

**MASTER DATA MANAGEMENT- MALLIEN VERTAILU HAMKIN
TIETOHALLINNOLE**



Ammattikorkeakoulututkinnon opinnäytetyö

Hämeenlinnan korkeakoulukeskus, tietojenkäsittelyn koulutusohjelma

Kevät, 2019

Annamari Mäenhovi

Tietojenkäsittelyn koulutusohjelma
Hämeenlinnan korkeakoulukeskus

Tekijä	Annamari Mäenhovi	Vuosi 2019
Työn nimi	Master data management- mallien vertailu HAMKin tietohallinnolle	
Työn ohjaaja/t	Lauri Salminen, Tero Keso	

TIIVISTELMÄ

Tämän työn tarkoituksena oli tehdä kvalitatiivista vertailevaa analyysiä toimeksiantajan, Hämeen ammattikorkeakoulun tietohallinnon, antamien master data management- mallien pohjalta. Tavoitteena oli etsiä HAMKin TKI- hankesalkun hallintajärjestelmässä liikkuvan datan hallintaan parhaiten sopiva malli tai osia malleista. Mallien tutkimisessa keskityttiin ensisijaisesti datan laadullisiin ominaisuuksiin ja toissijaisesti datan omistajuuskysymyksiin.

Tutkimusosiossa tarkastellaan toimeksiantajan järjestelmässä liikkuvaa dataa ja määritellään datan pohjalta tarpeita ja kysymyksiä, joihin master data management- mallien tulisi antaa vastauksia. Kvalitatiiviseen vertailuun valikoitui neljä mallia, joita tarkastellaan ja pisteytetään kolmiportaisella asteikolla laadun, omistajuuden, tietoturvan ja käytettävyyden kannalta.

Vertailun tuloksista käy ilmi, että master datan hallintaa voidaan tarkastella useasta eri näkökulmasta ja ettei yhtä ainoaa mallia ole tarkoituksenmukaista käyttää sellaisenaan. Todettiin myös, että sekä korkealla abstraktitasolla liikkuville, että konkreettisemmille malleille on tarvetta. Abstraktitasoisemmat mallit esittävät tarpeellisia kysymyksiä ja huomioita organisaation datanhallinnasta, kun taas konkreettisemmat mallit tarjoavat käytännönläheisempiä lähestymistapoja omistajuus- ja laatuksymyksiin.

Avainsanat Master data, master data management, data

Sivut 42 sivua, joista liitteitä 1 sivua

Degree Programme in Business Information Technology
Hämeenlinna University Center

Author	Annamari Mäenhovi	Year 2019
Subject	Comparative analysis of master data management models for IT-administration of HAMK	
Supervisors	Lauri Salminen, Tero Keso	

ABSTRACT

The aim for this thesis was to make qualitative comparative analysis of master data management- models provided by the commissioner, IT administration of Häme University of Applied Sciences. The aim was to search the most appropriate model or parts of the models for managing the master data in R&D project data management system. The focus of the study of the master data management models were firstly in the qualitative attributes of the data and secondly regarding the questions of the ownership of the data.

The research part discusses the needs and essential questions regarding the commissioner's data. This is the theoretical base which the master data management models are compared against. Four models were chosen to be analyzed and compared via three step scoring scale. Criteria for comparative analysis consists of questions regarding the quality, ownership, security and usability of the models.

The results show that master data management can be viewed from several viewpoints and using only one individual model is not justifiable. It was also concluded that there are a case sensitively appropriate uses for both highly abstract and concrete models.

Keywords Master data, master data management, data

Pages 42 pages including appendices 1 pages

SISÄLLYS

1	JOHDANTO.....	1
2	DATASTA.....	2
2.1	Mitä datalla tarkoitetaan ja miksi se on tärkeää	2
2.2	Master data ja muut datatyypit	4
2.3	HAMKin TKI-hankesalkun hallintajärjestelmässä liikkuva data	4
3	MASTER DATA MANAGEMENT JA KOKONAISARKKITEHTUURI.....	8
3.1	Kokonaisarkkitehtuuri HAMKissa	8
3.2	Master data management.....	9
4	MALLIT	11
4.1	Kerrostettu datavarasto	11
4.2	Datan laatu: organisationaalisten toimintatopojen määrittely	12
4.3	Muutokset rooleissa, vastuissa ja omistajuudessa master data managementia organisoidessa	14
4.4	Tarve data governancelle: tapaustutkimus.....	15
4.5	Data governancen suunnittelu.....	20
5	TUTKIMUKSEN PERUSTA	24
5.1	Laadullisesta tutkimuksesta	24
5.2	Pohdintaa datasta ja omistajuudesta aineiston pohjalta	25
6	VERTAILUTUTKIMUS.....	30
6.1	Vertailu	30
6.2	Vertailun tulokset.....	36
7	JOHTOPÄÄTÖKSET	37
7.1	Tutkimuksen tuloksista	37
7.2	Suosituksia datan hallintaan	38
7.3	Pohdintaa lopputuloksesta	39
	LÄHTEET.....	40

Liitteet

Liite 1 Arter Framework metamalli

1 JOHDANTO

On yleisesti tiedossa, että laadukas data tarjoaa kilpailullista etua organisaatioille ja tämä korostuu etenkin suurten toimijoiden tapauksessa. Tällöin on käytettävissä valtaisa määrä dataa erilaisista lähteistä, mutta asianmukainen strategia tämän hyödyntämiseen saattaa puuttua. Master data management, ydintiedon hallinta, onkin harkinnan arvoinen menetelmä omaksuttavaksi organisaatioihin, alati kasvavien data- ja informaatiovirtojen hallintaan. (Spruit & Pietzka, 2015.)

Tämän opinnäytetyön toimeksiantajana on Hämeen ammattikorkeakoulun, HAMKin, tietohallinto. Tavoitteena on tutkia korkeakoulun TKI-hankesalkun hallintajärjestelmässä liikkuvan datan pohjalta master data management- malleja ja sitä, miltä osin tai mikä niistä olisi sopivin järjestelmässä liikkuvan datan laadulliseen parantamiseen. TKI- hankkeilla tarkoitetaan tutkimus, kehitys ja innovointihankkeita. Tämän opinnäytetyön toisena tavoitteena on myös kartoittaa sitä, millä tavoin datan omistajuutta voitaisiin määrittellä, lähtökohtana data itsessään, ei järjestelmä, johon se on tallennettu.

Tutkimusaineistona on toimeksiantajan antamat materiaalit erilaisista master data management- malleista ja tapauskohtaisista tutkimuksista, sekä kuvaukset TKI-hankesalkun hallintajärjestelmässä liikkuvasta datasta. Tutkimus tapahtuu kvalitatiivisen, eli laadullisen, vertailevan analyysin keinoin, vertailemalla aineistosta valikoituja malleja keskenään ja suhteessa datan ja korkeakouluympäristössä toimimisen asettamiin vaatimuksiin. Näitä edellä mainittuja, huomioon otettavia, vaatimuksia on muun muassa Hämeen ammattikorkeakoulussa käytetty kokonaisarkkitehtuurin viitekehys Arter Framework.

Tutkimuskysymykset ovat:

1. Miltä osin tai mikä valituista master data management- malleista olisi sopivin TKI- hankesalkun hallintajärjestelmässä liikkuvan datan laadulliseen parantamiseen?
2. Miltä osin tai mikä valituista master data management- malleista olisi sopivin TKI- hankesalkun hallintajärjestelmässä liikkuvan datan omistajuuden määrittämiseen perustuen dataan itseensä, ei järjestelmään?

2 DATASTA

Tässä luvussa käsitellään sitä, mitä datalla tarkoitetaan, miksi se on tärkeää ja minkälaista erityyppistä dataa on olemassa. Alaluvussa 2.3 käydään läpi HAMKin dataa, johon tämä opinnäytetyö pohjaa.

2.1 Mitä datalla tarkoitetaan ja miksi se on tärkeää

Informaatiotieteissä kolme yleisesti tunnustettua, perustavanlaatuaista ja toisiinsa tiiviisti linkittyneitä ydinkäsitettä ovat data, informaatio ja tietämys. Data on tässä hierarkkisessa jäsenyyksessä alimpana, informaation ja lopulta tietämyksen rakennuspalikkana. Dataa ovat esimerkiksi aistinelimemme havaitsemat ärsykkeet ulkoisesta maailmasta ja symbolisten merkkien sarjat, kuten yksittäiset numeeriset elementit. Eli kaikenlainen tieto, millä ei ole mitään tulkittavissa olevaa merkitystä ilman kontekstia. (Zins 2007, s. 479–493.)

