

Sari Lappalainen

**KETTERÄN DATA-ANALYYSIN HYÖDYNTÄMINEN
SOSIAALI- JA TERVEYDENHUOLLON TIETOJOHTA-
MISEN TUKENA**



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2024

TIIVISTELMÄ

Lappalainen, Sari

Ketterän data-analyysin hyödyntäminen sosiaali- ja terveydenhuollon tietojoh-
tamisen tukena

Jyväskylä: Jyväskylän yliopisto, 2024, 51 s.

Tietojärjestelmätiede, pro gradu -tutkielma

Ohjaaja: Taipalus, Toni

Sosiaali- ja terveydenhuollon digitalisoituminen on herättänyt laajaa kiin-
nostusta datan hyödyntämisestä palveluiden kehittämisessä. Palveluissa kertyy
monipuolista dataa ja siitä voidaan jalostaa tietoa tietojoh-
tamisen tueksi. Tekno-
logian kehittyminen ja sen mahdollistama tietopohja ovat kiinnittyneet tietojoh-
tamisen paradigmaan. Datan hyödyntämisessä on kuitenkin omat haasteensa,
kuten datan monipuolisuus, hitaus, eheys ja turvallisuus. Terveydenhuollossa
datan tarjoamaa lisäarvoa ei myöskään aina tunnisteta, joten sitä ei osata aina
käyttää. Ketterällä data-analyysillä pyritään vastaamaan näihin ongelmiin ja
tuottamaan nopeasti ajankohtaista tietoa toiminnan kehittämiseksi. Ketterän
data-analyysin asema sosiaali- ja terveydenhuollossa vaikuttaa epäselvältä.

Tämä tutkielman tarkoituksena on tutkia, miten sosiaali- ja terveydenhuol-
lossa hyödynnetään ketterää data-analyysia tietojoh-
tamisen tukena. Tutkimus
on toteutettu systemaattisena kirjallisuuskartoituksena. Aineiston haku on suo-
ritettu tietokantahakuina kolmesta eri tietokannasta: Scopus, ScienceDirect ja
ACM Digital library. Aineistoon valikoitui lopulta 26 tieteellistä artikkelia tai
konferenssijulkaisua.

Tuloksissa tuli ilmi, että ketterää data-analyysia ei ole vielä kovin paljon
tutkittu sosiaali- ja terveydenhuollon kontekstissa. Vain yksi artikkeli käsitteli
ketterää data-analyysia ja tutkittu menetelmä oli DataOps. Laajemmin katsot-
tuna data-analyysi on kuitenkin ollut melko suosittu tutkimuskohde. Aiheen
kiinnostavuus on kasvanut 2010-luvun loppupuolella ja aihetta on tutkittu glo-
baalisti. Tulosten perusteella aihe on vielä vähänlaisesti tutkittu ja sitä olisikin
syytä tutkia lisää.

Avainsanat: sosiaali- ja terveydenhuolto, ketterä data-analyysi, tietojoh-
taminen, systemaattinen kirjallisuuskartoitus

ABSTRACT

Lappalainen, Sari

Utilization of agile data analysis to support knowledge management in social and healthcare

Jyväskylä: University of Jyväskylä, 2024, 51 pp.

Information Systems Science, master's thesis

Supervisor: Taipalus, Toni

The digitalisation of social and healthcare has generated widespread interest in utilising data in the development of services. The services accumulate diverse data and can be refined into information to support knowledge management. The development of technology and the knowledge base it enables are anchored in the knowledge management paradigm. However, the utilisation of data has its own challenges, such as the versatility, slowness, integrity and security of data. In healthcare, the added value from the data is not always recognised, so managers do not usually know how to use it. Agile data analysis aims to respond to these problems and quickly produce up-to-date information for the development of operations. The status of agile data analysis in social and healthcare seems unclear.

The purpose of this thesis is to study how agile data analysis is utilized in social and healthcare to support knowledge management. The study has been carried out as a systematic scoping survey. The data has been searched as database searches in three different databases: Scopus, ScienceDirect and ACM Digital library. In the end, 26 scientific articles or conference proceedings were selected for the study material.

The results revealed that agile data analysis has not yet been studied very much in the context of social and healthcare. Only one article was about agile data analysis and the method in the article was DataOps. More broadly, however, data analysis has been a fairly popular research topic. The topic has become more interesting in the late 2010s and the topic has been studied globally. Based on the results, the topic has not yet been studied to a limited extent and should be studied further.

Keywords: social and healthcare, agile data analysis, knowledge management, systematic scoping review

KUVIOT

KUVIO 1	DataOps elinkaari.....	13
KUVIO 2	Datan matka DataOps -putken läpi.....	14
KUVIO 3	Integroitu tietojohdamisen syklimalli	16
KUVIO 4	Tiedon jalostusprosessi.....	19
KUVIO 5	Kirjallisuuskartoituksen vaiheet	23
KUVIO 6	Aineiston valikoituminen	26
KUVIO 7	Julkaisumäärät	28
KUVIO 8	Julkaisutyypit.....	31
KUVIO 9	Julkaisut jaoteltuina maanosittain	31
KUVIO 10	Mitä data-analyysiin liitetään.....	33
KUVIO 11	Tuki tietojohdamiselle	34

TAULUKOT

TAULUKKO 1	Hakutulokset.....	25
TAULUKKO 2	Hyväksymis- ja hylkäämiskriteerit.....	25
TAULUKKO 3	Julkaisufoorumit.....	29

SISÄLLYS

KUVIOT

TAULUKOT

1	JOHDANTO.....	6
1.1	Tutkimuksen tarkoitus.....	6
1.2	Tutkimuskysymys ja tutkimuksen rakenne.....	7
2	TUTKIMUSKONTEKSTI JA KÄSITTEET	9
2.1	Monimutkainen data ja käyttämisen haasteet.....	9
2.2	DataOps.....	11
2.3	Tietojohdaminen sosiaali- ja terveydenhuollossa	14
2.4	Tiedon jalostumisprosessi	18
2.5	Yhteenveto teoriasta	20
3	TUTKIMUSMENETELMÄ JA -AINEISTO	22
3.1	Tutkimuksen toteutus	24
4	TULOKSET.....	27
4.1	Hakutulosten vaihtelevuus ja julkaisumäärät.....	28
4.2	Julkaisujen jakaumat ja data-analytiikan käsite	29
4.3	Hyödyntäminen tietojohdamisen tukena.....	34
5	POHDINTA	36
5.1	Johtopäätökset.....	36
5.2	Suhde muuhun tutkimukseen	39
5.3	Tutkimuksen toteutus, havaitut rajoitteet ja jatkotutkimusaiheet	39
6	YHTEENVETO	42
	LÄHTEET	44
	LIITTEET.....	49

1 JOHDANTO

1.1 Tutkimuksen tarkoitus

Sosiaali- ja terveydenhuolto on digitalisoitunut voimakkaasti. Digitalisoituminen on mahdollistanut esimerkiksi erilaiset etäpalvelut, sähköiset rekisterit ja pilvipohjaiset palvelut (Marques ym., 2019). Muun muassa nämä edellä mainitut teknologiset ratkaisut ovat kerryttäneet ja kerryttävät edelleen valtavia määriä dataa. Tämä on herättänyt mielenkiinnon data-analyysin hyödyntämiseen sosiaali- ja terveydenhuollon palveluissa (emt. 2019). Data-analyysin avulla voidaan tuottaa tietoa tietojohdantamisen tueksi ja palveluita voidaan esimerkiksi kohdentaa asiakasryhmille sekä tuottaa palveluita kustannustehokkaasti.

Datasta jalostettu informaatio, tietämys ja viisaus voivat toimia pohjana tietojohdantamiselle (Rowley, 2007). Itse tietojohdantaminen on käsitteenä laaja ja sen voi ajatella olevan kattokäsite, joka sisältää eri osa-alueita (Laihonen ym., 2013). Tietojohdantamisen käsite on kehittynyt teknologian kehittymisen myötä, kun tiedon tallentaminen, säilyttäminen ja analysointi helpottuivat (Colnar, 2022). Tässä tutkielmassa tietojohdantaminen kytkeytyy vahvasti teknologiaan ja sen sisältämän datan hyödyntämiseen johtamisen tukena.

Tämä lisäksi datan hyödyntämiseen yhdistyy tässä tutkielmassa ketteruus. Ketterä data-analyysi on vielä melko uusi malli datan analysoinnin hyödyntämisessä ja aihe on vielä melko vähän tutkittu akateemisen tutkimuksen parissa. Ketterän data-analyysin malli DataOps sai osakseen laajempaa mielenkiintoa 2010-luvun loppupuolella, joka kertoo osaltaan aiheen tuoreudesta. DataOps on ketterään data-analyysiin tehty menetelmä, joka soveltaa DevOpsista tuttuja menetelmiä, mutta DataOps keskittyy pelkästään data-analyysien tekemiseen (Ereth, 2018).

Data itsessään ei ole tietoa, mutta tiedon jalostumisprosessin kautta datasta voidaan jalostaa tietoa ja viisautta päätöksenteolle. Tiedon jalostumisprosessi on tietojenkäsittelytieteelle ominainen malli tiedon muodostumisen

ymmärtämiseen ja se kytkeytyy teknologian kehittymiseen sekä datan lisääntyneeseen määrään. (esim. Crémier ym., 2019.)

Data-analytiikka on sosiaali- ja terveydenhuollossa mielenkiintoa herättänyt tutkimusaihe ja aiheesta on tehty systemaattisia kirjallisuuskatsauksia. Useat kartoitukset keskittyvät data-analytiikkaan yleisellä tasolla sekä sen menetelmiin ja hyötyihin (esim. Khanra ym., 2020; Salazar-Reyna ym., 2020; Mehta ym., 2018). Epäselvää kuitenkin on, onko ketterää data-analyysiä tutkittu tietojohdamisen tukena sosiaali- ja terveydenhuollon kontekstissa. Tässä tutkielmassa tartutaan tähän tutkimusaukkoon ja tarkoituksena on tuottaa aiheesta ajankohtaista tietoa.

1.2 Tutkimuskysymys ja tutkimuksen rakenne

Sosiaali- ja terveydenhuollon datan käyttöä sekä tietojohdamista pyritään vahvistamaan. Dataa on olemassa paljon, sitä kerätään jatkuvasti lisää ja sen käyttäminen nähdään mahdollisuutena muun muassa palveluiden kehittämiseksi (Lehtonen & Kalliola, 2023). Datan hyödyntämiseksi tarvitaan uusia keinoja, jotta datasta saataisiin jalostettua tietoa nopeammalla tahdilla. Tietoa tarvitaan tietojohdamisen tueksi, kun sosiaali- ja terveydenhuollon palveluita kehitetään ja tuotetaan kustannustehokkaasti. Tämän tutkielman tarkoitus on tutkia ketterää data-analyysiä sosiaali- ja terveydenhuollon kontekstissa.

Tutkielman tutkimuskysymys on:

Miten sosiaali- ja terveydenhuollossa hyödynnetään ketterää data-analyysiä tietojohdamisen tukena?

Tutkimus on toteutettu systemaattisena kirjallisuuskartoituksena, jossa on hyödynnetty kolmea eri tietokantaa tutkimusartikkelien hakemiseen. Käytetyt tietokannat ovat: Scopus, ScienceDirect ja ACM Digital library. Systemaattisen kirjallisuuskartoituksen avulla voidaan selvittää tutkittavasta aiheesta yleiskuva sekä tehdä yhteenveto sen hetkisestä tutkimustilanteesta ilmiön parissa (Taipalus, 2023). Tutkielman tarkoitus on tuottaa systemaattisen kirjallisuuskartoituksen avulla ajankohtaista tietoa tehdyistä tutkimuksista ja miten ketterää data-analyysiä hyödynnetään tietojohdamisen tukena.

Tutkielman rakenne lähtee liikkeelle ilmiöön liittyvien käsitteiden tarkastelusta. Käsitteet kiinnittyvät tässä tutkielmassa sosiaali- ja terveydenhuollon kontekstiin. Tärkeitä käsitteitä ovat data-analytiikka sekä ketterä data-analytiikka, jota käsitellään konkreettisesti DataOps käsitteen avulla. Tämän jälkeen tutkielmassa siirrytään käsittelemään tietojohdamisen käsitettä, joka on laaja käsite ja sitä on lähestytty erilaisista tutkimuksellisista tulokulmista. Tietojohdamiseen liittyy vahvasti tiedon jalostumisen prosessi, jota käsitellään viimeisenä teoreettisena käsitteenä. Tiedon jalostumisen vaiheita käsitellään vaihe vaiheelta aina datasta muodostettavaan viisauteen saakka. Tämän jälkeen on

tutkimusmenetelmän sekä tutkimuksen toteutuksen esittelyiden vuoro. Tutkimuksen tulokset esitellään omassa luvussa, jonka jälkeen tuodaan esille tulosten johtopäätökset sekä käydään läpi tutkimuksen rajoitteita ja jatkotutkimusaiheita.

2 TUTKIMUSKONTEKSTI JA KÄSITTEET

Tässä luvussa esitellään tarkemmin tutkimuksen kontekstia sekä tutkimusaiheeseen kytkeytyvät käsitteet. Tutkimus käsittelee sosiaali- ja terveydenhuollon tietojohdamisen tukena toimivaa ketterää data-analyysia. Näin ollen tärkeät käsitteet tässä tutkielmassa ovat data, DataOps, tietojohdaminen sekä tiedon jalostuminen. Nämä käsitteet yhdessä luovat tutkimuksen teoreettisen kokonaisuuden, jonka kautta tutkimusongelmaa lähestytään. Seuraavaksi näihin käsitteisiin syvennytään yksi kerrallaan.

2.1 Monimutkainen data ja käyttämisen haasteet

Dataa kerätään nykyaikana useista eri lähteistä. Dataa saadaan esimerkiksi käyttäjätiedoista, sensoreista ja tietojärjestelmistä (Crémier ym., 2019; Parma ym., 2017). Dataa kerääntyy paljon ja sitä jää myös paljon hyödyntämättä. Tämä ongelma on tunnistettu myös sosiaali- ja terveydenhuollossa, jossa keskustelu datan tehokkaammasta käytöstä on ajankohtainen (esim. Lehtonen & Kalliola, 2023). Sosiaali- ja terveydenhuollon data koostuu muun muassa potilastiedoista, lääkitystiedoista, hoitotoimenpiteistä, käyntitiedoista, arvioinneista ja saaduista palveluista. Myös erilaiset teknologiat, joita käytetään diagnoosien teossa sekä hoitamisessa, keräävät dataa. Myös asiakkaat itse ovat arvokkaita datan lähteitä ja asiakkaita kannustetaan tuottamaan nykyaikana dataa erilaisten sovellusten avulla. Tätä dataa voidaan hyödyntää asiakkaan hoitoprosessissa. (Choi ym., 2022.) Sosiaali- ja terveydenhuollossa kerääntyvä data on hyvin monipuolista ja sitä kerätään useiden eri väylien kautta.

Terveydenhuollon datan määrä on valtava ja monimutkainen kokonaisuus. Dataa myös tuotetaan kiihtyvällä määrällä lisää. Datan hyödyntämisellä on terveydenhuollon näkökulmasta useita etuja. Data-analyysilla voidaan esimerkiksi hallita ja ennakoida väestön terveyttä, seurata sairaalakäyntien syitä, seurata sairauksien etenemistä tai vaikka arvioida leikkausten hyötyjä tietyille asiakasryhmille. Datan analysoinnilla on paljon annettavaa myös alan tutkimukselle ja palveluiden kehittämiseksi. (Raghupathi ym., 2014.)

Valtavia datamääriä kuvataan massadatan (big data) määritelmällä. Massadataa puolestaan voidaan kuvata erilaisilla ominaisuuksilla. Massadatan ominaisuuksiksi määritellään useimmiten määrä (volume), valikoima (variety), nopeus (velocity) sekä todenmukaisuus (veracity) (Raghupathi ym., 2014). Nämä ominaisuudet tulee ottaa huomioon, kun data-analyysiä suunnitellaan ja toteutetaan. Ominaisuuksilla on myös vaikutusta, siihen millaisilla työkaluilla analyysia tehdään. Data-analyysin tekemiseen sisältyy useita erilaisia haasteita ja sosiaali- ja terveydenhuollon toimintaympäristö tuo mukanaan omat lisähaasteensa.

