Tampereen yliopisto

Kiia Keski-Sikkilä

TIETOTURVAHAASTEET TERVEYDEN-HUOLLON SUURISSA DATAKOKONAI-SUUKSISSA

Kirjallisuuskatsaus terveydenhuollon big datan tietotur-vallisuushaasteisiin

Johtamisen ja talouden tiedekunta
Kandidaatin tutkielma
Elokuu 2025

TIIVISTELMÄ

Kiia Keski-Sikkilä: Tietoturvahaasteet suurissa datakokonaisuuksissa terveydenhuollossa

Kandidaatin tutkielma

Tampereen yliopisto

Tietojohtamisen tutkinto-ohjelma

Elokuu 2025

Terveydenhuollossa syntyy ja hyödynnetään valtavia määriä arkaluonteista dataa, kuten potilastietoja, genomidataa ja IoT-laitteiden mittaustuloksia, joiden turvallinen käsittely on kriittistä potilasturvallisuuden ja yksityisyyden suojan kannalta. Tässä kandidaatintutkielmassa tarkastellaan suurten datakokonaisuuksien tietoturvahaasteita terveydenhuollon kontekstissa. Tutkielma toteutettiin systemaattisena kirjallisuuskatsauksena, ja sen tavoitteena oli tunnistaa keskeisimmät tietoturvahaasteet terveydenhuollon suurissa datakokonaisuuksissa ja arvioidaan nykyisten sääntelykehyksien vaikutuksia.

Keskeisiksi haasteiksi nousivat kyberuhat, tietovuodot, anonymisoinnin riittämättömyys sekä teknologiset ja organisatoriset haavoittuvuudet. Tutkielmassa analysoidaan myös EU:n GDPR-asetuksen ja Suomen kansallisen lainsäädännön roolia tietoturvan varmistamisessa. Johtopäätöksissä korostetaan muun muassa henkilöstön koulutuksen, pääsynhallinnan ja teknisten suojaustoimien merkitystä. Tutkielma tarjoaa ajankohtaisen katsauksen siihen, miten terveydenhuollon organisaatiot voivat kehittää tietoturvakäytäntöjään suurten datamäärien hallinnassa.

Avainsanat: data, suuret datakokonaisuudet, tietoturva, tietoturvallisuus, GDPR, terveydenhuolto, tietoturvahaaste

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla

ABSTRACT

Kiia Keski-Sikkilä: Data security challenges within large data sets in healthcare

Bachelor’s thesis

Tampere University

Information Knowledge Management

August 2025

Healthcare generates and utilizes enormous amounts of sensitive data, such as patient information, genomic data, and measurement results from IoT devices, the secure processing of which is critical for patient safety and privacy protection. This bachelor's thesis examines the security challenges of big data in the context of healthcare. The study was conducted as a systematic literature review, and its aim was to identify the most important security challenges in big data in healthcare and assess the impacts of current regulatory frameworks.

The key challenges were cyber threats, data leaks, insufficient anonymization, and technological and organizational vulnerabilities according to the literature review. The study also analyzes the role of the EU GDPR regulation and Finnish national legislation in ensuring data security. The conclusions emphasize, among other things, the importance of personnel training, access control, and technical protection measures. The study provides a timely overview of how healthcare organizations can develop their data security practices in managing large amounts of data.

Keywords: data, big data, information security, data security, GDPR, healthcare, information security challenge

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TEKOÄLYN KÄYTTÖ OPINNÄYTTEESSÄ

Opinnäytteessäni on käytetty tekoälysovelluksia:

☐ Ei

☒ Kyllä

Ilmoitukseni mukaan olen käyttänyt opinnäytteessäni tutkielmaprosessin aikana seuraavia tekoälysovelluksia:
Microsoft Copilot, Scopus AI

Tekoälysovellusten nimet ja versiot:
GPT-4

Käyttötarkoitus:
Työn rakenteen parantelussa, joidenkin lauseiden uudelleen muotoilussa, alustavassa lähteiden etsinnässä

Osiot, joissa tekoälyä on käytetty:
Luvut 3, 4 ja 5

Olen tietoinen siitä, että olen täysin vastuussa koko opinnäytteeni sisällöstä, mukaan lukien osat, joissa on hyödynnetty tekoälyä, ja hyväksyn vastuun mahdollisista eettisten ohjeiden rikkomuksista.

SISÄLLYSLUETTELO

1. JOHDANTO 2
1.1 Tutkimuksen tausta 2
1.2 Tutkimusongelma ja tavoitteet 3
1.3 Tutkimuksen rakenne 4
2. TUTKIMUSMENETELMÄT JA -AINEISTO 5
2.1 Tutkimusmenetelmät 5
2.2 Tutkimusaineisto 5
3. TIETOTURVALLISUUS 8
3.1 Tietoturvan keskeiset periaatteet 8
3.2 Tietoturvahaasteet 9
3.3 Tietoturvariskien hallinta 10
4. BIG DATA JA SEN HALLINTA TERVEYDENHUOLLON KONTEKSTISSA 12
4.1 Suuret datakokonaisuudet 12
4.2 Terveydenhuollon datan erityispiirteet 13
5. TIETOTURVALLISUUS TERVEYDENHUOLLON BIG DATASSA 15
5.1 Tietoturvan merkitys potilasturvallisuudelle ja terveydenhuollon toiminnalle 15
5.2 Euroopan unionin asetukset 16
5.3 Suomen lainsäädäntö 17
6. JOHTOPÄÄTÖKSET 18
6.1 Yhteenveto 18
6.2 Käytännön suositukset tietoturvan kehittämiseksi 19
6.3 Tutkimuksen rajoitukset ja jatkotutkimusaiheet 19
LÄHTEET 21
LAIT, ASETUKSET JA STANDARDIT 25

LYHENTEET JA MERKINNÄT

AI	Artificial Intelligence
CIA	Confidentiality, Integrity, Availability
EU	Euroopan unioni
GDPR	General Data Protection Regulation
IoT	Internet of Things
MFA	Multi-Factor Authentication
EKG	Elektrokardiogrammi (sydänsähkökäyrä)

1. JOHDANTO

Terveydenhuollossa tuotetaan ja hyödynnetään valtavia määriä dataa erilaisista lähteistä. Sähköiset potilastiedot, lääketieteelliset kuvantamistulokset, laboratoriotulokset, genomidata sekä IoT-laitteiden keräämät mittaustiedot muodostavat laajan datakokonaisuuden, jota käytetään potilaiden hoidon tukena, lääketieteellisissä tutkimuksissa ja terveydenhuollon palveluiden kehittämisessä.

Vaikka terveydenhuollon big data eli suuret datakokonaisuudet tarjoavat merkittäviä mahdollisuuksia hoidon tehostamiseen ja ennakoivaan analytiikkaan, sen hyödyntämiseen liittyy myös tietoturvahaasteita. Arkaluonteisten potilastietojen hallinta edellyttää tiukkoja tietoturvatoimia, sillä väärinkäytökset, kyberhyökkäykset ja tietovuodot voivat vaarantaa sekä yksilön yksityisyyden että terveydenhuollon toimintakyvyn. Suomessa erityisesti Vastaamon tietomurto (Tietosuojakeskus, 2020) nosti esiin terveydenhuollon tietoturvan haavoittuvuudet ja herätti laajaa keskustelua siitä, miten potilastietoja tulisi suojata entistä paremmin.

1.1 Tutkimuksen tausta

Terveydenhuollossa käsitellään ja hyödynnetään laajoja tietomassoja, joihin sisältyy usein arkaluonteisia tietoja, kuten potilaiden henkilökohtaisia tietoja. Suurten datakokonaisuuksien, eli niin sanotun big datan, analysointi on keskeinen osa modernia data-analytiikkaa. Big data tarkoittaa valtavia, nopeasti kasvavia ja monimutkaisia tietomääriä, jotka vaativat erityisiä tallennus-, käsittely- ja analyysimenetelmiä (Badman & Kosinski, 2024). Suurten datakokonaisuuksien hyödyntäminen tuo mukanaan merkittäviä mahdollisuuksia terveydenhuoltoalan liiketoiminnalle, potilaiden hoitotyöhön, tutkimukselle ja päätöksenteolle, mutta samalla se asettaa uusia haasteita tietoturvalle.