Informaatiota syntyy, kun dataa tulkitaan ja se vaatii interaktiota ihmismielen kanssa. Eli sitä on tulkittu ja se on saanut muodon, jolla on jokin tietty tulkittavissa oleva merkitys. Informaatiolle ominaista on se, että se syntyy aina yksilön kognitiivisten prosessien tuloksena. (Zins 2007, s. 479-493.)

Samoin kuin informaatio, tietämys on myös ihmismielen prosessoinnin lopputuote. Informaation, tai tiedon, voidaan ajatella sijaitsevan eri paikoissa; informaatio on olemassa yksilön ulkopuolella ja tietämys on osa yksilön omaa kokemusmaailmaa ja sijaitsee vain hänen mielessään. Kummakin ovat kuitenkin valideja representaatioita ympäröivästä todellisuudesta. (Kari 2010, s. 116.)

Tiedonkäyttö ohjaa kaikkea tekemäämme ja teemme päätökset sen pohjalta, niin arjessa kuin työympäristössä. Informaatio ei sinällään ole kuitenkaan minkään arvoista, vaan se saa merkityksen vasta käyttäjien kautta. Tästä syystä hyödyllisyys ja käyttöasteet määrittävät tietojärjestelmien todellisen hyötyarvon. (Kari 2010, s. 116-117 ; Savolainen 2009.)

Yritysten ja organisaatioiden keräämät datamäärät ovat kasvaneet niin nopeasti, että kehityksen mukana on vaikea pysyä. Dataa on hyödynnetty ja varastoitu vaihtelevalla menestyksellä, vaikka tehokkaasti hyödynnetystä datasta on mahdollista saada merkittävää liiketoiminnallista hyötyä. Tallennettävien tietomäärien kasvaessa myös datan varastoinnin kustannukset ovat kasvaneet ja tämän myötä tallennetun datan hyödyntämisen rooli korostuu entisestään. (MIT Sloan 2012, s. 18-19.)

Yrityksille tai organisaatioille elintärkeä ja kriittinen data tulisi pyrkiä tunnistamaan, jotta datasta saataisiin maksimaalinen liiketoiminnallinen hyöty. Esimerkiksi järjestämättömän datan tapauksessa tämä vaatii dataa

käyttävien työnkulkujen selvittämistä, eli mitkä työnkulut luovat tai muokkaavat, hakevat ja uudelleenkäyttävät erilaista tietojärjestelmiin tallennettua järjestämätöntä dataa, kuten esimerkiksi kuvia ja muistikirjoja. (MIT Sloan 2012, s. 19.)

Tällaista kriittistä dataa määriteltäessä tullaan samalla määritelleeksi parametrejä kokonaisarkkitehtuurille, kun johto selvittää tapoja hallinnoida liiketoimintaa. Tämän jälkeen tietohallinto voi alkaa rakentamaan tarvittavia järjestelmiä tai toteuttamaan vaadittavia muutoksia. Dataa tulisi käyttää hyödyksi myös liiketoimintaprosessien parantamisessa, mikä tosin vaatii yritykseltä tai organisaatiolta jatkuvaa datan analysointia ja aktiivista työtä prosessien kehittämiseksi. (MIT Sloan 2012, s. 20.)

Datan tallennuskustannusten vähentämiseen tehokas keino on kustannusten läpinäkyväksi tekeminen. Informaation tallentamiselle ei osata ajatella vielä kovin luontevasti kustannuksia, koska tietomäärien kasvu on tapahtunut niin nopeasti. Ei ole kulunut kauaa siitä, kun tallennettavat tietomäärät ja niihin linkittyvät kustannukset olivat pieniä. Kun datan varastointiin liittyvät kustannusrakenteet tehdään läpinäkyviksi, johtajat ja esimiehet suhtautuvat kriittisemmin turhan datan tallentamiseen ja siihen, minkälainen data on tärkeää. (MIT Sloan 2012, s. 20.)

Valtaosa kerätystä datasta on järjestelemätöntä, koska sitä tuotetaan esimerkiksi Microsoftin toimisto-ohjelmilla. Tällöin data koostuu muun muassa kuvista, erilaisista muistikirjoista, asiakirjoista ja viesteistä. Järjestelmätön data on luonteeltaan vaikeammin hyödynnettävää kuin järjestetty data ja sen käyttäminen vaatii usein manuaalista työtä. (MIT Sloan 2012, s. 19.)

Siinä, missä monissa yrityksissä IT-osasto on sitoutunut ja järjestänyt datan turvallisen varastoinnin, sen hyödyntämiseen ei ole välttämättä käytetty tarpeeksi resursseja. Osa tästä ongelmasta johtuu siitä, että It-osastolla on päävastuu datasta, mutta tietomäärien hyödyntämiseen tarvittaisiin eri liiketoiminta- tai organisaatioyksikköjen rajoja ylittävää yhteistyötä ja sitoutumista johdolta. Tämänkaltainen yhteistyö, sekä datan hyödyntämisen vaatimien kustannusten minimointi, sekä hyötyjen maksimointi, vaatii taas vastuunjaon selvittämistä. Samaiset organisaatiotason esteet tai muurit saattavat myös vaikeuttaa jonkin tietyn yksikön tuottaman tai hallinnoiman tiedon saavuttamista. (MIT Sloan 2012, s. 19.)

Paremmalla dokumenttien hallinnalla, sekä sosiaalisesta mediasta kerätyllä järjestelemättömällä datalla on mahdollista saavuttaa merkittäviä hyötyjä. Tällöin voidaan parantaa esimerkiksi asiakaspalvelua ja tehostaa tiedon jakamista. Järjestelemättömän datan hyödyntäminen vaatii kuitenkin datan uudelleenjärjestämistä, sen kategorisointia lisäämällä esimerkiksi erilaisia metatietoja tai niin sanottuja tageja, että indeksointia. Järjestämätön data ei palvele järjestämättömiä liiketoimintaprosesseja, vaikka niin saatetaankin ajatella joissain tapauksissa. (MIT Sloan 2012, s. 19.)

Datan huono laatu voi myös olla syyllisenä organisaation prosesseissa tapahtuviin menetyksiin. Eli datan laatu ei ole pelkästään tietojärjestelmätasolla huomioon otettava ja tärkeä asia, vaan sillä on suoraa merkitystä organisaatiotasolla. Yrityksessä tehtävä päätöksenteko vaatii datan analyysiä ja datan huonolla laadulla on myös tältä osin suora vaikutus liiketoimintaprosesseihin. (Storey, Dewan, Freimer 2012, s. 434.)

2.2 Master data ja muut datatyytit

Metatiedolla tarkoitetaan joko dokumentteja tai dataa kuvailevia tietoja. Metadata kertoo tallennetun tiedon, datan tai vaikkapa asiakirjan syntyyn, ymmärtämiseen, tekijään, aikaan ja paikkaan liittyviä asioita, eli toisin sanoen, se antaa datalle kontekstin. (Henttonen & Valtonen 2010, s. 214.)

Master datalla, ydintiedolla, tarkoitetaan nimenmukaisesti tietojärjestelmiin tallennettua ydininformaatiota. Se on dataa, millä on suuri merkitys sitä käyttävän tahon toimintaan. On myös huomionarvoista, että master data sisältävää myös historia- ja esiintymistiheys-tietoja. (Pekkola & Vilminko-Heikkinen, 2019, s. 76 ; Van Der Lans 2012, s. 219-220.).

Mitä master dataan tulee, niin datan oikeellisuus on äärimmäisen tärkeää datan käytettävyyden takia. Englanninkielinen termi master viittaa siihen, että data indikoi mikä tietojärjestelmiin tallennetuista historiatiedoista tai datan versioista on ajantasainen ja oikea. Samaa tietoa voi siis olla myös muissa järjestelmissä eriävinä versioina, mutta master data on loppukäyttäjän näkökulmasta ajankohtainen ja oikea. (Van Der Lans 2012, s. 220.)