Terveydenhuollon organisaatioilla on tutkimusten mukaan ollut vaikeuksia tunnistaa data-analyysin tuottamaa lisäarvoa omalle toiminnalle. (esim. Wang ym., 2018.) Toimintaympäristöt muuttuvat yhä monimutkaisemmaksi ja arvaamattommammaksi. Useiden asioiden huomioiminen yhtä aikaa päätöksenteossa on yhä haasteellisempaa. Tähän data-analyysi voisi tuoda helpotusta. Sousa ja kollegat (2019) tuovat esille tutkimuksessaan, että data-analytiikkaa hyödyntävät terveydenhuollon organisaatiot ovat kyenneet tekemään kattavampia päätöksiä organisaation toiminnan näkökulmasta. Data-analytiikan avulla päätökset ovat useammin kustannustehokkaita.

Sosiaali- ja terveydenhuoltoon liittyvät tiedonhallinnan ja tiedon hyödyntämisen haasteet ovat tieteellisessä tutkimuksessa tunnistettu ja tutkimuksessa painopistettä on siirretty erityisesti terveydenhuollon teknologioiden tutkimisesta yhä enemmän data-analytiikan pariin. Terveydenhuollon data-analytiikka sisältää erilaisia menetelmiä, joita voidaan hyödyntää tietojen käsittelyssä. Näitä menetelmiä ovat esimerkiksi datan louhinta ja tilastotietojen hyödyntäminen. (Wan ym., 2020). Sosiaali- ja terveydenhuollon data-analytiikka on kehittyvä osa-alue ja sen hyödyntämisen sekä mahdollisuuksien tutkiminen on vielä kesken. Kyseessä on laaja ja monimutkainen kokonaisuus, mutta sen hyödyllisyys on tunnistettu.

Datan jalostuksessa on yleisesti olemassa paljon haasteita, jotka vaikuttavat tiedon jalostusprosessiin. Datan skaalautuvuus, heterogeenisyys, integraatio, turvallisuus ja eheys ovat muun muassa Parmarin ja Yadavin (2017, 166-167) esille nostamia ongelmia massadatan hyödyntämisessä. Lisäksi tutkijat korostavat, että oma haasteensa on päättää mitä dataa käytetään ja mitä hylätään. Myös datan puhdistaminen tuo omat haasteensa analyysin teolle. Myös valtavan datamäärän säilyttäminen on kallista organisaatioille. Massadata sisältää hyvin eri muodoissa olevaa dataa ja datan järjestäminen käsiteltävään muotoon voi olla erittäin haastavaa. Yu kollegoineen (2023) ovat omassa tutkimuksessaan tuoneet esille datan monipuolisuuden ja siihen liittyvät laatuongelmat. Datan puhdistaminen ja järjestäminen käsiteltävään muotoon vaatii paljon aikaa sekä erilaisia työväihteitä, joita voisi automatisoida. Datan käsittelytekniikat vaativat kehittämistä, jotta käsittely ja analyysin tekeminen olisi nopeampaa ja tehokkaampaa.

Organisaatioiden ohjelmistoekojärjestelmät ovat muuttuneet yhä kompleksisemmiksi kokonaisuuksiksi, ja organisaation järjestelmien sisältämistä tiedoista on vaikea saada selkeää kokonaiskuvaa. Organisaatiot ovat joutuneet siirtämään ja muokkaamaan dataa varastoinnin yhteydessä, koska usein dataa säilytetään eri järjestelmissä ja näiden välillä ei aina ole yhteyttä toisiinsa. Datan siirtäminen

ja muokkaaminen ovat organisaatiolle ylimääräistä työtä. Nämä vaiheet vaativat myös tarkkaa dokumentointia, jotta datan muokkaaminen on yhtenäistä ja laadukasta. (Pinkel ym., 2016.) Koska data on arvokasta, niin datan hyödyntämistä on pyritty helpottamaan. Monimuotoista dataa voidaan säilyttää esimerkiksi NoSQL-tietokannoissa, joka mahdollistaa hajautetun datan säilyttämisen perinteisen relaatiotietokannan sijasta. Datan säilytykseen on myös kehitetty useilla eri alustoilla toimivia tietokantaohjelmistoja, kuten esimerkiksi MongoDB. (Wang ym., 2018.)

Datan siirtäminen paikasta toiseen on yhä helpompaa ja sen säilyttäminen on pyritty tekemään helpoksi. Ongelmana on kuitenkin edelleen datan ketterä ja nopea hyödyntäminen organisaation muuttuviin tiedon tarpeisiin. Tarve on tunnistettu ja ketterään data-analyysiin on kehitetty DataOps -malli, joka on vielä melko tuore menetelmä.

2.2 DataOps

Haasteista huolimatta data on tänä päivänä organisaatioille arvokas tiedon lähde. Datan avulla voidaan kehittää organisaation sisäisiä prosesseja, luoda asiakkaille uusia palveluita sekä parantaa palveluiden laatua. Datan tehokas käyttäminen vaatii kuitenkin oikeat teknologiat ja työkalut sekä standardit. Nopeasti muuttuva toimintaympäristö tuottaa jatkuvasti uutta dataa ja datan tehokas hyödyntäminen on haasteellista. Datan tehokkaaseen hyödyntämiseen on etsitty ketteriä ja automatisoituja menetelmiä, joilla pystytään analysoimaan nopeasti muuttuvaa dataa. DataOps menetelmä on kehitetty ketterän data-analytiikan tarpeeseen. DataOpsissa voidaan nähdä samankaltaisuuksia kuin mitä DevOpsissa, mutta DataOps perustuu pelkästään data-analyysiin. (Ereth, 2018, 1-2). DevOps on puolestaan ohjelmistokehitykseen, testaamiseen ja ylläpitoon painottuva toimintamalli, jossa painottuu ketterät ja automatisoidut työskentelytavat. (Gill ym., 2017). Kehittäminen (*development*, Dev) sekä toiminta (*operations*, Ops) muodostavat DevOpsin. DevOps noudattaa jatkuvan kehittämisen ideaa. Ihanteellisessa tilanteessa automatisoitua prosessia ohjaa viitekehys, joka huomioi niin teknisen toteutuksen, kuin laadun hallinnan ja kehityksen kustannusten hallinnan. (Gill ym., 2017.)

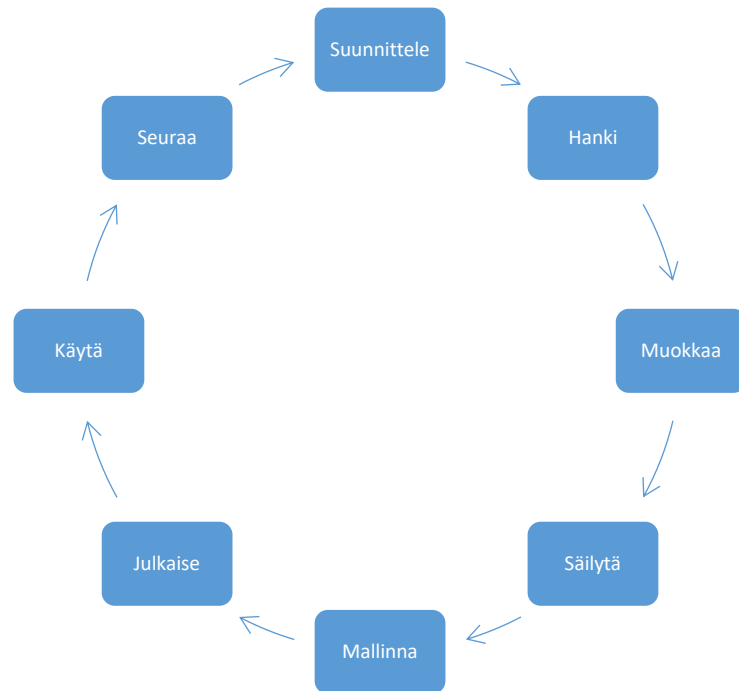
Erethin (2018, 5) mukaan DataOpsissa yhdistyy erilaiset käytänteet ja prosessit, jotka luovat integroidun ja prosessilähtöisen näkökulman dataan. Tähän kytkeytyy mukaan myös automaatio sekä ketterät ohjelmistosuunnittelun menetelmät, joilla saadaan aikaiseksi laatua, nopeutta ja yhteistyön parantumista sekä jatkuvan kehityksen kulttuurin korostamista. Myös Patterson (2019) määrittelee DataOpsin DevOpsin toimintaperiaatteita noudattavaksi menetelmäksi, joka kuitenkin keskittyy vain data-analytiikkaan. DataOps on automatisoitu prosessi, jonka tavoitteena on lyhentää analytiikan toimintasykliä ja tuottaa ajankohtaista tietoa organisaatiolle tavoitteiden saavuttamisen tueksi. DataOpsin avulla organisaatiot voivat integroida tietovirran suunnittelun ja toiminnan jatkuvaksi sekä ketteräksi prosessiksi.

Gartnerin (2023) määrittelee DataOpsin yhteiskäyttöiseksi tiedonhallinnan käytännöksi, joka sisältää viestinnän, integroimisen, tietovirtojen automatisoinnin ja datan käyttämisen kehittämisen. Gartnerin (2023) määritelmässä DataOpsin tavoitteena on tuottaa organisaatiolle nopeammin arvoa tietojen ja tietomalleihin liittyvien artefaktien hallinnan avulla. DataOps hyödyntää teknologiaa, jotta tietojen toimittaminen, suunnittelu ja käyttöönotto voidaan automatisoida ja myös metadata voidaan hyödyntää prosessissa. Tämä lisää datan arvoa ja tuottaa organisaatiolle lisäarvoa muuttuvissa ympäristöissä.

Capizzi kollegoineen (2018) tuovat esille artikkelissaan, että DataOps data-analyysi prosessina voidaan määritellä kolmen laajan vaiheen avulla.

1. **Rakenna.** Tässä vaiheessa tietoja haetaan erilaisista tietokannoista ja lähteistä. Usein dataa voi joutua muokkaamaan tai siistimään käyttökelpoiseksi. Käyttökelpoisen datan siirtoa varten rakennetaan DataOps-putkia, joiden avulla data siirretään kohdepaikkaan odottamaan analyysin tekoa.
2. **Suorita.** Tässä toisessa vaiheessa rakennetut dataputket siirretään haluttuun tuotantoympäristöön, joka voi olla esimerkiksi pilvipalvelu. Huomattavaa on, että organisaation olemassa oleva infrastruktuuri dataputkien käyttämiseen hillitsee lisäkulojen syntymistä.
3. **Käytä.** Tässä vaiheessa dataputket ovat käytössä ja niiden toimintaa tulisi seurata sekä mahdollisiin toiminnan muutoksiin tulisi reagoida nopeasti. Esimerkiksi dataputkeen tulevien datamäärien muutoksiin tulisi reagoida mahdollisimman nopeasti.

Rodriguez ja hänen kollegansa (2020) ovat lähestyneet DataOpsia menetelmää noudattaen tarkemmin DevOpsin menetelmää. DataOpsin elinkaari noudattaa DevOpsin periaatteita, mutta se keskittyy dataan ja datan hyödyntämiseen. Elinkaari kuvattuna kuviossa 1.



Kuvio1. DataOps elinkaari (Rodriguez ym. 2020)

Datan elinkaaren lisäksi datan matka tulee rakentaa ja suunnitella, jotta datan käyttäminen olisi tehokasta. Kuviossa 2 on kuvattuna dataputki sekä datan kulkema matka DataOps-putken läpi. Rodriguez kollegoineen (2020) kuvaavat datan matkan hyvin samankaltaisena hyvin kuin aiemmin mainittu Capizzi työryhmineen (2018). Erona Capizzin ja kumppaneiden (2018) malliin Rodriguez ja hänen kollegansa (2020) korostavat datan matkassa viimeisenä vaiheena arvon luomisen organisaatiolle. Datan matkan kuvattuna alla olevassa kuviossa (kuvio2).



Kuvio2. Datan matka DataOps -putken läpi. (Mukaillen Rodriguez ym., 2020)

Erethin (2018) mukaan DataOpsin hyödyntäminen antaa organisaatioille mahdollisuuden vastata nopeammin muuttuviin vaatimuksiin. Muuttuvat vaatimuksen ovat tänä päivänä yhä kiinteämmin osana myös sosiaali- ja terveydenhuollon palveluita. DataOps -menetelmä tarjoaa organisaatioille keinon tuoda nopeasti pohjatietoa päätöksenteon tueksi. Tietojohdamisen paradigma on kehittynyt organisaatioiden tiedon hyödyntämisen tarpeesta johtamisen tukitoimena. Seuraavaksi tarkastellaan tarkemmin tietojohdamisen käsitettä.

2.3 Tietojohdaminen sosiaali- ja terveydenhuollossa

Tiedon voidaan katsoa olevan organisaation yksi tärkeimmistä strategisista voimavaroista (Grant, 1996). Tiedon hyödyntämisestä johtamisen tukena on puhuttu voimistuvasti viimeisten vuosikymmenten aikana. Suomessa puhe tietojohdamisesta (knowledge management) alkoi 1990-luvulla. Kehityksen taustalla ovat vaikuttaneet teknologian kehittyminen sekä datan ja informaation säilyttämisen, analysoinnin ja hyödyntämisen kehittyminen. (Laihonen ym., 2013, 6). Tiedon rooli on korostunut myös julkisen hallinnon parissa yhä voimakkaammin. Tiedon hyödyntämisestä johtamisen tukena puhutaan sekä tietojohdamisen että tiedolla johtamisen käsitteiden avulla. (Laihonen & Ahlgrén-Holappa, 2020.)

Tietojohdaminen voidaan käsittää kokonaisuudeksi, joka sisältää sekä tiedolla johtamisen, että tiedon johtamisen. Tiedolla johtamisessa organisaatio on kyvykäs oppimaan ja uusiutumaan. Organisaatio kykenee luomaan uutta tietoa ja hallitsemaan tietovarantoja sekä -virtoja. Tiedon johtamisessa sen sijaan organisaatiolla on kyky jalostaa tietoa sekä hyödyntää tietoa organisaation johtamisen tukena. (Laihonen ym., 2013.) Tässä tutkielmassa käytetään tietojohdamisen

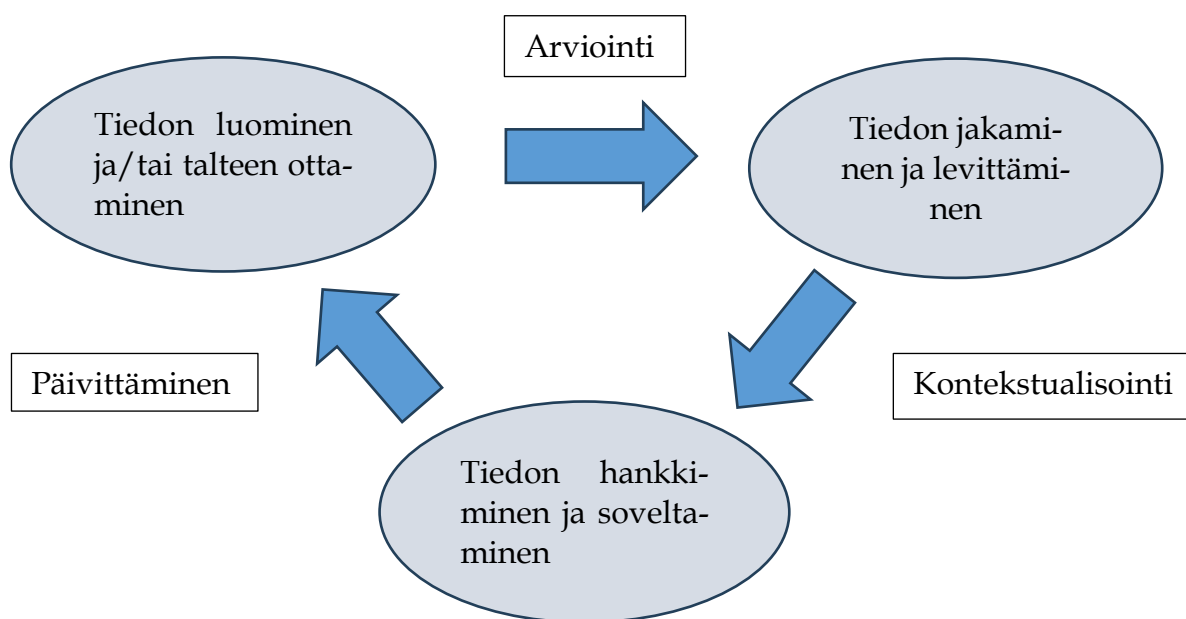
käsitettä, koska se käsittää sen kokonaisuuden, miten tietoa organisaatiossa tuotetaan, hallitaan, jalostetaan sekä hyödynnetään.