Tässä tutkielmassa tarkastellaan terveydenhuollon suurten datakokonaisuuksien tietoturvahaasteita keskittyen erityisesti tietovuotoihin, lainsäädännön vaatimuksiin sekä teknologisiin ja organisatorisiin ratkaisuihin. Tämä aihe tarjoaa laajan ja ajankohtaisen tutkimuskentän, jossa voidaan tarkastella tietoturvan ja suurten tietomassojen monimutkaista suhdetta. Tutkimuksen tavoitteena on selvittää, mitkä ovat keskeisimmät tietoturvahaasteet, miten nykyinen sääntely vaikuttaa terveydenhuollon tietoturvaan ja millä keinoilla tietoturvaa voitaisiin parantaa.

1.2 Tutkimusongelma ja tavoitteet

Kandidaatintyön lähtökohtana on tekijän kiinnostus tietoturvasta ja siitä mitä on otettava huomioon tietoturvallisuuden näkökulmasta suuria datakokonaisuuksia käsiteltäessä. Kiinnostus on saanut pohjaa yliopiston kursseilta, työelämästä ja omasta henkilökohtaisesta harrastuneisuudesta. Aihe rajautui terveydenhuoltoalalle, sillä aihetta on käsitelty useilla kursseilla harkkatöiden ja esimerkkien muodossa, ja on hyvin ajankohtainen.

Kandidaatintyön tutkimusongelmana on tietoturvahaasteet suurten datakokonaisuuksien käsittelyssä terveydenhuollossa. Käytännössä tämä voisi auttaa terveydenhuoltoalan organisaatioita ja niiden suuria datakokonaisuuksia käsitteleviä työntekijöitä suojaamaan herkkää ja arkaluontoista dataansa, varmistamaan liiketoimintansa jatkuvuuden ja säilyttämään potilaiden luottamuksen.

Päätutkimuskysymyksenä kandidaatin työssä on: *mitkä ovat keskeisimmät tietoturvahaasteet suurten datakokonaisuuksien hallinnassa?* Päätutkimuskysymystä tukevia alatutkimuskysymyksiä on listattuna taulukossa 1.

Taulukko 1: Tutkimuskysymykset

Päätutkimuskysymys	Mitkä ovat keskeisimmät tietoturvahaasteet terveydenhuollon suurten datakokonaisuuksien hallinnassa?
Alatutkimuskysymys 1	Mitä suuret datakokonaisuudet ovat ja mitä niiden hallintaan kuuluu?
Alatutkimuskysymys 2	Millaisia ongelmia mahdolliset tietoturvahaasteet aiheuttavat terveydenhuollon datakokonaisuuksissa, ja miten niitä voidaan ehkäistä?
Alatutkimuskysymys 3	Miten nykyinen sääntely ja lainsäädäntö vaikuttavat terveydenhuollon datan tietoturvaan?

Tutkimuksen tavoitteena on kartoittaa keskeisimpiä tietoturvahaasteita suurten datakokonaisuuksien käsittelyssä sekä arvioida nykyisten tietoturvamenetelmien tehokkuutta ja syventyä siihen millaisia uhkia tietovuodot aiheuttavat terveydenhuollossa.

Aihetta on rajattu tarkastelemaan erityisesti terveydenhuollon datakokonaisuuksia. Tässä kirjallisuuskatsauksessa ei käsitellä pienimuotoisen datan tietoturvaa tai data-analytiikan laatuun liittyviä kysymyksiä.

1.3 Tutkimuksen rakenne

Tutkielma jakautuu kuudesta pääluvusta, jotka etenevät loogisesti tutkimusongelman jäsentämisestä kohti johtopäätöksiä ja kehitysehdotuksia. Ensimmäisessä luvussa esitellään tutkimuksen taustat, rajaukset, määritellään tutkimusongelma ja tavoitteet sekä hahmotellaan tutkielman rakenne. Toisessa luvussa kuvataan tutkimusmenetelmät ja aineiston valinta.

Kolmas luku käsittelee tietoturvallisuutta yleisellä tasolla. Luvussa esitellään tietoturvan keskeiset periaatteet. kuten luottamuksellisuus, eheys ja saatavuus, sekä tarkastellaan tietoturvahaasteita ja riskienhallinnan keinoja. Neljäs luku keskittyy big datan käsitteeseen ja sen hallintaan terveydenhuollon kontekstissa. Siinä määritellään suurten datakokonaisuuksien ominaispiirteet ja terveydenhuollon datan erityisluonne.

Viidennessä luvussa tietoturvaa tarkastellaan nimenomaan terveydenhuollon big datan näkökulmasta. Luvussa käsitellään tietoturvan merkitystä potilasturvallisuudelle, teknologisia ja organisatorisia ratkaisuja sekä Suomen ja EU:n lainsäädännön vaikutuksia. Kuudes luku kokoaa yhteen tutkimuksen keskeiset havainnot, esittää käytännön suosituksia terveydenhuollon tietoturvan kehittämiseksi ja tuo esiin tutkimuksen rajoituksia sekä jatkotutkimusaiheita. Viimeisessä, eli kuudennessa, luvussa kootaan kirjallisuuskatsauksen havainnot yhteen, ehdotetaan kehitysehdotuksia ja annetaan jatkotukimusaiheita.

2. TUTKIMUSMENETELMÄT JA -AINEISTO

2.1 Tutkimusmenetelmät

Tutkimus toteutetaan kirjallisuuskatsauksena Finkin (2014) seitsemänvaiheisen systemaattisen kirjallisuuskatsauksen mallin mukaisesti. Ensimmäisessä vaiheessa asetetaan tutkimuskysymys, jonka jälkeen valitaan tietokannat ja kirjallisuus (Fink, 2014). Näitä esitellään tässä tutkimussuunnitelmassa myöhemmin. Kolmannessa vaiheessa valitaan hakusanoja ja -lauseita (Fink, 2014), joita on myös hieman pohdittu tässä tutkimussuunnitelmassa. Neljäntenä valitaan käytännön hakukriteerit ja tehdään metodologinen rajaus. Toiseksi viimeisessä vaiheessa tehdään itse kirjallisuus katsaus ja viimeisenä syntetisoidaan kirjallisuuskatsauksen tulokset. (Fink, 2014)

1. Tutkimuskysymyksen asettaminen
2. Kirjallisuuden ja tietokantojen valitseminen
3. Hakutermien valitseminen
4. Haku/rajauskriteerien valitseminen
5. Metodologinen rajaaminen
6. Kirjallisuuskatsauksen tekeminen
7. Tulosten syntetisointi (Fink, 2014)

Ennen varsinaisen tutkimuksen aloittamista tehtiin aiheanalyysi, jonka avulla perehdyttiin aiheeseen etukäteen. Aiheanalyysissä käsiteltiin tutkimusongelmaa ja tutkimusaineistoa alustavasti.

2.2 Tutkimusaineisto

Pääasiallisina tiedonhakualustoina tulee olemaan Tampereen yliopiston kirjaston tietokanta Andor sekä kirjaston libguides.tuni.fi sivustolta löytyvät sekä tietojohtamisen, tietotekniikan ja informaatiotutkimuksen aloilla käytetyt tietokannat. Lisäksi aineistoa haetaan myös muista tieteellisen kirjallisuuden hakupalveluista, esimerkiksi Scopus. Lähtökohtaisesti käytettävän tutkimusaineisto on akateemista, vertaisarvioitua kirjallisuutta, mutta tutkielmassa on käytetty myös luotettavien tai arvostettujen lähteitä. Tietotekni-

kan osa aineistoista on sellaisia, joita ei ole vertaisarvioitu vielä, mutta niissä on kaikkien tuorreimat tutkimukset, joista voi olla hyötyä tutkielman aiheen takia. Hakusanoina toimii mm. seuraavat hakusanat: information security, cyber security, big data, health care sekä muut vastaavat hakulausekkeet. Hakusanojen avulla löytyy yksittäin todella paljon kirjallisuutta, esimerkiksi hakusanalla ”cyber security” löytyy Andorista 138 117 tulosta. Suuren aineistomäärän myötä on siis tärkeää löytää sopivin aineisto rajaamalla hakuja riittävästi esimerkiksi aineistotyypin perusteella. Kuitenkin tekemällä kunnon hakulausekkeet tuloksia on huomattavasti vähemmän. Taulukossa 2 on esitelty tutkimuksen keskeisimpiä hakulausekkeita ja niiden tuottamien tulosten määriä.