Henkilötietoja sisältävää dataa määrittää GDPR, yleinen tietosuojasetus (engl. General data protection regulation), joka on Euroopan Unionin henkilötietojen käsittelyä säätelevä laki. Henkilötietojen käsittelyä säätelevät periaatteet on määritelty artiklassa viisi. Henkilötietoja tulee tietosuojasetuksen mukaan käsitellä ”lainmukaisesti, asianmukaisesti ja rekisteröidyn kannalta läpinäkyvästi”, henkilötietoja ei tule kerätä tarpeettomasti. Laki määrää keräämään vain käyttötarkoitukseen nähden tarpeellisia tietoja, joiden on oltava täsmällisiä ja jotka on voitava päivittää tarpeen vaatiessa. Tietoja saa kerätä vain tiettyyn ja lailliseen tarkoitukseen, eikä niitä saa käyttää muuhun. Tietojen eheys ja luottamuksellisuus ovat vaatimuksia säilyttämiselle ja rekisterinpitäjällä on vastuu, sekä tarpeen vaatiessa todistustaakka siitä, että tietoja käsitellään lain vaatimilla tavoilla. (EU:n yleinen tietosuojasetus 2016/679 §5.)

2.3 HAMKin TKI-hankesalkun hallintajärjestelmässä liikkuva data

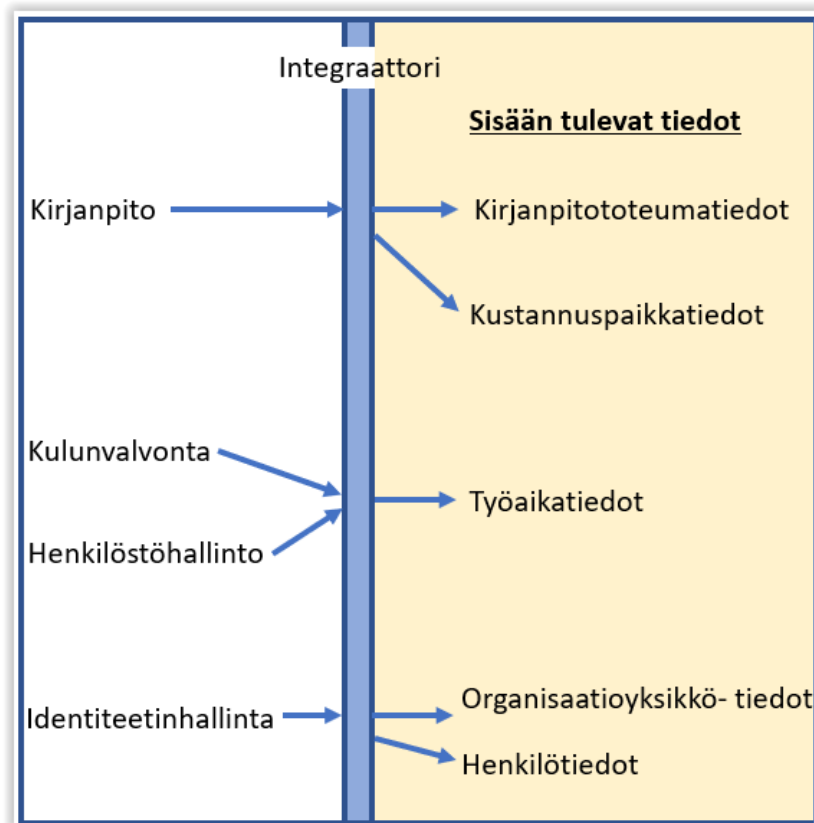
Hämeen ammattikorkeakoulussa käytettävässä TKI-hankesalkun hallintajärjestelmässä liikkuu kirjanpitoon, kulunvalvontaan ja identiteetin hallin-

taan liittyvää dataa. Yksityiskohtaisemman seurannan tasolla järjestelmään tallennetaan monenlaista projektikohtaista tietoa, kuten esimerkiksi projektin perustiedot, käsittäen muun muassa ajankohtatietoja, määrittelyn salaustason, nimen, lyhenteen, hankkeen tyyppin ja projektin tilan. Perustietojen lisäksi järjestelmään tallennetut tiedot on jäsennetty muilla määreillä erilaisiksi tietotyypeiksi, kuten osapuoliin liittyväksi tiedoksi tai yhteenvetotiedoksi, joka on koostettu muiden tietojen pohjalta. Osa tietotyypeistä on jäsennetty vielä alemmalla tasolla omiksi alatietotyypeikseen, kuten esimerkiksi projektin sisältö, joka sisältää jaottelun ideointiin, valmisteluun, toteutukseen ja projektin hylkäämistä koskeviin tietoihin.

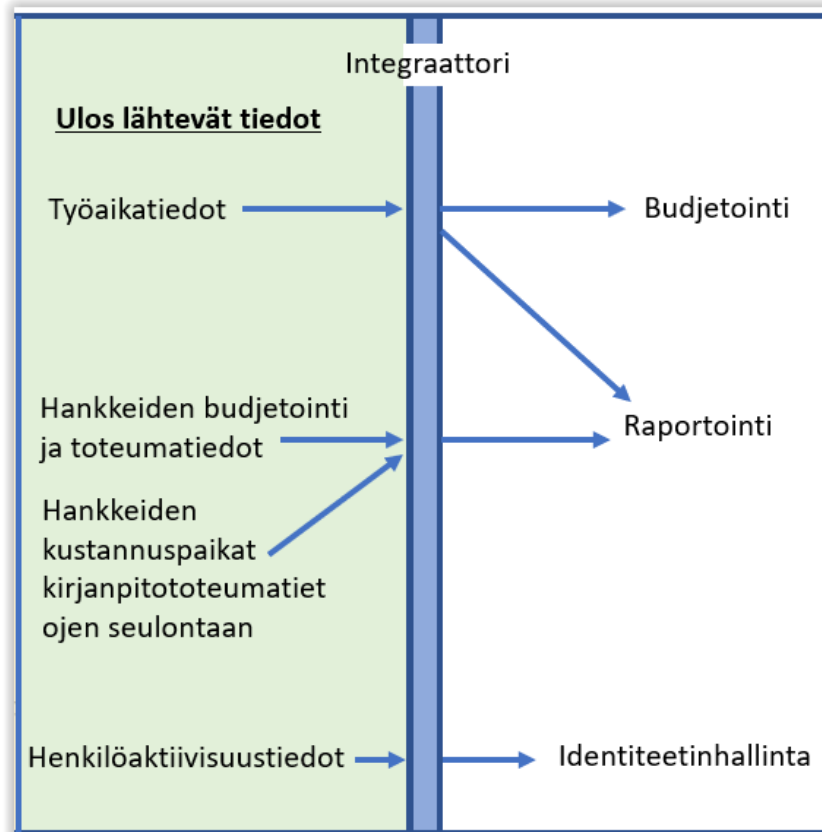
Esimerkiksi työajan kohdennus perustuu kulunvalvonnasta ja henkilöstöhallinnosta saataviin tietoihin, jotka kohdennetaan asianmukaisille kustannuspaikoille tuntimääräisinä ja joita käytetään muun muassa raportointiin. Järjestelmään tulevat työaikatiedot ovat kerran vuorokaudessa päivittyviä, budjetointiin ja raportointiin käytettäessä, eli järjestelmästä ulos lähtiessä ne päivittyvät 15 minuutin välein.

Projektikohtaisten tietojen, eli projektikohtaisesti määriteltyjen kustannuspaikkojen osalta järjestelmään tulee kirjanpilototeumatietoja, eli hankkekohtainen euromääräinen toteuma kirjanpidosta. Tämä on kerran vuorokaudessa toimitettava tieto. Kirjanpilototeumatiedot seulotaan projektitietojen avulla ja lähetetään kerran vuorokaudessa raportointijärjestelmään.

Järjestelmään kirjanpidosta, kulunvalvonnasta, henkilöstöhallinnosta ja identiteetinhallinnasta TKI- hankesalkun hallintajärjestelmään tulevaa tietoa havainnollistetaan kuvassa 1. Integraattori yhdistää tulevan datan tietyn tietotyypin dataksi. Kuvassa 2. havainnollistetaan järjestelmästä budjetointiin, raportointiin ja identiteetinhallintaan lähtevää dataa.



Kuva 1. TKI- hankesalkun hallintajärjestelmään tuleva data.



Kuva 2. TKI-hankesalkun hallintajärjestelmästä lähtevä data.

HAMKissa eri järjestelmissä liikkuvasta datasta koostetaan raportteja ja katsauksia muun muassa Power BI:llä. TKI-hankesalkun hallintajärjestelmästä, sekä muista lähteistä kerätystä datasta koostettua informaatiota käytetään siis erilaiseen raportointiin ja seurantaan.