Tietojohdaminen voidaan ymmärtää sekä liikkeenjohdollisen tai teknisen näkökulman kautta. Liikkeenjohdollisessa näkökulmassa painotetaan organisaation sosiaalisia prosesseja ja niiden johtamista. Teknisessä näkökulmassa painottuvat puolestaan teknologioiden hyödyntäminen tiedonhallinnassa ja erityisesti tietohallinnon toiminta. (Laihonen ym., 2013, 32.) Colnar kollegoineen (2022) tuovat esille artikkelissaan, että teknologia on tunnistettu yhdeksi tärkeimmistä tietojohdamisen mahdollistajaksi. Tietojohdamisen tarve korostuu nykyajan tietointensiivisessä maailmassa ja organisaation on kyettävä oppimaan sekä muuttamaan toimintaansa tiedon pohjalta, jotta organisaation toiminta on tehokasta ja se kehittyy.

Tiedon luonnetta voidaan tarkastella hiljaisen ja selkeän (eksplisiittisen) tiedon kautta. Eksplisiittinen eli selkeä tieto on tallennettavissa ja dokumentoitavissa olevaa tietoa. Hiljainen tieto sen sijaan on yksilön omaa tulkintaa käsitelystä tiedosta. (Collins, 2010.) Nämä tiedon muodot tarjoavat molemmat päätöksenteolle merkityksellistä tietoa, jota johtaja punnitsee päätöksenteossa.

Nonaka ja Takeuchi (1995) ovat painottaneet tiedon muodostamisen prosesseja johtamisen tueksi. Heisig (2009) on puolestaan omassa tutkimuksessaan löytänyt tietojohdamiselle neljä merkittävää mahdollistajaa. Nämä ovat: ihmislähtöiset tekijät, kuten kulttuuri, ihmiset, johtajuus. Organisaatio ja sen prosessit sekä rakenne. Teknologia ja sen arkkitehtuuri sekä sovellukset. Neljäntenä ovat johtamisen prosessit sisältäen strategian, tavoitteet ja mittaamisen. Tietojohdamista on tutkittu paljon ja erilaisia tulkintoja tietojohdamisesta on olemassa paljon erilaisia. Tutkimuskirjallisuus on jakautunut eri suuntiin tutkijoiden intressien mukaisesti ja kokonaiskuvan hahmottaminen voi olla haastavaa (Laihonen & Ahlgrén-Holappa, 2020.)

Eräs keino määritellä tietojohdamista yksinkertaisemman mallin kautta on Dalkirin (2005) esittelemä tietojohdamisen syklimallin, jossa yhdistyvät useat eri tietojohdamisen määritelmät. Mallissa toistuvat sykleittäin tiedon luominen ja tallentaminen, tiedon jakaminen ja levittäminen sekä tiedon hankkiminen ja soveltaminen.



Kuvio3. Integroitu tietojohdamisen syklimalli. (Dalkir 2005, 43)

Dalkirin (2005) esittelemä malli jättää mallin käyttäjälle mahdollisuuden tulkita tiedon merkityksellisyyttä ja sen hyödyntämistä organisaatiossa. Mallia voidaan käyttää laajasti eri yhteyksissä ja tietojohdamisen tarpeissa, koska malli jättää käyttäjälle mahdollisuuden kontekstualisoida saatua tietoa ja sen soveltamista.

Tietojohdaminen on monipuolisista määrittelyistä huolimatta jalkautunut vahvasti mukaan julkiseen hallintoon. Julkisen hallinnon lisääntynyt kompleksisuus on tuonut mukanaan omia haasteita päätöksenteolle, joihin pyritään vastaamaan tehokkaammin muun muassa tiedon avulla. (Wiig, 2000.)

Sosiaali- ja terveydenhuollon palveluissa on tunnistettu tiedon merkitys johtamisen tukena. Tietojohdaminen on noussut vahvasti mukaan hallinnolliseen keskusteluun, kun palveluita pyritään kehittämään ja tuottamaan kustannustehokkaasti. (Laihonen ym., 2021, 8.) Julkisen sektorin toimijan, kuten sosiaali- ja terveydenhuollon palveluntuottajien tietojohdaminen tukee vaikuttavien ja laadukkaiden palveluiden tuottamista kansalaisille (Laihonen & Ahlgrén-Holappa, 2020). Myös asiakkaiden vaatimukset parempien palveluiden muodossa haastaa sosiaali- ja terveydenhuollon toimialaa. Colnar kollegoinee (2022, 2) painottavat tietojohdamisen sekä laadukkaan tiedonhallinnan merkitystä asiakkaiden tarpeisiin vastaamisessa. Tutkijat myös korostavat, että tiedon tuottamisen ja jakamisen merkitykset korostuvat, kun palveluita kehitetään.

Sosiaali- ja terveydenhuollon palveluissa tietojohdamista on kehitetty viimeisen vuosikymmenen aikana yhä voimakkaammin. Sosiaali- ja terveydenhuollon tietojohdamista on kuitenkin edelleen vähänlaisesti tutkittu. Laihonen ja Saranto (2022) tuovat omassa tutkimuksessaan esille, että tietojohdamisen kompetenssit sosiaali- ja terveydenhuollossa vaihtelevat muuhun tietojohdamisen

tutkimukseen verrattuna. Esimerkiksi vahva lainsäädäntö sekä poliittinen ohjaus ohjaavat omalta osaltaan päätöksentekoa sosiaali- ja terveydenhuollon palveluissa. Lisäksi tietojohdantamiseen liittyviä osaamistarpeita tulisi kehittää tällä palvelusektorilla. Sosiaali- ja terveydenhuollossa teknologian kehittyminen, resursien niukkuus sekä asiakkaiden odotukset ja vaatimukset asettavat paineita tietojohdantamiselle ja palveluiden kehittämiseksi yhä nopeammalla tahdilla.

Palveluiden digitalisoituminen on ollut vahvaa sosiaali- ja terveydenhuollon palveluissa viime vuosina. Digitalisoituminen on mahdollistanut esimerkiksi erilaiset etäpalvelut, sähköiset rekisterit, pilvipohjaiset palvelut sekä data-analyysien hyödyntämisen. (Marques ym., 2019, 575.) Sähköisten potilastietojen lisääntymisen myötä erityisesti terveydenhuollossa on herännyt mielenkiinto tietojen hyödyntämiseen. Asiakkaille voisi tarjota kohdennettuja palveluita tai organisaatioiden sisäistä toimintaa voitaisiin tehostaa olemassa olevan datan avulla. (Li ym., 2016; Sharma ym., 2014.) Myös Sanatarsiero kollegoinee (2023) ovat tunnistaneeet terveydenhuollon data-analyysin kohteiksi palvelun tarpeen ennakoinnin, eri asiakasryhmien toiveiden ja tarpeiden tunnistamisen sekä uudenlaisten palveluiden kehittämisen. Tutkijat tuovat esille, että sähköisten palveluiden tallentama data on johtajille merkityksellistä, koska se tarjoaa tietoa päätöksenteon tueksi ja mahdollistaa päätösten paremman perustelun.

Datan hyödyntämisen yhteydessä tulee muistaa, että sosiaali- ja terveydenhuollon toimintaympäristö sisältää paljon arkaluontoista dataa asiakkaista. Tämä tuo datan hyödyntämiseen omat haasteensa ja hyödynnettävä data ei saa sisältää asiakkaiden henkilökohtaisia tietoja, joista heidät voisi tunnistaa (Li ym., 2016). Myös toimintaympäristön monimutkaisuus tuo omat haasteensa datan käytölle. Sosiaali- ja terveydenhuollossa on mukana monia toimijoita, kuten julkinen sektori ja yksityinen sektori. Alalle on ominaista useat eri ammattiryhmät, jotka katsovat asioita omasta professioasemastaan. Myös tutkimustoiminta tuo mukanaan oman lisänsä datan hyödyntämisessä. (Santarsiero ym., 2023.) Datan hyödyntämisen tulee olla eettisesti kestävää ja pohjata siihen ajatukseen, että palveluita kehitetään tuottamaan kansalaisille parempia palveluita.

Suomessa tiedon hyödyntämistä päätöksenteon tukena on pyritty ohjaamaan valtakunnallisesti erilaisten hankkeiden ja lainsäädännön avulla. Esimerkiksi sote-tieto-hyötykäyttöön-strategia, Sote-tiedolla johtamisen, ohjauksen ja valvonnan toimeenpano-ohjelma (Toivo-ohjelma), vähimmäistietosisältöasetus (STM 2022) ja laki sosiaali- ja terveydenhuollon järjestämisestä (612/2021, §29, §30) ovat viimeaikaisia keinoja vahvistaa sosiaali- ja terveydenhuollon tietojohdantamista. Nämä ohjelmat ja lait ovat osoitus siitä, että datan merkitys on ymmärretty sosiaali- ja terveydenhuollon palveluiden kehittämiseksi ja asiaa edistetään valtiontasolla saakka.

Tahtotila datan hyödyntämiselle vaikuttaa olevan vahva sosiaali- ja terveydenhuollon palveluissa. Data itsessään ei kuitenkaan ole vielä tietoa. Jotta datasta saadaan tietoa päätöksenteon tueksi, niin dataa tulee analysoida ja tulkita. Seuraavaksi syvennyttään tarkastelemaan, miten datasta jalostetaan tietoa ja miten tiedon muodostuminen käsitetään tässä tutkielmassa.

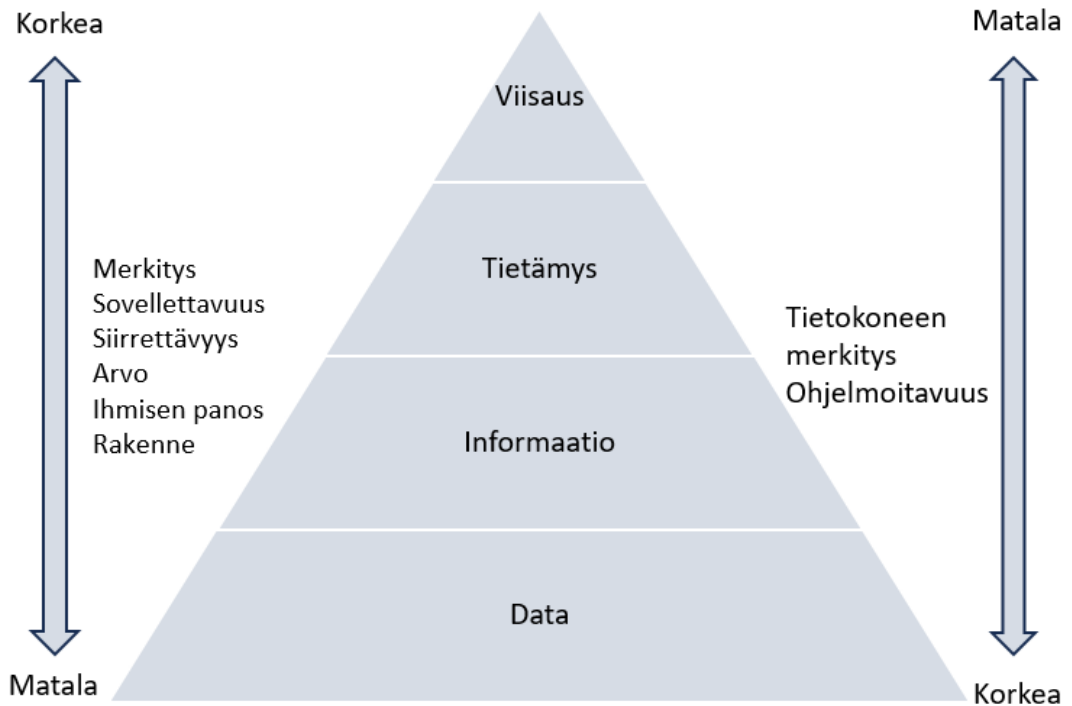
2.4 Tiedon jalostumisprosessi

Mitä tieto on? Tämä kysymys on aiheuttanut historiassa paljon väittelyitä ja pohdintoja. Yksiselitteistä vastausta tähän kysymykseen ei ole tarjolla. (Grant 1996, 110.) Tiedon määritelmästä sekä tiedon eri lajeista on käyty paljon keskustelua. Tiedon määritelmät ovat liittyneet ihmisen antamaan tulkintaan. Tiedon määritelmille sekä pohdinnoille on kuitenkin usein yhteistä se, että tieto vaatii totuutta. (Lammenranta, 2022.) Tiedon määritelmässä havainnot ovat usein liittyneet ihmisten omiin havaintoihin sekä niiden tulkintaan. Nyky-yhteiskunnassa teknologian kehittyminen ja datan kerääminen ovat monipuolistaneet tiedon määrittelyä ja tämä kehitys ohjaa tässä tutkielmassa tiedon muodostumisen käsitystä.

Teknologian kehittyminen on mahdollistanut yhä laajemmat datan keräämisen mahdollisuudet ja datasta muodostettavan tiedon laajuuden. Tämä on samalla haastanut perinteisiä käsityksiä tiedon luonteen epistemologiasta sekä tiedon olemuksen ontologisista kysymyksistä. Data on jopa nähty vapauttavana tekijänä, joka vapauttaa tutkijat perinteisen tieteellisen tutkimuksen lainalaisuuksista. Datan avulla voidaan tarkastella maailmaa ja algoritmit voivat löytää malleja, joita ei muuten välttämättä löydettäisi. Tämä on myös nähty puhtaan objektiivisena tiedon tuottamisena. (Crémier ym., 2019, 191–192.) Tämä on kuitenkin herättänyt myös kritiikkiä, koska datasta muodostettava tieto on yleensä markkinoita palvelevaa eikä algoritmien toiminta ole aina avointa. Myös eettiset ongelmat ja inhimilliset näkökulmat ovat nousseet esille kriitikoiden puheenvuoroissa (esim. Metcalf & Crawford, 2016.) Dataa hyödyntäessä onkin syytä muistaa, että data tarjoaa tiedon muodostumiselle vasta pohjan. Tiedon muodostuminen aina viisaudeksi asti vaatii ihmisen antamia tulkintoja ja kykyä tarkastella dataa suhteessa kontekstiin.

Tässä tutkielmassa tiedon muodostuminen käsitetään tiedon jalostumisprosessin avulla. Vuonna 1989 julkaistussa artikkelissa ”*from data to wisdom*” Akchoff (1989) esitteli tiedon jalostumiselle hierarkkisen mallin, jossa tieto jalostuu datasta aina viisaudeksi asti. Tiedon jalostumisprosessissa data jalostuu informaatioksi, joka puolestaan tuottaa tietämystä. Tietämys jalostuu edelleen viisaudeksi, ja viisaus toimii lopulta päätöksenteon tukena.

Rowley (2007) tuo esille artikkelissaan ja tiedonjalostuksen pyramidissa vahvemmin tietokoneen roolin tiedon jalostumisessa. Tietokoneella on isompi rooli datan ja informaation tasoilla, kun taas ihmisen rooli korostuu tietämyksen sekä viisauden tasoilla. Ihminen ymmärtää tiedon merkityksen ja kykenee soveltamaan saatua tietoa kontekstissaan sekä luomaan tiedon turvin lisäarvoa toiminnalle. Tämä tiedon jalostuminen ihmisen ja teknologian välisenä prosessina on kuvattu alla olevassa kuviossa (kuvio 4).