Taulukko 2: Hakulausekkeet sekä niiden tuottamien tulosten määrät.

Hakulauseke	Andor	Scopus
“cyber security” AND ("big data” OR “large datasets”)	3 503	1 691
("big data” OR “large datasets”) AND "security challenges"	908	449
("data breach" OR “data leakage” OR “information breach”) AND ("big data” OR “large datasets”)	878	530
"healthcare" AND ("big data” OR “large datasets”) AND ("information security" OR “cyber security”)	535	211
"EU data protection" AND "healthcare" AND ("big data” OR “large datasets”)	503	150

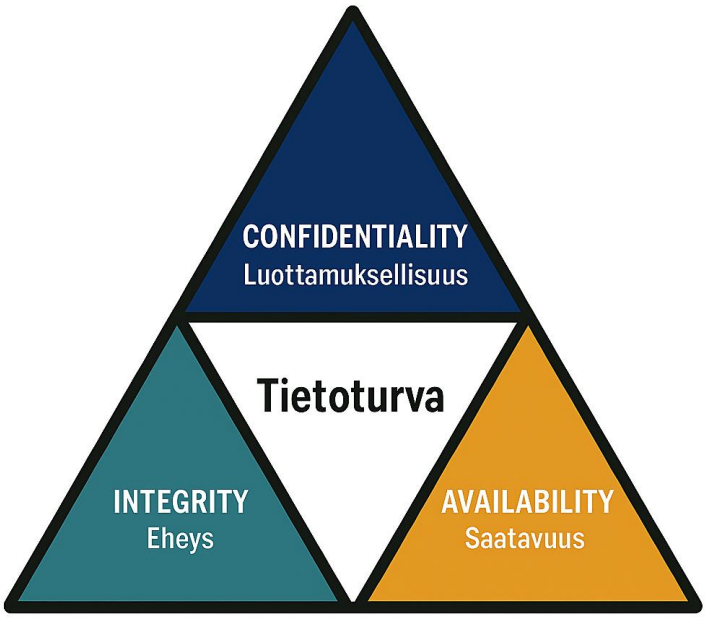
Hakutulosten määrissä ilmeni selkeitä eroja eri tietokantojen välillä. Andor-tietokanta tuotti huomattavasti enemmän osumia kuin Scopus, mikä saattaa johtua Andorin laajemmasta aineistopohjasta ja sen sisältämistä kotimaisista julkaisuista. Scopus puolestaan keskittyy kansainvälisiin, vertaisarvioituihin artikkeleihin, mikä rajaa sen tulosjoukkoa tiukemmin. Vaikka Scopuksen hakutulokset olivat määrällisesti vähäisempiä, ne osoittautuivat laadullisesti erittäin hyödyllisiksi tutkielman kannalta. Suomenkielisiä lähteitä löytyi vain rajallisesti, mutta nekin käytiin läpi, sillä ne tarjosivat kontekstuaalista ymmärrystä erityisesti kansallisen lainsäädännön ja käytäntöjen osalta. Lisäksi hakusanojen synonyymien ja lyhenteiden huomioiminen oli tärkeää, jotta aineistoa saatiin mahdollisimman kattavasti ja monipuolisesti. Tämä korostui erityisesti englanninkielisessä kirjallisuudessa, jossa terminologia voi vaihdella esimerkiksi "data protection", "information security" ja "cybersecurity" välillä.

Lähteiden käsittelyssä etenin temaattisen analyysin kautta, jossa jaottelin aineiston tutkimuskysymysten mukaisiin kokonaisuuksiin. Ensin luin läpi valitut lähteet ja arvioin niiden relevanssia suhteessa tutkielman tavoitteisiin. Tämän jälkeen ryhmittelin lähteet keskeisiin teemoihin, kuten kyberuhat, anonymisointi, lainsäädäntö ja teknologiset ratkaisut, jotta tutkimuksen rakenne tukisi selkeästi päätutkimuskysymykseen vastaamista. Erityistä huomiota kiinnitin lähteiden ajankohtaisuuteen ja siihen, miten ne käsittelivät nimenomaan terveydenhuollon kontekstia. Lähteiden avulla muodostin tutkielman teoreettisen viitekehyksen ja perustelin keskeiset käsitteet, kuten tietoturva, big data ja riskienhallinta, sekä analysoin niiden yhteyksiä tutkielman aihepiiriin.

3. TIETOTURVALLISUUS

3.1 Tietoturvan keskeiset periaatteet

Tietoturvalla tarkoitetaan niitä periaatteita ja organisaation laajuisia toimenpiteitä, joilla pyritään suojaamaan tietojärjestelmiä, tietoa ja siihen liittyviä prosesseja väärinkäytöksiltä ja tuhoutumiselta. Se muodostuu teknisistä, hallinnollisista ja organisatorisista keinoista, joiden yhteisenä tavoitteena on varmistaa tietoturvan keskeisten periaatteiden toteutuminen. (Bhaskar & Kapoor, 2013) Tietoturvan keskeiset periaatteet voidaan tiivistää kolmeen pääperiaatteeseen, joita ovat tiedon luottamukselisuus, eheys ja saatavuus. (Suomi.fi, 2025)



Kuva 1. CIA-kolmio (Keski-Sikkilä, 2025)

Kuva 1 kuvaa CIA-kolmiota, joka on tietoturvan perusmalli. Se koostuu tietoturvan kolmesta pääperiaatteesta: luottamuksellisuudesta (Confidentiality), eheydestä (Integrity) ja saatavuudesta (Availability) (Suomi.fi 2025). CIA-malli toimii perustana tietoturvan suunnittelulle, arvioinnille ja toteutukselle eri ympäristöissä, kuten yrityksissä, julkisissa organisaatioissa ja terveydenhuollossa. Luottamuksellisuus tarkoittaa, että tieto on suo-

jattu luvattomalta pääsyltä. Vain valtuutetut henkilöt tai järjestelmät saavat käsitellä tai nähdä tiedon. (Shabtai et al. 2012) Tämä voi tarkoittaa esimerkiksi salasanalla suojattua järjestelmää, salattua tiedonsiirtoa (esim. HTTPS) tai käyttöoikeuksien hallintaa. Shabtai et al. (2012) kertovat eheyden tarkoittavan tiedon tarkkuuden ja luotettavuuden säilyttämistä. Tieto ei saa muuttua vahingossa tai tahallisesti ilman asianmukaista valvontaa. Eheys voi sisältää esimerkiksi versionhallintaa, tarkistussummia tai digitaalista kirjoitusta. Lisäksi Shabtai et al. (2012) sanovat saatavuuden tarkoittavan tiedon ja järjestelmien käytettävyyttä silloin kun niitä tarvitaan. Tämä edellyttää esimerkiksi varmuuskopiointia, palvelinten redundanssia ja suojautumista palvelunestohyökkäyksiltä. (Shabtai et al. 2012)

Tietoturvaan kuuluvat keskeisesti myös seuraavat käsitteet: skaalautuvuus, anonymisointi ja pseudonymisointi. Skaalautuvuudella tarkoitetaan järjestelmän tai tietoturvaratkaisun kykyä mukautua kasvaviin tietomääriin ilman, että suorituskyky kärsii (Orellana et al. 2019). Terveydenhuollossa, jossa datamäärät voivat olla huomattavia, skaalautuvuuteen liittyy erityisiä haasteita. Esimerkiksi pilvipalveluiden käyttö mahdollistaa suurten tietomassojen hajauttamisen eri palvelimille, mutta samalla se voi lisätä tietovuotojen riskiä, mikäli suojaus ei ole riittävällä tasolla. (Koumaditis, 2015) Lisäksi tietoturvan ja käytettävyyden välinen jännite on jatkuva haaste: mitä tiukemmat suojaustoimet, sitä suurempi riski, että järjestelmien käytettävyys heikkenee – ja tämä voi puolestaan hidastaa kliinisiä prosesseja ja vaikuttaa potilastyöhön.