3 MASTER DATA MANAGEMENT JA KOKONAISARKKITEHTUURI

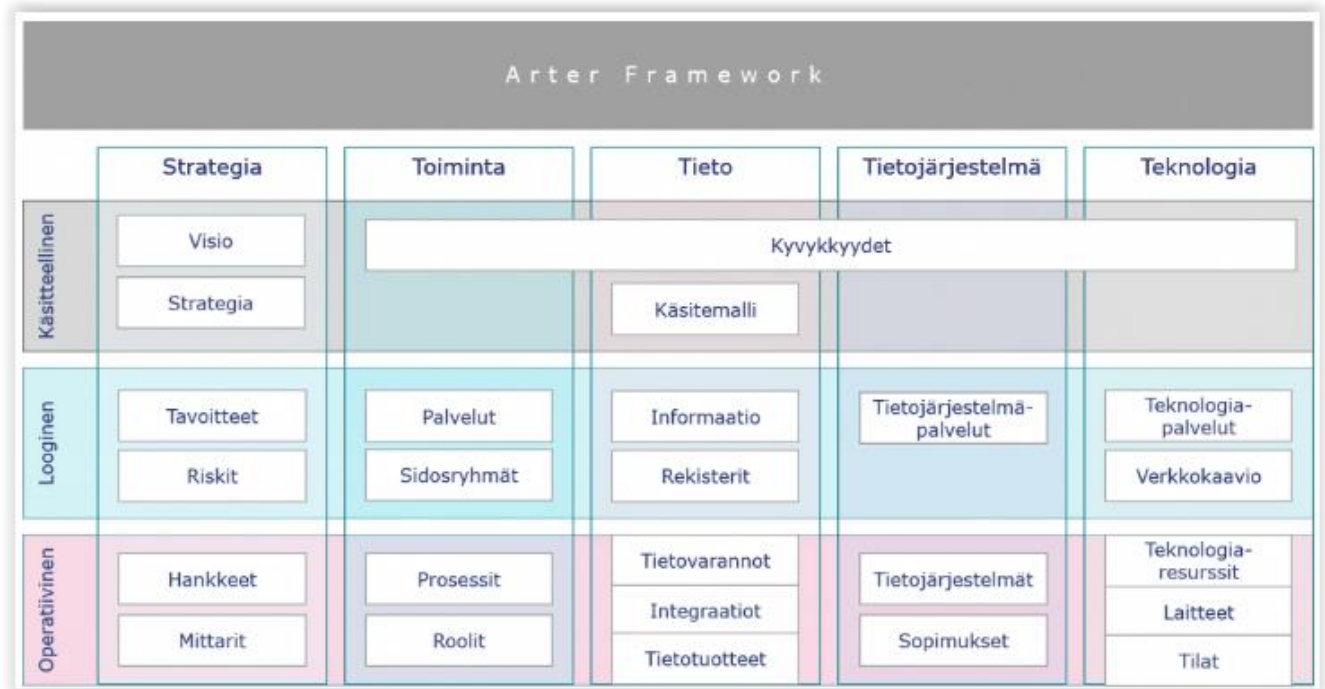
HAMKissa on käytössä suomalaisen Arterin laatima Arter Framework- kokonaisarkkitehtuurin viitekehys. Kokonaisarkkitehtuuri määrittelee raamit myös datan hallinnalle korkeakouluympäristössä. Tässä luvussa käydään myös läpi sitä, mitä master data management on ja mihin sitä käytetään.

3.1 Kokonaisarkkitehtuuri HAMKissa

Kokonaisarkkitehtuurilla tarkoitetaan kokonaisvaltaista toiminnan mallia, jonka tarkoituksena on varmistaa yrityksen tai organisaation toiminnan yhdenmukaisuus suhteessa sen kaikkiin osa-alueisiin tai komponentteihin. Tarkoituksena on luoda kuvaukset nykytilasta ja niiden suhteesta tulevaisuuden tavoitteisiin, eli dokumentoida nykyhetki ja minkälainen tahtotila yrityksellä tai organisaatiolla on toimintansa kehittämisen suhteen. (Kartturi 2013, s. 16.)

HAMKissa käytetty Arter Framework pohjaa ARC- ohjelmistoon, jonka ominaispiirteinä on joustavuus ja hankkeissa kehitettävän muutoksen selkeä havainnollistaminen. Viitekehyksessä keskitytään myös kyvykkyyksiin kokonaisarkkitehtuurin näkökulmasta. Arter Frameworkin metamalli kaikine komponentteineen on liitteessä 1. (Arter n.d.)

Arterin viitekehyksessä strategiaa painotetaan nostamalla se omaksi näkökulmaksi. Muita frameworkin näkökulmia ovat toiminta, tieto, tietojärjestelmä ja teknologia, jotka ovat kaikki nähtävissä kuvassa 3. Arterilla kokonaisarkkitehtuuri on määritelty strategisen suunnittelun osana ja hankkeet työkaluiksi, joilla strategiaa toteutetaan. Kyvykkyydet ovat toteutettavien hankkeiden sivutuote ja toiminnan kehittämiseen käytettävä resurssi. (Aunola 2019.)



Kuva 3. Arter Framework (Aunola 2019).

3.2 Master data management

Kuten luvussa kaksi käsiteltiin, informaatiolla ja datalla on suuri merkitys faktoihin perustuvassa päätöksenteossa. Tämä perustuu erityisesti käytävissä olevan datan laatuun, koska käytettävän datan tulee olla oikeaa, eheää ja käytettävää. Datan laadun hallintaan on kehitetty oma menetelmänsä, master data management, ydintiedon hallinta. (Pekkola & Vilminho-Heikkinen 2018, s. 76; Van Der Lans 2012, s. 217.)

Datan hallinta on elintärkeää yrityksille ja organisaatioille niiden kaikissa toiminnoissa ja prosesseissa. Paradoksaalisesti dataa ei sen tärkeydestä huolimatta hyödynnetä aina niin tehokkaasti kuin olisi mahdollista. Järjestelmiin tallennettu data tai siihen liitetyt metatiedot tai tunnisteet eivät ole välttämättä keskenään yhteensopivia tai datasta tallennetaan keskenään ristiriitaisia versioita. (Van Der Lans 2012, s. 217-218.)

Master data management on menettelytapa, prosessi, joka on kehitetty pitämään data ajantasaisena, yhtenäisenä, eheänä ja yleisesti ottaen käytettävänä. Se on jatkuvaa, aktiivista toimintaa datan laadun ja käytettävyyden parantamiseksi ja vaatii yritykseltä tai organisaatiolta sitoutumista ja tarpeeksi resursseja. (Van Der Lans 2012, s. 222.)

Master data managementissa tarkoituksena on muun muassa tunnistaa master datan lähteet, varmistaa onko järjestelmiin syötetty uusi data uutta ydintietoa vai vanhan datan uusia versioita. Se on virheellisen ja vanhentuneen datan poistoa järjestelmistä, sekä erilaisten taksonomisten datan jäsenyystapojen luomista ja ylläpitämistä. Master data managementiin on

myös käytettävissä erilaisia työkaluja, ydintiedon hallintajärjestelmiä (eng. master data management systems, MDMS). (Van Der Lans 2012, s. 222.)

Master data managementissa keskitytään yrityksen tai organisaation master datan, laadulliseen parantamiseen. Datan korkea laatu on elintärkeää, jotta sitä voitaisiin käyttää koko yrityksen tai organisaation laajuisesti. Tämän saavuttamiseksi on usein välttämätöntä tehdä muutoksia ja parannuksia käytössä oleviin prosesseihin tai teknologiseen arkkitehtuuriin tai laitteistoihin. (Pekkola & Vilminko-Heikkinen 2019, s. 76.)

Master data management on kokonaisvaltainen, koko organisaation laajuinen, lähestymistapa dataan. Tämä tarkoittaa sitä, että menettelytavassa ei keskitytä pelkästään teknisiin ratkaisuihin, vaan myös esimerkiksi ihmisiin ja heidän rooleihinsa, työnkulkuihin, prosesseihin ja hallintorakenteisiin. (Pekkola & Vilminko-Heikkinen 2019, s. 76.)

Data governance, eli datan hallinnointi, on olennainen osa master data managementia. Data governancessa tarkastellaan organisaation tasolla sitä, miten dataa hallinnoidaan, sekä dataa käyttäviä prosesseja ja rakenteita. Tämä lähestymistapa on tekninen ja peräänkuuluttaa johtajuutta ja omistajuutta datan hallinnoinnissa. Rooleihin ja vastuisiin keskittymisen vuoksi data governance kytkeytyy tiiviisti organisaation sosiaalisiin prosesseihin, kuten esimerkiksi liiketoiminnan ja IT:n väliseen yhteistyöhön, joka on välttämätöntä, jotta datan asema strategisena resurssina saadaan turvattua. (Pekkola & Vilminko-Heikkinen 2019, s. 77.)