Kuvio4. Tiedon jalostusprosessi (mukaillen Rowley 2007, 176)

Data on yksinkertaisimmillaan sarja 1:itä ja 0:ia. Nämä ykköset ja nollat kulkevat sähköisessä muodossa merkkijonossa ja ne tallennetaan tietokantoihin. Tämä data tuotetaan laitteen käyttöliittymän avulla ja se siirtyy kohteesta toiseen. Dataa voidaan myös tuottaa ja kerätä erilaisten digitaalisten työkalujen, kuten esimerkiksi antureiden tai käyttötietojen avulla. (Crémier ym., 2019, 195–196.) Dataa säilytetään tietokannoissa, joissa hyödynnetään erilaisia tallennusratkaisuja tallennettavan datan muodon mukaan (Strohbach ym., 2016).

Datan tuottamisen määrät ovat kasvaneet ja on arvioitu, että datan määrä tulee edelleen kasvamaan eksponentiaalisella kasvulla tulevina vuosina. Teknologian kehittyminen lisää datan tuottamisen määrää. Tätä valtavan datan määrää kuvataan massadatan käsitteellä. Massadataa kertyy hyvin erilaisista lähteistä ja se voi olla strukturoimatonta, puoli strukturoitua tai strukturoitua dataa. Strukturoimatonta dataa ei ole pakattu tiettyyn muotoon ja sitä säilytetään siinä muodossa kuin se on kerätty. Esimerkiksi sähköpostit tai kuvat voivat olla strukturoimatonta dataa. Puoli strukturoitu data on jossain määrin käsiteltyä, mutta ei ole selkeästi organisoitua. Strukturoitu data on sen sijaan käsiteltyä ja organisoitua. (Parmar ym., 2017, 166.) Datalla itsessään ei ole vielä arvoa tai merkitystä.

Näiden saamiseksi data tarvitsee kontekstia ja tulkintaa. Tämä tapahtuu tiedon jalostusprosessissa informaation muodostamisen tasolla. (Rowley, 2007.)

Informaatio (information) voidaan määritellä eri tavoin. Rowley (2007) tuo esille artikkelissaan, että informaatio voidaan määritellä kokoelmaksi dataa, jolle on annettu merkitys kontekstin kautta. Toisaalta informaatio voidaan käsittää dataksi, joka on muotoiltu niin merkitykselliseen ja hyödylliseen muotoon. Informaatio voi myös olla dataa, joka on järjestetty niin, että se tuottaa arvoa vastaanottajalle. Informaatioon näyttäisi liittyvän aina ihmisen datalle antama tulkinta ja konteksti. Floridi (2002) nostaa esille, että tietokoneiden kehitys, laskennallinen tiede ja tietojenkäsittely tuovat uusia haasteita informaation määrittelylle. Perinteisen informaation muodostuksen rinnalle on noussut teknologialähtöinen tiedon muodostus.

Tietämys (knowledge) on käsitteenä monitahoinen ja hankala määritellä tarkasti. Käsitteen määrittelyä on käsitelty paljon filosofian historiassa ja hyvin yleinen määritelmä on ”perusteltu tosi uskomus.” (Nonaka, 1994.) Tietämyksen voidaan katsoa olevan informaatiosta jalostunutta tietoa, joka on saanut merkityksiä. Informaatiota voi ajatella tietämyksen alikomponenttina, josta tietämys kehittyy (O’Grady, 2012, 952). Myös Rowleyn (2007) tutkimuksessa nousee esille, että tietämyksen määrittelyssä ihmisen antamat merkitykset muodostavat datasta ja informaatiosta tietämystä. Tietämys on oleellista ja tarpeellista ihmiselle ja sen avulla ihminen voi toimia.

Viisaus (wisdom) sijaitsee tiedon jalostumisen pyramidin huipulla. Rowley ja Slack (2009) ovat tutkineet viisautta omassa tutkimuksessaan ja tuovat esille, että viisaus määrittellään tietojenkäsittelytieteissä useimmiten tiedon jalostusprosessin kautta, jossa viisaus saavutetaan pyramidin huipulla. Viisautta voidaan myös määritellä kykynä ajatella asioita pidemmälle sekä kriittistä arviointi kykyä. Viisauden voidaan myös ajatella olevan kyky tehdä oikeita päätöksiä tiedon ja sen muodostaman kokonaisuuden pohjalta. (Rowley, 2006). Johtajan viisaus on tärkeää, jotta johtaja voi tehdä päätöksiä, joilla on hyödyllisiä vaikutuksia organisaatiolle.

Tiedon jalostusprosessissa data toimii pyramidin pohjalla tarjoten yksittäisiä tiedon palasia. Nämä palaset muotoutuvat parhaimmillaan viisaudeksi, jossa ihmisen omilla tulkinnoilla on tärkeä osuus. Ihmisen osallisuus tiedon jalostumisessa viisaudeksi on tärkeää, jotta päätöksenteossa voidaan ottaa huomioon ihmilliset näkökulmat. Erityisesti sosiaali- ja terveydenhuollossa tämä on tärkeää, koska kyseessä ovat ihmisten hyvinvointia ja terveyttä edistävät ja tukevat palvelut.

2.5 Yhteenveto teoriasta

Tässä osiossa on esitelty data-analyysin haasteita, ketterää data-analyysia sosiaali- ja terveydenhuollon viitekehityksessä sekä miten se kytkeytyy tietojohdantamiseen tiedon jalostumisenprosessin avulla. Ketterän data-analyysin viitekehystä on käsitelty DataOps viitekehityksen avulla, joka on DevOpsin jaloanjalkia

seuraava ketterän kehittämisen malli, mutta se keskittyy data-analyysiin. Sosiaali- ja terveydenhuollon runsaan datan määrän vuoksi ketterälle datan käytölle on tilausta ja DataOps on yksi vaihtoehto ketterään data-analyysiin.

Tietojohtamisen tueksi tarvitaan tietoa ja data-analyysin avulla voidaan tuottaa tietoa johtamisen tueksi. Data itsessään ei ole vielä tietoa vaan on olennaista huomata, että tieto jalostuu datasta aina viisauteen saakka, kuten kuviossa 4 asia on kuvattu. Tiedon jalostumisessa ihmisen antaman kontekstin merkitys kasvaa mitä ylemmäksi tiedonjalostumisen pyramidissa edetään. Sosiaali- ja terveydenhuollon toimintaympäristössä kontekstin ymmärtäminen on tärkeää, jotta tietoon pohjautuva päätöksenteko olisi inhimillistä ja yhteistä hyvää tuottavaa.

Ketterä data-analyysi liitetään tässä tutkielmassa sosiaali- ja terveydenhuollon tietojohtamisen kontekstiin. Tietojohtaminen on julkisessa hallinnossa kasvattanut merkitystään erityisesti viimeisen vuosikymmenen aikana ja kehitykseen on vaikuttanut muun muassa teknologian kehittyminen ja datan yhä laajempi kerääminen. Aihetta on tutkittu erilaisista tulokulmista ja tutkimuskentän kokonaisuus on hajanainen. Kokonaiskuvan muodostaminen ja ketterän data-analyysin merkitys sosiaali- ja terveydenhuollon tietojohtamisen tukena on epäselvä. Tämä tutkielma tarttuu tähän haasteeseen kirjallisuuskartoituksen avulla. Seuraavassa luvussa esitellään tutkielman aineisto sekä systemaattisen kirjallisuuskartoituksen menetelmä, jonka avulla aihetta tutkitaan.

3 TUTKIMUSMENETELMÄ JA -AINEISTO

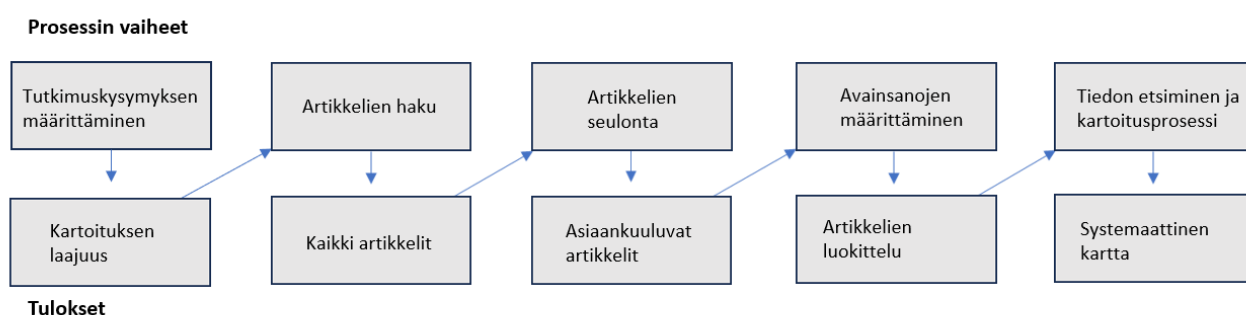
Kirjallisuuskatsaukset ja näyttöön perustuva tutkimus ovat olleet erityisesti lääketieteessä käytetty tutkimusorientaatio. Tutkimustietoon perustuvat ohjeet ovat osoittautuneet luotettavammiksi verrattuna pelkkiin asiantuntijalausuntoihin. Tämän havainnon jälkeen kirjallisuuskatsaukset ovat saaneet vahvemmin jalansijaa useilla eri tieteenaloilla, kuten sosiaalitieteissä, taloustieteissä sekä tekniikan eri tieteenaloilla. Ohjelmistosuunnittelussa kirjallisuuskatsaukset tarjoavat tutkimustuloksia käytännön toiminnan tueksi. (Kitchenham ym., 2009.)

Systemaattisen kirjallisuuskartoituksen avulla voidaan selvittää tutkittavasta aiheesta yleiskuva ja tehdä yhteenveto sen hetkisestä tutkimustilanteesta aiheen parissa. Systemaattinen kirjallisuuskartoitus on läheinen menetelmä systemaattisen kirjallisuuskatsauksen kanssa. Näiden kahden menetelmän välillä on kuitenkin eroja menetelmän syvyydessä ja laajuudessa. Kirjallisuuskatsaus pyrkii syvälliseen ymmärrykseen ja tietoon tutkittavasta aiheesta, kun taas kirjallisuuskartoitus tarjoaa kuvan aiheesta yleisemmällä tasolla. Kirjallisuuskartoitus on toimiva menetelmä, kun tutkittavan aiheen lähestymistavoista tieteellisessä tutkimuksessa halutaan saada käsitys yleistilanteesta. (Taipalus, 2023; Petersen ym., 2015.) Kitchenham ja Charters (2007) tuovat esille, että systemaattinen kirjallisuuskartoitus on toimiva menetelmä, mikäli tutkittava aihe on vielä tuore ja tieteellistä tutkimusta on tehty vielä vähänlaisesti. Toisaalta kirjallisuuskartoitus on toimiva menetelmä myös silloin, kun aihe on hyvin laaja ja tutkijalla on tarve saada rajattua tutkimuskirjallisuutta sekä selvittää millaista tutkimusta aiheesta on tehty.

Kitchenham ja Charters (2007) listaavat systemaattisen kirjallisuuskatsauksen sekä -kirjallisuuskartoituksen välisiä eroja: kartoitustutkimuksessa tutkimuskysymykset ovat yleensä laajempia ja mahdollistavat tietokantahauille laajemat artikkelitulokset. Kartoitustutkimuksen hakutermit ovat vähemmän foku-soituja, jolloin hakutulokset artikkeleista ovat kirjallisuuskatsausta suuremmat. Tarkoituksena on saada laaja katsaus aiheesta. Kartoitustutkimuksen tulokset antavat yleiskuvaa ja tulosten avulla voidaan ohjata tulevaa tieteellistä tutkimusta huomaamalla tutkimusaukkoja sekä tutkimuksen tämänhetkiset painopisteet.

Näiden kahden menetelmän välisissä tavoitteissa on siis selkeät erot ja perusteet menetelmien käyttämiselle ovat selkeät. Systemaattinen kirjallisuuskartoitus on tässä pro gradu -tutkielmassa toimiva tutkimusmenetelmä, koska tutkittava aihe on vielä melko uusi. Tutkielman tarkoituksena on tarjota aiheesta yleiskuva ja tunnistaa meneillään olevat tutkimussuunnat ja aiheen sovelluskohteet tällä hetkellä. Tulokset tarjoavat tietopohjaa tulevaisuuden tutkimukselle.

Systemaattinen kirjallisuuskartoitus etenee suunnitelmallisesti. Petersen tutkimuskumppaneineen (2008) ovat mallintaneet kartoituksen vaiheet visuaalisesti. Tämä malli on kuvattuna alla olevassa kuviossa.



Kuvio 5. Kirjallisuuskartoituksen vaiheet (Petersen ym., 2008)

Kun kirjallisuuskartoituksen tarve on todettu ja tutkimuskysymys on määriteltä, niin tutkija siirtyy prosessissa artikkelien hakuvaiheeseen. Artikkelien haku varten määritellään hakusanat tietokantahakuja varten. Myös sopivat tietokannat tulee määritellä ja arvioida niiden soveltuvuutta tutkittavan aiheen kannalta. Hakusanojen muodostama hakulauseke tulisi olla sopivan väljä, jotta se ei rajaa liikaa artikkeleita haun ulkopuolelle. Artikkelien läpikäymiseen puolestaan tulee määritellä sisäänotto- ja poissulkukriteerit, joiden mukaan tutkija valikoi sopivat artikkelit mukaan lopulliseen aineistoon. (Petersen ym., 2008.) Kriteerit voivat olla monenlaisia ja hyvin yleisesti kriteereissä on määriteltä artikkelien kieli, saatavuus, julkaisuvuodet ja julkaisuforumit. Kriteerit voivat vielä muuttua tietokantahakujen aikana, mikäli hakutulokset osoittautuvat haasteellisiksi kriteereitä ajatellen. (Taipalus, 2023).

Kerätyn aineiston läpikäyminen tapahtuu vaiheittain (Taipalus, 2023; Petersen ym., 2008). Ensimmäinen vaihe on artikkelin otsikon läpikäyminen. Tämän jälkeen siirrytään abstraktien ja tutkimuksen avainsanojen lukemiseen. Tämän perusteella tutkija pystyy yleensä arvioimaan artikkelin soveltuvuuden omaan tutkimukseen. Myös koko artikkelin tai pelkkien johtopäätösten läpikäyminen voi tulla kyseeseen, jos abstrakti jättää tutkijan epätietoisuuteen artikkelin soveltuvuudesta.

Tämän jälkeen aineiston sisältö on hahmottunut tutkijalle. Aineiston käsitelyyn helpottamiseksi aineisto luokitellaan eri ryhmiin: validointitutkimukset

(*validation research*), arviointitutkimukset (*evaluation research*), ratkaisuehdotukset (*solution proposal*), filosofiset julkaisut (*philosophical paper*), mielipidekirjoitukset (*opinion papers*) sekä kokemukseen perustuvat kirjoitukset (*experience papers*). (Petersen, 2008.) Nämä luokittelut auttavat tutkijaa tunnistamaan tutkittavan aiheen painopisteet tutkimuskentällä. Joidenkin aiheiden kohdalla nämä luokittelut ovat kuitenkin osoittautuneet haasteellisiksi ja luokitteluita voidaan joutua har-kitsemaan uudelleen, mikäli tutkimusaihe tätä vaatii. (Taipalus, 2023.)

On hyvä huomata myös, että tietokantahaut voivat jättää joitakin tutkimuk-selle oleellisia artikkeleita pois aineistosta. Tämä voi olla merkki hakulausekkeen heikkoudesta tai tietokannan omista toiminnallisista tekijöistä. Tutkijan on mah-dollista soveltaa lumipallotekniikkaa aineiston läpikäymisessä ja lisäaineiston keruussa. Lumipallotekniikassa tutkija tutustuu relevantin artikkelin lähdeluet-teloon ja valitsee luettelosta soveltuvan tai soveltuvia artikkeleita lisättäväksi ai-neistoon. Tämä voidaan toistaa useita kertoja. (Taipalus, 2023.)

Systemaattisessa kirjallisuuskartoituksessa koko prosessi tulee kuvata sel-keästi lukijalle auki, jotta prosessi on toistettavissa ja luotettava. Tutkijan tulee jatkuvasti arvioida prosessin etenemistä ja aineiston soveltuvuutta. Kirjallisuus-kartoituksessa voidaan kuvata tuloksia auki visuaalisia keinoja hyödyntäen. Esi-merkiksi erilaiset taulukot ja kuvaajat ovat yleisesti käytettyjä tulosten läpikäy-misessä. (Taipalus, 2023.)

