeää selvittää, kuka on tietolähdettä ylläpitävä rekisteriviranomainen, joka voi hyväksyä tiedon hyödyntämisen tutkimuskäytössä. Lisäksi rekisteriviranomaisen pitäisi ilmoittaa missä määrin tietolähteen sisältö on luottamuksellista, jotta yksilöiden henkilökohtaiset tiedot eivät vuoda valtuuttamattomille osapuolille. Rekisterinpitäjän tulisi ilmoittaa myös tiedon hyödyntämisen kustannukset. Tiedon formaatti voi vaikeuttaa tutkimustyötä, jos rekisteri

tai sen muuttujat ovat sellaisessa muodossa, joita on vaikeaa hyödyntää tutkimuskäytössä. Näitä voivat olla esimerkiksi diagnostiset kategoriat tai populaation ikähaitarit. Tutkimuksen kannalta on usein hyödyllistä yhdistää eri tietolähteitä ja siksi tiedon hyödynnettävyys tutkimuskäytössä riippuu myös siitä, kuinka helposti yksittäinen henkilö on identifioitavissa eri tietolähteissä. (Sørensen ym. 1996.)

Janssen ym. (2012) tutkivat avoimen tiedon hyötyjä ja hyödyntämisen esteitä. Haastattelututkimuksen pohjalta identifioitiin useita avoimen tiedon hyötyjä, joista tärkeimpänä pidettiin poliittisia ja yhteiskunnallisia hyötyjä. Muita avoimen tiedon hyötyjen kategorioita ovat taloudelliset, operationaaliset ja tekniset hyödyt. Useista hyödyistä huolimatta haastatteluissa identifioitiin myös useita avoimen tiedon hyödynnettävyyden esteitä. Janssen ym. (2012) jaottelivat esteet kuuteen kategoriaan: institutionaaliset, tiedon käsittelyn monimutkaisuus, tiedon käyttö ja osallistuminen, lainsäädäntö, tiedon laatu ja tekniset esteet.

Institutionaalisia esteitä ovat esimerkiksi tiedonlähteen ylläpidon motivaatioon ja resursseihin liittyvät asiat. Tiedon käsittelyn monimutkaisuuteen liittyviä esteitä ovat puolestaan kaikki mahdolliset tekijät, jotka saattavat vaikeuttaa ja monimutkaistaa tiedon löytämistä, käsittelyä ja ylläpitoa. Näitä tekijöitä voivat olla esimerkiksi kompleksiset järjestelmät, alkuperäisen muokkaamattoman tiedon puuttuminen, tukipalveluiden puuttuminen tai tiedon metatietojen puuttuminen. Käytön ja osallistumisen esteet liittyvät esimerkiksi tiedon käytön kannustimien puuttumiseen, kykenemättömyyteen perehtyä syvällisesti ja yksityiskohtaisesti tietolähteeseen ajanpuutteen vuoksi ja tilastotieteellisen tietämyksen puuttumiseen. Lainsäädäntöön liittyvät esteet voivat olla muun muassa yksityisyyteen liittyvät loukkaukset, turvallisuus ja tiedon käytön luvat. (Janssen ym. 2012.)

Janssen ym. (2012) mukaan myös tiedon laatu aiheuttaa sen hyödyntämiselle esteitä, jos esimerkiksi tieto ei ole kyllin tarkkaa, tieto on vaillinaista, tieto on vanhentunutta tai ei muuten pidä paikkansa. Teknisten esteiden kategoriaan lukeutuu puolestaan standardisoitujen käytäntöjen puuttuminen, esimerkiksi metatietojen esittämisen suhteen, ohjelmistojen ja sovellusten pirstaloituminen ja teknisen tuen puute sekä tietojen julkaisemista monimutkaistava lainsäädäntö. Janssen ym. (2012) kuitenkin korostavat, että avointa tietoa ei tulisi kohdella homogeenisena asiana vaan erityyppisillä avoimilla tiedonlähteillä on erilaisia hyötyjä ja esteitä. Janssen ym. (2012) tutkimuksen tuloksia voidaan kuitenkin soveltaa avoimeen tietoon tapauskohtaisesti.

Máchová & Lněnička (2017) esittelivät viitekehyksen avoimien tietoportaalien vertailuanalyysille kansallisella tasolla. Vertailuanalyysissä huomioidaan sekä tietoportaalin, että sen sisältämän tiedon ominaisuuksia. Tietoportaalin ominaisuudet jaotellaan vertailuanalyysissä teknisen, saatavuuden sekä viestinnän ja osallistumisen ulottuvuuksiin. Teknisen ulottuvuuteen sisältyy esimerkiksi informaatio portaalia ylläpitävästä viranomaisesta sekä portaalin käyttöön saatavilla olevat kielet. Saatavuuden kannalta viitekehysessä huomioidaan esimerkiksi tietoaineistojen määrä, ohjelmointirajapinnan olemassaolo kehittäjille sekä temaattiset kategoriat ja avainsanat. Viestinnän ja osallistumisen ulottuvuudessa vertaillaan esimerkiksi mahdollisuuksia antaa palautetta portaalista ja keskustella siihen liittyvistä asioista muiden käyttäjien kanssa sekä tukipalveluiden saatavuutta. (Máchová & Lněnička 2017.)

Tietoaineistojen ominaisuuksien osalta Máchová & Lněnička (2017) vertailuanalyysissä huomioidaan esimerkiksi metatietojen saatavuus, informaatio lisenseistä, maantieteellinen kattavuus ja tietoaineiston koko. Máchová & Lněnička (2017) mukaan avoimet tietoportaalit ovat rajapintoja valtiollisen tiedon ja toisiokäyttäjien välillä. Heidän mukaansa hyvä tietoportaali on ulkoasultaan puhdas, sisältää hakupalkin sekä tiedot portaalia ylläpitävästä viranomaisesta. Lisäksi tietojen pitäisi olla esitetty yksinkertaisesti sekä jaoteltu luokkiin ja tunnisteisiin. Vaikka portaalien tärkein tehtävä on tarjota tietoa, sen tulisi myös olla vastaanottavainen palautteelle ja uusille ideoille. Máchová & Lněnička (2017) mukaan tietoportaalien tehokas kehittäminen edellyttää laadun systemaattista arviointia, jotta niitä ja niiden tuottamaa arvoa voidaan ymmärtää.

Kučera ym. (2013) mukaan valtioiden avoimet tietoaineistot edistävät valtioiden läpinäkyvyyttä ja lisäksi niiden toisiokäytöstä voi olla myös taloudellista hyötyä. Valtiolliset tietoaineistot ovat usein hajautettu useisiin eri lähteisiin, mikä saattaa tehdä oikean tiedon löytämisestä toisiokäyttäjille vaikeaa (Kučera ym. 2013). Kučera ym. (2013) analysoivat datacatalogs.org-sivustolle rekisteröityjä avoimien tietokatalogien laatua. Analyysin pohjalta he esittävät tekniikoita, joiden avulla tietokatalogien laatua voidaan parantaa.

Kučera ym. (2013) mukaan tietokatalogien laatua voidaan arvioida neljässä eri ulottuvuudessa: tarkkuus, täydellisyys, johdonmukaisuus ja ajantasaisuus. Kučera ym. (2013) mukaan kaikkien katalogitietojen täytyy olla tarkkoja. Täydellisyys puolestaan tarkoittaa sitä, että kaiken julkaistun avoimen tiedon pitäisi olla rekisteröity katalogiin, mutta tiedoista ei saa kuitenkaan olla kaksoiskappaleita (Kučera ym. 2013).



Johdonmukaisuuden saavuttamiseksi katalogitiedoissa ei saisi olla ristiriitaisuutta. Lisäksi samaan kategoriaan kuuluvia tietoja pitäisi luokitella käyttäen samoja termejä ja konsepteja. Myös tuntematonta tai puuttuvaa informaatiota tulisi käsitellä samalla tavalla koko katalogissa. Kaiken kaikkiaan katalogitietojen tulisi olla johdonmukaisia. Ajantasaisuus puolestaan vaatii tietojen päivittämistä ajan tasalle. (Kučera ym. 2013.) Kučera ym. (2013) mukaan tietokatalogien laatua voidaan parantaa sekä data- että prosessilähtöisillä tekniikoilla. Näitä tekniikoita ovat muun muassa käyttäjäpalautteen kerääminen ja analysointi, laatuongelmien manuaalinen ja automaattinen havaitseminen sekä korjaaminen, uusien tai päivitettyjen tietojen manuaalinen tai automaattinen validointi ja metatietojen automaattinen määrittely.

## 4 Tutkimusasetelma

### 4.1 Tutkimuksen lähestymistapa

Jotta tutkimus voidaan tehdä, on päätettävä ja määriteltävä millä tapaa tutkimusta lähestytään. Kvalitatiivinen lähestymistapa valikoitui tämän tutkimuksen suuntaukseksi. Kvalitatiivinen suuntautuminen sopii tilanteisiin, joissa on tarpeellista ymmärtää ilmiöön vaikuttavan ympäristön yksityiskohtia ja sen vaikutusta itse ilmiöön. (Ghauri ym. 2020.) Koska tässä tutkimuksessa tutkitaan THL:n terveysrekistereiden tiedon tämän hetken hyödynnettävyyttä toisiokäytössä, on terveysrekistereiden ympärillä olevan arkkitehtuurin ymmärtäminen tutkimuksen kannalta hyvin oleellista. Terveysrekistereiden tiedon ja sen järjestelmien toiminnot ja niiden kehitys on aktiivisesti tapahtuva prosessi. Tästä syystä rekistereiden toimintoihin ja kehitykseen liittyvät kokemukset ja tieto on rekistereiden kehittäjä- ja käyttäjäsidonnaista (Eriksson & Kovalainen 2015). Kvalitatiiviset tutkimus ja tiedonkeruumenetelmät ovat soveltuvia tähän tutkimukseen ja niiden avulla on mahdollista vastata tutkielman tutkimuskysymyksiin.

### 4.2 Tutkimusmenetelmät

Tutkimuksessa tutkimusmenetelminä ovat kirjallisuuskatsaus ja tapaustutkimus. Kirjallisuuskatsauksen myötä koottu teoreettinen tausta ja viitekehys tutkimukselle ovat esitelty luvuissa 2 ja 3. Kirjallisuuskatsaus suoritettiin tietokantahakujen, sen myötä löydettyjen artikkeleiden ja niiden lähteiden avulla.

Tapaustutkimus suoritettiin haastattelemalla THL:n rekisterivirkahenkilöitä ja THL:n terveysrekistereitä hyödyntäviä tutkijoita. Haastattelut toteutettiin puolistrukturoituina teemahaastatteluina. Tapaustutkimuksen pääpiirteisiin kuuluu laaja-alaisen ilmiön yksittäisen alueen syvälinen ja läheinen tutkiminen, jotta aiheesta löydetäisiin sitä selittäviä syy-seuraus -suhteita. Tapaustutkimuksessa luonnollista on tutkia nykyhetken ilmiöitä tutkittavan aiheen omassa ympäristössä. (Zainal 2007.) Tapaustutkimuksessa todisteet ja aineisto kerätään tutkittavan tapauksen lähellä olevien lähteiden (tässä tutkimuksessa THL:n rekisterivirkahenkilöiden ja rekisteriä hyödyntävien tutkijoiden) avulla, joiden myötä itse aihetta on mahdollista ymmärtää paremmin (Runeson & Höst 2009). Tutkimuksen teoreettisen viitekehysten myötä

tehdyn arvioinnin avulla muodostetaan ymmärrettäviä tuloksia aiheesta, joita voitaisiin hyödyntää tulevaisuudessa muihin tapauksiin tai aiheen kehittämiseen. (Zainal 2007.)

Tämän tutkielman teema vaatii sen ympäristön yleisen kuvan ja tilanteen syvällistä ymmärtämistä, jotta tutkielman tutkimusosuus on mahdollista tehdä tuottaen johdonmukaisia johtopäätöksiä. Tutkielman tutkimuksen kohteena olevat THL:n terveysrekisterit ja niiden sisältämä tieto luovatkin oman yksittäisen, mutta monimutkaisen tapauskokonaisuutensa. Rekistereiden tämänhetkistä ympäristöä, tiedon lähteitä, ajankohtaisuutta, relevanttiutta ja muita attribuutteja on tarkkailtava nykyhetken puitteiden mukaan. Terveysrekistereiden tiedon ja sen hyödynnettävyyden arkkitehtuurin nykytilan ymmärtämisen lisäksi tapaustutkimus sopii tutkimusstrategiaksi, koska sen metodien avulla on mahdollista kuvata monimutkaisia ympäristöjä selkeällä ja ymmärrettävällä tavalla rajoittamatta itse tutkittavaa aihetta (Eriksson & Kovalainen 2015). THL:n terveysrekistereiden nykyisten tiedon hyödynnettävyyteen vaikuttavien tekijöiden ympäristö on erittäin monimutkainen kokonaisuus. Tapaustutkimusstrategian avulla on mahdollista luoda selkeä kuva tästä monimutkaisesta ympäristöstä ja tuoda ilmi siihen eniten vaikuttavat ongelmakohdat.

### **4.3 Tiedonkeruumenetelmät**

#### **4.3.1 Menetelmien valinta**

Tutkielman kirjallisuuskatsauksessa luodaan teoreettinen tausta ja viitekehys aiemmasta potilastietorekistereitä koskevasta kirjallisuudesta ja sen datan hyödynnettävyyden attribuuttien käsitteistä. Kirjallisuuskatsaukseen kerätään kirjallisuutta ja tutkimuksia Turun yliopiston Volter-palvelussa olevista tietokannoista (EBSCO, Scopus & ProQuest). Tutkimuksia ja raportteja kerätään myös kansallisilta asiantuntijalaitoksilta (THL ja Kela).

Tutkielman tapaustutkimuksen aineistonkeruu suoritettiin puolistrukturoituina teemahaastatteluina. Tämä aineistonkeruutapa on sopiva tutkimuksen mahdollisten tulosten vertailtavuuden takia, vaikka haastattelu voidaan pitää joustavana ja keskustelu luontaisena (Kallio, Pietilä, Johnson & Kangasniemi 2016). Teemahaastattelu antaa osallistujalle tilaa esittää henkilökohtaisia mielipiteitä ja ajatuksia suunniteltujen kysymysten ulkopuolella (Cohen & Crabtree 2006). Koska haastattelu ei ole kaavamainen, mahdollistaa se syvällisen keskustelun haastateltavan ja haastattelijan välillä avaamalla mahdollisuuden laajemmille tutkimuslöydöille (Galletta 2012).