Datan omistajuuden määrittäminen on avainasemassa master data managementissa. Omistajuudella tarkoitetaan sitä, että on määritelty taho, joka tekee päätöksiä datan ominaisuuksiin, saavutettavuuteen, turvallisuuteen ja päivitettävyyteen liittyen, eikä sitä, että data olisi konkreettisesti omistettu ja vain tietyn tahon käytettävissä. (Pekkola & Vilminko-Heikkinen 2019, 77.) Data governancessa kyse on siitä, kenellä on oikeus ja vastuu tehdä päätöksiä, koskien yrityksen tai organisaation data-resursseja (Khatri & Brown 2010, s. 149).

Master data managementissa painotus ei ole tekniikassa, vaan organisaation tasolla, koska asianmukaisen, eheän, käytettävän ja yhtenäisen master datan kerääminen, yhdistäminen ja jakaminen vaatii erilaisia käytänteitä, palveluita ja infrastruktuureja. Teknologia on kuitenkin myös elintärkeä osa organisaation master data managementia, mutta se pitäisi hahmottaa osaksi kokonaisuutta, eikä erilliseksi osa-alueekseen. (Pekkola & Vilminko-Heikkinen 2019, s. 77-78.)

4 MALLIT

Tässä opinnäytetyössä tehtävän vertailun keskiössä ovat master data management mallit. Tässä luvussa esitellään tiivistelmät toimeksiantajan aiheistosta valikoiduista malleista.

4.1 Kerrostettu datavarasto

Yksi kustannustehokkuuteen tähtäävä ja tieteellisesti tutkittu datanhallinnan malli on kerrostetun datanhallinnan malli (tiered storage). Malli on kuvattu hyvin yleisellä ja abstraktilla tasolla, mutta tarjoaa hyvin yleishyödyllisiä menetelmiä etenkin järjestelemättömän datan hyödyntämiseen ja ottaa kantaa siihen, millaisissa tilanteissa koko organisaation tai liiketoimintakonaisuuden käsittäviä vastuunjakokysymyksiä tulisi esittää. (MIT Sloan 2012, s. 19-20.)

Kerrostetussa datavarastossa on nimenmukaisesti kolme erilaista tasoa, jolla dataa tallennetaan. Tasot on määritelty sen mukaan, miten tärkeää, ajankohtaista ja eheää niihin tallennettavan datan tulee olla, sekä kuinka usein kutakin tasoa käytetään tai kuinka vaivattomasti ne tulee olla saavutettavissa. (MIT Sloan 2012, s. 19-20.)

Ensimmäisellä tasolla datavarastossa säilytetään dataa, joka on kaikista kriittisintä yritykselle tai organisaatiolle. Ensimmäisen tason datan tulee olla kaikista ajantasaisinta, paikkansapitävintä, eheintä ja helpoiten käytävissä olevaa. Tämän tason dataa käytetään siis eniten ja sen säilyttäminenkin on kaikista kalleinta. (MIT Sloan 2012, s. 19-20.)

Toisen tason varasto tarjoaa verkkotallennuspaikan järjestämättömälle datalle ja verkottaa järjestetyn datan. Tämän tason datan luotettavuudella tai ajantasaisuudella ei ole niin suurta merkitystä kuin ensimmäisellä tasolla, sekä sen ylläpitokustannukset ovat huokeammat. (MIT Sloan 2012, s. 20.)

Kolmannen tason datavarasto on edullisin ylläpitää, mutta datan hakeminen sieltä on kaikista kalleinta. Kolmannella tasolla data voidaan tallentaa esimerkiksi nauhatallennusta tai optical jukebox- järjestelmiä käyttäen. Kolmannelle tasolle siirtyvää ja siellä säilytettävää dataa saattaa määrittää maakohtaiset lait ja säädökset datan ja informaation säilyttämisestä, sekä yrityksessä pitää määritellä se, minkälaisissa tapauksissa esimerkiksi vähän käytetty data siirtyy kolmannen tason varastoon. (MIT Sloan 2012, s. 20.)

4.2 Datan laatu: organisationaalisten toimintatopojen määrittely

Storey, Dewan & Freimer (2012, s. 434, 436-441) tutkimuksen keskiössä on yksittäisten ihmisten vaikutus organisaatiossa käytettävään dataan, niin osana prosesseja kuin yksittäisinä työntekijöinä. Artikkelissa esitellään lukuisia matemaattisia malleja tehtyjen havaintojen ja väitteiden tueksi.

Storey ym. nostavat esille sen, että datan laatu nähdään usein lopputavoitteena, vaikka sen tulisi olla organisaatiotason tavoitteisiin pääsyn mahdollistaja. Datan laadun parantamiseen tarvittavien toimien on määritelty käsittävän kustannusten arvioinnin ja prosessivastuiden määrittelyn. Data quality: Setting organizational policies- tutkimuksen päämäärinä oli organisaatioissa usein ilmenevien datan laadullisten ongelmien tunnistaminen, dataan liittyvien käytänteiden ehdottaminen organisaatioille ja datan laadunhallinnallisten ratkaisujen esittäminen. Storey ym. esittävät mallin, jolla voidaan ennustaa yksilöiden ja organisaatioiden tekemiä laadullisia valintoja. (Storey ym. 2012, s. 434-435.)

Tutkimuksessa kuvataan prosesseja ja tapahtumasarjoja, jotka vaikuttavat datan laatuun ja loppukäyttöön. Dataa tuottavien tahojen toimet, jotka määrittävät datan laatua tai rajoittavat sen käyttöä vaikuttavat myös muiden tahojen mahdollisuuksiin käyttää tätä nimenomaista dataa. Eli datan luomisessa tai tallentamisessa tapahtuvat virheet vaikuttavat datan laatuun ja käyttöön sen syntyhetkestä lähtien. Storeyn ym. esittämän mallin kuvataan olevan tarpeeksi joustava, jotta muut organisaatiot voivat soveltaa ja muokata sitä, kuten esimerkiksi lisäämällä omia työnkuluja ja prosessejaan. Malli keskittyy mallintamaan tärkeimpiä dataan liittyviä toimia ja huomioita. (Storey ym. 2012, s. 435.)

Mallissa tarvittavissa oletuksissa on otettu huomioon se, että valitut muuttujat ovat enemmän tai vähemmän mitattavissa, joten mallin käyttöön ottavien organisaatioiden on mahdollista mitata sitä, minkälainen vaikutus mallin käyttöönotolla on. Storeyn ym. mukaan näitä oletuksia on kolme. Datan laatutyön hinta, eli korkeampi laatu maksaa enemmän kuin huono. Laadun arvo, eli päätöksien arvo voi nostaa datan laatua, mutta ei suorassa suhteessa, vaan kuperana käyränä. Kolmanneksi: työntekijän tuottaman datan arvo on komplementaarista, eli yhden aktiviteetin laadullinen paraneminen nostaa muidenkin aktiviteettien laadullista arvoa. (Storey ym. 2012, s. 436.)

Organisaatiotasolla keskinäisten riippuvuuksien tunnistaminen informaatioteknologian, organisaation ja prosessien suhteen on seuraava askel. Artikkelissa esitetään kaksi vaihtoehtoista olettamusta tämän kaltaisista organisationaalisisista riippuvuussuhteista. Nämä olettamukset ovat keskinäistä riippuvuutta, mutta ei synergiaa, sekä keskinäistä riippuvuutta ja synergiaa. (Storey ym. 2012, s. 436.)

Ensin mainitussa tapauksessa, eli ei-synergisessä keskinäisessä riippuvuussuhteessa, toimien arvo organisaatiolle on yksittäisten osa-alueiden summa. Eli toimien arvo ei ole synergistä, mutta keskinäiset riippuvuussuhteet pohjaavat yhteiseen datan käyttöön. Toisessa tapauksessa, eli synergisen keskinäisen riippuvuuden suhteessa, toimien arvo organisaatiolle on suurempi kuin niiden arvo olisi yksilölle. Eli tässä tapauksessa toimen arvo on organisaatiolle huomattavasti suurempi, koska se vaikuttaa muihin organisaation prosesseihin, vaikka sillä ei olisikaan kuin vähäistä arvoa yksittäiselle työntekijälle. (Storey ym. 2012, s. 436.)