Anonymisointi tarkoittaa prosessia, jossa yksilöivä tieto poistetaan tai muunnetaan siten, ettei yksittäistä henkilöä voida enää tunnistaa suoraan tai epäsuorasti. Anonymisointi on erityisen tärkeää suurten tietomassojen käsittelyssä, joissa yksityisyyden suojaaminen on keskeinen eettinen ja oikeudellinen kysymys. (Mimoto et al., 2017) Pseudonymisointi puolestaan viittaa henkilötietojen käsittelyyn siten, että tiedot voidaan yhdistää alkuperäisiin tietoihin vain erillisen, suojatun lisäinformaation avulla (Basdekis et al., 2023). Vaikka sekä anonymisointi että pseudonymisointi ovat tehokkaita keinoja tietosuojan vahvistamiseksi, ne eivät täysin poista uudelleenidentifioinnin riskiä etenkään tilanteissa, joissa anonymisoitua dataa yhdistetään muihin tietolähteisiin (Mimoto et al., 2017; Basdekis et al., 2023).

3.2 Tietoturvahaasteet

Tietoturvan periaatteiden ymmärtäminen luo perustan, mutta käytännön tasolla terveydenhuollon tietojärjestelmät kohtaavat monia konkreettisia ja jatkuvasti kehittyviä uhkia. Näiden haasteiden tunnistaminen ja jäsentäminen on keskeistä, jotta tietoturvaa voi-

daan kehittää systemaattisesti ja kestävästi. Terveydenhuollon tietojärjestelmiin kohdistuvat tietoturvahaasteet ovat monimuotoisia ja lisääntyneet merkittävästi viime vuosina. Haasteet voivat olla teknologisia, organisatorisia tai inhimillisiä, ja niiden vaikutukset ulottuvat potilasturvallisuudesta organisaation maineeseen ja taloudelliseen tilanteeseen. Ne liittyvät erityisesti kyberhyökkäyksiin, tietovuotoihin, sisäisiin uhkiin sekä teknologian nopeasta kehityksestä johtuviin riskeihin. (Bogdan, 2024)

Kyberhyökkäykset ovat yksi merkittävimmistä ulkoisista uhkista. Ne voivat ilmetä esimerkiksi tietomurtoina, palvelunestohyökkäyksinä tai kiristysohjelmien (ransomware) muodossa. Tietovuoto puolestaan tarkoittaa tilannetta, jossa arkaluontoiset tiedot päätyvät luvattomille tahoille joko tahattomasti tai tahallisesti. Vuotojen taustalla voivat olla hakkerointi, inhimilliset virheet tai sisäpiiririkokset. Näiden seuraukset voivat olla vakavia: potilasturvallisuus voi vaarantua, organisaatio kärsiä mainehaittaa tai kohdata huomattavia taloudellisia menetyksiä. (Prajapat et al., 2024)

Myös sisäiset uhat muodostavat merkittävän tietoturvariskin. Näihin kuuluvat muun muassa työntekijöiden huolimattomuus, liian laajat käyttöoikeudet ja puutteellinen tietoturvakoulutus. Inhimillisten virheiden osuus tietoturvaloukkauksista on huomattava, ja siksi organisaation sisäinen tietoturvakulttuuri on keskeisessä asemassa riskien hallinnassa. (Ahmed et al., 2020)

Yhä useammin myös anonymisoinnin riittämättömyys on noussut esiin ongelmana. Vaikka tiedot olisi muodollisesti anonymisoitu, ne voidaan joskus yhdistää muihin tietolähteisiin ja näin yksilöidä uudelleen (Pirha, 2024). Tämä tuo erityisiä haasteita erityisesti suurten datakokonaisuuksien käsittelyyn, joissa tietosuoja ja tietoturva limittyvät tiiviisti.

3.3 Tietoturvariskien hallinta

Tietoturvariskien hallinta muodostaa keskeisen osan organisaation laajempaa riskienhallintastrategiaa (Rajathi & Rukmani, 2023). Kyse ei ole vain teknisestä suojauksesta, vaan järjestelmällisestä prosessista, jossa tunnistetaan, arvioidaan ja käsitellään uhkia, jotka voivat vaarantaa tiedon luottamuksellisuuden, eheyden tai saatavuuden. Tavoitteena on minimoida haitallisten tapahtumien todennäköisyys ja vaikutukset sekä varmistaa toiminnan jatkuvuus myös poikkeustilanteissa olipa kyseessä julkinen palvelu tai yksityinen liiketoiminta. Riskienhallinnassa pyritään löytämään tasapaino turvallisuuden ja toiminnan tehokkuuden välillä. Täydellinen riskittömyys ei siis ole realistinen tavoite, mutta hallittu riskitaso on (Rajathi & Rukmani, 2023).

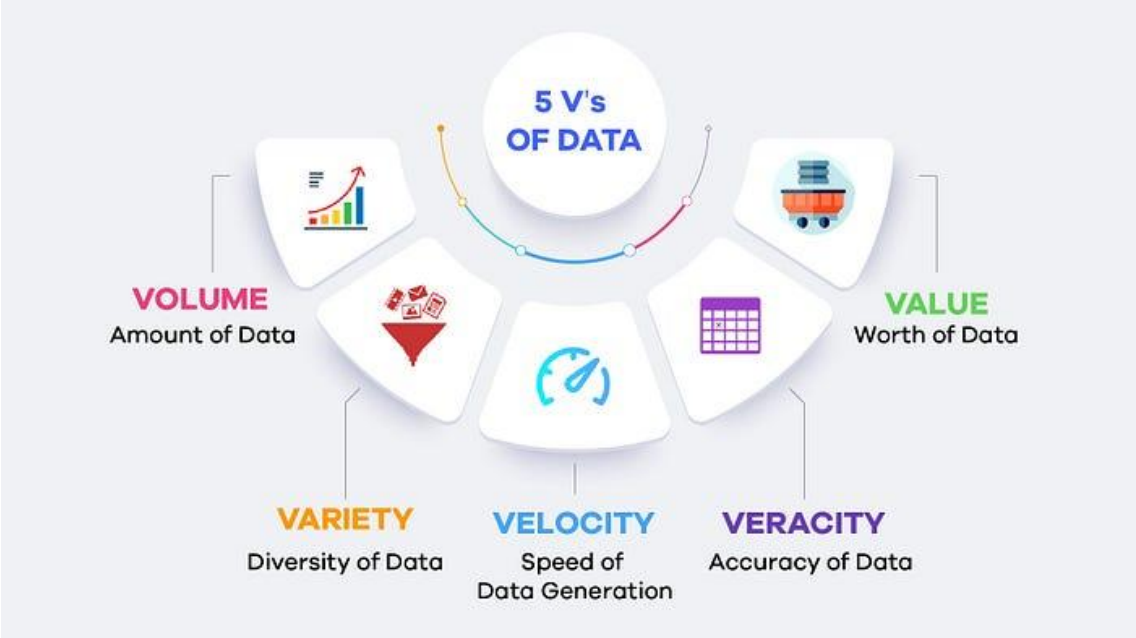
Riskienhallintaprosessi rakentuu neljästä keskeisestä vaiheesta: riskien tunnistaminen, arviointi, hallinta ja jäännösriskien hyväksyminen (Tiedonhallintalautakunta, 2021). Ensimmäisessä vaiheessa kartoitetaan, millaisia uhkia organisaation tietoihin ja tietojärjestelmiin kohdistuu. Tämä voi sisältää niin ulkoisia kuin sisäisiä riskejä. Seuraavaksi arvioidaan tunnistettujen riskien todennäköisyys ja mahdollinen vaikutus. Arvioinnin pohjalta voidaan priorisoida toimenpiteet. Kolmannessa vaiheessa päätetään konkreettisista toimista riskien pienentämiseksi tai poistamiseksi. Toimenpiteet voivat olla teknisiä, organisatorisia tai koulutuksellisia. Lopuksi tunnustetaan ne riskit, joita ei ole tarkoituksenmukaista poistaa esimerkiksi kustannusten tai toiminnallisten vaikutusten vuoksi. Näiden kanssa on opittava elämään hallitusti. Jäännösriski viittaa siis siihen osaan riskeistä, joka jää jäljelle hallintatoimien jälkeen. (Tiedonhallintalautakunta, 2021)