Tutkielman tekijän puolesta haastattelun keskustelua on hyvä pitää yllä tutkielman aihetta mukailevalla teemalla ja niihin liittyvillä kysymyksillä. Haastattelutyyppejä ei kuitenkaan rajoita kysymysten esittämistä vain ennalta määriteltyihin kysymyksiin, vaan tärkeää on keskustelunomaisesti löytää uusia näkökulmia tutkittavaan aiheeseen ja siten mahdollisesti improvisoida uusia kysymyksiä. Tutkielman tutkimuksessa otollista on, jos haastateltavat pystyvät kertomaan omia kokemuksiaan ja ymmärryksiään tutkimuksen aiheesta vapaasti, kuitenkin siten, että tuloksia on mahdollista järjestelmällisesti verrata keskenään (Gill ym. 2008). Tämä mahdollistaa aiemman kirjallisuuden ja koosteen vertailtavuuden käytännön näkökulmiin. Teemahaastattelut toteutettiin puolistrukturoituina, koska haastatteluissa haluttiin kerätä tietoa THL:n terveysrekistereiden tiedon hyödynnettävyyteen vaikuttavista attribuuteista. Haastatteluiden pohjana toimi tutkielman kirjallisuuskatsausosiossa koostettu tieteellinen viitekehys terveystiedon laadun ja hyödynnettävyyden attribuuteista. Koostettujen attribuuttien avulla luotiin alustavia kysymyksiä ja teemoja haastatteluihin. Kysymysten ja teemojen myötä syntyneiden keskusteluiden avulla haettiin vastauksia tiedon hyödynnettävyyden attribuuttien tilasta koskien THL:n terveysrekistereiden tietoa.

Tutkimuksen aiheen takia on oleellista, että haastateltavat henkilöt ovat terveysrekistereiden tiedon ja sen järjestelmien kehityksen parissa työskenteleviä ammattilaisia. Tämän lisäksi tutkimukselle on oleellista kerätä tietoa terveysrekistereiden tiedon hyödynnettävyyteen vaikuttavista attribuuteista.

Gioian ja kollegojen (2013) mukaan organisaatiossa tapahtuvat ilmiöt ja muutokset ovat parhaiten selitettävissä organisaatiossa toimivien asiantuntijoiden avulla. Nämä muutokset ja ilmiöt tapahtuvat Gioian ym. (2013) mukaan sellaisten ”asiantuntija-agenttien” (eng. knowledgeable agents) toimesta, jotka parhaiten ymmärtävät toimiansa tarkoitukset ja siten pystyvät parhaiten kertomaan niistä. Näiden syiden takia haastateltaviksi valikoitui rekisterivirkahenkilöitä ja terveysrekistereiden tietoa ammatikseen hyödyntäviä tutkijoita, joilla on asiantuntijuutta eri osa-alueilta.

Haastateltujen rekisterivirkahenkilöiden valinta suoritettiin olemalla yhteydessä THL:n data-analytiikkayksikön päällikköön, joka tarjosi yhteystiedot THL:n eri osa-alueiden asiantuntijoille, jotka ovat käytettävissä haastatteluihin. Haastateltavat valittiin asiantuntijoista niin, että rekistereiden tiedon hyödynnettävyydestä saataisiin tietoa mahdollisimman kattavasti. Rekisterivirkahenkilöiden lisäksi tutkimuksessa haastateltiin tutkijoita, jotka tilaavat ja hyödyntävät työssään THL:n terveysrekistereiden tietoa. Nämä tutkijat löydettiin ottamalla yhteyttä suomalaiseen lääke- ja terveydenhuollon

markkinatutkimusyrytykseen. Tutkijat valittiin haastatteluihin heidän kokemuksensa mukaan. Kokemuksen mittarina oli työvuodet sellaisten projektien parissa, joissa käsiteltiin THL:n terveystietorekistereiden tietoa.

Teemahaastattelut nähtiin oleelliseksi toteuttaa niin terveystietorekistereiden kehitykseen vaikuttaville rekisterivirkahenkilöille kuin terveystietorekistereiden asiakkaina toimiville tutkijoille. Koska rekisterivirkahenkilöt työskentelevät samassa organisaatiossa, jonka maksullisena palveluna tietoa terveystietorekistereistä voi tilata, saattaa heillä virkahenkilöinä olla eri näkemyksiä tiedon hyödynnettävyydestä kuin itse palvelun asiakkailla. Tietoa hyödyntävät tutkijat saattavat kokea tiedon hyödynnettävyyden eri attribuutit tärkeämmäksi, kuin itse kehittäjät rekistereiden tiedon hyödynnettävyyden nykytilan takia. Koska rekisterivirkahenkilöt kuitenkin työskentelevät erittäin läheisesti rekistereiden tiedon kanssa, on heillä viimeaikaisinta ja tarkinta tietoa rekistereiden tiedon hyödynnettävyyden ominaisuuksista ja kehitysaskeleista.

#### 4.3.2 Haastateltavat

Teemahaastatteluja toteutettiin yhteensä kahdeksan kappaletta. Tutkielmaa varten haastateltiin viittä THL:n rekisterivirkahenkilöä ja kolmea markkinatutkimusyrytyksessä työskentelevää tutkijaa. Haastateltavien tunnukset on jaettu niin, että THL:n asiantuntijat ja virkahenkilöt ovat merkitty VH-tunnuksella ja haastatellut tutkijat TH-tunnuksella.

Taulukko 2 Tutkimuksen haastattelut

Tunnus	Haastateltava	Kesto	Ajankohta
VH1	THL, Tiedonhallinnan asiantuntija	52 min	3.8.2021
VH2	THL, Tietopoimintojen asiantuntija	48 min	5.8.2021
VH3	THL, Laaturekisterit, Lääkäriasiantuntija	1 h 2 min	20.8.2021
VH4	THL, Aineistokatalogi, metatietoasiantuntija	1 h 17 min	30.8.2021
VH5	THL, Tietopyyntöjen lupa-asiantuntija	55 min	31.8.2021
TH1	Tutkimusyrytytys, Myyjä, Markkinatutkija, Toimitusjohtaja	38 min	2.3.2022
TH2	Tutkimusyrytytys, Markkinatutkija, Data-analyytikko	36 min	3.3.2022
TH3	Tutkimusyrytytys, Markkinatutkija, Data-analyytikko	32 min	3.3.2022

Haastateltavat valittiin sen perusteella, jotta terveystietorekistereiden tiedon hyödynnettävyyden ominaisuuksista saataisiin kehittäjien näkökulmasta mahdollisimman laaja kuva. Haastateltavat olivat työskennelleet THL:n datayksikössä eri

rooleissa useita vuosia. Haastateltava VH1 oli toiminut THL:n tietojärjestelmien teknisessä kehityksessä aiemmassa työnkuvassaan, mutta nykyään johti kehitystiimiä. Haastateltava VH2 on työskennellyt tietopyyntöjen käsittelijänä ja tiedon toimittajana. Hänen tehtävänä on luovuttaa luvalliset tietopyynnöt tietojen käsittelijöille ja on täten THL:n tietopoimintojen asiantuntija. Haastateltava VH3 on THL:n lääkäriasiantuntija, joka on ollut läheisesti mukana kehittämässä THL:n koodistopalvelua lääketieteellisenä asiantuntijana. Haastateltava VH4 on THL:n rekistereiden asiantuntija, joka on koordinoanut THL:n aineistokatalogin kehitystä. Hän on asiantuntija THL:n terveysrekistereiden tiedon metatiedon kehittämässä. Asiantuntija VH5 on vastuuasemassa THL:n puolella tietopyynnöistä, jotka tulevat suoraan THL:lle Findatan sijaan. Haastateltujen virkahenkilöiden avulla terveysrekistereistä ja niiden tiedon tilasta oli mahdollista saada kokonaisvaltainen kuva palveluntarjoajan näkökulmasta.

Haastateltujen rekisterivirkahenkilöiden lisäksi tutkielmaan haastateltiin lääke- ja terveydenhuollon alan markkinatutkimusyhtiön tutkijoita. Markkinatutkimusyhtiön edustajien haastattelut nähtiin oleelliseksi tehdä, koska tutkimuksen kattavuuden takia on olennaista ymmärtää tiedon hyödyntäjien näkökulma. Tutkijahaastateltava 1 on markkinatutkimusyhtiön perustaja, toimitusjohtaja, myyjä ja toimii konsulttina markkinatutkimuksissa. Hänellä on pitkäaikaisin kokemus yrityksessä THL:n tietojen tilaamisesta ja analyysien muodostamisesta. TH2 on yrityksen toiseksi pitkäaikaisin markkinatutkija ja toimii yrityksen päädata-analyytikkona. TH3 on yrityksen pitkäaikaisin työntekijä, markkinatutkija ja data-analyytikko. TH2 ja TH3 ovat työskennelleet useita vuosia THL:n tietopyyntöjen ja tietojen parissa tehden satoja analyyseja eri terveydenhuollon alalla toimiville yrityksille.

#### 4.3.3 Tutkimusyhtiön esittely

Tutkimusyhtiö, jonka alaisena työskenteleviä tutkijoita on haastateltu tätä tutkielmaa varten, on suomalainen vuonna 2005 perustettu yritys. Yritys tuottaa pohjoismaisille lääke- ja terveydenhuoltoalan yrityksille muun muassa markkinatutkimuksia, rekisterianalyyskejä, E-koulutuksia ja konsultointia. Yhtiön pääasiallinen toimialue on Suomi ja osa sen liiketoiminnasta tulee rekisterianalyysistä, joiden tieto on tilattu THL:n terveysrekistereistä. Yhtiö on tehnyt yhteistyötä THL:n kanssa jo usean vuoden ajan. Yhtiöllä on tutkimuksen aikana ollut keskiarvolta 14 työntekijää, joista neljä ovat olleet myyjiä ja loput markkinatutkijoita. Tutkimukseen valittiin yhtiön asiantuntijoita,

joilla on pisin kokemus THL:n terveystietorekistereiden tiedon tilaamisen ja analyysien parissa.

#### 4.3.4 Haastatteluiden sisältö

Kaikki haastattelut toteutettiin ja nauhoitettiin haastateltavien luvalla videopuheluna Microsoft Teams -palvelun välityksellä. Kaikille haastateltaville lähetettiin etukäteen tietoa haastattelun aiheesta, tyypistä ja tutkielman viitekehyksen attribuutit. Tämä mahdollisti sen, että haastateltavat pystyivät miettimään etukäteen, mitä olisi oleellista ottaa esille itse haastattelussa. Haastateltaville ilmoitettiin etukäteen myös siitä, että heidän identiteettiään ei julkaista tutkimuksessa. Tällä haluttiin myötävaikuttaa siihen, että haastateltavat voivat keskustella haluamistaan asioista niin vapaasti ja syvällisesti kuin haluavat.

Haastattelut toteutettiin perinteisten puolistrukturoitujen teemahaastattelujen tapaan. Haastatteluiden alussa haastateltavan annettiin esitellä itsensä sekä oma työkuvansa ja käytiin läpi haastattelun tavoitteet. Jokaiselta haastateltavalta kysyttiin haastatteluiden alkupuolella heidän näkemyksensä terveystiedon hyödynnettävyyden tärkeimmistä attribuuteista ja siitä, ovatko tutkimuksen viitekehyksen attribuutit tutkimukseen soveltuvia. Haastatteluissa haastattelija, eli tutkimuksen toteuttaja, kysyi satunnaisessa järjestyksessä ennalta harkittuja kysymyksiä. Kysymykset kysyttiin satunnaisessa järjestyksessä, koska haastatteluilla haluttiin kannustaa mahdollisimman avoimeen ja luontevaan keskusteluun aiheista, jotka haastateltava koki tärkeäksi sillä hetkellä täysin kaavamaisen haastattelun sijaan. Räätelöityjä kysymyksiä kysyttiin myös haastateltavan asiantuntijuuden mukaan, koska eri rekisterivirkahenkilöt olivat eri osa-alueiden asiantuntijoita ja siten tiesivät tietyistä teemoista enemmän kuin toiset haastateltavat. Tutkijoille toteutetut haastattelut erosivat rekisterivirkahenkilöiden haastatteluista siten, että heiltä kysyttiin samoja kysymyksiä palvelujen asiakkaan näkökulmasta. Täten kysymyksiä rekistereiden tiedon järjestelmien kehityksestä ei kysytty heiltä lainkaan. Haastatteluiden päätteeksi kävimme haastateltavien kanssa yhdessä läpi haastattelussa pohditut aiheet, löydökset aiheen tiimoilta ja haastattelija kysyi palautetta käydystä keskustelusta. Haastatteluiden keskimääräinen runko ja ennalta määritellyt kysymykset ovat nähtävillä liitteessä 1.

#### 4.4 Tutkimusaineiston analyysi

Tämä tutkimus noudattaa litteroitujen haastattelujen analyysissa teemoittelua. Teemoittelua lähestytään deduktiivisesti ja latenttisesti perustamalla haastatteluista kooditetut teemat tutkielman viitekehyksen tiedon hyödynnettävyyden attribuuteille. Teemoitteluanalyysissä litteroitua kvalitatiivista dataa jaotellaan teemoihin haastatteluiden tarkemman koodaamisen avulla (Kiger & Varpio 2020). Tämän tutkimuksen teemoitteluanalyysissä käytettiin Clarcken ja Braunin luomaa metodia teemoitteluun, joka sisältää kuusi rekursiivista vaihetta tiedon analysointiin (Braun & Clarke 2006, Kiger & Varpio 2020). Analyysin vaiheet on esitelty Taulukko 3.