Tältä pohjalta Storey ym. ovat laatineet kaksi edellisiin tapauksiin pohjaavaa teoreemaa. Ensimmäisen tapauksen, eli ei-synergisessä keskinäisessä riippuvuussuhteessa, yksittäiset työntekijät eivät panosta laatuun niin paljon kuin pystyisivät, koska työntekijät luontaisesti valitsevat laatutason, joka on alempi kuin organisaatiolle ihanteellinen taso. Tätä tilannetta verrataan yhteismaan ongelmaan, eli skenaarioon, missä yhteiskäytössä olevaa resurssia käytetään liikaa, koska yksilöt toimivat omien yksittäisten intressiensä mukaan. Toisessa teoreemassa, eli synergisen keskinäisen riippuvuussuhteen tapauksessa, työntekijät panostavat laatuun vielä edellisen teoreeman kuvaamaa tapausta vähemmän. (Storey ym. 2012, s. 437.)

Tutkimuksessa on esitetty työntekijä- ja organisaatioteoriatason käytänteitä edellä kuvattujen datan laatuongelmien ratkaisuksi. Työntekijälähtöisiä ratkaisuja on henkilökohtainen tietojenkäsittely, eli toimien konsolidointi, jolloin datan kaappaaminen ja prosessointi on yhdistetty päätöksentekoon. Muita työntekijätason toimia on työntekijöiden tukitoimet, jolloin prosessissa ylimpänä oleva, lähimpänä asiakasta toimiva, työntekijä tekee päätökset ja moniammatilliset ja monitaitoiset tiimit. Organisaatioteoriaan pohjaavat ratkaisut ovat ohjelma prosessien avaintyöntekijöiden kannustimille, datan laadulle, sekä laadun mittaukselle ja datan omistajuudelle, eli muutoksille datan omistajuudessa. (Storey ym. 2012, s. 437)

Datan omistajuuskäytänteeseen kuuluu organisaation informaatioteknisten resurssien, kuten osastokohtaisen tietojenkäsittelyn, hajauttaminen. Tässä tapauksessa näiden resurssien hajauttamisella, tai epäkeskittämisellä, tarkoitetaan sitä, että keskitetyn tietokannan sijaan dataa hallinnoi osastojen omistamat järjestelmät. Storeyn ym. mukaan Coase (1960) määrittelee, että siinä tapauksessa, kun datan laadusta on kiistatonta hyötyä tietylle osastolle, on suositeltavaa, että kyseinen osasto myös omistaa tämän datan. Tällöin osasto, jolle datan laadulla on eniten arvoa, voi määrittellä datan käyttöön liittyviä käytänteitä vapaasti ja parhaiten tarkoitusta palvelevaksi. Jos datan laatu ei ole olennaisessa asemassa millekään tietylle taholle, organisaatio tarvitsee omistajuus- ja kannustinkäytänteiden yhdistelmän datan laadun hallinnointiin. (Storey ym. 2012, s. 440.)

4.3 Muutokset rooleissa, vastuissa ja omistajuudessa master data managementia organisoidessa

Riikka Vilminko-Heikkinen ja Samuli Pekkola (2018, s. 76,79.) käsittelevät artikkelissaan, sitä, miten omistajuus ja vastuut muuttuivat kahden master data management- projektin edetessä. Master data managementin kehittämisen motivaattorina oli laatuongelmat master datan kanssa. Ongelmien tiedostettiin johtuvan niin prosesseista kuin sovelluksista. Muita tunnistettuja tavoitteita oli integroinnit ja konsolidoinnit, datakokoelmaprosessien tarjoaminen, laadun varmistus ja informaation käytön kontrollointimekanismien käyttäminen. Tutkimuksessa peilattiin sosiaalisten ja materiaalisten tekijöiden limittyneitä vaikutuksia päällekkäisyyden ja limittyneisyyden metaforan (engl. metaphor of imbrication) kautta.

Master data management- toimintojen rakentaminen aloitettiin kartoittamalla organisaation tarpeet ja niistä muodostettiin tavoitteet ensimmäisenä tehdylle master data management- projektille. Samalla tunnistettiin myös tarve datan omistajuudelle. Datan hallintaan ja ylläpitoon tarvittavien roolien sisältöä ei kuvattu tai määritelty. Datan ylläpidon määrittely tapahtui tietojärjestelmä- ei datalähtöisesti. (Vilminko-Heikkinen & Pekkola 2018, s. 79.)

Tarvittavien määrittelyiden jälkeen tutkimuksessa tunnistettiin organisaation master data, mitä varten laadittiin kriteeristö ja määriteltiin toimialue pilottihankkeelle. Tämän jälkeen luotiin yleinen hallinnoinnin malli, joka sisälsi erilaisia käytänteitä ja toimia, säädöksiä ja vastuuta ja rooleja kuvauksineen. Tämä hallinnointi määriteltiin kolmelle tasolle, jotka olivat organisaatiotasoa, tukitoimitaso ja data-toimialuetaso. Yksityisyys, tietoturva, laadunhallinta ja informaatiojärjestelmien integraatiot tunnistettiin tukitoimitasolla. (Vilminko-Heikkinen & Pekkola 2018, s. 79.)

Yleisen hallinnoinnin mallin luomisen jälkeen tutkimuksessa kuvattiin organisaatiossa virinneen keskustelua datan omistajuudesta ja siihen liittyvistä vastuista. Informaatiojärjestelmien takia nähtiin prosessien omistajien olevan vastuussa luomastaan datasta ja useampiin prosesseihin liittyvän datan omistajuuden suhteen oli sekavuutta. Tätä ei kuitenkaan tapahtunut silloin kun data tuotettiin yhden toiminnon piirissä, näissä tapauksissa datan omistajuus henkilöityi kyseisen toiminnon johtajaan. Datan laadun huononeminen huonosti määriteltyjen vastuiden ja omistajuuden takia nähtiin IT-osaston ongelmana ja liiketoimintaprosessien omistajat eivät ymmärtäneet rooliaan prosesseissaan liikkuvan datan parantamisessa. Data-toimialueen omistajuus katsottiin erittäin tärkeäksi asiaksi. Omistajuus käsitti yhden toimialueen datan kehittämisen ja yhden master datan toimialueen ylläpidon, mutta kävi ilmi, että projektiin osallistuneet ja nimittetyt omistajat eivät ymmärtäneet omaa rooliaan. (Vilminko-Heikkinen & Pekkola 2018, s. 80.)

Datan omistajuuden määrittelyn jälkeen määriteltiin tavoitteet pilottidatalle. Eli tunnistettiin ja arvioitiin sellaisia prosesseja, jotka liittyivät pilotissa käytettävään dataan. Artikkelin mukaan dataa seurattiin minimivaa- timuksilla. Datan itsensä perusteella ei myöskään saanut määriteltyä omistajuutta, vaan se viittasi useisiin eri tahoihin organisaatiossa ja se sijaitsi useissa eri informaatiojärjestelmissä. IT-osastoa ehdotettiin omistajaksi, koska se vastasi kyseistä dataa sisältävistä informaatiojärjestelmistä. Tutkimuksen mukaan oli selkeää, että datasta ei haluttu ottaa vastuuta, eikä se olisi mahdollistakaan ilman omistajuutta. (Vilminko-Heikkinen & Pekkola 2018, s. 81.)

Master data management- mallin implementointivaiheessa oli ensimmäisen projektin tapauksessa vastahakoisuutta omistajuusasioissa ja listausta datan omistajuutta koskevia vastuista ei koskaan hyväksytty. Koko master data management- projektin omistajaksi nimettiin tietohallintojohtaja. Toisessa projektissa vastuut ja omistajuus jaettiin useille eri tasoille. Teh- tyjen muutosten seurauksena datan omistajuutta saatettiin osoittaa esi- miestasolle ja omistajia koskevia vastuuta uudelleen määriteltiin muun mu- assa erilaisten datatyyppeiden vastuualueeseen sisällyttämisen osalta. (Vil- minko-Heikkinen & Pekkola 2018, s. 81, 83.)

Tekninen ratkaisu master datan hallinnalle ja suunnitelma pilotille on ar- tikkelissa kuvattu viimeinen vaihe. Tutkitussa organisaatiossa siirrettiin da- tan ylläpitovastuuta talousosastolta, jossa se oli ollut aiemmin, IT:lle. Ylä- pitovastuu siirrettiin, koska ylläpidolliset toimet olivat niin monimutkaisia ja sitoutettuja työnkulkuihin. Datan omistajuutta rajoitettiin myös käsittä- mään kaksi vaihtoehtoa; henkilöstöresursseihin ja taloushierarkioihin. Il- meni myös, että datan omistajien nimitys ja vastuiden määrittäminen ko- ettiin monimutkaiseksi ja epäselväksi. Erilaisia sovelluksia master data ma- nagementista tunnistettiin ensimmäisen projektin aikana ja otettiin käyt- töön toisen projektin aikana. (Vilminko-Heikkinen & Pekkola 2018, s. 83.)