3.1 Tutkimuksen toteutus

Tutkimusta varten suoritettiin koehakuja eri tietokantoihin hakulausekkeen tes-taamisen ja arvioinnin vuoksi. Myös hakutulosten määrää sekä artikkeleita arvi-oiitiin silmämääräisesti ja tuloksien mielekkyys todettiin soveltuvaksi suhteessa tutkimuskysymykseen.

Tutkimuskysymys on: miten sosiaali- ja terveydenhuollossa hyödynnetään ket-terää data-analyysiä tietojohdamisen tukena?

Tutkimusta varten valikoitui kolme eri tietokantaa, joissa aineistohaut teh-dään. Käytettävät tietokannat ovat: Scopus, ScienceDirect ja ACM Digital library. Nämä tietokannat tarjoavat sekä tietojärjestelmätieteen, että terveystieteen tie-teenalojen artikkeleita, jolloin tietokantojen käyttäminen on perusteltua tässä tut-kimuksessa.

Hakulauseke muodostui teoreettisen viitekehyksen sekä testihakujen pe-rusteella. Hakulauseke mahdollisti hyvin laajat hakutulokset, jotta aineistosta ei vahingossa sulkeutuisi relevantteja artikkeleita pois. Tutkimusaiheen uutuuden vuoksi haku suoritettiin laajalla skaalalla. Hakulausekkeeksi muodostui:

("knowledge management") AND ("healthcare") AND ("data analytics")

Tietokantahaut tehtiin samalla hakulausekkeella kaikkiin kolmeen tietokantaan syyskuussa 2023. Hakutulokset olivat lukumääriltään vaihtelevat eri tietokannoissa. Hakutulokset vietiin Microsoft Excel-taulukkoon aineiston läpikäymistä varten. Alla olevassa taulukossa (taulukko 1) hakutulokset on esitelty tietokantojen mukaisesti.

Tietokanta	Hakutulokset
Scopus	25
ScienceDirect	540
ACM Digital library	187
Yhteensä	752

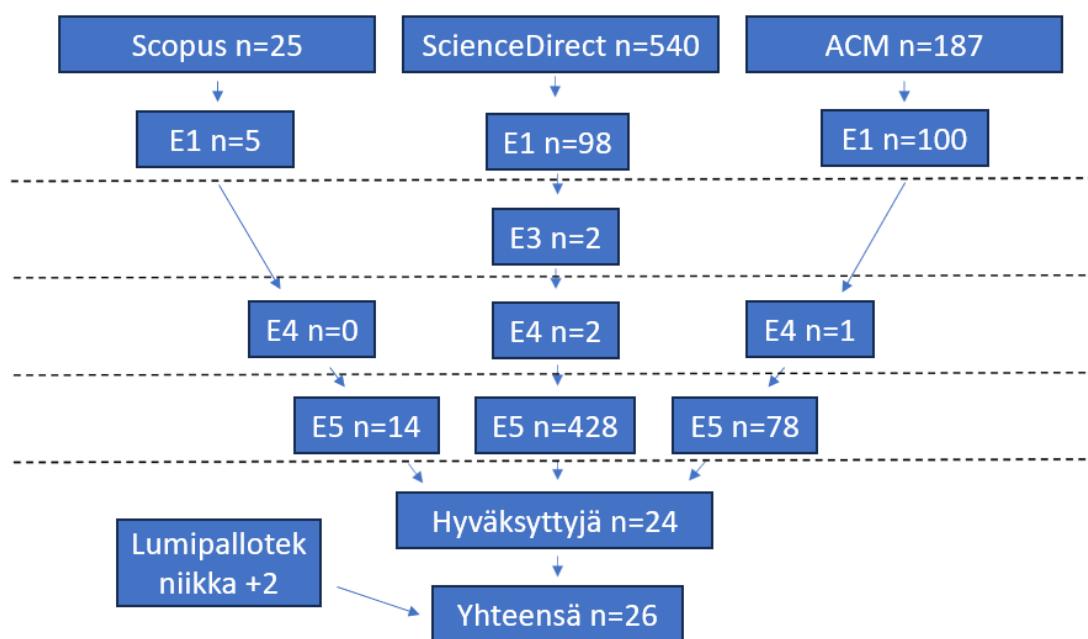
Taulukko 1. Hakutulokset

Aineiston läpikäymistä varten muodostettiin artikkeleiden hyväksymis- ja hylkäämiskriteerit. Näiden avulla hakutuloksista pystyttiin valikoimaan lopullisen aineiston muodostaneet artikkelit. Kriteerit esitelty alla olevassa taulukossa (taulukko 2).

Hyväksytty	Hylätty
K1 Akateeminen artikkeli tai konferenssi julkaisu	E1 Ei-akateeminen artikkeli sekä ei-akateeminen konferenssijulkaisu
K2 Kirjoitettu suomen tai englannin kielellä	E2 Julkaistu muilla kuin suomen tai englannin kielellä
K3 Vapaasti saatavilla digitaalisessa muodossa	E3 Duplikaatti
K4 Julkaistu vuosien 2013–2023 välillä	E4 Julkaistu ennen vuotta 2013
K5 Liittyy tutkimusaiheeseen	E5 Ei liity tutkimusaiheeseen

Taulukko 2. Hyväksymis- ja hylkäämiskriteerit

Tietokantahakujen tulosten läpikäyminen suoritettiin hyväksymis- ja hylkäämiskriteerien avulla. Alla olevassa kaaviossa kuvataan vaiheittain aineiston läpikäyminen ja lopullisen aineiston muodostuminen.



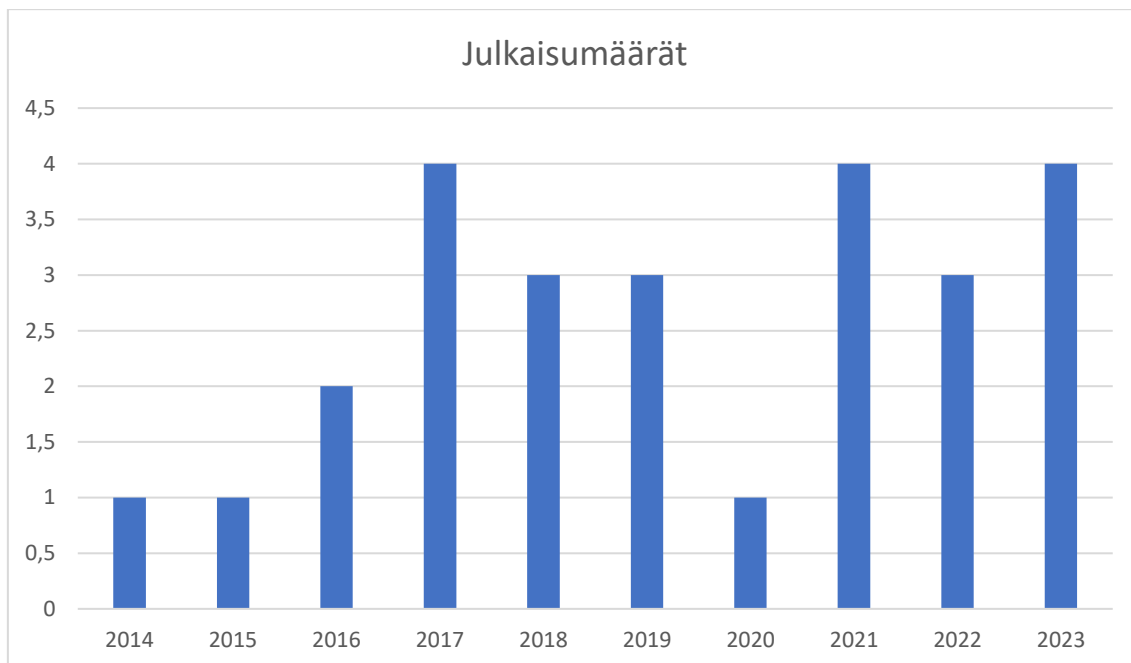
Kuvio 6. Aineiston valikoituminen

Aineistosta suodattuiivat pois ei-akateemiset artikkeli tai konferenssijulkaisut, duplikaatit, ennen vuotta 2013 julkaistut artikkelit sekä tutkimusaiheen ulkopuolelle menevät artikkelit. Kaaviossa on kuvattu kaikkien kolmen tietokantahaun kohdalla tapahtuneet artikkelit valikoitumiset hylkäämiskriteereittäin. Kuviossa jokainen hylkäämiskriteeri on kuvattuna omina tasoinaan, ja ne vähentävät joka tason kohdalla aineiston kokoa n:nä ilmoitetun määrän verran. Hylkäämiskriteerien jälkeen jäljelle jäi lopulta 24 artikkelia. Tämän jälkeen aineistoon lisättiin lumipallotekniikkaa hyödyntäen 2 artikkelia lisää, jolloin aineiston lopulliseksi kooksi muodostui 26 artikkelia. Seuraavassa luvussa aineistoa käydään läpi tarkemmin ja esitellään tulokset.

4 TULOKSET

Tässä luvussa esitellään tutkielman tulokset. Aineistoon valikoituneita artikkeleita oli yhteensä 26 kappaletta. Tutkimuskysymys oli: *miten sosiaali- ja terveydenhuollossa hyödynnetään ketterää data-analyysiä tietojohdamisen tukena?* Tutkimuskysymys on toiminut ohjaavana tekijänä, kun aineistoa on kerätty, käyty läpi sekä analysoitu. Tässä luvussa tulokset esitellään omissa alaluvuissa. Ensimmäiseksi käydään läpi tutkimusjulkaisujen määriä aikavälillä 2013–2023. Tämä kuvaa aiheen kiinnostusta akateemisessa tutkimuksessa. Tämän jälkeen tarkastellaan tarkemmin julkaisujakaumia ja miten data-analytiikan käsite on ymmärretty tutkimuksissa. Näiden jälkeen tuloksissa käydään tarkemmin läpi, miten ketterää data-analyysiä on hyödynnetty tietojohdamisen tukena sosiaali- ja terveydenhuollossa.

4.1 Hakutulosten vaihtelevuus ja julkaisumäärät



Kuvio 7. Julkaisumäärät

Kuvio 7 kuvaa artikkelien vuosijakauma julkaisumäärien mukaan data-analytiikkaa on tutkittu jo 2010-luvulla, mutta aiheen suosio on hieman kasvanut 2020-luvulle tultaessa. Vuoden 2023 artikkelien osalta lukumäärä voisi olla vielä suurempi, mutta aineistonhaku on tehty syyskuussa 2023, joka rajaa loppuvuoden artikkelit pois tuloksista. Julkaisumäärät olivat kuitenkin jo syyskuussa suuremmat kuin vuoden 2022 tasolla, joten aiheen kiinnostavuus vaikuttaisi olevan kasvava akateemisessa tutkimuksessa. Kuvio 7 kuvaa tutkimusaiheeseen soveltuvien julkaisujen määrää kokonaisuudessaan eri vuosien aikana. Seuraavassa osiossa hakutuloksia tarkastellaan hieman tarkemmin ja artikkeleita jaotellaan julkaisufoorumien mukaisesti sekä sukellaan syvemmälle mitä data-analytiikka oikeastaan tarkoittaa aineiston artikkeleiden mukaan.

4.2 Julkaisujen jakaumat ja data-analytiikan käsite

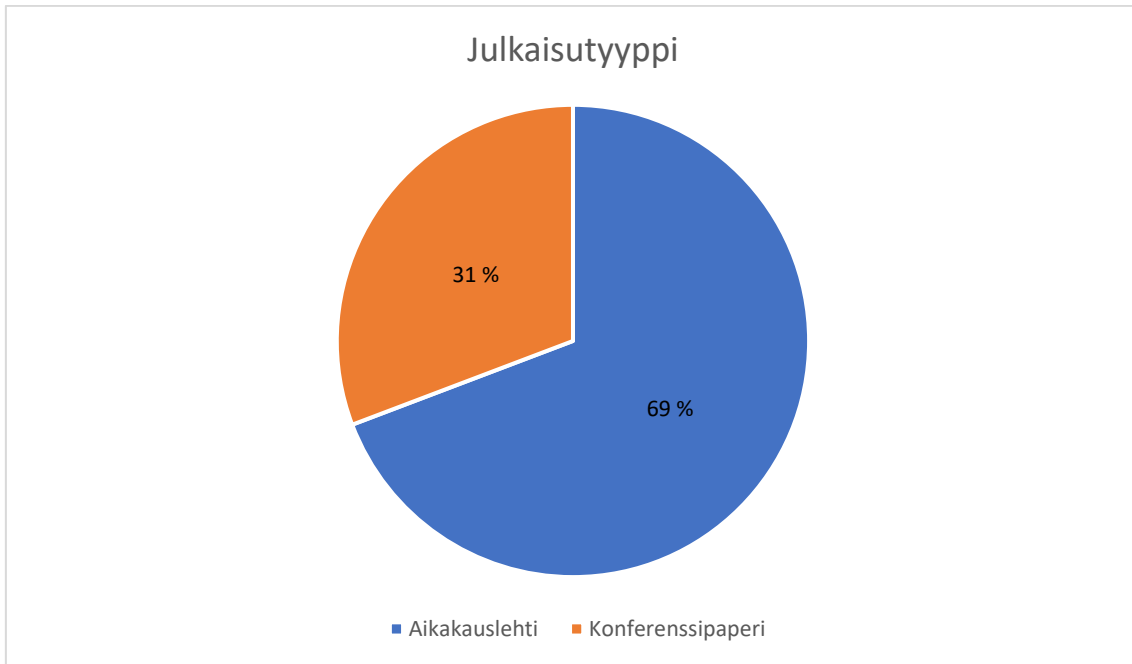
Julkaisut hajautuivat hyvin laajasti eri aikauslehdille sekä konferensseille. Aikauslehtien kohdalla vain *Technological Forecasting and Social Change* -lehti tarjosi kaksi artikkelia aiheeseen liittyen. Muuten tulokset jakautuivat eri foorumeille. Julkaisuforumien laatu vaihteli hyvin suuresti. Osa foorumeista oli hyvin laadukkaita ja arvostettuja. Sen sijaan joidenkin foorumien kohdalla artikkelin laadun kohdalla on syytä olla kriittinen. Alla olevassa taulukossa julkaisuforumit on jaoteltu aikauslehtien, konferenssijulkaisujen sekä näiden lukumäärien mukaan. Julkaisuforumien suosiossa ei ollut havaittavissa selkeitä eroja vuosien välillä eikä aiheen tutkimukset ole selkeästi painottuneet tiettyihin julkaisuforumeihin.