Taulukko 3 Teemoitteluanalyysin vaiheet (mukaillen Braun & Clarke 2006; Kiger & Varpio 2020)

Analyysin vaihe	Selitys
1: Dataan tutustuminen	Toistuva ja aktiivinen datan läpi lukeminen, jotta siihen on perehdytty kunnolla. Tämä on oleellista, jotta tutkija ymmärtää, mistä aiheista koodaus kannattaa aloittaa. Dataan tutustuminen on perusta kaikille muille vaiheille analyysissä.
2: Koodaus	Koodauksessa data jaotellaan toisistaan erottuviin, selkeisiin ja perustavanlaatuisiin aiheisiin. Aiheille annetaan oma aihetta selittävä koodi, tunnussana tai -lause, jolla ne voidaan tunnistaa ymmärrettävästi. Koodauksen tarkoituksena on järjestellä data niin, että koodien avulla datasta voidaan myöhemmissä vaiheissa ylemmän tason laajempia teemoja.
3: Teemojen haku	Alustavat ylemmän tason teemat haetaan analysoimalla ja vertailemalla koodeja ja niiden yhteyttä toisiinsa. Koodeista olisi tarkoitus löytää yhteyksiä ja niiden avulla ymmärtää pinnan alla olevaa laajempaa kuvaa. Deduktiivisessa analyysissä teoreettinen viitekehys ohjeistaa teemojen kehityksessä.
4: Teemojen arviointi	Teemojen arvioinnissa on kaksi osaa. Ensimmäisessä osassa tarkastellaan koodattua dataa ja varmistetaan, onko data jaoteltu sopiville teemoille. Toisessa osassa arvioidaan tarkemmin teemoja ja varmistetaan, että ne kattavat kaiken datan.
5: Teemojen määrittely ja nimeäminen	Teemat nimetään ja niille luodaan selitys. Teemojen selityksellä on tuotava ilmi, miksi teema on tärkeä kyseisen tutkimuksen kannalta.
6: Johtopäätösten raportointi	Lopullinen analyysi ja sen löydökset raportoidaan johtopäätöksinä. Tässä raportissa on tärkeää tuoda ilmi, miksi kyseiset teemat on valittu ja kuinka ne vastaavat tutkimuskysymyksiin. Johtopäätösten tulisi loogisesti vastata siihen, kuinka tutkija on tulkinnut dataa ja päätyntä kyseisiin löydöksiin.

Teemoittelulla tarkoitetaan tiedon analyysitapaa, jossa datasta etsitään, analysoidaan ja raportoidaan toistuvia aiheita (Braun & Clarke 2006). Tällöin tietoa jaotellaan ja



lokeroidaan monivaiheisesti eri aiheisiin tavoitteena hahmottaa toistuvia teemoja useasta eri lähteestä. Analyysin lopullisena tarkoituksena on luoda teemojen avulla selkeitä johtopäätöksiä ja tuloksia monimutkaisesta aineistosta (Kiger & Varpio 2020). Kyseinen analyysitapa nähdään perustana rajatumille kvalitatiivisille analyysitavoille, kuten ankkuroitu teoria (eng. grounded theory) (Kiger & Varpio 2020; Watling & Lingard 2012). Teemoittelussa tiedon johtamisessa koodeista teemoihin tutkijan on käytettävä omaa harkintaansa, jolloin tutkijan tulkinta tiedosta vaikuttaa tutkimustuloksiin. Koska tutkija voi käyttää omaa harkintaansa, teemoittelu on analyysitapana kuitenkin hyvin joustava ja sopiva monimutkaisten aiheiden analyysiin. Tutkijan omaa tulkintaa voidaan myös hallita pohjaamalla tiedon analyysi deduktiivisesti kirjallisuudesta luotuun viitekehykseen (Kiger & Varpio 2020.) Deduktiivisessa lähestymistavassa analyysiin käytetään ennalta määritettyä teoriaa tai viitekehystä pohjana kiinnostuksen kohteena oleville teemoille (Braun & Clarke 2012; Varpio ym. 2020). Deduktiivisen lähestymistavan vastakohtana pidetään induktiivista lähestymistapaa, jossa teemat juonnetaan suoraan tutkijan datasta (Varpio ym. 2020).

Teemoittelussa teemalla tarkoitetaan datasta johdettua abstraktia aihetta tai tarkoitusta, joka on yhdistettävissä moniin datassa ilmeneviin datapisteisiin (Braun & Clarke 2006; Kiger & Varpio 2020). Teemojen avulla tiedosta voidaan tehdä johtopäätöksiä ja raportoida ne tuloksina. Teemoja johdetaan datasta luotujen koodien avulla. Koodeilla tarkoitetaan raaka-asta datasta juonnettuja, eri aiheita linkittäviä huomioita (Kiger & Varpio 2020). Koodi yleensä siis summaa ja selkeyttää sitä, mitä kyseinen raakadata käsittelee (Boyatzis 1998).

Braunin ja Clarcken (2006) kehittämä kuuden vaiheen metodi teemoitteluun alkaa tietoon tutustumisella. Tiedon tyyppistä riippumatta tutkijan tulisi toistuvasti käydä tieto läpi, jotta tutkija on tutustunut tutkittavaan dataan tarpeellisesti ennen koodien luontia (Braun & Clarke 2006). Huolellinen dataan tutustuminen avaa teemoitteluanalyysin ja on perusta muille vaiheille tässä analyysimetodissa (Kiger & Varpio 2020). Metodoin toisessa vaiheessa data koodataan yksilöllisiin, toisistaan selkeästi erottuviin koodeihin siten, että koodit ovat yhdistettävissä niin sanottujen koodausohjeistusten kanssa. Nämä koodausohjeet ovat jo valmiiksi olemassa olevia isompia ohjeistuksia koodausprosessiin. (Attride-Stirling 2001; Braun & Clarke 2006; Crabtree & Miller 1999; Kiger & Varpio 2020; King 2004; Nowell ym. 2017.) Tämän tutkimuksen koodit ovat luotu latenttisesti ja deduktiivisesti tutkielman viitekehyksen pohjalta. Latenttisella jaottelulla tarkoitetaan sitä, että teemoista haetaan syvällisempiä tarkoituksia, oletuksia ja ideologioita, kun

verrataan semanttiseen jaotteluun. Semanttisessa jaottelussa teemat jaotellaan koodeihin tarkastelemalla pinnallisesti niiden tarkoituksia. Deduktiivisella lähestymistavalla tarkoitetaan sitä, että tutkimuksen kiinnostuksen alla olevia teemoja tarkastellaan jo olemassa olevan teorian tai viitekehyksen pohjalta. (Braun & Clarke 2006.) Kolmannessa vaiheessa datan avulla luoduista koodeista ja datakoosteista haetaan laajempia merkittäviä teemoja (Braun & Clarke 2006). Deduktiivisessa analyysissä tutkimuksen viitekehys ohjaa teemojen luontia, jolloin teemat keskittyvät tiettyihin näkökulmiin datassa (Braun & Clarke 2006). Analyysin neljännessä vaiheessa luodut teemat tarkastellaan kahdessa eri osassa. Ensimmäisessä osassa tarkastetaan koodatun datan sopivuus niille jaettuihin teemoihin. Neljännen vaiheen toisessa osassa tutkijan tulisi tarkastaa, kattavatko teemat kaiken tutkimuksessa mukana olevan datan loogisesti. Teemoista on selkeästi nähtävä, kuinka ne viittaavat toisiinsa ja itse tutkielman tutkimuskysymyksiin. (Braun & Clarke 2006.) Teemoittelun viidennessä vaiheessa teemat nimetään ytimekkäästi. Kuudennessa vaiheessa teemat ja niihin liittyvät analyysit koostetaan ja raportoidaan tutkimuksen johtopäätöksinä. Tässä vaiheessa tutkijan on oleellista kertoa, kuinka hän on tulkinut dataa ja siten myös osoittaa, miksi valitut teemat ovat tärkeitä kyseisen tutkimuksen kannalta. Tämän lisäksi johtopäätösten tulisi kuvailla dataa ja siten argumentoida, miten kyseiset teemat vastaavat tutkielman tutkimuskysymyksiin. (Braun & Clarke 2006; Kiger & Varpio 2020.)

Teemoitteluanalyysi nähtiin sopivaksi tutkimukseen haastatteluiden ja haastateltavien asiantuntijaluonteen takia. Aihe on laaja ja monimutkainen. Jotta tutkielman tutkimuskysymyksiin on mahdollista vastata, on analyysin oltava joustava ja teemaan soveltuva. Teemoittelun etuna onkin, että se antaa joustavuutta ja tilaa analysoida kompleksista aihetta tutkielman viitekehyksen mukailemana. Kun tutkimuksen aineisto kerätään asiantuntijoilta, jotka työskentelevät aiheen parissa, heidän omat näkemyksensä vaikuttavat koodeihin. Teemoitteluanalyysin joustavuuden ansiosta tämänkaltaisen tutkimus on mahdollista tehdä mahdollisimman selkeästi. (Clarke & Braun 2013, Kiger & Varpio 2020.)

#### **4.5 Tavoite ja ennakoitujen tulokset**

Tutkimuksen ensimmäisenä tavoitteena on saada yleinen kuva ja ymmärrys THL:n terveysrekistereiden reaali maailman tiedon hyödynnettävyydestä toissijaisena aineistolähteenä tutkimuskäyttöön kansallisesti. Nämä tutkimustulokset

potilastietorekistereiden piirteistä ja niiden mahdollisesta vertailusta rakentavat pohjaa myös tulevaisuuden tutkimustyölle.

Tutkielman johtopäätökset tavoitetilassaan tuovat ilmi Suomen terveysrekistereiden reaalimaailman tiedon hyödynnettävyyteen liittyviä mahdollisia haasteita ja piirteitä. Syntyneitä huomioita voidaan hyödyntää potilastietorekistereiden ja niiden tiedon hyödynnettävyyden kehittämisessä, täten edistäen tulevaa tutkimustyötä potilastietorekistereiden reaalimaailman tiedon parissa.

## 5 Tutkimustulokset

### 5.1 Laatu

Haastateltujen THL:n asiantuntijoiden ja tutkijoiden mukaan toisiokäytössä hyödynnettävä THL:n rekistereiden tieto on pääasiassa hyvin laadukasta. Rekistereiden tiedon laadun tarkkailua tehdään THL:n toimesta pääosin manuaalisesti. Tiedon laatuun vaikuttavat myös tiedon välitykseen liittyvät automaattiset järjestelmät ja tiedon tallennuksen periaatteet, joilla pyritään varmistamaan, että tieto kirjataan järjestelmiin oikealla tavalla ja laadukkaasti. THL:n tiedon laadun tarkkailu tapahtuu pääosin vertailemalla eri tietolähteitä.

Laaturekistereissä validoimme dataa eli vertaamme eri tietolähteitä. Esimerkiksi, kun olemme päässeet katsomaan potilastiedon arkiston dataa, niin vertaamme sitä HILMO:n kanssa. Ja sitten meillä on pienemmät aineistot, esimerkiksi diabetesrekisteri sote-organisaatioista ja me vertaamme niitä keskenään. Mutta tähän ei ole tällä hetkellä automaattista, vaan sitä tehdään manuaalisesti ja sieltä nousee esiin erinäköisiä asioita. (VH3)

Tiedon laadun tarkkailua tehdään eri rekistereille eri aikaväleihin. Mitä ripeämmällä tahdilla tietoa julkaistaan, sitä vähäisempää tiedon laadun tarkistus on. Tämä tarkoittaa sitä, että pikaisella aikataululla ja reaaliajassa luovutettu tieto ei ole välttämättä tarkistettu laadultaan yhtä kattavasti kuin pidemmällä aikavälillä luovutettava tieto.

HILMO-rekistereissä tarkistuksia tehdään kattavasti. [...] AvoHILMOssa data tulee päivittäin ja sitä voidaan luovuttaa reaaliajassa. Toisaalta reaaliaikaisesti tietoa luovutettaessa se ei ole niin laadukasta. Mutta jos asiakas priorisoi aikaa laadun sijaan, niin sen luovuttaminen on mahdollista. [...] TerveysHILMOssa tiedon tarkistus on kuukausittaista. Aiemmin se oli vuosittaista, kun olimme tarkistaneet tiedot tarpeeksi laadukkaaksi. Nykyään tarkistus tehdään juoksevasti. (VH2)

Tiedon laadun tarkistusta tapahtuu vertailun lisäksi jälkijättöisesti tiedon hyödyntäjien ja rekisterivirkahenkilöiden tekemällä raportoinnilla. Tietoa käsittelevät osapuolet saattavat huomata tiedossa virheitä tai puutteita, jolloin heidän on mahdollista raportoida tästä rekisteriviranomaisen suuntaan. Tällöin rekisteriviranomainen, kuten THL, pystyy tarkistamaan puutteet tiedoissa ja selvittämään syyn, miksi kyseinen puute on rekisterissä ollut.

Laatuongelmat yleensä havaitaan seuraamalla tietoa. Me esimerkiksi tiedämme, paljonko tekonivelleikkauksia keskimäärin tapahtuu kuukaudessa. Ja jos tieto ei yhtäkkiä vastaa tulevan datan määrän kanssa, niin siihen kiinnitetään huomiota (VH3)

Ekosysteemiajattelun avulla saadaan datan laatua parannettua. Se siis, että me saamme tietoa ulkopuolisilta toimijoilta, datan käyttäjiltä, jotka saattavat huomata siellä laatupoikkeamia. (VH4)

Tiedon laadun ylläpitoa tapahtuu osittain myös automaattisesti ennalta määriteltyjen vähimmäisdatapisteiden ja metatiedoissa asetetun tiedon oikeanlaisen, hyväksytyyn formaatin avulla. Tiedon kirjaamisvaiheessa tietoa on oltava vaadittu määrä ja sen on oltava oikeanlaisessa formaatissa, jotta tallentaminen on mahdollista.

Metatiedon avulla tehdään laaduntarkistusta. Esimerkiksi vastaanotettava data validoidaan metatiedon avulla. Sinne ei pysty ottamaan vastaan tietoa, joka on väärässä formaatissa kirjattu. (VH4)

Laatuun liittyviä haasteita aiheutuu laadun tarkkailun ja tiedon vertailun manuaalisuudesta sekä lääkäreiden tiedon kirjaamishaasteista. Lisäksi haasteita aiheuttaa laadun tarkkailuun käytettävien resurssien määrä verrattuna tarpeelliseen määrään. Tiedon laadun tarkkailuun ei ole olemassa automaattisia järjestelmiä, vaan laadun tasoa ylläpidetään tarkkailemalla poikkeavuuksia tiedoissa manuaalisesti.

Tiedon tarkistuksessa ei ole automaattista validointia. Tiedon tarkastaminen tulee viiveellä. Siinä kohtaa, kun menemme vaikka Helsingin terveyskeskukselle esittämään, että meillä on tietynlainen datasetti missä diabeetikoita on tietyn verran. Niin he saattavat huomata, että luvut eivät pidä paikkansa. Täten pääsemme tarkastelemaan kirjauksia (VH3)

Terveysrekistereiden tiedon alkuperäisessä kirjaamisvaiheessa koetaan olevan tiedon laatuun negatiivisesti vaikuttavia tekijöitä. Terveystieteiden työntekijöiden kirjaamisen ei koeta olevan aina tarpeeksi laadukasta. Syynä tiedon kirjaamisen laadun vaihtelulle pidetään tietojen kirjaamisjärjestelmiä ja sitä, että tietojen kirjaaminen on terveydenhuollon työntekijöiden työnkuvassa toissijaista potilaiden hoidon rinnalla.