4.4 Tarve data governancelle: tapaustutkimus

Cheongin ja Changin (2007, s. 999) The Need for Data Governance: A Case Study- tutkimus käsittelee yrityksen informaationhallintaa keskittyen It- hallinnoinnin ja data governancen (datan hallinnointi) väliseen suhtee- seen. Tutkimusta käsittelevässä artikkelissa todetaan, että dataa voidaan hallita paremmin käyttämällä data governancea ja data governancen viite- kehystä.

Tutkimusmenetelmänä käytettiin case-, eli tapaustutkimusta ja tutkitta- vana yrityksenä, Yritys A:na, oli yritys, jossa useista erilaisista lähteistä ja järjestelmistä koostetulla informaatiolla on yli 100 käyttäjää. Asiakaskun- taa, jolle tutkittu yritys tuottaa palveluita, on 890 000 ja ydinhenkilöstöä 1850. Kyseisessä organisaatiossa datan laatu oli määritelty tärkeäksi prio- riteetiksi. Yritys A:ssa on myös monimutkaisia ongelmia datan integraation suhteen. Tutkimusdataa kerättiin haastatteluilla, joissa painotettiin IT- ja

liiketoimintanäkökulmaa dataan liittyvissä asioissa ja ongelmissa. (Cheong & Chang 2007, s. 1002-1003.)

Cheongin ja Changin tutkimus keskittyy Yritys A:n datan hallintatiimiin, jonka jäsenet toimivat datan suhteen tietovastaavina (engl. Data steward). Tiimi toimii tietovastaavan roolissa, koska heillä on syvällistä tietoa yrityksen resursseihin (eng. Asset) liittyvistä informaatiojärjestelmistä ja liiketoimintaprosesseista. Tiimin ongelmana on hallita resursseihin liittyvää dataa yrityksellisenä resurssina. Kyseinen tiimi kamppaili monenlaisten dataan liittyvien ongelmien kanssa, ennen kuin se päätti aloittaa datan hallinnointimenetelmien käytön. Näitä ongelmia oli datan reaktiivinen ja ad hoc-tyylinen hallitseminen, tiimillä ei ollut pääsyä alkuperäiseen tietokantaan, dataan liittyvissä ongelmissa ei tahdottu päästä yhteisymmärrykseen, datan laatuun tähtäävät projektit törmäsivät budjettiongelmiin ja muut liiketoiminnalliset yksiköt ja It-osasto eivät ottaneet vakavasti tiimin yrityksiä laatia standardeja datan suhteen. (Cheong & Chang 2007, s. 1003.)

Yritys A:n Datan hallintatiimin toimet ongelmien parantamiseksi ovat jaettavissa viiteen kohtaan:

1. Reaktiivisen datanhallinnan ongelma: proaktiivisuuden lisäämiseksi tiimi määritteli strategian datan laadun parantamiseksi.
2. Resurssidatan vaikea saavutettavuus: dataa on säilytetty tietokannassa, johon tiimillä ei ole ollut pääsyä. Pääsy hoidetaan joltain osin kopioidulla datalla ja data on levittäytynyt erilaisiin tietokantoihin ja järjestelmiin.
3. Datanparannusprojektien budjettiongelmat: ongelmat ratkaistaisiin sillä, että nimitetään projektipäällikkö yhteyshenkilöksi liiketoiminnan ja IT:n väliin.
4. Data-standardien määrittelyn vaikeus: datan hallintatiimi päätti alkaa työstämään metadatan säilytyspaikan hankkimista avuksi datastandardi-ongelmaan.
5. Vaikeus päästä yhteisymmärrykseen dataan liittyvissä ongelmissa: yhteisymmärryksen puutetta yritettiin parantaa datan omistajuuteen liittyvillä käytänteillä, mutta se osoittautui vaikeaksi ottaa käyttöön, koska tarvittava hallinnollinen rakenne puuttui. (Cheong & Chang 2007, s. 1004.)

Artikkelin mukaan Thomas (2006) kutsuu tämänkaltaista epävirallista lähestymistapaa nimellä hallinnoinniksi tietovastaavuuden kautta (eng. Governance via stewardship). Tämänkaltaisen lähestymistavan todetaan kuitenkin olevan toimimaton, koska data-tietovastaavien roolia ja omistajuutta ei määritetty selkeästi ja johdolta ei saatu riittävästi valtuuksia. Cheong ja Chang kertovat Thomasin (2006) korostavan, että yleisimmät syyt tämänkaltaisen mallin kaatumiseen ovat selkeiden roolien ja auktoriteetin puute. Tämän takia datan hallintatiimi koosti suunnitelman ”hallin-

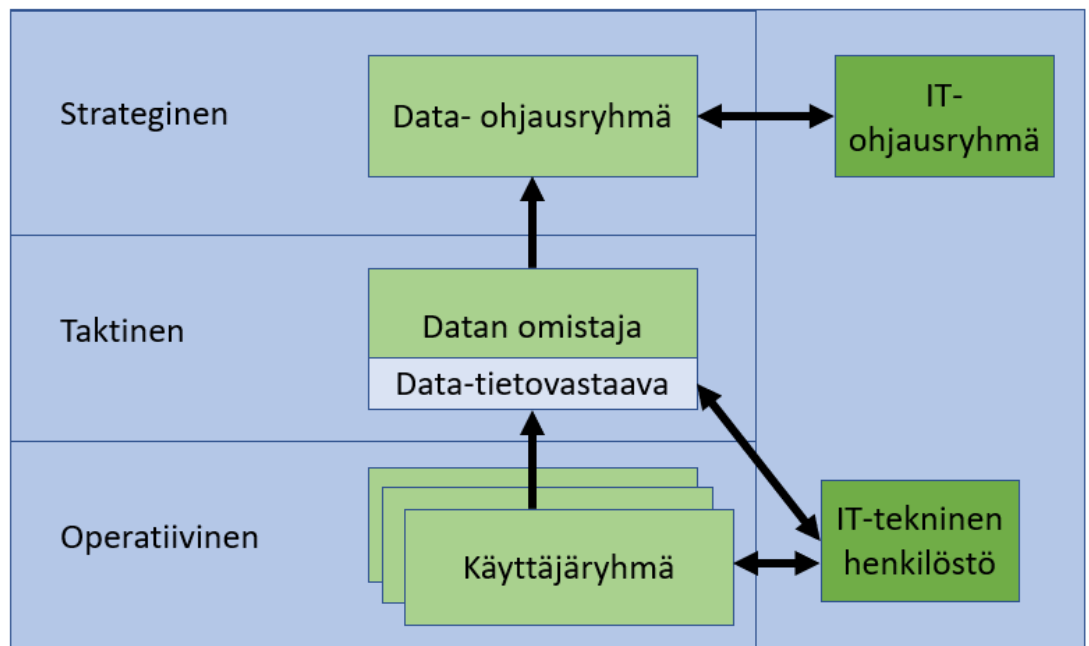
nointia hallinnoinnin kautta”- datan hallinnoinnin mallin laatimiseksi kyseiselle organisaatiolle. Näin ollen taulukossa 1 esitetty datan hallinnoinnin viitekehys otettiin käyttöön Yritys A:ssa. (Cheong & Chang 2007, s. 1004.)

Taulukko 1. Datan hallinnointi- viitekehysten osat (Cheong & Chang 2007, s. 1004)

Organisationaaliset kokonaisuudet ja politiikat	Standardit ja prosessit	Datan hallinnointiteknologiat
Hallinnointirakenne	Datan määrittely ja standardointi (metadatan hallinta)	Metadatan säilytyspaikka (repositorio)
Datavastaaajuus	Kolmannen osapuolen datan keräys	Datan profilointityökalu
Käyttäjärhymävuokra	Mittareiden kehittäminen ja seuranta	Datan puhdistustyökalu
Päätöksenteko oikeudet	Datan profilointi	
Ongelman eskaloitumisprosessi	Datan puhdistus	

Datan hallinnoinnin rakenne on esitetty kuvassa 4. Data- ohjausryhmän jäsenet ovat johtajia eri osastoilta ja heillä on kiinnostusta resursseihin liittyvään dataan ja sen hallintaan. Tämä ryhmä on vastuussa muun muassa uusista käytänteistä ja he ratkovat ongelmia, jotka käsittävät useampia sektoreita yrityksessä ja ovat yhteydessä IT-ohjausryhmään strategisella tasolla. Datan omistajat hallitsevat resursseihin liittyvää dataa ja ovat vastuussa sen laadusta. (Cheong & Chang 2007, s. 1005.)