Julkaisufoorumi	Tyyppi	Lukumäärä
Australasian Journal of Information Systems	Aikauslehti	1
Procedia Computer Science	Aikauslehti	1
South Eastern European Journal of Public Health	Aikauslehti	1
Journal of Open Innovation: Technology, Market, and Complexity	Aikauslehti	1
International Journal of Intelligent Networks	Aikauslehti	1
Journal of Medical Systems	Aikauslehti	1
Technological Forecasting and Social Change	Aikauslehti	2
Procedia Computer Science	Aikauslehti	1
International Journal of Medical Informatics	Aikauslehti	1
Computers & Electrical Engineering	Aikauslehti	1
International Journal of Information Management	Aikauslehti	1
Journal of Business Research	Aikauslehti	1
Procedia Computer Science	Aikauslehti	1
Neurocomputing	Aikauslehti	1
Decision Support Systems	Aikauslehti	1
International conference on knowledge management in organizations	Konferenssi	1

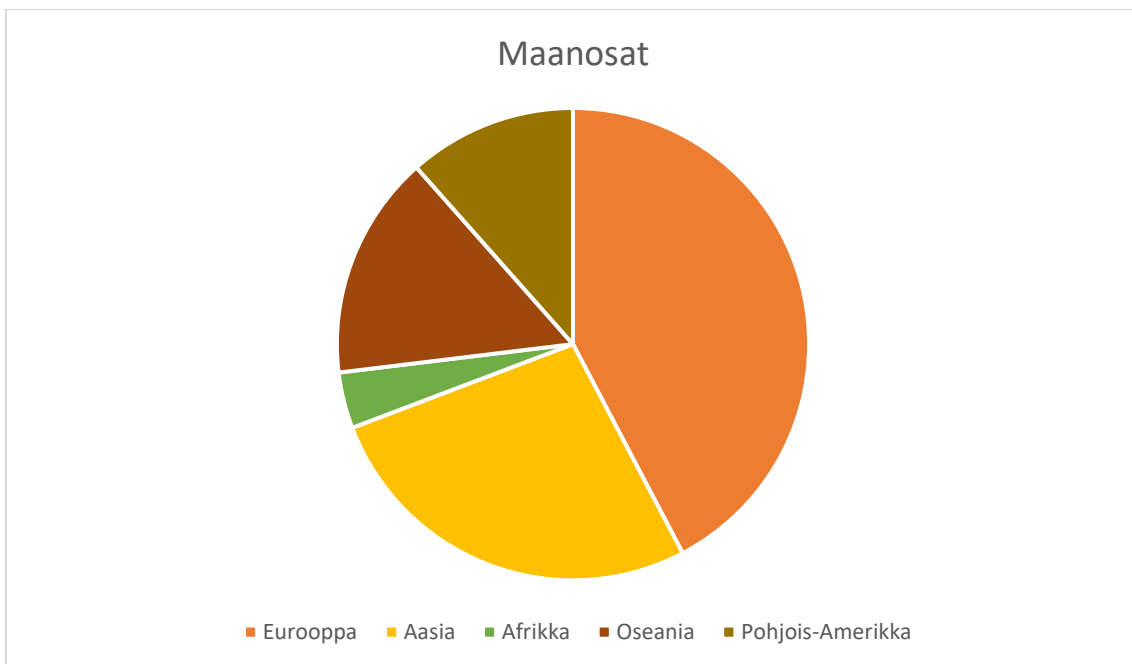
Proceedings of the Australasian Computer Science Week Multi-conference	Konferenssi	1
Proceedings of the 45th ACM technical symposium on Computer science education	Konferenssi	1
Proceedings of the 21st International Academic Mindtrek Conference	Konferenssi	1
ACM Transactions on Computing for Healthcare	Aikakauslehti	1
6 th International Conference on Medical and Health Informatics	Konferenssi	1
Proceedings of the 2018 International conference on digital health	Konferenssi	1
In 24th Pan-Hellenic Conference on Informatics	Konferenssi	1
International Journal of Advanced Computer Science and Applications	Aikakauslehti	1
Journal of Big Data	Aikakauslehti	1
Yhteensä		26

Taulukko 3. Julkaisufoorumit

Jakautuminen julkaisufoorumeissa on ollut tasaista aiheen tutkimuksen parissa. Kuvio 8 kuvaa julkaisufoorumien julkaisujen jakaumaa visuaalisesti konferenssipaperien sekä aikakauslehtien välillä. Aikakauslehtien suosio on ollut suurempi aineiston mukaan, mutta aihe on kiinnostanut tutkijoita myös konferenssipapereissa.



Kuvio 8. Julkaisutyytit

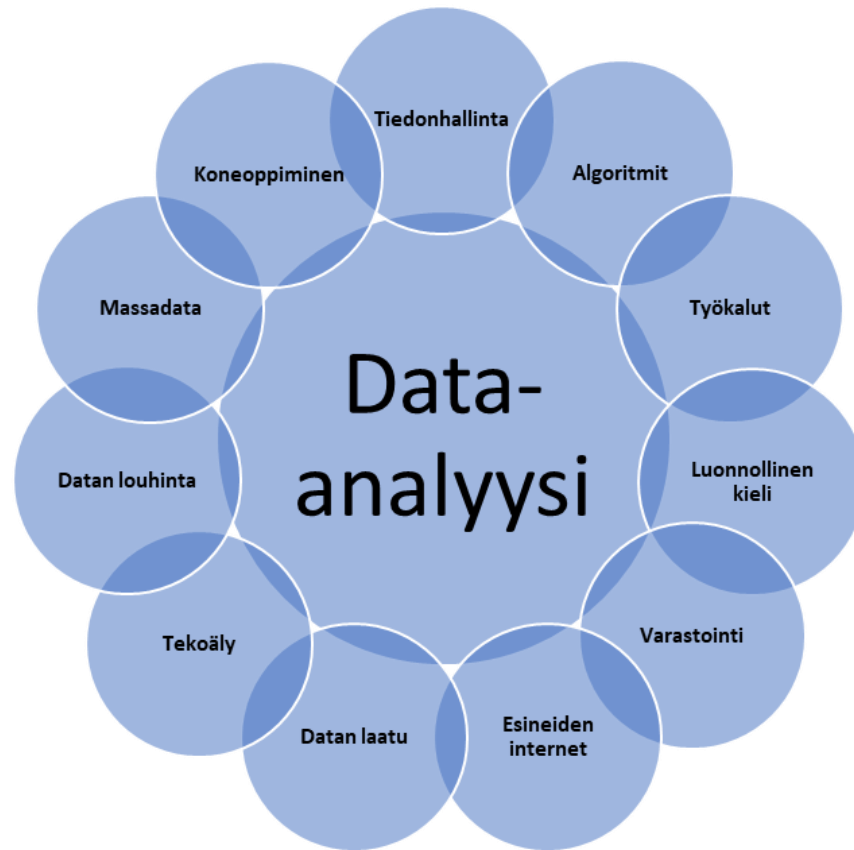


Kuvio 9. Julkaisut jaoteltuina maanosittain

Kuviossa 9 aineiston artikkelit on jaoteltu artikkelien ensimmäisen kirjoittajan maanosan mukaan. Aihe on kiinnostanut tutkimusmaailmaa hyvin globaalisti. Aineiston perusteella vaikuttaa siltä, että data-analytiikka sekä sen hyödyntäminen johtamisen tukena ovat ajankohtaisia tutkimusaiheita laajasti. Aineistosta on kuitenkin havaittavissa eroja tutkimusten määrissä maanosien välillä. Euroopassa tutkimuksia on julkaistu aineiston mukaan eniten, mutta Aasiassa aihe kiinnostaa myös paljon. Oseaniassa oli kolmanneksi eniten julkaisuja. Pohjois-Amerikka oli neljänneksi suurin maanosa ja Afrikka sijoittu viidenneksi.

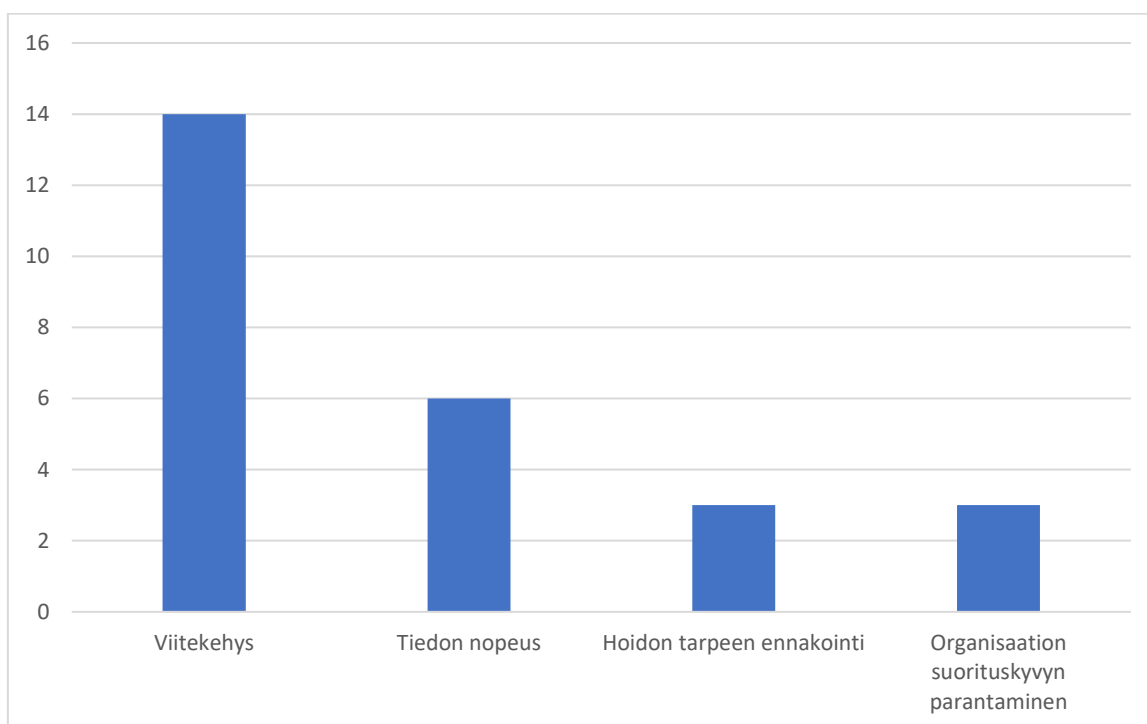
Globaalisti sen sijaan data-analytiikka kiinnostaa tutkijoita sosiaali- ja terveydenhuollon kontekstissa. Aineistosta nousi kuitenkin selkeästi esille, että data-analyysi voi tarkoittaa hyvin erilaisia asioita tutkimuksissa. Asiaa on lähestytty niin algoritmien, tiedonhallinnan, datan louhinnan kuin vaikkapa datan laadun kautta. Tutkimusala onkin hyvin monipuolinen ja osittain hankalasti hahmotettava. Data-analyysin tarkka määrittely osoittautuu hyvin hankalaksi. Käsitteenä data-analyysia voi määritellä parhaiten sateenvarjokäsitteen avulla. Eri tutkimusaloilla keskitytään eri osa-alueisiin, joka tuottaa data-analyysistä hyvin erilaisia näkökulmia. Alla olevassa kuviossa (kuvio 10) data-analyysi on kuvattu keskiöön ja sitä ympäröi aineistosta esiin nousseet näkökulmat, jotka data-analyysiin liitettiin.

Data-analyysin empiirinen tutkimus painottui algoritmien, työkalujen, koneoppimisen, tekoälyn, datan louhinnan sekä luonnollisen kielen kehittämiseen. Tiedonhallinta, massadata, datan laatu, esineiden internet ja varastointi olivat puolestaan data-analyysin teoreettisen näkökulman painopisteitä. Tämä osoittaa konkreettisesti sen, että data-analyysin hyödyntämistä voi tutkia erilaisista näkökulmista ja alan tutkimusta tehdään hyvin laajasti.



Kuvio 10. Mitä data-analyysiin liitetään

4.3 Hyödyntäminen tietojohdamisen tukena



Kuvio 11. Tuki tietojohdamiselle

Kuviossa 11 aineisto on sisällöllisesti luokiteltu niiden tuottaman hyödyn mukaan. Luokittelun ohjaavana tekijänä toimi tuki sosiaali- ja terveydenhuollon tietojohdamiselle. Pääluokkien alaisuuteen on yhdistetty sisällöllisesti samankaltaiset artikkelit, jolloin tulosten tarkasteleminen on mielekkäämpää ja analyysi tuottaa selkeät kokonaisuudet.

Viitekehyksiin keskittyviä artikkeleita oli aineistossa 14 kappaletta. Viitekehykset sisälsivät teoreettisia näkökulmia, teknologian hyödyntämisen näkökulmia, data-analytiikan teknisten vaiheiden kehittämistä sekä tiedon varastointia. Viitekehyksissä data-analyysin käsitteen laajuus oli huomattavissa selkeästi, kun aihetta käsiteltiin eri näkökulmista.

Tiedon saamisen nopeuteen algoritmien, tekoälyn, koneoppimisen, ketterän data-analyysin sekä organisaation data-analyysi kyvykkyksiä käsiteltiin puolestaan 6 artikkelissa. Tähän kategoriaan sijoittui aineiston ainoa artikkeli, joka käsitteli selkeästi ketterää data-analyysia sosiaali- ja terveydenhuollossa. Artikkelit käsitteli DataOps -mallin hyödyntämistä.

Hoidon tarpeen ennakointi johtamisen näkökulmasta oli 3 artikkelin aiheena. Data-analyysin vaikutuksia organisaation suorituskyvyille ja johtamiselle käsiteltiin myös 3 artikkelissa. Nämä luokat muodostavat vastauksen siihen, miten ketterää data-analyysia hyödynnetään tietojohdamisen tukena.

5 POHDINTA

5.1 Johtopäätökset

Sosiaali- ja terveydenhuollon data-analyysille on asetettu isoja odotuksia (kts. esim. Lehtonen ym., 2023). Tiedon merkitys organisaatiolle on tunnistettu strategisesti (Grant, 1996). Teknologian kehitys ja datan keräämisen helpottuminen ovat mahdollistaneet tietojohdamisen kehittymisen (Laihonen ym., 2013). Julkinen hallinto on myös herännyt datan hyödyntämisen mahdollisuuksiin ja tietojohdaminen on vankistanut asemansa julkisessa hallinnossa (Laihonen & Ahlgrén-Holappa, 2020). Tietojohdamisella on kytkös teknologian kehittymiseen sekä data-analyysiin. Tietojohdamisessa tiedon nähdään jalostuvan datasta aina viisauteen saakka (Rowley, 2007). Tiedon jalostumisessa teknologian merkitys vaihtelee tiedon jalostumisen eri asteilla ja datasta saatu informaatio tulee asettaa kontekstiinsa, jotta siitä olisi hyötyä. Tämä on erityisen tärkeää sosiaali- ja terveydenhuollossa, jonka toimialalla ratkotaan ihmiselämään liittyviä haasteita.

Vaikka tietojohdaminen on vahvistanut asemansa yhtenä johtamisparadigmana, niin käsitteenä tietojohdaminen on edelleen hyvin monitulkintainen. Tutkimuskirjallisuutta on monelta eri tieteenalalta ja tulokulmia on näin useita erilaisia tietojohdamiseen. (Laihonen & Ahlgrén-Holappa, 2020.) Tämä hankaloittaa osaltaan tietojohdamisen tutkimista.

Tietojohdamisen käsite on osoittautunut laajaksi, eikä tilanne ole sen helpompia data-analytiikan käsitteen kohdalla. Data-analytiikan määrittämisen vaikeus lähtee liikkeelle jo datan valtavasta määrästä ja sen muotojen vaihtelemisesta. Massadataa on olemassa erilaisissa muodoissa ja sen hyödyntämiseen tarvitaan erilaisia teknologisia ratkaisuja. Valtavan datamäärän kohdalla ongelmaksi nousevat myös laatuongelmat, puhdistaminen ja järjestäminen (Yu ym., 2023.) Myös datan skaalautuvuus, heterogeenisuus, integraatio, turvallisuus ja eheys ovat osa data-analytiikan sisältöä (Parmar & Yadav, 2017). Datan

hyödyntämisen kohdalla tehokkuus ja ketteruus ovat nousseet yhä enemmän keskustelun aiheeksi. Ketterään data-analyysia kuvataan esimerkiksi DataOpsin käsitteellä, joka on DevOpsista kehitetty datan analysointiin keskittyvä menetelmä. Gartnerin (2023) määritelmä DataOpsista on, että se on yhteiskäyttöinen tiedonhallinnan käytäntö, joka sisältää viestinnän, integroimisen, tietovirtojen automatisoinnin ja datan käyttämisen kehittämisen. Ketterän menetelmän avulla datan käyttöä halutaan tehostaa.

Useissa tutkimuksissa on tuotu esille, miten erilaiset ongelmat ja haasteet vaikuttavat data-analytiikan käsitteen määrittelyyn. Tämän tutkimuksen tuloksissa on noussut esille, että jo pelkästään sosiaali- ja terveydenhuollon data-analytiikkaan liitetään useita erilaisia tekijöitä (kts. kuvio 10). Tämä osaltaan selittää sitä, kuinka monesta eri tulokulmasta data-analytiikkaa on tutkittu sosiaali- ja terveydenhuollon kontekstissa ja tutkimusaihe on laaja.

Tutkielman tarkoituksena on vastata kysymykseen: *miten sosiaali- ja terveydenhuollossa hyödynnetään ketterää data-analyysia tietojohdamisen tukena?* Aineiston perusteella tutkimuskysymykseen ei ole tarjolla yksinkertaista vastausta, vaan vastaus jakautuu neljään eri kategoriaan, jotka on muodostettu aineistosta (kts. kuvio 11). Näiden kategorioiden avulla voidaan tarkastella, miten data-analytiikka tukee tietojohdamista sosiaali- ja terveydenhuollossa erilaisten käyttötarkeitusten kautta.