Luulen, että lääkäreiden kirjaaminen on yksi iso tekijä virheiden syntymisessä tietoihin. [...] Kyseinen työ on nykyään aikamoista rutiinia lääkäreille, ja kirjaaminen ei välttämättä ole ollut aina kovin heidän ydinhommaansa. (VH2)

Me oikeastaan toivomme, että mahdollisimman paljon pystyisi tekemään niin, että ammattilaisen ei tarvitse erikseen miettiä, että nyt kirjaan tiedon järjestelmään. Tämä polku vaatii sen, että se kirjaaminen on laadukasta ja mehän tiedetään, että se ei aina sitä ole. (VH3)

Osasyynä terveydenhuollon työntekijöiden tiedon kirjaamiseen liittyviin haasteisiin pidetään sitä, että tietoja kirjattaessa on useita eri järjestelmiä, joihin samat tiedot tulisi kirjata moneen kertaan. Kyseistä haastetta on lähdetty ratkaisemaan kertakirjaamisen periaatteella. Tällä tarkoitetaan sitä, että eri järjestelmien tietojen kirjaamisvaiheita pyritään vähentämään ja itse tiedon kirjaamista virtaviivaistamaan järjestelmäintegraatioilla niin, että terveydenhuollon työntekijöiden tarvitsee kirjata hoitotilanteessa tiedot vain kerran yhteen järjestelmään. Kertakirjaamisperiaatteen lisäksi tietojen kirjaamisjärjestelmien käyttöliittymiä on pyritty parantamaan niin, että niiden käyttämiseen ei kuluisi liikaa aikaa.

Kertakirjaamisen periaatteella haetaan sitä, että samaa tietoa ei tarvitsisi kirjata useaan eri järjestelmään. Vaan, että tieto kirjataan kerran laadukkaasti ja se jakautuisi sieltä muihin järjestelmiin. (VH1)

Osa näistä erillisistä tietojärjestelmistä on sellaisia, että lääkärin tallentaessa tiedot kirjaukset siirtyvät suoraan potilastietojärjestelmään. Ja se on tietysti hyvä, että ei tarvitse tuplakirjata kyseisiä tietoja. Mutta sitten on erillisiä tietojärjestelmiä, joissa lääkäri joutuu tekemään normaalin potilaskertomustekstin toiseen kertaan ja siitähän me haluamme päästä eroon. (VH3)

## 5.2 Tarkkuus

Tiedon tarkkuus ja datapisteet ovat haastateltavien mukaan rekisterikohtaisia. THL:n laaturekisterit on luotu ja kehitetty yhteistyössä terveysalojen laaturekistereiden ammattilaisten kanssa. Nämä ammattilaiset määrittelevät, kuinka tarkalla tasolla tietoa kerätään.

Meillä on erilaisia asiakas- ja potilasryhmäkohtaisia laaturekistereitä. Niihin kerätään vähän erityyppistä tietoa. Hoidon laatua mitataan hieman eri tavalla eri sairauksissa ja silloin tiedon ja muuttujien tarkkuus ja laajuus räätälöidään jokaiseen rekisteriin erikseen. (VH3)

Laaturekistereiden datapisteet vaihtelevat riippuen siitä, mistä tietoa kerätään. Esimerkiksi tekonivelrekisterissä tietoa kerätään muihin laaturekistereihin verraten useammalla datapisteellä, koska kyseisellä alalla on ammattilaisten mukaan tarpeellisten analyysien ja vertailujen tekemiseksi oleellista, että datapisteitä on monia. Tiettyjen hoitoalojen tiedonkeruumenetelmät ovat kuitenkin vielä kehitystyön alla, minkä vuoksi laaturekistereihin ei ole saatavilla kaikkea tietoa, mitä ammattilaiset ehkä toivoisivat. HILMO-rekistereissä tietoa kerätään hoitotapahtumakohtaisesti yksilötasolla.

Laaturekistereihin voidaan kerätä jo olemassa olevaa tietoa, koska kerätty tieto on jo laadukasta. Mutta joskus tarvitsee kerätä yksityiskohtaisemmin tietoa. Esimerkiksi tekonivelrekisterissä kerätään tarkkaa nippelitietoa. [...] Diabetesrekisterissä tiedon tarkkuustaso on riittävä tiedon arviointiin. Mutta reumarekisterin tietosisältöä ajateltaessa siellä on yksityiskohtaisia asioita, joita reumalääkärit haluaisivat kerätä. Semmoista tietoa ei juurikaan rakenteisesti kirjata vielä minnekään ja siinä tarvitaan kehittämistä. (VH3)

Haastateltavien mukaan THL:n terveysrekisterit sisältävät pääasiallisesti tietoa tarpeellisen tarkalla tasolla. Vaikka laaturekisterit ovat vielä kehitystyön alla, sisältävät toisiokäyttöön tarkoitettut, yleisimmin tutkimuskäyttöön hyödynnetyt rekisterit niin haastateltujen THL:n ammattilaisten kuin tutkijoidenkin mukaan kattavasti ja tarpeeksi tarkalla tasolla tietoa aihekohtaisesti tutkimuskäyttöä varten.

Olemme tällä hetkellä saaneet tietoa oikein hyvin ja tarkalla tasolla. Toki jos verrataan Kelan ylläpitämiin tilastoihin, jotka siis sisältävät absoluuttista myyntitietoa. Niin niihin verrattuna siellä on toki puutteita, koska kaikki yksityiset toimipaikat eivät edelleenkään ole siellä THL:n rekistereissä täydellisesti. Mutta THL:n rekistereistä saatava tieto on hyvin hyvällä tasolla kaiken kaikkiaan mielestäni. (TH2)

Kattavuudesta tarvitaan usein jokin vertailukohta. Usein muita rekistereitä verrataan HILMO:on, koska luotetaan että HILMO:ssa tiedot ovat tarkasti ja kattavasti. (VH1)

### 5.3 Täydellisyys

THL:n terveysrekistereiden sisältämää tietoa pidetään haastateltavien mukaan hyvin kattavana ja täydellisenä. Täydellisyyteen vaikuttavat datapisteet, joihin kerätään tietoa

ja se, kuinka laadukkaasti tieto on kirjattu. THL:n terveystietorekistereihin kerätään sairaanhoidon yksiköiden ja organisaatioiden alun perin tuottamaa ja kirjaamaa tietoa. Tämä tarkoittaa sitä, että THL:n asiantuntijat eivät itse määrittele aihealueittain dataa, jota terveystietorekistereihin kerätään. Kerätyn tarpeellisen tiedon ovat alun perin määritelleet kyseisen aihealueen asiantuntijat sairaanhoidon organisaatioissa, jotka ovat päättäneet, mitä tietoa on olennaista kerätä kunkin aiheen tiimoilta.

Tietääkseni HILMO:ssa on äärimmäisen kattavasti tietoa. [...] Monissa rekistereissä kenttä on aktiivisesti mukana kehityksessä ja ne on kehitetty asiantuntijaryhmien kanssa, jotka ovat operatiivisessa toiminnassaan töissä kentällä. Usein lääkärit ovatkin kiinnostuneita olemaan mukana alansa rekistereidensä kehityksessä. (VH1)

Vaikka terveystietojen kattavuuden koetaan pääasiassa olevan hyvällä tasolla, on haastateltava VH3:n mukaan tietojen kattavuudessa eroja THL:n eri rekistereissä. Koska THL:n rekisterit koostuvat tiedoista, jotka on kerätty eri tavoin, keräämistapa vaikuttaa rekistereiden tietojen täydellisyyteen. Joidenkin rekistereiden tiedon keräys saattaa olla vielä kehitystyön alla, ja rekistereitä kehittäessä THL ei alkujaan kerää tietoa kaikilta sairaanhoitopiireiltä.

Esimerkiksi hiv-rekisteriin kerätään erikoissairaanhoidosta tiedot kaikista hiv-potilaita hoitavista yksiköistä. Tietoa kerätään THL:lle erillisillä tietojärjestelmillä tai sitten otamme vastaan ihan Excel-tiedostoja manuaalisesti lähetettyinä. Siinä on päästy 100 % -kattavuuteen eli siellä on kaikki yksiköt mukana. Ja sitten jos ajatellaan diabetesrekisteriä, joissa kerätään tietoa HILMO:on pohjautuen, niin tietysti HILMO on hyvin kattava. Mutta sitten laboratoriotiedot ei ole meillä kattavasti, koska siinä olemme vielä rekisterin perustamisvaiheessa. Kyseiseen rekisteriin keräämme tietoa vain kolmen sairaanhoitopiirin alueelta ja jatkossa sopivien toteutustapojen löydyttyä laajennamme tiedonkeruuta. (VH3)

Tiedon täydellisyyteen (kuten myös luvussa 5.1 käsiteltyyn tiedon laatuun) vaikuttaa se, kuinka kattavasti tieto on kirjattu tiedon lähteessä. THL:n asiantuntijoiden mukaan tietojen alkuperäinen kirjaaminen ei välttämättä aina ole laadukasta, koska tietojen kirjaaminen ei ole hoitotilanteessa terveydenhuollon työntekijän ydintehtävä. Kirjaamiseen liittyvät tutkimustulokset on käsitelty tarkemmin luvussa 5.1.



## 5.4 Johdonmukaisuus

Terveystietoja on kerätty rakenteisesti THL:n nykyään ylläpitämiin tietokantoihin jo 60-luvulta alkaen. Täten tiedon johdonmukaisuutta, vertailukelpoisuutta ja eheyttä on mahdollista tarkkailla vertailemalla vuosien saatossa kerättyjä tietoja. Haastateltavien mukaan THL:n terveystiedon ja sen rekistereiden johdonmukaisuuteen ja eheyteen vaikuttavat pääasiassa rekistereihin ja niiden tietojen keräykseen ajan myötä tapahtuneet muutokset ja kehitysaskeleet. Suomessa on useita eri tietojärjestelmiä, joiden kautta tietoa välitetään THL:lle. Niissä tapahtuvat muutokset vaikuttavat myös rekisterikehitykseen THL:n sisällä.

Suomessa on käytössä paljon eri potilastietojärjestelmiä ja saman verran on eri ratkaisuja. Nämä tuottavat tietenkin hieman päänvaivaa. Aina, kun jossain vaihtuu järjestelmä, niin se tuottaa harmaita hiuksia. (VH2)

Eri rekistereissä on ajan myötä saatettu alkaa käyttää erilaisia merkintätapoja kerätyille tiedoille. Tiedot ovat voineet olla eri rekistereissä esimerkiksi tekstimuodossa, numeraalisessa muodossa tai tietoa kerättyäessä on tieto kysytty lähteestä eri muodossa. Tämä on johtanut siihen, että rekistereiden tietoa on ollut hankala yhdistää ja täten kokonaiskuvan muodostaminen ei ole ollut mahdollista. Tätä ongelmaa on lähdetty ratkaisemaan kansallisella koodistopalvelulla ja metatietojen aineistokatalogilla. Nämä on luotu yhtenäistämään kansallisesti tapoja, joilla eri aiheista kerätään tietoa eri rekistereihin.

Tupakointitieto on hyvä esimerkki tietojen keruun epäjohdonmukaisuudesta. Niistä tiedoista on huonosti saatavilla kuvauksia kansallisesti, koska tupakointitiedot kirjataan eri tavalla eri organisaatioissa. Joissain rekistereissä tieto on ”kyllä tai ei” – muodossa ja joissain kirjattuna esimerkiksi 1–5 savuketta päivässä tai vastaavaa. Näiden vertailu ja yhdistäminen on vaikeaa. Tarkoituksena olisi harmonisoida tiedonkeruuta ja kerätä tutkittavat tiedot yhtenäisessä muodossa terveystietorekistereihin. (VH3)

THL:n metatietoaineistojen avulla on mahdollista nähdä, miten tiedon keruu on muuttunut ajan myötä. Vaikka itse tietoaineisto poistettaisiin, on aineistokatalogin periaatteena säilyttää tietoaineiston kuvailu pysyvästi. Tämän periaatteen avulla

tietoaaineistojen kuvailuista on mahdollista nähdä ja päätellä millä tapaa tieto on verrattavissa aiemmin kerättyyn tietoon.

Haastateltujen tutkijoiden mukaan tietoaaineistojen johdonmukaisuuteen voi vaikuttaa esimerkiksi aineistotilausta käsittelevän rekisterivirkahenkilön vaihtuminen. Uusi rekisterivirkahenkilö ei esimerkiksi välttämättä tiedä, mitä muokkauksia tietopyyntöön on tehty, jolloin hän saattaa tehdä omia tulkintoja siitä, mitä tietoa on tarpeellista lähettää tutkijalle. Nämä johdonmukaisuusvirheet kuitenkin yleensä korjaantuvat tutkijan ja tiedon toimittajan välisellä kommunikaatiolla.

Noin kaksi vuotta sitten tiedon toimittajana oli joku toinen tuuraamassa meille yleensä tietoa toimittavaa rekisterivirkahenkilöä. Toimitetuissa tiedoissa ei ollut päätä eikä häntää, koska se ei vastannut yhtään aikaisempaa dataa, jota olimme tottuneet saamaan. (TH3)

Myös tutkijahaastateltava TH1 huomioi sen, että tietojärjestelmien ja siten tiedonlähteiden muuttuminen saattaa vaikuttaa etenkin AvoHILMO-rekisterin tietojen vaihteluun ja johdonmukaisuuteen. Hän koki, että tietojen vaihtelu on enemmänkin tietoaaineistojen ominaisuus kuin suoranainen puute. Tässä hän näkee hyvänä puolena sen, että tiedon laatua korjataan jatkuvasti.