Suomessa tietoturvariskien hallintaa ohjaavat erityisesti Tiedonhallintalaki (906/2019) sekä EU:n yleinen tietosuoja-asetus (GDPR). Näiden säädösten mukaisesti rekisterinpitäjien, erityisesti julkisen hallinnon, on huolehdittava siitä, että henkilötietojen käsittely tapahtuu turvallisesti ja että riskit on huomioitu koko tietojärjestelmien elinkaaren ajan. Tiedonhallintalain 13 § korostaa, että tietojärjestelmien turvallisuus on varmistettava suunnittelusta aina käytöstä poistoon saakka. Tämä tarkoittaa, että tietoturva ei voi olla kertaluonteinen toimenpide, vaan sen on oltava jatkuva osa järjestelmien ja prosessien ylläpitoa. Lain 5 § edellyttää myös tiedonhallintamallien laatimista. Tiedonhallintamalli toimii dokumentoituna kokonaisuutena, jossa kuvataan organisaation tietoaineistot, tietovarannot, järjestelmien käyttötarkoitukset, tietojen säilytysajat sekä tietoturvatoimenpiteet ja tiedonsiirtotavat. Ajantasainen tiedonhallintamalli parantaa näkyvyyttä organisaation tietojärjestelmäkokonaisuuteen ja tukee riskien arviointia sekä hallintaa. (Tiedonhallintalautakunta, 2021) Lainsäädäntö tarjoaa siis selkeät raamit, mutta käytännön toteutus vaatii systemaattista työskentelyä, organisaation johdon sitoutumista ja koko henkilöstön osallistumista.

4. BIG DATA JA SEN HALLINTA TERVEYDENHUOLLON KONTEKSTISSA

4.1 Suuret datakokonaisuudet

Big data eli suuret datakokonaisuudet tarkoittavat suuria, monimuotoisia ja nopeasti kasvavia tietomääriä, joiden analysointi edellyttävät kehittyneitä tietoteknisiä menetelmiä (Hung, 2016). Suuriin datakokonaisuuksiin liittyy viisi keskeistä ominaisuutta, joita voidaan hyödyntää kontekstista riippumatta. Nämä käsitteet auttavat ymmärtämään, miksi big data eroaa perinteisestä datankäsittelystä ja millaisia haasteita ja mahdollisuuksia se tuo mukanaan. Big datan 5 V:tä ovat määrä (eng. volume), nopeus (eng. velocity), monimuotoisuus (eng. variety), luotettavuus (eng. veracity) ja arvo (eng. value) (Badman & Kosinski, 2024).



Kuva 2. Big datan 5 V:tä. (Excelsior, 2022)

Ensimmäinen V, määrä (eng. volume), viittaa valtavien tietomäärien käsittelyyn (Badman & Kosinski, 2024). Esimerkiksi terveydenhuollossa sähköisiin potilastietojärjestelmiin tallennetaan valtavasti tietoa sairaalalaitteiden, laboratoriotutkimusten ja terveysteknologian avulla kerätystä datasta (Talman, 2024). Suomessa esimerkiksi Kansalli-

nen terveysarkisto Kanta tallentaa miljoonia potilastietoja, kuten reseptejä ja laboratoriotuloksia, mikä edellyttää tehokasta datan hallintaa ja suojaamista.

Toinen V, nopeus (eng. velocity), korostaa datan jatkuvaa tuottamista ja analysointia reaaliajassa tai lähes reaaliajassa (Badman & Kosinski, 2024). Tämä on erityisen tärkeää kriittisissä hoitotilanteissa, kuten tehohoidossa, jossa potilaan elintoimintoja on seurattava tauotta (Reijula, 2007). Etämonitorointi on esimerkki siitä, kuinka terveydenhuollossa käytetään reaaliaikaisesti päivittyvää dataa. Esimerkiksi potilaan EKG- tai verenpainemittauksia voidaan seurata etänä, jolloin hälytysjärjestelmä voi varoittaa lääkäreitä mahdollisista riskeistä. (Lumikari et al., 2021; Reijula, 2007)

Monimuotoisuus (eng. variety), eli kolmas V, kuvaa datan syntymistä monessa eri muodossa (Badman & Kosinski, 2024). Osa tiedosta on rakenteellista, kuten laboratoriotulokset ja lääkemääräykset, osa puolirakenteellista, kuten lääkäreiden kirjoittamat potilaskertomukset, ja osa täysin rakenteetonta, kuten röntgen- ja MRI-kuvat (Eastwood, 2023). Tämä tekee tietojen yhdistämisestä ja analysoinnista haastavaa, mutta samalla se voi mahdollistaa entistä tarkempia diagnooseja ja yksilöllistettyä hoitoa.

Neljäs V, luotettavuus (eng. veracity), viittaa datan laatuun ja oikeellisuuteen (Badman & Kosinski, 2024). Terveydenhuollon datan on oltava mahdollisimman virheetöntä ja kattavaa, sillä epätarkka tai puutteellinen tieto voi johtaa vakaviin seurauksiin, kuten väärään diagnoosiin tai hoitovirheisiin (Lääkäriliitto, 2021). Potilastiedot saattavat sisältää virheitä, esimerkiksi tietojen kirjausvaiheessa tai eri järjestelmien välisissä siirroissa, mikä korostaa tietojen validoinnin ja lähteiden tarkistamisen merkitystä.

Viimeinen V, arvo (eng. value), määrittää datan hyödyntämisen perimmäisen tavoitteen (Badman & Kosinski, 2024). Terveydenhuollossa big dataa voidaan käyttää esimerkiksi sairauksien ennustamiseen, hoitoprosessien tehostamiseen ja yksilöllisen lääketieteen kehittämiseen (Awrahman et al., 2022). Esimerkkinä tästä ovat tekoälyyn pohjautuvat diagnostiikkasovellukset, jotka analysoivat röntgenkuvia ja tunnistavat syövän varhaisessa vaiheessa (Sahoo et al., 2025). Tämä voi parantaa hoitotuloksia merkittävästi ja vähentää terveydenhuollon kustannuksia.

4.2 Terveydenhuollon datan erityispiirteet

Terveydenhuollon suurilla datakokonaisuuksilla tarkoitetaan laajoja tietomassoja, joita kerätään ja käsitellään potilaiden hoidossa, tutkimuksessa ja terveydenhuollon kehittämisessä. Näitä ovat esimerkiksi sähköiset potilastiedot, genomitieto, lääketieteellinen tutkimusdata, lääketieteellinen kuvantaminen, IoT- ja etämonitorointilaitteet, kuten äly-

kellot ja veronsokerimittarit, sekä tekoäly ja koneoppiminen (Tiedonhallintalautakunta, 2021).

Terveydenhuollon suurten datakokonaisuuksien erityispiirteisiin kuuluvat sen arkaluonteisuus, korkea sääntely (esim. GDPR) sekä laaja hyödyntämispotentiaali hoidon parantamisessa ja lääketieteellisessä tutkimuksessa (Price & Cohen, 2019). Suurten datamäärien hallinta tuo mukanaan sekä mahdollisuuksia että riskejä. Datan avulla voidaan tehostaa terveydenhuollon prosesseja ja kehittää yksilöllisempää hoitoa. mutta samalla arkaluonteisten tietojen käsittely vaatii erityistä huolellisuutta (Hendolin & Hämäläinen, 2022).

5. TIETOTURVALLISUUS TERVEYDENHUOLLON BIG DATASSA

5.1 Tietoturvan merkitys potilasturvallisuudelle ja terveydenhuollon toiminnalle

Kun tietoturvaa tarkastellaan terveydenhuollon näkökulmasta, sen merkitys korostuu entisestään. Kyse ei ole pelkästään teknisestä vaatimuksesta, vaan tietoturva kytkeytyy suoraan potilasturvallisuuteen, järjestelmien luotettavuuteen ja kansalaisten luottamukseen koko palvelujärjestelmää kohtaan. Terveydenhuollon tietojärjestelmissä käsitellään erityisen arkaluonteisia henkilötietoja, kuten potilaskertomuksia, diagnoositietoja ja hoitohistoriaa, joiden suojaaminen on elintärkeää sekä yksityisyyden että hoidon laadun näkökulmasta. (Khera et al., 2020)

Mikäli tietoturva ei toteudu asianmukaisesti, seuraukset voivat olla vakavia. Tietovuodot, palvelunestohyökkäykset tai käyttöoikeuksien väärinkäytöt voivat estää hoitohenkilöstöä pääsemästä käsiksi kriittisiin tietoihin oikea-aikaisesti, mikä puolestaan voi viivästyttää hoitoa tai johtaa virheellisiin päätöksiin. Samalla potilaiden luottamus järjestelmään voi kärsiä merkittävästi. Vastaamon tietomurto vuonna 2020 toimii konkreettisena esimerkkinä siitä, kuinka puutteellinen suojaus ja käyttöoikeushallinta voivat johtaa laajamittaisiin seurauksiin yksilöille ja koko yhteiskunnalle (Traficom, 2021).