Analyysissa mielenkiintoiseksi havainnoksi osoittautui, että data-analytiisin tutkiminen sosiaali- ja terveydenhuollon tietojohdamisen kontekstissa oli paljon data-analytiikan viitekehysten kehittämistä. Löydöstä saattaa selittää ainakin osin se, että käsitteinä tietojohdaminen sekä data-analyysi ovat vaikeasti hahmotettavissa tarkasti ja niitä tutkitaan hyvin erilaisista tulokulmista. Erilaisen viitekehysten kehittäminen on edelleen tarpeellista, jotta prosessit kehittyisivät tehokkaammiksi.

Ketterä data-analyysi osoittautui hyvin vähän tutkituksi tämän tutkimuksen toimintaympäristön kontekstissa, eli sosiaali- ja terveydenhuollossa. Aineiston artikkeleista ketterään data-analyysiin keskityttiin yhdessä artikkelissa (Baha ym., 2023). Artikkelissa tuotiin ketterä data-analyysi sosiaali- ja terveydenhuollon kontekstiin testaamalla Data-Ops viitekehystä datan analysoinnissa. Tutkimuksessa viitekehukseen yhdistettiin koneoppimista, jonka avulla ketterää data-analyysia kehitettiin tehokkaammaksi ja tarkemmaksi. Ketterä data-analyysi ei ole vielä ollut tutkimuksellinen painopiste sosiaali- ja terveydenhuollossa. Bahaan ja kumppaneiden (2023) tutkimus oli hyvin tuore ja oletettavaa on, että aihe voi nousta lähitulevaisuudessa vahvemmin mukaan akateemiseen tutkimukseen.

Aineistosta nousi melko selkeästi esille kuva siitä, miten data-analyysi ylipäätään nähdään isona mahdollisuutena sosiaali- ja terveydenhuollon johtamiselle. Data-analyysin avulla voidaan tutkimusten mukaan varautua palveluntarpeeseen sekä ennakoita mitkä ihmisryhmät palveluita tulevat käyttämään (esim. Bag ym., 2023 ; De Silva ym., 2015 ; Khan ym., 2017). Palveluiden ennakoiminen ja palveluiden suuntaaminen ennakoivasti oikealle väestöryhmälle ovatkin olleet esillä julkisessa keskustelussa sekä tutkimuskirjallisuudessa.

Ennakoivalla toiminnalla on todettu olevan lopullisia kustannuksia laskeva vaikutus palvelujärjestelmäkokonaisuudessa (kts. esim. Lehtonen ym., 2023).

Sosiaali- ja terveydenhuollon tietojohdaminen näkyi aineiston artikkeleissa yleisenä puheena päätöksenteon tuesta, jota data-analyysi tarjoaa. Kaivo-Oja kollegoineen (2016) käsittelivät artikkelissaan tarkemmin tietojohdamista sosiaali- ja terveydenhuollon kontekstissa. Artikkelissaan kirjoittajat tunnistivat digitalisaation ja datan laajan keräämisen mukanaan tuomat mahdollisuudet tietojohdamiselle sosiaali- ja terveydenhuollossa. Kirjoittajat toivat myös esille digitalisaation mukanaan tuomat haasteet tietojohdamiselle. Artikkelissa tuodaan esille tietojohdamisen teorian kehittämisen tarve toimintaympäristön kehityksen mukana. Tälle työlle olisi tarve, jotta tietojohdamisen käsite ja sen sisältö pysyvät mukana toimintaympäristön kehityksessä ja vaatimuksissa.

Tiedon lisääntyminen sekä tiedon hyödyntämisen paineiden kasvaessa herää pohdinta sosiaali- ja terveydenhuollon tietojohdamisen merkityksestä organisaation suorituskyvyn parantamiseksi. Jo pelkästään yksittäisen organisaation tietovarannot mahdollistavat tiedon monenlaisen hyötykäytön, mutta tämä vaatisi tietovarastojen kokonaiskuvan sekä hyödyntämismahdollisuuksien kokonaisvaltaisen hahmottamisen. Tutkielman aineistossa nousi esille yksittäisten toimintojen kohdalla datan hyödyntämisen kehittämistä (esim. Gonçalves ym., 2017; Reda ym., 2018), mutta myös kokonaisvaltaisen kuvan hahmottamista oli artikkeleissa havaittavissa. Esimerkiksi Ravikumar kollegoineen (2022) ovat artikkelissaan hahmotelleet massadatan hyödyntämistä niin organisaation kuin asiakkaankin hyötyjen näkökulmasta. Ratia ja kumppanit (2017) nostavat puolestaan esille Business Intelligence (BI) työkalujen käytön mahdollisuudet organisaation kokonaisvaltaisen näkökulman kannalta. Wang tutkimusryhmineen (2022) ovat puolestaan tutkimuksessaan tutkineet eri järjestelmien integrointia, jonka avulla eri järjestelmistä voidaan saada koostetusti tietoa päätöksenteon tueksi. Sosiaali- ja terveydenhuollon hajautuneet järjestelmät ovat osoittautuneet ongelmaksi tietojohdamisen kannalta. Dataa on saatavilla, mutta se pitää kerätä eri järjestelmistä ja viedä yhteen järjestelmään. Tämän jälkeen datasta voidaan muodostaa kokonaiskuvaa organisaation toiminnoista ja esimerkiksi palveluiden tarpeesta.

Kaiken kaikkiaan aineiston artikkelit jakautuivat eri vuosille melko tasaisesti. Aiheen kiinnostavuudessa oli havaittavissa nousua vuoden 2020 jälkeen (kts. kuvio 7). Aineiston artikkelien julkaisukanavat olivat jakautuneet eri foorumeille, joka omalta osaltaan osoitti aiheen kiinnostavuutta eri tutkijoiden ja foorumien kesken (taulukko 3). Tutkimusaiheen kiinnostavuus akateemisessa tutkimuksessa vaihteli hieman eri maanosien välillä. Mielenkiintoisesti Pohjois-Amerikka jäi Euroopan, Aasian ja Oseanian jälkeen neljännelle sijalle julkaisumäärissä (kts. kuvio 9). Aineiston keruun aikana oletuksena oli, että tutkimusaineistoon kertyisi paljon tutkimuksia juuri Pohjois-Amerikasta, jossa datan käyttäminen esimerkiksi yritysten liiketoiminnoissa on yleistä. Sosiaali- ja terveydenhuollon kontekstissa sen sijaan tulokset olivatkin jopa yllättävät.

5.2 Suhde muuhun tutkimukseen

Sosiaali- ja terveydenhuollon data-analyysin kirjallisuuskartoitukset sekä -katsaukset ovat terveydenhuolto painotteista. Terveydenhuoltoon painottuvat tutkimukset tarkastelevat data-analytiikan menetelmiä ja hyötyjä organisaatiolle yleisellä tasolla (esim. Khanra ym., 2020; Salazar-Reyna ym., 2020; Mehta ym., 2018). Terveydenhuollon trendi näkyi myös tämän tutkielman aineistossa.

Tarkka tulokulma nimenomaan tietojohdamiseen puuttui aiemmista kartoittavista tutkimuksista. Tietojohdamisen puuttuminen data-analyysin yhteydessä näkyi myös tämän tutkimuksen aineistossa. Tämä puute oli jopa hieman yllättävä, koska tietojohdaminen johtamisen paradigmana on linkittynyt teknologian kehittymiseen (kts. esim. Laihonen ym. 2013).

Tämä tutkielma kartoitti ketterän data-analyysin ja siihen liittyvää tietojohdamista sosiaali- ja terveydenhuollon kontekstissa. Tutkimus sijoittuu aiempien kartoittavien tutkimusten välimaastoon tuoden uudenlaisen tarkastelukulman data-analyysin hyödyntämiseen ketteryys näkökulman kautta. Tämä tutkimus on kytköksissä aiempiin data-analyysia kartoittaviin tutkimuksiin tuoden esille samoja havaintoja kuin aiemmissä tutkimuksissa esimerkiksi data-analyysin tutkimuksellisten näkökulmien runsaudesta ja todetuista hyödyistä sosiaali- ja terveydenhuollon toimialalle. Kuitenkin tämä tutkielma nostaa esille myös havainnon siitä, että ketterä data-analyysi on vielä melko tuntematon asia akateemisessa tutkimuksessa sosiaali- ja terveydenhuollon kontekstissa.

5.3 Tutkimuksen toteutus, havaitut rajoitteet ja jatkotutkimusaiheet

Sosiaali- ja terveydenhuoltoon liittyvää data-analyysi tutkimusta on tehty jonkin verran, mutta mielenkiintoinen havainto jo tietokantahakujen suunnittelussa ja testaamisessa oli, että tietojohdamiseen kytkettyä data-analyysi tutkimusta ei löytynyt kovin paljoa. Toinen mielenkiintoinen havainto oli, että sosiaalipuolen data-analyysitutkimus oli määrältään hyvin vähäistä. Terveydenhuoltoon liittyvää tutkimusta löytyi runsaasti.

Tutkimus toteutettiin kirjallisuuskartoituksena ja aineiston keruuta varten muodostettiin hakulauseke tietokantahakuja varten. Hakulauseketta testattiin ja muokattiin ennen varsinaisia tietokantahakuja. Hakutulosten runsaan saamisen turvaamiseksi lausekkeesta tuli yksinkertainen, joka mahdollisesti laajat hakutulokset. Lopullinen hakulauseke oli: ("knowledge management") AND ("healthcare") AND ("data analytics").

Mielenkiintoinen havainto liittyi sosiaalihuollon osa-alueen lisäämisyritykseen lausekkeessa. Mikäli lausekkeen muutti muotoon: ("knowledge management") AND ("social and healthcare ") AND ("data analytics"), niin

hakutulokset romahtivat. Tämä oli mielenkiintoinen havainto ja tämä voi olla viite sosiaalihuollon vähäisestä tutkimuksesta data-analytiikasta. Hakutulosten laajuus haluttiin tutkielmassa tarpeeksi laajaksi ja tästä syystä social sana jäi pois hakulausekkeesta. Tämä on otettava huomioon tuloksien tarkastelussa ja huomattava, että valtaosa artikkeleista edustaa terveydenhuollon näkökulmaa. Lopullisessa aineistossa sosiaalihuolto on aliedustettuna, mutta analyysi on tehty sosiaali- ja terveydenhuollon kokonaiskuvaa ajatellen. Tämä tulokulma on relevantti, koska suomalainen palvelujärjestelmä on historiallisen sote-uudistuksen myötä viety yhä vahvemmin kohti sosiaali- ja terveydenhuollon integraatiota ja hallinnon tasolla näitä osa-alueita tulisi katsoa kokonaisuutena. Tämä on ohjannut analyysia tarkastelemaan ketterää data-analytiikkaa sekä sosiaali- ja terveydenhuollon tietojohdamista koko palvelujärjestelmän tasolla.

Petersen kollegoineen (2015) tuovat esille artikkelissaan tutkijan mahdollisten harhojen ja ajatusvääristymien vaikutukset koko kartoitusprosessille sekä erityisesti tulosten analysointivaiheessa. Ajatusvääristymät voivat ohjata tulkintaa ja voi ohjata analyysia. Samoin artikkelien valintavaiheessa joitain olennaisia artikkeleita voi jäädä analyysin ulkopuolelle. Tätä riskiä voidaan pienentää suorittamalla kirjallisuuskartoitus useamman tutkijan yhteistyöllä. Tällöin yksittäisen tutkijan ajatusvääristymät eivät ohjaa tutkimusprosessia ja erilaiset näkökulmat tulevat huomioiduksi. Tämä tutkielma on toteutettu yksilötyönä ja yhteistyön hyödyt jäivät tässä tutkielmassa saavuttamatta. Tutkimusprosessin aikana tutkija on ollut tietoinen niistä riskeistä, joita yksintehtävä kirjallisuuskartoitus sisältää. Tietoisuuden avulla ajatusvääristymiä on pyritty vähentämään prosessin aikana.

Tutkimus on toteutettu kirjallisuuskartoituksen prosessia noudattaen (Taipalus, 2023 ; Petersen ym., 2015) ja vaiheet on dokumentoitu kirjalliseen muotoon tähän pro gradu -tutkielmaan. Tämä tukee prosessin luotettavuutta, toistettavuutta ja läpinäkyvyyttä. Kuitenkin on syytä huomata, että artikkelien haut, läpikäyminen sekä analyysin tekeminen on tapahtunut yksilötyönä, joka on osaltaan vaikuttanut siihen, mitkä artikkelit ovat valikoituneet aineistoon, joka puolestaan on vaikuttanut analyysin lopputulokseen. Tätä ongelmaa on pyritty hallitsemaan tutkimuksen hyvällä suunnittelulla, teoriaan perehtymisellä sekä vertailemalla tuloksia aiempiin tutkimuksiin.

Kartoituksen tulokset osoittivat, että ketterää data-analyysiä ei ole vielä tutkittu kovin paljoa sosiaali- ja terveydenhuollon kontekstissa. Tuloksissa ilmeni myös tietojohdamisen ja ketterän data-analyysin välisen yhteyden heikkous akateemisen tutkimuksen osalta. Tietojohdamisen paradigman voimistuminen sosiaali- ja terveydenhuollossa olisi hyvä tunnistaa ja aihetta olisi syytä tutkia lisää. Samoin ketterän data-analyysin merkitys sosiaali- ja terveydenhuollossa on mahdollisesti kasvava osa-alue ja myös tätä olisi syytä tutkia.

Ketterä data-analyysi voi tuottaa merkittäviä hyötyjä sosiaali- ja terveydenhuollon tietojohdamiselle, mutta toisaalta tämä voi tuoda mukanaan uudenlaisia haasteita. Tieteellisen tutkimuksen avulla näitä haasteita ja hyötyjä voidaan tuoda ilmi ja arvioida niiden merkittävyyttä sosiaali- ja terveydenhuollon kontekstissa. Tämä on samalla osa tärkeää julkista keskustelua, jota on syytä käydä

tämän tutkimuskontekstin osalta. Digitalisoituva sosiaali- ja terveydenhuolto mahdollistavat meitä kaikkia koskevien palveluiden kehittämisen laajan tietopohjan avulla. Tässä ketterä data-analyysi ja tietojohdaminen nousevat merkittävään asemaan ja nämä ovatkin mielenkiintoisia tutkimuskohteita.

6 YHTEENVETO

Tämän tutkielman tarkoituksena oli kartoittaa ketterän data-analyysin hyödyntämistä sosiaali- ja terveydenhuollon tietojohdamisen tukena. Tutkimus toteutettiin systemaattisena kirjallisuuskartoituksenan jossa on hyödynnetty kolmea eri tietokantaa tutkimusartikkelien hakemiseen. Tutkielmassa käytetyt tietokannat ovat: Scopus, ScienceDirect ja ACM Digital library.

Kartoitus valikoitui tutkielman menetelmäksi, koska tutkimusilmiöstä haluttiin saada ajantasainen kuva. Sosiaali- ja terveydenhuollon datan hyödyntäminen palveluiden kehittämisessä ja tehostamisessa on herättänyt viime vuosina laajaa kiinnostusta ja aihe on noussut julkiseen keskusteluun (Lehtonen ym., 2023). Sosiaali- ja terveydenhuolto on digitalisoitunut voimakkaasti ja dataa on kerääntynyt paljon ja sitä kerätään edelleen kasvavissa määrin. Tietojohdaminen on kytköksissä datan hyödyntämiseen. Tietojohdamisessa datasta jalostuu informaatiota, tietämystä ja viisautta (Rowley, 2007). Datasta saatua tietoa hyödynnetään johtamisen tukena sekä suoritettavien toimien perusteluna. Tietojohdaminen on puolestaan tullut mukaan yhä vahvemmin julkiseen hallintoon ja sosiaali- ja terveydenhuollon toimialueelle (Laihonen & Saranto 2022).

Data ja sen hyödyntäminen ovat olleet akateemisessa tutkimuksessa suosittuja tutkimusaiheita. Dataa on saatavilla valtavia määriä ja siitä voidaan jalostaa useanlaista tietoa. Valtavan datamäärän kohdalla hyödyntämisessä on myös erilaisia ongelmia ja analysointi on usein hidasta. (esim. Yu ym., 2023; Parmar ja Yadav, 2017). Tehokkuus ja ketteryys ovatkin nousseet yhä enemmän keskustelun aiheeksi data-analyysin kohdalla. DataOps on ketterän data-analyysin menetelmä. DataOps on DevOpsista kehitetty datan analysointiin keskittyvä menetelmä. Gartnerin (2023) määritelmä DataOpsista on, että se on yhteiskäyttöinen tiedonhallinnan käytäntö, joka sisältää viestinnän, integroimisen, tietovirtojen automatisoinnin ja datan käyttämisen kehittämisen.