AvoHILMO:n puolella on vaihtelua ja tiedon keräys on paljon epävakaampaa kuin esimerkiksi HILMO:ssa. Se on itse asiassa enemmänkin aineiston ominaisuus kuin suoranainen puute. Jos esimerkiksi vuoden 2021 tuloksia tarkastellaan myöhemmin, niin ne päivittyvät ja muuttuvat matkan varrella, koska tietoja korjataan tai täydennetään täysin uudella tiedolla. Vaihtelu saattaa syntyä esimerkiksi siitä, jos jokin alue on jättänyt tietoja toimittamatta. (TH1)

## **5.5 Saatavuus**

Saatavuus oli yksi attribuuteista, johon liittyvää pohdintaa haastatteluissa nousi eniten esille. Haastateltavien mukaan terveystiedon saatavuuteen toisiokäyttöä varten vaikuttavat useat eri tekijät tiedon tilausprosessissa. Isoimpina vaikuttajina nähtiin metatiedot, tiedon tietolupaprosessit ja itse tietopyynnön käsittelyprosessiin liittyvät vaiheet.

THL:n rekisteriasiantuntijoiden mukaan saatavuuteen vaikuttaa positiivisesti pääasiassa tietoaaineistojen tilausprosessin suoraviivaisuus, asiantuntijoiden kanssa

yhteistyössä kehitetyt metatietoaineistot (aineistokatalogi) ja tietolupaviranomaisen Findatan toiminta tietolupien käsittelijänä ja tietoaineistojen välittäjänä.

Itse tietoaineistojen tilausprosessin koetaan olevan suoraviivainen sen selkeiden vaiheiden takia. Tämä vaikuttaa rekisterivirkahenkilöiden mukaan positiivisesti tietojen saatavuuteen. Tietoaineistojen tilausprosessia helpottaa nykyään Findata, joka osaa kohdentaa laajat, moneen rekisteriin kohdistuvat tietopyynnöt oikeille rekisteriviranomaisille.

Tiedon tilaaminen on Suomessa tällä hetkellä aika suoraviivaista. Tilausta varten on Findata, johon tutkija lähettää tietopyynnön. (VH3)

Ensimmäiseksi tutkijalta edellytetään yhteyden ottamista tietolupaviranomaiseen, eli Findataan. Findata antaa luvat tietojen luovuttamiseen ja käyttöön. Riippuu tietopyynnöstä, mihin aineistoon se kohdistuu, tietolähteestä, poiminnan laajuudesta ja monimutkaisuudesta, kauanko tietopyynnön käsittelyyn menee. Mutta tietopyyntöjen käsittelyssä on tietyt aikakriteerit, joiden perusteella pyritään vastaamaan kaikkiin pyyntöihin. -- Tietolupaviranomainen on olemassa juuri sen takia, että tietävät mihin suuntaan pyynnöt on suunnattava. (VH1)

Tietoaineistojen toimitusta sujuvoittaa tutkijoiden suora kommunikointi tietopyyntöä käsittelevien rekisterivirkahenkilöiden kanssa. THL:n asiantuntijoiden mukaan tietopyynnön määrittelyä tapahtuu usein rekisterivirkahenkilön ja tutkijan välillä sen jälkeen, kun itse alkuperäinen tilaus on tehty. Kommunikointi tapahtuu rekisterivirkahenkilön aloitteesta, kun he huomaavat tarvitsevansa tarkempia tietoja tietopyynnön käsittelemistä varten. Tietopyyntöjen tarkennuksen ja määrittelyn tarpeeseen vaikuttaa se, kuinka hyvin itse tietopyyntö on kuvattu tutkijan toimesta.

Kun pyyntö tulee THL:lälle, niin tämän jälkeen on yleensä myös tarve tarkentaa tutkijan kanssa jälkeen päin miten tietopyynnön määrittely tulee tehdä. Yhdessä tutkijan kanssa käydään läpi tarkemmin ja määritellään ja suunnitellaan tilaus. Tämän jälkeen THL tekee teknisen työn ja luovuttaa tilauksen tietolupaviranomaiselle, joka sitten välittää tutkijalle tiedon omaan etäkäyttöympäristöönsä, jossa voi tehdä oman työnsä. (VH1)

Rekisterivirkahenkilöiden ja tutkijahaastateltavien mukaan tietopyyntöjä on myös mahdollista edelleen tehdä suoraan SOTE-organisaatiosta ja rekisteriviranomaisilta ohittamalla Findata kokonaan. Tämä on hyödyllistä, mikäli tutkimukseen on tarpeellista

saada dataa vain kyseisestä terveystietorekisteristä. THL täten toimittaa myös tietoaineistopyyntöjen myötä tietoa tutkijoille ja organisaatioille suoraan.

Findata on aika uusi viranomaisen ja THL:llä on paljon lupia, joita he [THL] ovat luvanneet ennen heitä [Findata]. Luvat ovat olleet voimassa viisi vuotta. Uudet luvat menevät Findatan kautta. Kun luvat on, niin asiakas ottaa yhteyttä rekisteriviranomaiseen/THL:ään. Viranomaisen katsoo, kuka heillä olisi paras tekemään tilauksen. Tietojen käsittelijä jumppaa aineiston kasaan asiakkaan kanssa. Parhaassa tapauksessa tilaus on hyvin kuvattu, mutta se on vaihtelevaa. (VH2)

Jos tutkija haluaisi tehdä yhden rekisterinpitäjän tai SOTE-organisaation datalla tutkimusta, niin riittää että hän hakee tutkimusluvan sieltä paikalliselta organisaatiolta. (VH3)

Tutkijahaastateltavat kokevat THL:n tietoaineistojen tietojen tilausprosessin olevan yksinkertainen. Tosin tutkijahaastateltavat myös ilmaisivat, että tilausprosessi ja siten tiedon saatavuus on parantunut ajan myötä, kun heille on itselleen kertynyt kokemusta tilausprosessista. Ensikertalaiselle tiedon tilaus ei välttämättä ole yhtä virtaviivaista. Syyksi tutkijahaastateltavat näkevät tähän sen, että he ovat vuosien varrella normaalin oppimisen myötä oppineet, minkälaista tietoa he tarvitsevat ja minkälaista tietoa on saatavilla. Tämän lisäksi tutkijoiden tilauksia käsittelevällä rekisterivirkahenkilöllä on pitkäaikaista kokemusta tiedoista, joita he ovat tilanneet. Parhaimmassa tapauksessa tiedon tilaukseen tarvitaan vain yksi sähköposti rekisterivirkahenkilölle tietotilausprosessin saattamiseksi valmiiksi.

Tilausprosessi on kehittynyt opetteluun ja erehdyksen kautta. Tiedon saanti on nykyään paljon helpompaa, kun tiedetään suunnilleen, miten THL:n päässä dataa lähdetään ajamaan ja minkälaisia tietoja siihen vaaditaan. -- Meidän tietopyyntömme ovat yleensä suhteellisen vakioita riippumatta siitä, mihin terapia-alueeseen tutkimus liittyy. Tällöin THL tietää valmiiksi suunnilleen, minkälaista tietoa halutaan. Pieniä poikkeuksia joskus on, mutta ylipäätään tiedon tilaus ja saaminen on helppoa nykyään. (TH3)

Kun tilaus tehdään, niin kaikki tilaukseen liittyvät tiedot lähetetään kootusti samaan osoitteeseen datan toimittajalle tai rekisterivirkahenkilölle. Tämä henkilö on meidän tapauksessamme aina sama henkilö käytännössä. Sen takia yhteistyö onkin minun mielestäni aika sujuvaa. (TH2)

Tutkijahaastateltavien mukaan Findata toimii hyvänä tiedonvälittäjänä silloin, kun tutkijalla on tarve saada tietoa useasta eri järjestelmästä. Tällöin tutkijan ei tarvitse perehtyä useaan eri rekisteriin, niiden tietosisältöihin, erillisiin tiedonhakuprosesseihin eikä hakea lupaa erikseen jokaiselle eri rekisterin tietopyynnölle. Tutkijoille, joille suomalaisten terveystietojen rekisterit eivät ole niin tuttuja (kuten ulkomaalaiset toimijat tai uudet tutkijat), vaikuttaa Findata tietojen saatavuuteen positiivisesti.

Findatan erityinen hyöty on siinä, jos halutaan yhdistää tietoa monesta eri rekisteristä. Findata on kykeneväinen yhdistämään näiden rekistereiden tiedot ja muodostamaan niistä uuden aineiston. -- hyöty on nimenomaan esimerkiksi Suomen terveystietojen viennille. Eli kun esimerkiksi ulkomaalainen yritys, joka haluaa tehdä jonkinlaisen selvityksen Suomen terveystietoihin perustuen, niin silloin ei välttämättä ole mielekästä lähteä lähestymään kultakin rekisterin ylläpitäjää erillisesti kasaamaan aineistoja ja ylipäättänsä selvittämään, että mitä tietoa mahtaa missäkin rekisterissä olla tarjolla (TH1)

Tutkijat ja rekisterivirkahenkilöt kokevat metatietoaineistojen vaikuttavan positiivisesti terveystietojen saatavuuteen, sillä niiden avulla he voivat nähdä, mitä tietoa rekistereistä on saatavilla. Vaikka tietoaineistotilauksen yhteydessä rekisterivirkahenkilöt auttavat ja käyvät keskustelua tutkijan kanssa tilauksesta ja tarpeellisista tiedoista tutkimusta varten, metatietojen olemassaolo nopeuttaa tilausprosessia. Metatietoaineistoja on historiassa rakennettu vaihtelevasti eri rekistereille. Rekisterivirkahenkilöiden mukaan eri rekistereiden metatiedot on kehitetty vaihtelevalla laadulla.

Ennen eri rekistereillä on ollut omia metadata-aineistoja. Jotkut ovat olleet laadukkaampia kuin toiset. Mutta jokainen on määritellyt ne omalla tavallaan. Eli hyvin toimivia systeemeitä on ollut, mutta ne ovat olleet ripotellen siellä täällä. Yhtenäistä suunnittelua ei ole ollut. (VH4)

Isaacus-hankkeen (Sitra, 2017) myötä THL alkoi rakentaa aineistokatalogia, jonka tarkoituksena on toimia kansallisena metatietoaineistona. Aineistokatalogi on luotu haastateltava VH4:n mukaan kokoamalla aiemmin kehitettyjä rekistereiden omia metatietoaineistoja. Aineistokatalogi on myös edelleen kehityksen alla ja sitä kehitetään manuaalisesti asiantuntijoiden kanssa. Aineistokatalogin päivittämistä varten THL on kehittänyt Aineistoeditori-palvelun, jossa oikeudet saaneet viranomaiset voivat päivittää

metatietoaineistoja aineistokatalogiin. THL:n kaikki aineistot on kuvattu vuodesta 2018 alkaen aineistokatalogin mukaisesti.

Sitran Isaacus-hankkeen myötä aineistokatalogia alettiin ajattelemaan enemmän kansallisesti. Siinä kohtaa tiedettiin, että THL:stä tulisi tietolupaviranomainen (Findata). Aineistokatalogia lähdettiin tällöin kehittämään toisiokäytön tueksi. [...] Me ei siinä kohtaa keskitytty metadatan tuottamiseen. Sen sijaan me lähdimme jo tuotetun tiedon kuvailemiseen jälkijättöisesti. Tätä toimintatapaa kritisoiitiin paljon, mutta toisaalta jostain oli pakko aloittaa. Siten saimme aikaiseksi aineistokatalogin ja aineistoeditorin, jotka ovat toimivia, mutta eivät missään nimessä lähellekään täydellisiä. (VH4)

Metatietoaineistojen olemassaolo ja ominaisuudet ovat rekisterivirkahenkilöiden mukaan vaikuttaneet negatiivisesti terveystiedon saatavuuteen. Haastatteluissa nousivat usein esiin puutteelliset aineistotilaukset sekä metatietojen tärkeys tilausta tehtäessä. Koska kattavia aineistoja metatiedoista ei aina ole ollut olemassa, haastateltavat kokivat, että historiassa tutkijoiden on ollut hankalaa tietää, minkälaista tietoa rekistereissä on ylipäänsä ollut saatavilla uutta tietotilausta tehtäessä.

Tutkijan täytyy aika tarkkaan tietää tilausta tehtäessä, mitä tietoa tutkimukseensa haluaa. Se on ollut tähän mennessä iso ongelma, kun niitä tietoja ei ole ollut hyvin saatavilla. Tutkijat ovat joutuneet ottamaan yhteyttä rekistereiden asiantuntijoihin. Asiantuntijat ovat joutuneet selittämään, mitä kaikkea on saatavilla. (VH4)

Kokemukseni on, että aineistokuvaukset ovat olleet riittävän hyvät, mutta eivät koskaan täydelliset. THL on kehittänyt niitä vuosien varrella ja julkaissut aineistokatalogiin. [...] Metadatan tärkeys tulisi siis huomioida. Se auttaa, mitä enemmän tutkijalle ja asiakkaalle tarjotaan mahdollisuutta tutustua rekisteriin etukäteen. Eihän koko pyyntöä voi tehdä, ellei tiedä mitä tietoa on mahdollista saada. (VH2)

Metatietoaineistojen kehittämisen koetaan olevan hidasta ja hankerahoituksesta ja resursseista riippuvaista. Metatietoaineistojen päivitys aineistokatalogiin on tällä hetkellä täysin manuaalista. Täten metatietojen kehittämiseen vaaditaan kehityksen alla olevan aineiston alan asiantuntijoita, joka itsessään jyrkentää työresurssien puutetta aineistokatalogin kehityksessä.

Resurssit ovat siirtyneet tietoaaineistojen GDPR-mukaistamiseen, tietoturvaan ja säilyttämiseen liittyviin kehitysprojekteihin. Resurssiongelmat ovat olleet

pääsyy isoon backlogiin ja projektien hitauteen aineistokatalogin ja -editorin kehityksessä. [...] Myös Aineistoeditorin tämän hetken ongelma on sen manuaalisuus. Sinne ei voi tuoda automaattisesti muista järjestelmistä tietoa. Resurssiongelmassa tulisi huomioida, että rekistereitä kehittäessä substanssiasiantuntijat ovat todella keskeisiä. Metatietojen tuottamisessa tarvitaan metatietoasiantuntijaa, joka pystyy kyseistä tietoa tuottamaan. (VH4)

Metatiedot ovat kytköksissä toisiolakiin, koska siellä annetaan suoraan lakisääteiseksi tehtäväksi ylläpitää aineiston kuvauksia. Mutta se taas kytkeytyy resursseihin siten, että kun laki tuli voimaan, niin se ei tietenkään suoraan antanut toimijoille mitään uusia resursseja, jotka mahdollistaisivat näiden asioiden teon. (VH5)

Rekisterivirkahenkilön VH5 mukaan yksi saatavuuteen vaikuttava tekijä on ollut se, että aineistokuvauksia ei ole ollut saatavilla muuten kuin suomenkielisenä. Kansainvälisille tutkijoille on ollut vaikeaa esitellä aineistoja, joka on vähentänyt VH5:n mukaan kansainvälisiä asiakkaita.