Terveydenhuollon tietoturva perustuu monikerroksiseen suojaukseen, jossa yhdistyvät tekniset ja hallinnolliset ratkaisut. Näitä ovat muun muassa: pääsynhallinta, salaus, auditointi ja lokitus sekä hyökkäyksentorjunta. Pääsynhallinnassa järjestelmien käyttö rajataan valtuutetuille henkilöille esimerkiksi henkilökohtaisten tunnusten ja kaksivaiheisen tunnistautumisen avulla. (Georgiou & Lambrinoudakis, 2020) Tietojen salaaminen suojaa dataa sekä siirron että tallennuksen aikana, estäen luvattoman pääsyn myös teknisten haavoittuvuuksien kautta. Auditointi ja lokitiedot mahdollistavat järjestelmien käytön seurannan, jolloin voidaan jäljittää, kuka on käsitellyt mitä tietoa ja milloin, mikä on ratkaisevaa mahdollisten väärinkäytösten ja tietovuotojen selvittämisessä (LePar, 2004). Tietomurtojen ja kyberhyökkäysten torjunta on keskeinen osa terveydenhuollon tietoturvaa, ja sitä toteutetaan muun muassa palomuureilla, haittaohjelmien torjuntaohjelmistoilla sekä tunkeutumisen havainnointijärjestelmillä. Näiden teknisten ja hallinnol-

listen toimenpiteiden avulla luodaan turvallinen ja luotettava tiedonkäsittely-ympäristö, joka tukee potilaan oikeutta yksityisyyteen ja turvalliseen hoitoon (El Rob, 2023).

Yhdessä edellä mainitut toimenpiteet muodostavat turvallisen ja luotettavan tiedonkäsittely-ympäristön, joka tukee potilaan oikeutta yksityisyyteen ja laadukkaaseen hoitoon. Tietoturva ei siis ole irrallinen tekninen osa-alue, vaan olennainen osa terveydenhuollon toiminnan jatkuvuutta ja eettistä vastuuta.

5.2 Euroopan unionin asetukset

Terveydenhuollon tietoturvaa säätelevät sekä Euroopan unionin että Suomen kansallisen tason lait ja asetukset. Näiden säädösten ensisijaisena tavoitteena on varmistaa, että henkilötietojen käsittely tapahtuu turvallisesti, läpinäkyvästi ja yksilön oikeuksia kunnioittaen.

Keskeisin EU-tason säädös on vuonna 2018 voimaan tullut yleinen tietosuoja-asetus (General Data Protection Regulation, GDPR), joka koskee henkilötietojen käsittelyä kaikissa unionin jäsenvaltioissa. GDPR määrittelee, että henkilötietoja saa käsitellä vain laillisin perustein, ja että niiden suojaamisesta on huolehdittava asianmukaisin teknisin ja organisatorisin toimenpitein (Asetus (EU) 2016/679; GDPR). Erityisen tiukat vaatimukset koskevat niin sanottuja erityisiin henkilötietoryhmiin kuuluvia tietoja, kuten terveyteen liittyviä tietoja, joiden käsittelyssä korostuu yksityisyyden suojan merkitys (Ferrandu, 2018). GDPR:n keskeisiä periaatteita ovat lainmukaisuus, kohtuullisuus ja läpinäkyvyys tietojen käsittelyssä, tarkoitussidonnaisuus eli tietojen kerääminen tiettyä, nimenomaista tarkoitusta varten, tietojen minimointi, oikeellisuus ja säilytyksen rajoittaminen sekä vastuullisuus, joka edellyttää rekisterinpitäjältä osoitusvelvollisuutta tietosuojan toteutumisesta.

Terveydenhuollon kontekstissa GDPR:n soveltaminen kohtaa erityisiä haasteita, sillä terveysdata on lähes poikkeuksetta henkilötietoa ja usein hyvin arkaluonteista. Tällainen data voi sisältää tietoa yksilön terveydentilasta, diagnooseista, hoitohistoriasta, geneettisistä ominaisuuksista tai mielenterveyteen liittyvistä seikoista (Pirha, 2024). Tämän vuoksi sen käsittely edellyttää erityistä huolellisuutta ja tarkkaa sääntelyä.

Terveydenhuollossa syntyy jatkuvasti suuria ja monimuotoisia tietomassoja, eli niin sanottua big dataa (Price & Cohen, 2019). Tämä asettaa tietoturvalle ja tietosuojalle uusia vaatimuksia. Datan hajautuminen useisiin järjestelmiin voi johtaa tiedonhallinnan epäyhtenäisyyteen, ja eri lähteistä yhdistetty tieto voi altistaa anonymisoidun datan uudelleenidentifioinnille. Lisäksi datan käyttötarkoitusten laajeneminen esimerkiksi tutkimuksessa tai tekoälyn hyödyntämisessä edellyttää huolellista suunnittelua ja eettistä

arviointia. (Mimoto et al., 2017) GDPR:n vaatimukset on sovitettava yhteen sen kanssa, että terveysdataa halutaan hyödyntää laajasti diagnostiikan, hoidon laadun ja tutkimuksen tukena. Tämä edellyttää jatkuvaa tasapainottelua yksityisyyden suojan ja tiedon hyödyntämisen välillä.

5.3 Suomen lainsäädäntö

Suomen kansallisella tasolla terveydenhuollon tietojen käsittelyä säätelee erityisesti laki sosiaali- ja terveydenhuollon asiakastietojen käsittelystä (703/2023), joka täydentää ja täsmentää GDPR:n vaatimuksia. Lain tarkoituksena on edistää asiakastietojen ja asiakkaan itsensä tuottamien hyvinvointitietojen tietoturvallista käsittelyä terveydenhuollon ja sosiaalipalveluiden järjestämisen ja tuottamisen yhteydessä. Lisäksi laki velvoittaa palveluntarjoajia huolehtimaan siitä, että käytössä olevat tietojärjestelmät ja toimintatavat täyttävät tietoturvan ja tietosuojan vaatimukset.

Yksi keskeinen osa Suomen sähköistä terveydenhuoltoa on Kanta-palvelu, joka toimii valtakunnallisena potilastietojen tallennus- ja jakelualustana. Kanta-järjestelmä edellyttää tiukkoja tietoturvakäytäntöjä, kuten pääsynhallintaa, tietojen salausta ja auditointia, jotta potilastietojen käsittely olisi turvallista ja läpinäkyvää. Kanta-palvelussa käytössä olevat tietoturvavaatimukset määritellään tarkasti, ja ne koskevat muun muassa järjestelmien sertifiointia, sanomaliikenteen suojausta ja käyttöoikeuksien hallintaa. (Kanta, 2015)

Kansalliset säädökset täydentävät EU:n tietosuoja-asetusta tuomalla lisävaatimuksia ja rakenteita, jotka on sovitettu Suomen terveydenhuoltojärjestelmän erityispiirteisiin. Yhdessä nämä sääntelykehykset muodostavat perustan, jonka avulla voidaan turvata yksilön oikeudet ja samalla mahdollistaa terveysdatan eettinen ja turvallinen hyödyntäminen.