Tutkielman tulosten mukaan data-analyysi sosiaali- ja terveydenhuollossa on herättänyt paljon mielenkiintoa tutkijoiden keskuudessa globaalisti ja tutkimusten määrä on kasvanut 2020-luvulla. Data-analyysin käsite voi tarkoittaa erilaisia asioita ja aihetta on tutkittu hyvin erilaisista näkökulmista.

Aineistossa korostuivat viitekehysten kehittäminen data-analyysien osalta. Tietojohtamisen näkökulmasta data-analyysi tarjosi keinoja palvelun tarpeen ennakointiin, tiedon saannin nopeuteen sekä organisaation toimintakyvyn kehittämiseksi. Mielenkiintoinen huomio aineistosta oli se, ettei ketterän data-analyysin tutkimusta ole tehty paljoa sosiaali- ja terveydenhuollon kontekstissa. Aineistossa vain yksi tutkimusartikkeli käsitteli DataOps menetelmää sosiaali- ja terveydenhuollon data-analyysissa. Tulosten perusteella vaikuttaisi siltä, että ketterän data-analyysin tutkimus on vasta saapumassa akateemiseen tutkimukseen ja tällä hetkellä ilmiön kohdalla on havaittavissa selkeä tutkimusaukko.

LÄHTEET

- Abidi, S. S. R. (2001). Knowledge management in healthcare: Towards 'knowledge-driven' decision-support services. *International journal of medical informatics (Shannon, Ireland)*, 63(1), 5-18.
[https://doi.org/10.1016/S1386-5056\(01\)00167-8](https://doi.org/10.1016/S1386-5056(01)00167-8)
- Ackoff, R.L. (1989). From data to wisdom. *Journal of Applied Systems Analysis* 16, 3-9.
- Capizzi, A., Disfano, S. & Mazzara, M. (2018). From DevOps to DevDataOps: Data Management in DevOps processes. Teoksessa: Bruel J., Mazzara M. & Meyer B. (toim.) *Software Engineering Aspects of Continuous Development and New Paradigms of Software Production and Deployment*, 52-62
- Choi, W., Chang, SH., Yang, YS. Jung S., Lee DJ., Chun JW., Kim DJ., Lee W & Choi IY. (2022). Study of the factors influencing the use of MyData platform based on personal health record data sharing system. *BMC Med Inform Decis Mak* 22, 182. <https://doi.org/10.1186/s12911-022-01929-z>
- Collins, H. (2010). *Tacit and Explicit Knowledge*. University of Chicago Press: Chicago
- Colnar, S., Radevic, I., Martinovic, N., Lojpur, A., & Dimovski, V. (2022). The role of information communication technologies as a moderator of knowledge creation and knowledge sharing in improving the quality of healthcare services. *PloS one*, 17(8), e0272346.
<https://doi.org/10.1371/journal.pone.0272346>
- Crémier, L., Bonenfant, M. & Lafrance St-Martin, L. (2019). Raw data or hypersymbols? Meaning-making with digital data, between discursive processes and machinic procedures. *Semiotica*; 230: 189-212.
<https://doi.org/10.1515/sem-2018-0110>
- Dalkir, 2005. *Knowledge Management in Theory and Practice*.
<https://doi.org/10.4324/9780080547367>
- Ereth, J. (2018). *DataOps - Towards a Definition*.
- Floridi, L. (2002). What is the philosophy of information? *Metaphilosophy*, 33(1-2), 123-145. <https://doi.org/10.1111/1467-9973.00221>
- Gartner (2023). *DataOps*. <https://www.gartner.com/en/information-technology/glossary/dataops> Luettu 10.8.2023
- Gill, A., Loumish, A., Riyat, I. & Han, S. (2017). DevOps for information management systems. *E Journal of Information and Knowledge Management Systems*, 48(1), 122-139 DOI 10.1108/VJIKMS-02-2017-0007

- Grant, R. M. (1996). Toward a Knowledge-Based Theory of the Firm. *Strategic management journal*, 17(S2), 109-122.
<https://doi.org/10.1002/smj.4250171110>
- Heisig, P. (2009). Harmonisation of knowledge management - comparing 160 KM frameworks around the globe. *Journal of Knowledge Management*, 13(4), 4-31 DOI:[10.1108/13673270910971798](https://doi.org/10.1108/13673270910971798)
- Khanra, S., Dhir, A., Islam, A.K.M.N. & Mäntymäki, M. 2020. Big data analytics in healthcare: a systematic literature review. *Enterprise Information Systems*, 14(7), 878-912 <https://doi.org/10.1080/17517575.2020.1812005>
- Kitchenham, B. & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. EBSE Technical Report EBSE-2007-01
- Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J. & Linkan, S. (2009). Systematic literature reviews in software engineering - A systematic literature review. *Information and Software Technology*, 51(1), 7-15
- Laihonen, H., Hannula, M., Helander, N., Ilvonen, I., Jussila, J., Kukko, M., Kärkkäinen, H., Lönnqvist, A., Myllärniemi, J., Pekkola, S., Virtanen, P., Vuori, V. & Yliniemi, T. (2013). Tietojohdaminen. Tampereen teknillinen yliopisto - tiedonhallinnan ja logistiikan laitos. Tampere.
- Laihonen, H. & Ahlgrén-Holappa, J. (2020). Tieto päätöksenteossa - 10 vuotta tietojohdamista julkisella sektorilla, Focus Localis, teemanumero Tiedonkäyttö päätöksenteossa, 2.
- Laihonen, H. & Saranto, K. (2021). Tiedä ensin, johda sitten Sote-tietojohdamisen osaamistarpeet sekä kansallisen koulutuksen ja tutkimuksen nykytila. Sosiaali- ja terveysministeriön raportteja ja muistioita 2021:33
- Laihonen, H. & Saranto, K. (2022). Knowledge Management Competencies in Health and Social Care. Teoksessa: Cerchione, Roberto; Centobelli, Piera (toim.). Proceedings of the 23rd European Conference on Knowledge Management ECKM 2022, 715-722. Academic conferences international
- Laki sosiaali- ja terveydenhuollon järjestämisestä (612/2021).
<https://www.finlex.fi/fi/laki/alkup/2021/20210612> Luettu 31.5.2023
- Lammenranta, M. (2022). Johdatus tieto-oppiin. Gaudeamus.
- Lehtonen, K. & Kalliola, M. (2023). Datasta voimaa sote-järjestelmään. Sitra, työpaperi. <https://www.sitra.fi/julkaisut/datasta-voimaa-sote-jarjestelmaan/>
- Li, Y., Bai, C. & Reddy, C. K. (2016). A distributed ensemble approach for mining healthcare data under privacy constraints. *Information sciences*, 330, 245-259. <https://doi.org/10.1016/j.ins.2015.10.011>

- Marques, I. & Ferreira, J. (2019). Digital transformation in the area of health: systematic review of 45 years of evolution. *Health and Technology* 10, 575–586 <https://doi.org/10.1007/s12553-019-00402-8>
- Mehta, N. & Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. *International Journal of Medical Informatics*. 114, 57-65 <https://doi.org/10.1016/j.ijmedinf.2018.03.013>
- Metcalfe, J. & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society*, 3(1), 1-14.
- Nonaka, I. (1994). A dynamic Theory of Organizational Knowledge Creation. *Organization Science*, 5(1), 14-37
- Nonaka, I. & Takeuchi, H. (1995) *The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, New York.
- O'Grady, L. (2012). What is knowledge and when should it be implemented? *Journal of Evaluation in Clinical Practice*, 18, 951–953 doi:10.1111/j.1365-2753.2012.01899
- Patterson, P. (2019). DataOps: Modernizing BI With DevOps for Data Analytics. *Big Data Quarterly*, 5(2), 21-26
- Parmar, V. & Yadav, J. (2017). Big Data: Meaning, Challenges, Opportunities, Tools. *International Journal of Advanced Research in Computer Science*, 8 (1), 165-168
- Petersen, K., Robert, F., Shahid, M. & Michael, M. (2008). "Systematic mapping studies in software engineering". Teoksessa 12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12, 1-10.
- Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and software technology*, 64, 1-18. <https://doi.org/10.1016/j.infsof.2015.03.007>
- Pinkel, C., Schwarte, A., Trame, J., Nikolov, A., Bastinos, A. S., & Zeuch, T. (2016). DataOps: Seamless End-to-End Anything-to-RDF Data Integration. https://doi.org/10.1007/978-3-319-25639-9_24
- Raghupathi, W. & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(3), <http://www.hissjournal.com/content/2/1/3>
- Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of information science*, 33(2), 163–180. <https://doi.org/10.1177/0165551506070706>
- Rowley, J. (2006). What do we need to know about wisdom? *Management Decision*, 44(9), 1246-1257. DOI 10.1108/00251740610707712

- Rowley, J. & Slack, F. (2009). Conceptions of wisdom. *Journal of Information Science*, 35 (1), 110–119. DOI: 10.1177/0165551508092269
- Salazar-Reyna, R., Gonzalez-Aleu, F., Granda-Gutierrez, E.M.A., Diaz-Ramirez, J., Garza-Reynes, J.A. & Kumar, A. (2020). A systematic literature review of data science, data analytics and machine learning applied to healthcare engineering systems. *Management Decision*, 60(2), 300-319
<https://doi.org/10.1108/MD-01-2020-0035>
- Santarsiero, F., Schiuma, G., Carlucci, D., & Helander, N. (2023). Digital transformation in healthcare organisations: The role of innovation labs. *Technovation*, 122, 102640.
<https://doi.org/10.1016/j.technovation.2022.102640>
- Sharma, A., & Mansotra, V. (2014). Emerging applications of data mining for healthcare management - A critical review.
<https://doi.org/10.1109/IndiaCom.2014.6828163>
- STM (2022). Vähimmäistietosisältöasetuksen valmistelu.
https://api.hankeikkuna.fi/asiakirjat/91de9e36-8653-4e5d-8bb5-e5cc83032d30/92ccd7f0-1006-4af2-bc76-fbfb87036122/ASETTAMISPAATOS_20220113080822.PDF
- Sousa, M., Pesqueira, A., Lemos, C., Sousa, M. & Rocha, Á. (2019). Decision-Making based on Big Data Analytics for People Management in Healthcare Organizations. *Journal of Medical Systems*, 43,
<https://doi.org/10.1007/s10916-019-1419-x>
- Strohbach, M., Daubet, J., Ravkin, H. & Lischka, M. (2016). Big data storage. Teoksessa: Cavanillas J., Curry E. & Wahlster W. (toim.) *New Horizons for a Data-Driven Economy A Roadmap for Usage and Exploitation of Big Data in Europe*. Springer Open.
- Taipalus, T. (2023). Systematic Mapping Study in Information Systems Research. *Journal of the Midwest Association for Information Systems*, 1, DOI: 10.17705/3jmw.000079
- Yu, S., Chen, T., Han, L., Demartini, G., & Sadiq, S. (2023). DataOps-4G: On Supporting Generalists in Data Quality Discovery. *IEEE transactions on knowledge and data engineering*, 1.
<https://doi.org/10.1109/TKDE.2022.3151605>
- Wan, T. & Gurupur, V. (2020). Understanding the Difference Between Healthcare Informatics and Healthcare Data Analytics in the Present State of Health Care Management. *Health services research and managerial epidemiology*, 7, 2333392820952668.
<https://doi.org/10.1177/2333392820952668>
- Wang, Y., Kung, L. & Byrd, T. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations.

/Technological Forecasting & Social Change, 126, 3-13.

<http://dx.doi.org/10.1016/j.techfore.2015.12.019>

Wiig, K. 2000. Knowledge management in public administration. Journal of Knowledge Management 6(3), DOI:[10.1108/13673270210434331](https://doi.org/10.1108/13673270210434331)

LIITTEET**AINEISTO**

Addressing the complexities of big data analytics in healthcare: The diabetes screening case	De Silva D., Burstein F., Jelinek H. & Stranieri A.	2015
PHAS: An End-to-End, Open-Source, and Portable Healthcare Analytics Stack	Abbasian M., Khatibi E., Azimi I. & Rahmani A.M.	2023
The Impact of Big Data Quality Analytics on Knowledge Management in Healthcare Institutions: Lessons Learned from Big Data's Application within The Healthcare Sector	Ravikumar R., Kitana A., Taamneh A., Aburayya A., Shwedeh F., Salloum S. & Shaalan K.	2022
The effect of healthcare data analytics training on knowledge management: A quasi-experimental field study	Kim B.J. & Tomprou M.	2021
Exploring big data analytics in health care	Ramesh T. & Santhi V.	2020
Decision-Making based on Big Data Analytics for People Management in Healthcare Organizations	Sousa M.J., Pesqueira A.M., Lemos C., Sousa M. & Rocha Á.	2019
Big data analytics for clinical decision-making: Understanding health sector perceptions of policy and practice	Weerasinghe K., Scahill S. L., Pauleen D. J. & Taskin N.	2022
OpenEHR and Business Intelligence in healthcare: an overview	Cunha J., Duarte R., Guimarães T. & Santos M.	2023
Visualizing the knowledge structure and evolution of big data research in healthcare informatics	Gu D., Li J., Li X. & Liang C.	2017
Metaheuristic Algorithms for Healthcare: Open Issues and Challenges	Tsai C-W., Chiang M-C., Ksentini A. & Chen M.	2016

A health data analytics maturity model for hospitals information systems	Carvalho J.V, Ã Rocha Ã., Vasconcelos J. & Abreu A.	2019
Role of big data analytics capability in developing integrated hospital supply chains and operational flexibility: An organizational information processing theory perspective	Yu W., Zhao G., Liu Q. & Song Y.	2021
Big data analytics and artificial intelligence technologies based collaborative platform empowering absorptive capacity in health care supply chain: An empirical study	Surajit B., Pavitra D., Rajesh K. S., Muhammad S. R. & Sreedharan V.R.	2023
Towards of a Real-time Big Data Architecture to Intensive Care	GonÃalves A., Portela F., Santos M. & Rua F.	2017
Enhancing random forest classification with NLP in DAMEH: A system for DAta Management in eHealth Domain	Flora A., Luigi C., Giovanni C., Giovanni M., Francesco M., & Roberto N.	2021
Processing electronic medical records to improve predictive analytics outcomes for hospital readmissions	Hamed Z.M. & Dursun D.	2018
Future Prospects for Knowledge Management in the Field of Health	Kaivo-oja J. Virtanen P. Jalonen H. Stenvall J & Wallin J.	2016
Understanding Chronic Disease Comorbidities from Baseline Networks: Knowledge Discovery Utilising Administrative Healthcare Data	Khan A., Uddin S. & Srinivasan U.	2017
Bringing Business Intelligence to Healthcare Informatics Curriculum: A Preliminary Investigation	Zheng, G., Zhang C. & Li L.	2014
Benefits and Required Capabilities of BI-Tools in the Private Healthcare	Ratia M., Myllärniemi J. & Helander N.	2017
Federated Learning for Healthcare Domain - Pipeline, Applications and Challenges	Madhura J., Ankit P. & Sankarasubbu M.	2022
Gauging the Gaps for Decision Support - Data Integration in the	Wang W. Y. C., Jiang P. H. W., Tiong T.G. & Hsieh C-C.	2022

Hospital Information Systems with Machine Learning		
Towards Consistent Data Representation in the IoT Healthcare Landscape	Reda R., Piccinini F. & Carbonaro A.	2018
Towards the Design of a Conceptual Framework for the Operation of Intensive Care Units Based on Big Data Analysis	Markopoulos, D., Tsolakidis, A., Karanikolas, N. & Skourlas C.	2021
DataOps Lifecycle with a Case Study in Healthcare	Shaimaa B., Ghalwash A., Harb H.	2023
Big data in healthcare: management, analysis and future prospects	Dash S., Shakyawar, S. K., Sharma, M. & Kaushik S.	2019