THL:llä on aineistokuvauksia suomeksi ja joitain ruotsiksi. Siten kansainvälisille tutkijoille on ollut huonosti aineistokuvauksia tarjolla. Se on suoraan sanottuna hävettänyt, kun kuvauksia on kysytty eri kielillä. Tämä on saattanut vaikuttaa asiakkaiden lähtemiseen. (VH5)

Tutkijahaastatteluvien mukaan aineistotilausten käsittelyn ruuhkat ovat vaikuttaneet negatiivisesti terveystiedon saatavuuteen sekä seuraavassa alaluvussa käsiteltävään ajantasaisuuden attributtiin. Ruuhkat johtuvat aineistotilausten priorisoinnista viranomaisille sekä resurssien puutteista aineistotilausten käsittelyssä.

Aineistotilausten viivästymiset mielestäni johtuvat siitä, että THL:llä ei ole riittävästi osaavaa väkeä. Toki viimeiset kaksi vuotta koronan myötävaikuttama kuormitus viranomaisten tekemien tietopyyntöjen THL:lle on ollut merkittävä tekijä. (TH3)

Tyypillisesti aineistotilauksen jälkeen odotellaan omaa vuoroa siihen, että se aineisto toimitettaisiin. Saatamme saada jonkinlaisen arvion toimitusaikataulusta. Tyypillistä on, että joudumme useaan kertaan kyselemään aineistojen perään ennen kuin ne saadaan. – THL:n aineistotilausten käsittelyn työvoiman resursointi on kokemuksemme mukaan pullonkaula tilauksissa. (TH1)

## 5.6 Ajantasaisuus

Haastateltavien mukaan ajantasaisuuteen vaikuttaa pääasiassa kaksi eri tekijää. Näitä ovat tiedon välittämiseen liittyvät prosessit ja tietojen välittymisestä järjestelmään kirjaamishetkestä itse kansallisiin terveystietorekistereihin. Tietojen ajantasaisuus on myös rekisterikohtaista. Esimerkiksi THL julkaisee avoimiin rajapintoihin julkisia raportteja, joiden tieto välittyy automaattisesti. Täten julkisten tietojen välitys on ajantasaisinta. Manuaalisesti toimitettavien tietojen tietopyyntöjen käsittelyprosessissa on tietyt määritellyt vaiheet, jotka tapahtuvat tiettyjen aikamääreiden sisällä. Aineistojen toimittamiselle on siis asetettu tietyt aikamääreet, joiden sisällä tiedot tulisi käsitellä ja toimittaa tilaajalle. Nämä aikamääreet on asetettu toisiolain myötä.

Tietojen ajantasaisuuteen vaikuttavat haastateltavien mukaan negatiivisimmin tietopyyntöihin liittyvät prosessit ja resurssipuutteet. Tällä tarkoitetaan sitä, että tiettyinä aikoina tietopyyntöjen käsittelyssä voi olla ruuhkaa, eikä niiden käsittelyyn aina ole riittävästi työvoimaa. Tämän lisäksi tietoaineistojen toimitusten hitautta voivat aiheuttaa aineistojen koko ja niiden välittämiseen tarkoitettut järjestelmät. Mitä isommassa tiedostossa tietoaineistot ovat, sitä hitaampaa niiden toimittaminen on.

Ruuhkat ovat välillä pahoja. Toisiolaissa on deadline't. Tiettyssä määräajassa täytyisi pystyä toimittamaan data eteenpäin, ja se tuottaa organisaatiossa paineita. Datan koko myös vaikuttaa. Jos asiakas haluaa paljon dataa, syntyy tilanne, jossa on monen kymmenen gigan kokoisia tekstitiedostoja. Isojen tiedostojen liikuttelu on vaikeaa. Joskus on ollut yhteistyöhön biopankin kanssa liittyvää geenidataa monta teraa käsittelyssä. Siinä loppuu keinot tiedostojen käsittelyssä. Mutta kokoajanhan voidaan enemmän ja enemmän kapasiteettia tarjota. Isot tilaukset yleensä yritetään toimittaa pilvipalveluiden tai etäyhteyksien kautta. (VH2)

Haastateltujen tutkijoiden mukaan tietojen ajantasaisuuteen voivat vaikuttaa virheet itse aineistotilausten koostamisessa niissä tilanteissa, kun tilausta koostava rekisterivirkahenkilö vaihtuu aiemmasta. Tällöin alkuperäisellä rekisterivirkahenkilöllä on saattanut olla hiljaista tietoa tietopyyntöihin liittyen, ja tämä ei välttämättä aina siirry häntä tuuraavalle virkahenkilölle. Myös inhimillisiä virheitä tapahtuu aineistotilaustoimituksissa, vaikka rekisterivirkahenkilö ei olisi vaihtunut. Näissä tilanteissa aineistotilausprosessi pitkittyy tutkijan näkökulmasta, koska tietojen saapumista oikeassa muodossa joudutaan odottamaan.



Ja sitten kun aineistot saadaan, niin tyypillisesti tarkastetaan ne aineistot. Aika tyypillistä on, että sieltä löytyy puutteita tai vajavuuksia mitä sitten pyydetään tarkentamaan. (TH1)

Voisi sanoa, että pääasiallinen rakenne, missä datat tulee meille, on hyvinkin vakio. Mutta toki välillä on tapahtunut sitä, että on jäänyt joku tilattu osio tai tuote puuttumaan. Tällaisia inhimillisiä virheitä on tapahtunut, mitä voi tapahtua ihan missä tahansa työssä, mutta ei ole mitään sellaista johdonmukaista. (TH2)

Tutkijahaastateltava TH1:n mukaan aineistotoimituksissa tapahtuvat viivästymiset aiemmin mainituista syistä vaikuttavat negatiivisesti kyseisten rekistereiden tiedon toimituksen ennustettavuuteen. Tietojen toimituksen ennustettavuudella tarkoitetaan sitä, kuinka tarkasti tietopyynnön tekijä pystyy määrittelemään ajankohdan, jolloin tieto hänelle saapuu. Tietojen saapumisen ajantasaisuus ja ennustettavuus on tutkijoille tärkeää muun muassa pitkäaikaisissa tutkimuksissa, joissa tarkastellaan tietojen muutosta tasaisin aikaväleihin. Tämän lisäksi tietopyynnöissä, jotka on tehty päätöksentekoa varten, ennustettavuuden koetaan olevan tärkeä tekijä. Tutkijoiden mukaan tietojen saapumisen ennustettavuus THL:n toisiokäyttöön tarkoitettuista terveystietorekistereistä ei ole toivotulla tasolla.

Tietojen saapumisen ennustettavuus on suurin haaste. Tietyllä tavalla aineistotilausten yhteydessä arviointi, että milloin se aineisto olisi käytettävissä, on hyvin haastavaa ja yleensä arvio ei myöskään pidä paikkaansa kovin tarkasti. Ennustettavuus on tärkeä attribuutti, koska tyypillisesti tietoa käytetään jonkinlaisen päätöksen tekemiseen tai toiminnan seurantaan. Varsinkin jos kyse on jonkinlaisen päätösten tekemisestä, liittyy se sitten vaikka lääkkeiden korvattavuuteen tai toimintamallien käyttöönottoon tai mihin tahansa muuhun, niin niillä on jonkinlaiset aikataulut. Jos päätöksillä on aikataulut niin silloin tietysti tieto, mihin ne päätökset pohjaa, pitää aina valita sen aikataulun mukaisesti. Siinä helposti käy niin, että jos on sellaisia tietokanavia, missä se aikataulu ei ole arvioitavissa, niin silloin ei ole mielekäästä tehdä päätöksiä ja sen takia ennustettavuus on jopa tärkeämpi kuin itse toimitusaikataulun nopeus. (TH1)

Tutkijat kokevat rekisterivirkahenkilöiden resursoinnin olevan pääsyyntä ajantasaisuus- ja ennustettavuusongelmiin. Samat rekisterivirkahenkilöt on resursoitu sekä viranomaisten (esimerkiksi Suomen ministeriöt) tekemiin että muihin tietopyyntöihin, mutta viranomaisten tietopyynnot ylittävät prioriteetiltaan normaalit

tietopyynnöt. Eri maailmantilanteet ja vuodenaajat voivat vaikuttaa siihen, että viranomaisilta tulee enemmän tietopyyntöjä.

Se on hyvin pitkälle resursointikysymys ja siinä on useita eri haasteita. Yksi konkreettien haaste on, että analytiikan parissa työskenteleville ei ole tarpeeksi aikaa käsitellä niin montaa tietopyyntöä. Toinen haaste on se, että THL:llä on samat henkilöt resursoitu sekä viranomaisten tietopyyntöihin että sitten muihin tietopyyntöihin. Ja jos syntyy paljon viranomaisten tietopyyntöjä, niin kaikki muut tietopyynnöt pysähtyvät ja siitä syntyy tämä tiedon saapumisen ennustamattomuus. (TH1)

Suurin syy on resurssit. Sehän tarkoittaa juuri sitä, että rekisterivirkahenkilöillä on liikaa tekemistä. Ruuhkaa aiheuttavat THL:n sisäiset tai ministeriöiden tietopyynnöt, jotka menevät käsittelyssä muiden tietopyyntöjen edelle. Tämä johtaa resurssipulaan. (TH2)

Tutkijoiden mukaan myös rekisterivirkahenkilöiden sairastumiset ja lomakaudet vaikuttavat resurssipulaan. Virkahenkilöiden sairastuttua tai lomakauden alettua henkilöstöä tietopyyntöjen käsittelyyn on vähemmän. Resurssipula vaikuttaa siihen, kuinka kauan aineistotoimituksien käsittely kestää.

Yksi tekijä on tietysti lomakaudet. Sehän nyt on päivänselvää, että jos datan käsittelijät ovat lomalla niin ei niitä kukaan aja silloin. Tai ne hidastuvat, jos siellä yhdellä ihmisellä on enemmän vastuuta. (TH2)

THL:llä hyvin rajattu joukko ihmisiä keskittyy aineistoanalyysiin ja yksittäinen sairastuminen saattaa venyttää aineistotoimituksia jopa kuukausilla. (TH1)

## 6 Johtopäätökset ja keskustelu

### 6.1 Aiempi kirjallisuus

Terveydenhuollon yksiköissä ja organisaatioissa syntyneen sosiaali- ja terveystiedon tehokas hyödyntäminen edistää kansallista terveydenhuoltoa, hoiva-alaa, tutkimusta ja terveydenhuollon palveluiden ympärillä toimivaa liiketoimintaa (Lehto & Neittaanmäki 2018). Tutkimuksen tulosten mukaan THL:n terveystietorekistereiden tiedot ovat tutkijan näkökulmasta kokonaisvaltaisesti hyödynnettävissä toisiokäyttötarkoituksiin. THL:n keräämän terveystiedon toisiokäyttöön saamisen prosessissa on kuitenkin tekijöitä, jotka hidastavat tiedon saatavuutta sekä rekistereiden, järjestelmien ja tiedon laadun kehitystä. Tutkimuksessa nämä negatiivisesti ja positiivisesti tiedon toisiokäytön hyödynnettävyyteen vaikuttavat tekijät on jaettu terveydenhuollon rekistereiden ja niiden tarkkailuun sopiville attribuuteille, joiden avulla on mahdollista arvioida hyödynnettävyyden tasoa, tuoda ilmi huomioita tiedon hyödynnettävyydestä sekä ymmärtää, missä osa-alueissa on kehitettävää.

Aiemman kirjallisuuden mukaan Suomen kansallisten asiakas- ja potilastietorekistereiden tiedon toisiokäytön hyödynnettävyydessä on havaittu keskeisinä kehityskohteina tiedon laatuun ja keräyslaajuuteen liittyvät aiheet, usean eri rekisterin tiedon samanaikaisen hyödyntämisen vaikeudet, erillisten päällekkäisten järjestelmien kehittäminen organisaatiokohtaisesti, tutkijoiden ja osapuolien tarve saada hyödyntää tietoa nopeammin tehokkaampien järjestelmien avulla ja tiedon hyödyntämisen palveluiden siirtyminen digitaalisiin, kansallisesti yhtenäisten palveluiden pariin (Martti & Viitanen 2016). Näiden lisäksi terveystiedon hyödynnettävyyteen toisiokäyttöön on kirjallisuudessa havaittu vaikuttavan organisaatioiden resursseihin, asiantuntijoiden rekrytoimiseen, lupien ja tietoaineistojen muodostukseen ja tiedon kirjaamiseen terveydenhuollon organisaatioissa liittyvät haasteet (Rannanheimo & Jauhonen 2018). Tämän tutkimuksen tulokset ovat linjassa aiemman kirjallisuuden huomioiden kanssa ja ne valottavat nykyhetken tilannetta ja kehityksen suuntaa.

### 6.2 Keskeiset johtopäätökset

Tutkimuksen empiirisessä osiossa haastattelutulokset teemoiteltiin viitekehyksen avulla. Viitekehys koostuu terveystiedon rekistereiden hyödynnettävyyden attribuuteista, joita ovat tiedon laatu, tarkkuus, täydellisyys, johdonmukaisuus, saatavuus ja ajantasaisuus.

Pääteemat tiedon laatuun, tarkkuuteen, täydellisyyteen ja johdonmukaisuuteen liittyviä attribuutteja käsitellessä olivat tiedon validointi, vertailu ja tarkastus, tietojen kirjaaminen terveydenhuollon palvelutilanteissa ja metatietoaineistot. Pääteemat liittyen saatavuuden ja ajantasaisuuden attribuutteihin olivat THL:n aineistokäsittelyyn allokoitujen resurssit, tietopyyntöjen käsittelyn prosessit ja metatietoaineistojen kehitys. Keskeisimpinä havaintoina tutkimuksesta on mahdollista kiteyttää, että negatiivisesti terveystietojen hyödynnettävyyteen toisiokäytössä vaikuttavat tällä hetkellä eniten saatavuuteen ja ajantasaisuuteen liittyvät teemat.

Tietoaineistojen käsittelyyn allokoitujen työvoimaresurssit ja manuaaliset toimenpiteet vaikuttavat suurimpaan osaan hyödynnettävyyden attribuuteista. Tietojen laatuun ja tarkkuuteen vaikuttavat manuaaliset tietojen tarkistus- ja vertailuprosessit. Tietojen johdonmukaisuuteen, saatavuuteen ja ajantasaisuuteen vaikuttavat puolestaan tietopyyntöjen käsittelyyn resursoitujen rekisterivirkahenkilöt sekä heille työhön allokoitu aika.