6. JOHTOPÄÄTÖKSET

6.1 Yhteenveto

Tässä kirjallisuuskatsauksessa on tutkittu laajasti terveydenhuollon suuriin datakokonaisuuksiin liittyviä tietoturvahaasteita. Terveydenhuollon suuret datakokonaisuudet ovat välttämätön osa modernia potilashoitoa, diagnostiikkaa ja lääketieteellistä tutkimuksesta, mutta ne tuovat mukanaan merkittäviä tietoturvauhkia. Keskeisiä tietoturvauhkia ovat kyberhyökkäykset, tietovuodot, anonymisoinnin puutteet sekä järjestelmien tekniset ja organisatoriset haavoittuvuudet

Ensimmäiseksi alatutkimuskysymykseksi määriteltiin ”Mitä suuret datakokonaisuudet ovat ja mitä niiden hallintaan kuuluu?”. Kirjallisuustutkimuksen avulla määriteltiin, että suurilla datakokonaisuuksilla tarkoitetaan laajoja, monimuotoisa ja nopeasti kasvavia tietomassoja (Hung, 2016), joita terveydenhuollossa syntyy esimerkiksi potilastiedoista, kuvantamisesta, genomidatasta ja IoT-laitteista. Niiden hallintaan kuuluu datan kerääminen, tallentaminen, yhdistäminen, analysointi ja suojaaminen (Fashakh, & Abddulkader, 2022).

Toinen tutkimuskysymys oli ”Millaisia mahdolliset tietoturvahaasteet aiheuttavat terveydenhuollon datakokonaisuuksista, ja miten niitä voidaan ehkäistä?”. Kirjallisuuskatsauksen mukaan terveydenhuollon datan erityispiirteet, kuten arkaluonteisuus ja sääntely, tekevät hallinnasta erityisen vaativaa. Tietoturvahaasteet voivat johtaa tietovuotoihin, palvelunestohyökkäyksiin, potilastietojen väärinkäyttöön ja järjestelmien toimintahäiriöihin (Awrahman et al., 2022). Näitä voidaan ehkäistä muun muassa tehokkaalla pääsynhallinnalla, salauksella, anonymisoinnilla, henkilöstön koulutuksella, tekoälyn hyödyntämisellä poikkeavuuksien tunnistamisessa sekä jatkuvalla riskien arvioinnilla ja säännöllisillä auditoinneilla (Giuca, 2018). Havainnot osoittavat, että tietoturva ei ole vain tekninen kysymys, vaan osa laajempaa eettistä ja yhteiskunnallista vastuuta.

Kolmantena tutkimuskysymyksenä oli ”Miten nykyinen sääntely ja lainsäädäntö vaikuttavat terveydenhuollon datan tietoturvaan?”. EU:n yleinen tietosuoja-asetus (GDPR) ja Suomen kansallinen lainsäädäntö asettavat tiukat vaatimukset henkilötietojen käsittelylle. Ne ohjaavat organisaatioita huolehtimaan tietoturvasta koko tietojärjestelmien elinkaaren ajan. Kuitenkin käytännön toteutuksessa ilmenee edelleen puutteita, esi-

merkiksi anonymisoinnin riittävyydessä ja tietojärjestelmien yhteen toimivuudessa (Tikkinen-Piri, 2018). Tämä osoittaa, että pelkkä sääntely ei riitä, vaan tarvitaan myös toimivia teknisiä ja organisatorisia ratkaisuja.

Päätutkimuskysymykseksi muodostettiin ”Mitkä ovat keskeisimmät tietoturvahaasteet terveydenhuollon suurten datakokonaisuuksien hallinnassa?”. Tutkimuksessa havaittiin, että keskeisimmät tietoturvahaasteet terveydenhuollon suurten datakokonaisuuksien hallinnassa liittyvät kyberuhkiin, tietovuotoihin, anonymisoinnin riittämättömyyteen sekä järjestelmien teknisiin ja organisatorisiin haavoittuvuuksiin. Nämä haasteet voivat vaarantaa potilasturvallisuuden, yksityisyyden suojan ja terveydenhuollon toiminnan jatkuvuuden. Alatutkimuskysymykset tukivat hyvin päätutkimuskysymystä ja auttoivat havaintojen etsimisessä.

6.2 Käytännön suositukset tietoturvan kehittämiseksi

Tämän kirjallisuuskatsauksen perusteella voidaan esittää useita käytännön suosituksia terveydenhuollon organisaatioille tietoturvan kehittämiseksi. Ensinnäkin henkilöstön tietoturvakoulutusta olisi hyvä lisätä, jotta työntekijät osaavat tunnistaa kyberuhkia ja toimia tietoturvallisesti työtehtävissään. Esimerkiksi koulutuksissa voitaiosiin käsitellä Vastaamon tietomurron kaltaisia tapauksia, jotka havainnollistavat tietoturvan laiminlyönnin seurauksia ja korostaisivat yksilön vastuuta tietojen käsittelyssä.

Pääsynhallintaa voidaan tehostaa ottamalla käyttöön monivaiheinen tunnistautuminen (MFA) ja tarkentamalla käyttöoikeuksien rajauksia roolipohjaisesti. Näin voidaan varmistaa, että vain valtuutetut henkilöt pääsevät käsittelemään arkaluonteisia potilastietoja. Lisäksi säännölliset tietoturva-auditoinnit auttavat havaitsemaan järjestelmien haavoittuvuuksia ja varmistamaan, että käytännöt vastaavat ajantasaisia vaatimuksia.

Lainsäätäjien ja viranomaisten rooli korostuu tietoturvasääntelyn ja valvonnan kehittämisessä. GDPR:n soveltamista terveydenhuollon suurten datakokonaisuuksien hallintaan tulisi arvioida kriittisesti, ja ohjeistuksia anonymisointiin ja uudelleenidentifioinnin riskeihin tulisi tarkentaa. Kansallisten tietoturvastandardien selkeyttäminen tukisi organisaatioita sääntelyn käytännön toteutuksessa.

6.3 Tutkimuksen rajoitukset ja jatkotutkimusaiheet

Tämä tutkimus perustuu kirjallisuuskatsaukseen, joten empiiriset havainnot puuttuvat tutkimuksesta. Tämä rajaus vaikuttaa siihen, että havainnot perustuvat olemassa olevaan tutkimukseen, eivät suoraan terveydenhuollon organisaatioiden käytäntöihin. Li-

säksi osa lähteistä edustaa teknologian uusinta kehitystä, mutta ei ole vielä vertaisarvioitu, tai ne ovat ei-akateemisista lähteistä, mikä voi vaikuttaa niiden luotettavuuteen.

Nykyteknologiat mahdollistavat suurten datamäärien käsittelyn, mutta samalla tuovat uusia haasteita tietoturvan näkökulmasta, minkä vuoksi jatkotutkimus aiheen ympärillä on tärkeää (Fashakh & Abdudulkader, 2022). Jatkotutkimuksessa olisi hyödyllistä tutkia anonymisointimenetelmien tehokkuutta käytännön sovelluksissa ja niiden kykyä estää uudelleenidentifiointi. Lisäksi ajankohtaisena tutkimusaiheena tekoälyn roolia tietoturvassa tulisi tarkastella syvällisemmin, erityisesti sen soveltamista poikkeavuuksien tunnistamiseen ja uhkien torjuntaan. Olisi myös tärkeää selvittää, kuinka hyvin terveydenhuollon organisaatiot noudattavat tietoturvastandardeja ja mitä haasteita niiden käytännön toteutuksessa ilmenee. Empiirinen tutkimus, kuten kyselytutkimus tai haastattelut, voisi tarjota arvokasta tietoa tietoturvan toteutumisesta eri organisaatioissa ja täydentää kirjallisuuskatsauksen havaintoja.

LÄHTEET

Ahmed, M., Kambam, H.R., Liu, Y. & Uddin, M.N. (2020) Impact of Human Factors in Cloud Data Breach. Advances in Intelligent Systems and Computing. p. 568-577. DOI 10.1007/978-3-030-34387-3_70

Awrahman, B.J., Aziz Fatah, C. & Hamaamin M.Y. (2022) A Review of the Role and Challenges of Big Data in Healthcare Informatics and Analytics. Hindawi Limited. DOI 10.1155/2022/5317760

Badman, A. & Kosinski, M. (2024). What is big data? IBM. Saatavilla: https://www.ibm.com/think/topics/big-data?mhsrc=ibmsearch_a&mhq=big%20data

Basdekis, I., Kloukinas, C., Agostinho, C., Vezakis, I., Pimenta, A., Gallo, L. & Spanoudakis, G. (2023) Pseudonymisation in the context of GDPR-compliant medical research. 19th International Conference on the Design of Reliable Communication Networks, DRCN 2023. Institute of Electrical and Electronics Engineers Inc. DOI 10.1109/DRCN57075.2023.10108370

Bhaskar, R. & Kapoor, B. (2013) Information technology security management. Managing Information Security, 2nd Edition. Elsevier Inc. p. 57-74. DOI 10.1016/B978-0-12-416688-2.00003-9

Bogdan, I. C. (2024) Cybersecurity Assessment And Certification Of Critical Infrastructures. UPB Scientific Bulletin, Series C: Electrical Engineering and Computer Science. 86 (4), 151–166.

Eastwood, B. (2023). How to Navigate Structured and Unstructured Data as Healthcare Organization. HealthTech. Sivustolla: https://healthtechmagazine.net/article/2023/05/structured-vs-unstructured-data-in-healthcare-perfcon

El Rob, M.F.A. (2023) A narrative review of advantageous cybersecurity frameworks and regulations in the United States healthcare system. International Association for Computer Information Systems. p.339-359. DOI 10.48009/4_iis_2023_126

Excelsior. (2022) Big Data, Explained: The 5Vs of Data. Medium. Sivustolla: https://medium.com/@get_excelsior/big-data-explained-the-5v-s-of-data-ae80cbe8ded1

Fashakh, A. & Abddulkader, H. (2022) Big Data and Cybersecurity: A Review of Key Privacy and Security Challenges. Institute of Electrical and Electronics Engineers Inc. DOI 10.1109/ICAIoT57170.2022.10121822

Ferrandu, G. (2018) Control and protection tools of personal data in digital healthcare. IOS Press. Vol. 19 issue 3-4. p.209-218.

Fink, A. (2014) Conducting research literature reviews: from the Internet to paper. 4th ed. Thousand Oaks (Calif.): Sage

Georgiou, D. & Lambrinoudakis, C. (2020) GDPR Compliance: Proposed Guidelines for Cloud-Based Health Organizations. Springer Science and Business Media Deutschland GmbH. p. 156-169. DOI 10.1007/978-3-030-64330-0_10

Giuca, OI., Popescu, T.M., Popescu, A., Prostean, G. & Popescu D.E. (2018) A survey of cybersecurity risk management frameworks. Springer. p. 240-272. DOI 10.1007/978-3-030-51992-6_20

Hendolin, M. & Hämäläinen, H. (2022) Terveysdatan sujuva ja turvallinen käyttö – viisi askelta kohti reilua datataloutta 2030. Sitra. Sivustolla: https://www.sitra.fi/julkaisut/terveysdatan-sujuva-ja-turvallinen-kaytto/

Hung, P.C.K. (2016). Big Data Applications and Use Cases. Springer International Publishing. DOI 10.1007/978-3-319-30146-4

Kanta. (2015) Tieto- ja sanomaliikenteen tietoturvavaatimukset. Sivustolla: https://www.kanta.fi/jarjestelmakehittajat/tieto-ja-sanomaliikenteen-tietoturvavaatimukset

Khera, A., Singh, D. & Sharma, D.K. (2020) Information security and privacy in healthcare records: Threat analysis, classification, and solutions. Institution of Engineering and Technology. p. 223-247. DOI 10.1049/PBHE020E_ch9

Koumaditis, K., Pittas, G., Themistocleous, M., Vassilacopoulos, G., Prentza, A., Kyriazis, D. & Malamateniou, F. (2015) Cloud services for healthcare: Insights from a multidisciplinary integration project. IGI Global. p. 709-734. DOI 10.4018/978-1-4666-8756-1.ch036

LePar, K. (2004) Quality consciousness... Auditing for HIPAA privacy compliance. Lippincott Williams and Wilkins. p. 105-113. DOI 10.1097/00001786-200404000-00006

Lumikari, T., Nieminen, T. & Putaala, J. (2021). Onko verenpaineen etämonitoroinnissa itua? Duodecim 2021; 137 (16): 1601-2

Lääkäriliitto (2021). Potilasasiakirjat. Sivustolla: https://julkaisut.laakariliitto.fi/laakarin-etiikka/potilas-laakarisuhde/potilasasiakirjat/

Mimoto, T., Kiyomoto, S., Tanaka, K. & Miyaji, A. (2017) (p, N)-identifiability: Anonymity under practical adversaries. Institute of Electrical and Electronics Engineers Inc. p. 996-1003. DOI 10.1109/Trustcom/BigDataSE/ICESS.2017.343

Orellana, C., Villegas, M. & Astudillo, H. (2019) Architectural tactics for scalability. Ibero-American Conference on Software Engineering, CIbSE 2019. p. 128-140.

Pirha (2024). Henkilötietojen käsittely tieteellisessä tutkimuksessa Sivustolla: https://www.pirha.fi/fi/ammattilaiselle/tutkimus/tutkimuksen-hallinnointi/tutkimuksen-tietosuoja-gdpr-/henkilotietojen-kasittely-tieteellisessa-tutkimuksessa#12-sovellettava-lainsaadanto

Prajapat, S., Gaur, A.S., Soni, P., Nagar, N. &Goswami, P. (2024) Cyber Attacks and Review of Recent Datya Breach Incidents Their Trends and Measures. Springer Science and Business Media Deutschland GmbH. p. 549-576. DOI 10.1007/978-3-031-74443-3_33

Price, W. N. 2nd, & Cohen, I. G. (2019). Privacy in the age of medical big data. Nature medicine, 25(1), 37–43. DOI 10.1038/s41591-018-0272-7

Rajathi, C. & Rukmani, P. (2023) Investigation of Assessment Methodologies in Information Security Risk Management. Springer Science and Business Media Deutschland GmbH. p.358-400. DOI 10.1007/978-981-99-5166-6_26

Reijula, J. (2007). Älykäs potilasvuode tehohoitopotilaan seurannassa. Sähkö- ja tietoliikennetekniikan osasto, Teknillinen korkeakoulu.

Sahoo, P., Kundu, M. & Begum, J. (2025) Artificial Intelligence in Cancer Diagnosis: A Game-changer in Healthcare. Bentham Science Publishers. p. 1314-1330. DOI 10.2174/0113892010298852240528123911

Shabtai, A., Elovici, Y. & Rokach, L. (2012) Introduction to Information Security. Springer. DOI 10.1007/978-1-4614-2053-8_1

Suomi.fi (2025) Digitaalisen turvallisuuden riskienhallinta. Saatavilla: https://kehittajille.suomi.fi/oppaat/riskienhallinta/tietoturva/hanki-yleiskuva-tietoturvasta

Talman, K. (2024). Yrityssektorin sote-tekoälyratkaisuja Suomessa. Sailab MedTech Finland ry. Sivustolla: https://digifinland.fi/wp-content/uploads/2024/09/Sailab_sote_tekoaly_ekosysteemitilaisuus_syyskuu.pdf

Tiedonhallintalautakunta. (2021) Suosituskokoelma tiettyjen tietoturvallisuus-säännösten soveltamisesta. Valtiovarainministeriö. Sivustolla: http://urn.fi/URN:ISBN:978-952-367-897-2

Tietosuojakeskus (2020). Case Vastaamo. Saatavilla: https://tietosuojakeskus.fi/case-vastaamo/

Tikkinen-Piri, C., Rohunen, A. & Markkula, J. (2018) EU General Data Protection Regulation: Changes and implications for personal data collecting companies. Elsevier Ltd. p. 134-153. DOI 10.1016/j.clsr.2017.05.015

Traficom (2021) Tietoturvan vuosi 2020. Traficom, Liikenne- ja viestitävirasto, Kyberturvallisuuskeskus. Sivustolla: https://www.kyberturvallisuuskeskus.fi/sites/default/files/media/publication/Tietoturvan-vuosi-2020_210212_FIN.pdf

LAIT, ASETUKSET JA STANDARDIT

Euroopan unionin yleinen tietosuoja-asetus 27.4.2016/679

Laki sosiaali- ja terveydenhuollon asiakastietojen käsittelystä 14.4.2023/703

Laki julkisen hallinnon tiedonhallinnasta 9.8.2019/906

Tietosuojalaki 5.12.2018/1050