Tutkimuksen huomioita resurssiongelmista tukee aiempi kirjallisuus ja tutkimus aiheesta. Aiemman kirjallisuuden mukaan yksittäisissä organisaatioissa henkilöresurssit ovat pienet, keskittyneet ydintoimintaan ja viranomaisilla on vaikeuksia rekrytoida ammattilaisia (Martti & Viitanen 2016; Rannanheimo & Jauhonen 2018). Konkreettisesti THL:n resurssipula ilmenee tutkijalle esimerkiksi, kun tietopyyntöjen käsittelyssä on ruuhkaa tai kun viranomaisten tekemiä prioriteetiltaan tärkeämpiä tietopyyntöjä on enemmän. Tietopyyntöjen käsittelyprosesseihin ja tietojen kokoamiseen liittyvät työvaiheet ovat yksittäisten rekisterivirkahenkilöiden vastuulla. Tällöin tiedot saattavat tutkijoiden mukaan viivästyä jopa kuukausilla. Myös tietopyyntöä tehneen rekisterivirkahenkilön vaihtuessa muun muassa sairauden tai työnkuvan vaihtumisen takia voi hiljaista tietoa aiemmista aineistopöiminnöistä jäädä välittämättä uudelle rekisterivirkahenkilölle. Tällöin tietoaineiston toimitus viivästyy lisääntyneen aineiston käsittelyn ja tarvittavien muutosten takia. Työvoimaresurssiongelmat ovat haastateltujen mukaan korostuneet COVID-19-pandemian aikana, kun viranomaisten tekemien tietopyyntöjen määrä on kasvanut huomattavasti verrattuna aiempaan.

Resurssivaje ilmenee myös THL:n järjestelmien kehityksessä. THL:n ja terveystietojärjestelmien, rekistereiden ja kehitys ylipäätään ovat riippuvaisia hankerahoituksista, jotka pääasiassa saadaan julkishallinnolta. Toisiokäyttölain myötä tulleet muutokset ovat käynnistäneet useita eri hankkeita, mutta rekisterivirkahenkilöiden mukaan toisiokäyttölaki ei automaattisesti kasvattanut resurssien määrää kehitystä varten.

Tämä on näkynyt esimerkiksi metatietopalvelun aineistokatalogin kehityksessä, kun SOTE-tiedon toisiokäyttölain tietoturvaluudistusten myötä THL:n resurssit olivat siirtyneet tietoaaineistojen GDPR-mukaistamiseen, tietoturvaan ja tietojen säilyttämiseen liittyviin kehitysprojekteihin. Metatiedon ja niiden muokkausjärjestelmien kehitykseen liittyvät projektit ovat tämän takia edenneet hitaasti. Metatietoaineistojen olemassaolo ja laatu vaikuttavat terveystiedon saatavuuteen ja johdonmukaisuuteen. Työvoimaresurssiongelmia näkyvät myös siinä, että rekistereiden ja niihin liittyvien järjestelmien kehitykseen tarvitaan kyseisten alueiden lääkäriasiantuntijoita kertomaan näkökulmansa kehitykseen liittyvistä asioista. Haastateltavien mukaan lääkärit ja substanssiasiantuntijat ovat historiassa mielellään osallistuneet terveystietoon liittyvien järjestelmien ja rekistereiden kehitykseen, mutta heitä on rajallinen määrä käytössä projekteihin. Tämä hidastaa järjestelmien kehitysprojektien etenemistä.

Työvoimaresurssieihin liittyviin huomioihin vaikuttaa myös yksittäisten järjestelmien kehityksen, tietopyyntöjen käsittelyn ja laadun tarkkailuun liittyvien prosessien manuaalisuus. Kehitys on hankeriippuvaista ja täten edistyneempien järjestelmien kehitys ei välttämättä aina ole ollut mahdollista hankkeiden resurssien myötä (Martti & Viitanen 2016, Rannanheimo & Jauhonen 2018). Muun muassa aineistokatalogin kehitystä varten on rakennettu aineistoeditorijärjestelmä. Aineistokatalogi on kehitetty sillä periaatteella, että sinne on mahdollista koota jo olemassa olevia metatietoaineistoja, jotta kyseisten tietojen kerääminen olisi jatkossa kansallisesti yhtenäisempää. Jo olemassa olevien metatietoaineistojen tuonti aineistoeditoriin ei kuitenkaan ole automaattista. Järjestelmän tietojen päivitys tapahtuu manuaalisesti siten, että käyttöoikeuden saaneen asiantuntijan on mahdollista yksitellen ladata aineistolistauksia aineistoeditoriin. Tietojen laadun tarkkailussa manuaalisuus näkyy siinä, että laadun tarkkailu pääasiassa tapahtuu vertailemalla tietoja referenssiaineistoihin. Virheiden tarkastus ja huomiointi tapahtuu myös jälkijättöisesti niin, että itse huomiot virheistä saattavat tulla jälkeenpäin tietoa toisiokäyttötarkoituksiin hyödyntävältä THL:n asiakkaalta.

Työvoimaresurssiongelmien myötä syntyvät aineistotoimitusten satunnaiset viivästymiset vaikuttavat tutkijahaastateltavien mukaan tiedon saatavuuden ennustettavuuteen. Tutkijoiden mukaan tieto aineistotoimituksen saapumisen ajankohdasta on tärkeä tutkimuksiin ja päätöksentekoon liittyvien liiketoiminnallisten syiden takia. Tutkijahaastateltavien mukaan terveystietojen käsittelyä varten luodut etäkäyttöjärjestelmät auttavat tietopyyntöjen käsittelyyn liittyvissä ongelmissa.

Etäkäyttöjärjestelmät siirtävät tietojen käsittelyyn liittyviä työvaiheita rekisterivirkahenkilöiltä tiedon tilaajille.

Tutkijat ja rekisterivirkahenkilöt kokevat tiedon laadun olevan hyvällä tasolla, koska tiedot välittyvät THL:n terveysrekistereihin suoraan lähteestä eli sosiaali- ja terveydenhuollon yksiköiden ja organisaatioiden palvelutilanteista ja järjestelmistä. Tiedon laatuun kuitenkin vaikuttaa haastateltavien mukaan lääkäreiden tietojen kirjaaminen. Tietojen kirjaaminen, koodaaminen ja niissä tapahtuvat virheet vaikuttavat tietojen käytettävyyteen tutkimuksessa (Garrison ym. 2007, Sørensen ym. 1996). Tietojen kirjaamisen ei koeta olevan lääkäreiden ydintyötä, minkä takia sen laadun koetaan olevan vähemmällä huomiolla etenkin silloin, kun lääkäri tai terveydenhuollon työntekijällä on useita eri järjestelmiä, joihin samat tiedot on kirjattava. Lääkäreiden kirjaamishaasteeseen on lähdetty etsimään ratkaisua edistämällä kertakirjaamisperiaatetta järjestelmien kehityksessä. Tällä tarkoitetaan sitä, että tiedonvälitysjärjestelmiä pyritään yhtenäistämään niin, että tieto välittyisi niiden välillä. Tällöin lääkärin ei parhaimmassa tapauksessa tarvitse kirjata tietoja kuin vain yhteen järjestelmään. Aiemman kirjallisuuden mukaan tämä ei kuitenkaan ole toteutunut tarpeellisella tasolla (Rannanheimo & Jauhonen 2018).

Metatietoaineistojen olemassaolo vaikuttaa haastateltujen mukaan terveystietojen laatuun, saatavuuteen ja johdonmukaisuuteen. Aineistokatalogin avulla on mahdollista nähdä, minkälaista tietoa terveysrekistereistä on saatavilla ja missä muodossa. Osa rekisterivirkahenkilöiden työnkuvaa on keskustella tietopyyntöjä tehneiden asiakkaiden kanssa tarkennuksista ja määrittelyistä tietopyyntöihin. Haastateltujen rekisterivirkahenkilöiden mukaan metatietokuvausten olemassaolo vähentää aineistotoimituksiin liittyvien määrittelyihin kohdentuvaa työtaakkaa. Tutkijoiden näkökulmasta metatietoaineistot parantavat rekistereiden avoimuutta ja auttavat näkemään, minkälaisen tiedon kanssa tutkimusta on mahdollista tehdä. Aineistokuvaukset ovat tällä hetkellä saatavilla vain suomenkielisenä ja osittain ruotsinkielisenä, mikä vaikeuttaa kansainvälisten tutkijoiden tutkimustyötä. Aineistokatalogin perustamisen taustalla on ollut myös tarkoitus rakentaa kansallinen metatietorekisteri toisiokäytön tueksi. Tällä tarkoitetaan sitä, että kerättyjä tietoja ja niiden kuvauksia yhtenäistetään kansallisella tasolla. Vuodesta 2018 eteenpäin THL on kuvaillut rekisteriaineistonsa aineistoeditorin kuvausten mukaisesti. Metatietoaineistot vaikuttavat tulosten mukaan tietojen johdonmukaisuuteen, kun tiedoista on yksittäinen kuvausaineisto useiden sijaan. Metatiedot vaikuttavat tiedon laatuun positiivisesti

toisiokäyttölain myötä tulleiden muuttujatasoisten vähimmäistietosisältöjen avulla. Vähimmäistietosisällöillä tarkoitetaan datapistemäärää ja niiden formaattia, joka rekisteriin täytyy vähimmillään tallentaa. Tämä vaikuttaa siihen, että tarpeelliset tiedot kirjataan järjestelmiin oikeassa muodossa, ja siten itsessään parantaa rekistereiden tiedon laatua.

Tulevaisuuden kehityskohteena tuloksista ilmeni ja kiteytyi metatietojen, resurssien ja järjestelmien automatisoinnin painottamisen tärkeys. Koska aineistokatalogi on vielä kehityksen alla, THL:n rekisterivirkahenkilöt kokivat tietoaineistojen kuvailujen olevan tärkeitä työkaluja tutkimuksen aloittamisvaiheessa ja tutkimuksen tekemisessä ylipäätään. Järjestelmien kehitys on hanke- ja rahoitusriippuvaista, ja toisiokäyttölain myötä iso osa resursseista on suunnattu tietoaineistojen GDPR-mukaistamiseen sekä tietoturvaan liittyviin projekteihin. Täten tiettyjen järjestelmien kehityksen koetaan olevan hidasta. Tutkijoiden näkökulmasta rekisterivirkahenkilöt ja tietopyyntöjen käsittelijät koetaan manuaalisiksi välikäsiksi tietojen toimituksessa. Automaattisten tietojenvälitysrajapintojen, joissa tutkija voi itse käydä tarkastelemassa dataa ja muokata sitä tarpeidensa mukaan, koettiin olevan ratkaisu tietopyyntöihin ja -käsittelyyn liittyviin saatavuuden ja ajantasaisuuden haasteisiin.

### **6.3 Kontribuutio, tutkimuksen rajoitukset ja jatkotutkimus**

Tutkimus mukailee aiempaa Suomen kansallisten sosiaali- ja terveydenhuollon tietorekistereiden parissa tehtyä tutkimusta tiedon toisiokäytön hyödynnettävyyden osalta. Aiempi tutkimus on keskittynyt luomaan kokonaisvaltaisia kuvia sosiaali- ja terveydenhuollon rekistereiden tiedon toissijaisen käytön arkkitehtuurista kansallisella tasolla. Aiempi tutkimus on tarkastellut toisiokäytön arkkitehtuuria pääosin rekistereiden kehittäjien näkökulmasta. Tällöin käsittelyssä ovat olleet tietojärjestelmät, rekisterit ja tietoalttaat ylemmällä tasolla. Tässä tutkimuksessa on tarkasteltu tapaustutkimuksen avulla lähemmin THL:n terveysrekistereiden ja niiden hyödynnettävyyttä toisiokäyttötarkoituksiin. Tutkimus on tehty ottamalla huomioon niin THL:n rekisterivirkahenkilöiden kuin terveysrekistereitä toisiokäyttöön hyödyntävien tutkijoiden näkemykset. Tutkimuksen avulla on mahdollista yhdistää ylemmän tason toissijaisen käytön arkkitehtuurin huomiot tarkempaan tarkasteluun kansallisten terveysrekistereiden tämän hetken toisiokäytön hyödynnettävyyden haasteiden kanssa. Tutkimus on mahdollistanut tarkastelun nykyhetken haasteisiin toisiokäyttöön tarkoitettujen rekistereiden kehityksessä ja asettaa valoon käytännön haasteita aiheuttavia

syitä niin tutkijoiden kuin rekisterivirkahenkilöiden näkökulmasta. Tutkimuksen avulla on mahdollista ymmärtää, mihin aihealueisiin on hyvä panostaa terveysrekistereiden järjestelmien ja niiden toisiokäytön hyödynnettävyyden kehityksessä.

Tutkimus rajoittuu THL:n terveysrekistereiden tarkasteluun tietyinä hetkenä. Tutkimuksen tuloksia ei voida yleistää koskemaan kokonaisvaltaisesti Suomen toisiokäyttöön tarkoitettujen rekistereiden toisiokäytön hyödynnettävyyteen. Muiden kansallisten toimijoiden toisiokäytön rekistereiden kehityksessä voi ilmetä rekisterikohtaisempia kehityshaasteita. Tämän tutkimuksen tarkoituksena on ollut saada kuva tämän hetken THL:n tutkimukseen käytetyimmistä toisiokäytön terveysrekistereiden tilanteesta ja niiden kehityssuunnasta. Tutkimuksessa on myös haastateltu vain yhden terveystietoa hyödyntävän yrityksen tutkijoita. Tämä rajoittaa tutkijoiden näkemykset vain kyseisen yrityksen tutkijoiden kokemuksiin haasteisiin.

Jatkotutkimuksessa olisi hyvä tutkia useamman eri tutkimusosapuolen kokemuksia Suomen kansallisten terveystietorekistereiden hyödynnettävyydestä toisiokäyttöön. Jatkotutkimuksessa voitaisiin keskittyä laajemman populaation näkemysten selvittämiseen kvantitatiivisen tutkimuksen avulla esimerkiksi kyselylomaketta hyödyntäen. Tämä avaisi laajemmin näkökulmia toisiokäytön arkkitehtuurin kehitykseen kansallisesti. Tällöin voitaisiin myös selvittää, ovatko tässä tutkimuksessa tehdyt havainnot sovellettavissa myös tutkimukseen osallistuneiden haastateltujen ulkopuolelle. Toisaalta haastattelututkimus tarjoaa laajempia löydöksiä kompleksisesta asiasta. Myös näkemykset tutkijoilta, jotka eivät ole ennen olleet tekemisissä tietolupaviranomaisen tai rekisterivirkahenkilöiden kanssa, voisivat olla arvokkaita tiedon hyödynnettävyyden tutkimuksessa.

Toinen suunta jatkotutkimukselle olisi tutkia ja vertailla useampaa eri toisiokäytön rekistereiden palvelutarjoajaa ja heidän rekistereidensä kehitystä. Tällöin olisi mahdollista vertailla palvelutarjoajien välisiä eroja kansallisesti. Yksi mielenkiintoinen tutkimusaihe olisikin vertailla, kuinka eri Pohjoismaiden kansalliset terveysrekisterit ja niiden hyödynnettävyys toisiokäyttöön eroavat toisistaan. Kyseisen tutkimuksen avulla olisi mahdollista löytää rekistereiden kehityksen eri puolia ja mahdollisesti lähentää kehitystyötä kansainvälisesti Pohjoismaiden välillä.



## Lähteet

- Adams, J., Khan, H.T.A. & Raeside, R. (2013) *Research Methods for Business and Social Science Students*. SAGE Publications, New Delhi.
- Aineistokatalogi (s.a.) <https://aineistokatalogi.fi/catalog>, haettu 2.6.2022
- Attard, J., Orlandi, F., Scerri, S. & Auer S. (2015) A systematic review of open government data initiatives. *Government Information Quarterly*, 32. 399–418.
- Attride-Stirling, J. (2001) Thematic networks: An analytic tool for qualitative research. *Qualitative Research*, 1 (3), 385–405. <https://doi.org/10.1177/146879410100100307>, haettu 2.6.2022
- Bakken, I.J., Ariansen, A.M.S., Knudsen, G.P., Johansen, K.I., & Vollset, S.E. (2020) The Norwegian Patient Registry and the Norwegian Registry for Primary Health Care: Research potential of two nationwide health-care registries. *Scandinavian Journal of Public Health*, 48 (1), 49–55.
- Berger, M.L., Lipset, C., Gutteridge, A., Axelsen K., Subedi, P. & Madigan, D. (2015) Optimizing the Leveraging of Real-World Data to Improve the Development and Use of Medicines. *Value in Health*, 18 (1), 127–130.
- Bloiland, B. & MacNeil, A. (2019) Defining & assessing the quality, usability and utilization of immunization data. *BMC Public Health*, 19, 380–388.
- Boyatzis, R.E. (1998) *Transforming qualitative information: thematic analysis and code development*. SAGE Publications, Thousand Oaks.
- Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3 (2), 77–101.
- Braun, V. & Clarke, V. (2013) Teaching thematic analysis: overcoming challenges and developing strategies for effective learning. *The Psychologist*, 26 (2), 120–123.
- Cohen, D. & Crabtree, B. (2006) Semi-structured Interviews. *Qualitative Research Guidelines Project*. <http://www.qualres.org/HomeSemi-3629.html>, haettu 10.12.2020.
- Crabtree, B. & Miller, W. (1999) Using codes and code manuals: a template for organizing style of interpretation. Teoksessa: Crabtree, B. & Miller, W. (toim.) *Doing qualitative research* (s. 163–178). Sage, Newbury Park.
- De Groot, S., Van Der Linden, N., Franken, M.G., Blommestein, H.M., Leeneman, B., Van Rooijen, E., Van Der Hoeven, J., Wouters, M.W., Westgeest, H. & Groot, C.A.U.-D. (2016) Balancing the optimal and the feasible: A practical guide for

setting up patient registries for the collection of real-world data for health care decision making based on dutch experiences. *Value Health*, 20, 627–636.

Eriksson, P. & Kovalainen, A. (2015) *Qualitative Methods in Business Research: A practical guide to social research*. Sage.

Findata (8.6.2022a) *Hinnasto*. <https://findata.fi/hinnasto>, haettu 13.6.2022

Findata (22.4.2022b) *Luvat*. <https://findata.fi/luvat/>, haettu 13.6.2022

Galletta, A. (2012) *Mastering the Semi-structured Interview and Beyond: From Research Design to Analysis and Publication*. New York University Press, New York.

Garrison, L.P., Neumann, P.J., Erickson, P., Marshall, D. & Mullins, C.D. (2007) Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report. *Value in Health*, 10 (5), 326–335.

Ghauri, P., Grønhaug, K. & Strange R. (2020) *Research methods in business studies* (5. painos). Cambridge University Press.

Gill, P., Stewart, K., Treasure, E. & Chadwick, B. (2008) Methods of data collection in qualitative research: interviews and focus groups. *British Dental Journal*, 204, 291–295.

Gilstad, H., Brattheim, B., Faxvaag, A., Nøhr, C., Villumsen, S., Reponen, J., Andreassen, H., Kangas, M., Harardóttir, G.A., Koch, S. & Hyppönen, H. (2016) Comparability, availability and use of medication ehealth services in the Nordic countries. *International Journal on Advances in Life Sciences*, 8 (1). 112–121.

Gioia, D.A., Corley, K.G., & Hamilton, A.L. (2013) Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational Research Methods*, 16 (1), 15–31.

Gliklich, R.E. & Dreyer, N.A. (2010) *Registries for Evaluating Patient Outcomes: A User's Guide* (2. painos). Rockville, MD: Agency for Healthcare Research and Quality.

Hearst, N. & Hulley, S.B. (1988) Using secondary data. Teoksessa Hulley, S. B. & Cummings, S. R. (toim.) *Designing Clinical Research* (s. 53–62). Williams & Wilkins, Baltimore.

Máchová, R., Hub, M., & Lnenicka M. (2018) Usability evaluation of open data portals. Evaluating data discoverability from a stakeholders' perspective. *Aslib Journal of Information Management*, 70 (3). 252–268.

Huovila, M. & Lähesmaa, J. (2019) *Sosiaali- ja terveystietojen toisiokäytön kokonaisarkkitehtuuri*. Sosiaali- ja terveystieteiden ministeriö.

- Janssen, M., Charalabidis, Y. & Zuiderwijk, A. (2012) Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29 (4). 258–268.
- Jiang, L., Borgida, A. & Mylopoulos, J. (2008) Towards a Compositional Semantic Account of Data Quality Attributes. Teoksessa Li, Q., Spaccapietra, S., Yu, E. & Olivé, A. (toim.) *Conceptual Modeling - ER 2008. ER 2008. Lecture Notes in Computer Science, vol 5231* (s. 55–68). Springer, Berlin.
- Kallio, H., Pietilä, A.-M., Johnson, M. & Kangasniemi, M. (2016) Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, 72 (12), 2954–2965.
- Kiger, M.E. & Varpio, L. (2020) Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher*, 42 (8), 846–854.
- King, N. (2004) Using templates in the thematic analysis of text. Teoksessa Cassell, C. & Symon, G. (toim.) *Essential guide to qualitative methods in organizational research* (s. 257–270) Sage, Lontoo.
- Kučera, J., Chlapek, D. & Nečaský, M. (2013) Open Government Data Catalogs: Current Approaches and Quality Perspective. Teoksessa Kö, A., Leitner, C., Leitold, H & Prosser, A. (toim.) *Technology-Enabled Innovation for Democracy, Government and Governance, vol 8061* (s. 152–166). Springer, Berlin.
- Laki sosiaali- ja terveystietojen tutkimus- ja kehittämiskeskukseen tilastotoimista 409/2001. Annettu Helsingissä 23.5.2001
- Laki sosiaali- ja terveystietojen toissijaisesta käytöstä 552/2019. Annettu Helsingissä 26.4.2019
- Laki Terveystietojen ja hyvinvoinnin laitoksesta 31.10.2008/668. Annettu Helsingissä 31.10.2008
- Laki terveydenhuollon valtakunnallisista henkilörekistereistä 556/1989. Annettu Helsingissä 9.6.1989
- Lehto, M. & Neittaanmäki, P. (2018) Suomen kansalliset SOTE-tiedonlähteet ja tietojen hyödyntäminen. *Informaatioteknologian tiedekunnan julkaisuja Vol 49/2018*. Jyväskylän yliopisto.
- Máchová, R. & Lněnička, M. (2017) Evaluating the Quality of Open Data Portals on the National Level. *Journal of Theoretical and Applied Electronic Commerce Research*, 12 (1), 21–41.

- Martti, T. & Viitanen, J. (2016) *Asiakas- ja potilastietojen toissijaisen käytön kokonaisarkkitehtuurin nykytila*, Atrain.
- Niemelä, M., Lähteenmäki, J. & Pajula, J. (2021) *Hyvinvointitietojen toisiokäyttö – esiselvitys*. Teknologian tutkimuskeskus VTT Oy. VTT Tutkimusraportti No. VTT-R-172176-21.
- Nøhr, C., Faxvaag, A., Tsai, C.H., Harðardóttir, G.A., Hyppönen, H., Andreassen, H.K., Gilstad, H., Jónsson, H., Reponen, J., Kaipio, J., Øvlisen, M.V., Kangas, M., Bertelsen, P., Koch, S., Villumsen, S., Schmidt, T., Vehko, T. & Vimarlund, V. (2020). *Nordic eHealth Benchmarking*. Pohjoismainen ministerineuvosto, Kööpenhamina.
- Nøhr, C., Koch, S., Vimarlund, V., Gilstad, H., Faxvaag, A., Hardardottir, G.A., Andreassen, H.K., Kangas, M., Reponen, J., Bertelsen, P., Villumsen, S. & Hyppönen, H. (2018) Monitoring and Benchmarking eHealth in the Nordic Countries. *Student Health Technological Inform*, 247, 86–90.
- Nowell, L.S., Norris, J.M., White, D.E. & Moules, N.J. (2017) Thematic analysis: striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16 (1), 1–13.
- Rannanheimo, P. & Jauhonen, H.-M. (2018) Mihin reaaliaikaisen dataa tarvitaan? Näkökulmana lääkkeiden ja lääkinnällisten laitteiden valvonta, arviointi (HTA) ja kansallinen ohjaus. *Sosiaali- ja terveystieteiden ministeriön raportteja ja muistioita*, 44/2018. Helsinki.
- Runeson, P. & Höst, M. (2009) Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14 (2), 131–164.
- Schopf, T.R., Nedrebø, B. & Hufthammer, K.O. (2019) How well is the electronic health record supporting the clinical tasks of hospital physicians? A survey of physicians at three Norwegian hospitals. *BMC Health Services Research*, 19 (1), 934-934.
- Sitra (2017) *Isaacus-esituotantohankkeet*. <https://www.sitra.fi/hankkeet/isaacus-esituotantohankkeet/#tulokset>, haettu 24.5.2022.
- Sørensen, H.T., Rasmussen, H.H., Ejlersen, E., Moller-Petersen, J.F., Hamburger, H. & Olesen, F. (1992) Epidemiology of pain requiring strong analgesics outside hospital in a geographically defined population in Denmark. *Danish Medical Bulletin*, 39, 464-67.

- Sørensen, H.T., Sabroe S. & Olsen J. (1996) A Framework for Evaluation of Secondary Data Sources for Epidemiological Research. *International Journal of Epidemiology*, 25 (2), 435–442.
- THL (28.4.2022) *Tutkimuskäyttö ja tietoluvat*. <https://thl.fi/fi/tilastot-ja-data/aineistot-ja-palvelut/tutkimuskaytto-ja-tietoluvat>, haettu 2.6.2022
- THL (3.12.2020a) *Koodistopalvelu*. <https://thl.fi/fi/web/tiedonhallinta-sosiaali-ja-terveysalalla/koodistopalvelu>, haettu 2.6.2022
- THL (3.4.2020b) *Aineistokatalogi*. <https://thl.fi/fi/tilastot-ja-data/aineistot-ja-palvelut/aineistoluettelo>, haettu 2.6.2022
- Timonen, J., Kangas, S., Kauppinen, H. & Ahonen, R. (2018) Electronic prescription anomalies: a study of frequencies, clarification, and effects in Finnish community pharmacies. *Journal of Pharmaceutical Health Services Research*, 9, 183–189.
- Varpio, L., Young, M., Uijtdehaage, S. & Paradis, E. (2020) The distinctions between theory, theoretical framework, and conceptual framework. *Academic Medicine*, 95 (7), 989–994.
- Watling, C.J., Lingard, L. (2012) Grounded theory in medical education research: AMEE guide no. 70. *Medical Teacher*, 34 (10), 850–861.
- Zainal, Z. (2007) Case study as a research method. *Jurnal Kemanusiaan*, 9, 1–6.

## Liitteet

### Liite 1 Haastattelurunko

1. Minkä terveystietorekistereiden parissa olet työskennellyt?
2. Jos mietitään tutkijan näkökulmasta tietopyyntöprosessia, niin mitä asioita prosessi vaatii tutkijalta?
3. Onko tietämyksesi mukaan tietopyyntöprosessissa joitain vaiheita, jotka vaativat kehitystä?
4. Minkälaiset asiat vaikuttavat negatiivisesti tietopyyntöprosessiin?
5. Kuinka haastavaa tutkimusluvan saaminen on?
6. Kuinka saatavilla rekistereiden metatiedot ovat?
7. Miten datan välitys toimii datan lähteestä alkaen THL:le?
8. Kuinka kauan datan toimittamisessa sen syntyhetkestä lähtien tutkijalle yleensä menee?
9. Minkälaisia datan luovutustapoja on käytössä?
10. Kuinka kattavaa data on tuntemissasi rekistereissä?
11. Kuinka datan siirtyminen rekistereistä ja järjestelmistä toisiin vaikuttaa datan eheyteen?
12. Kuinka dataa prosessoidaan THL:län sisäisissä järjestelmissä?
13. Missä formaatissa data toimitetaan?
14. Millä periaatteella yksilöityä ja aggregoitua dataa jaetaan?
15. Kuinka datan otsikoinnit määritellään?
16. Tapahtuuko datan otsikoinnin ja muodon päivittämistä usein?
17. Miten ja kuinka usein dataa tarkistetaan eri rekistereissä?
18. Kuinka yleistä on, että datasta löytyy virheitä?
19. Mitkä vaiheet datan muodostumisessa ja siirtymisessä koet semmoiseksi, missä näet, että virheitä tietoihin voi syntyä todennäköisimmin?