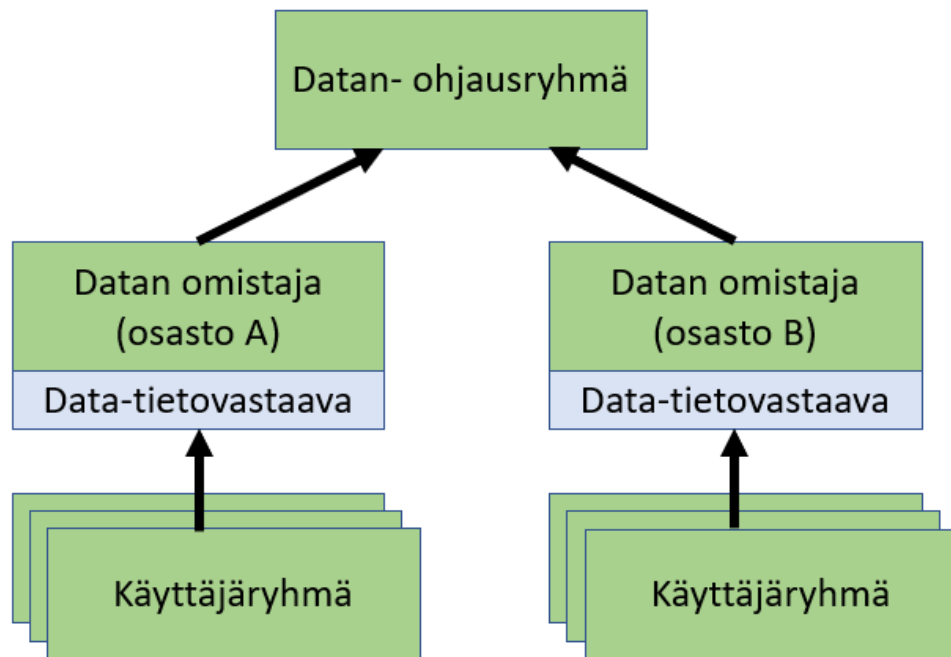
Omistajat myös ratkovat ongelmia, joita nousee esiin käyttäjäryhmien tapamisissa, mutta jos esiin nousseet ongelmat ovat vaikutukseltaan laajalaisempia, ne kuuluvat data- ohjausryhmän hoidettaviksi. Omistajat ovat myös vastuussa datan hallinnointisuunnitelmien täytäntöönpanosta. Data-tietovastaavat ovat taasen ryhmä, jolla on syvempää ymmärrystä yrityksen prosesseista ja datan asettamista vaatimuksista. He toimivat myös tekniikkaa ymmärtävänä ryhmänä tulkkavaana välikappaleena IT:n ja liiketoiminnan välissä. Tietovastaavat toimivat myös muutoksenhaallinnassa, vastaavat sidosryhmistä ja osakkaista datan omistajien puolesta, hoitavat resursseihin liittyvien informaatiojärjestelmien ja projektien hallinnoinnin. (Cheong & Chang 2007, s. 1005.)



Kuva 4. Datan hallinnan rakenne (Cheong & Chang 2007, s. 1005)

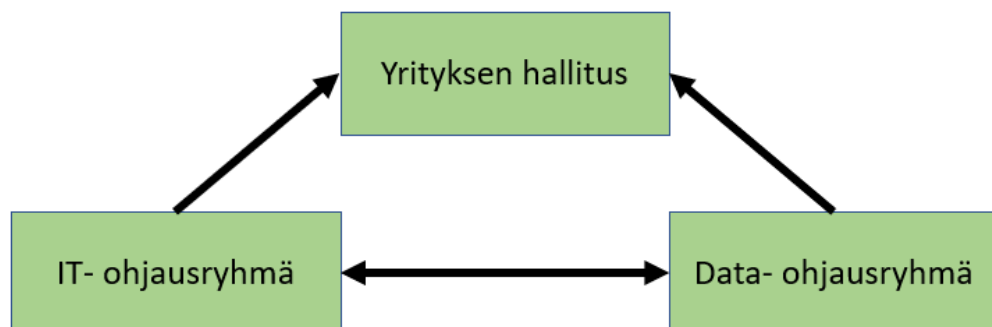
Kuvassa 4. esitetyt käyttäjäryhmät ovat koostettu eri liiketoimintayksiköistä, sekä osastoista ja heidät on kutsuttu käyttäjäryhmäkokouksiin. He ovat avainhenkilöitä, jotka ovat osallisina datan keräämisessä, sen prosessoinnissa ja dataan liittyvässä raportoinnissa. IT-tekniistä henkilöstöä osallistuu myös näihin tapaamisiin, jolloin heidän tietoteknistä osaamistaan voidaan hyödyntää. Tapaaminen on myös foorumi kiireellisten, dataan liittyvien asioiden tiettäväksi tuomiseen. Nämä dataa käyttävät ihmiset tuovat myös dataan liittyviä ongelmia yleiseen tietoisuuteen ja osallistuvat oman työnkuvansa kehittämiseen. (Cheong & Chang 2007, s. 1005.)

Tutkimuksessa esitelty datan hallinnoinnin malli on skaalautuva, kuten kuvassa 5. esitetään. Tällöin data-ohjausryhmä toimii ”organisationaalisenä liimana”, kuten Cheong ja Chang (2007, s. 1006) sen esittävät.



Kuva 5. Datan hallinointimallin skaalautuvuus (Cheong & Chang 2007, 1006)

Tutkimuksessa todetaan, että Yritys A:n tapaus osoittaa, että yrityksellisen datan tehokas hallinnointi vaatii yhteistyötä IT:ltä ja liiketoiminnalta. Kuvassa 6. esitetään, kuinka tämä yhteistyö toimii. Kun aloitettu datan hallinnointi alkaa edetä, sen pitäisi alkaa olla vuorovaikutuksessa yrityksellisen hallinnoinnin viitekehyksen kanssa. Data- ohjausryhmä ja IT- ohjausryhmät toimivat yhteistyössä ja raportoivat yrityksen hallitukselle. Artikkelissa kerrotaan Demberin (2006) määrittelevän datan hallinnoinnin ”tarjoavan viitekehyksen risteykseen, jossa IT:n ja liiketoiminta työskentelee yhdessä luodakseen luottamusta ja luotettavuutta yrityksen informaatioon”. (Cheong & Chang 2007, s. 1006.)



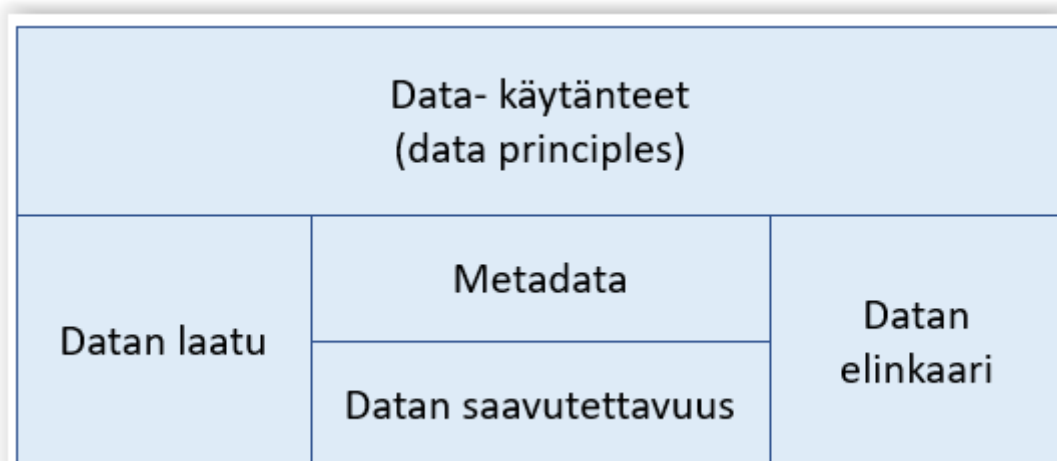
Kuva 6. Datan hallinnoinnin vaatima yhteistyö liiketoiminnan, data- ohjausryhmän ja IT- ohjausryhmän välillä (Cheong & Chang 2007, s. 1006)

4.5 Data governancen suunnittelu

Khatri ja Brown (2010, 148) esittelevät data governance- viitekehysten, jonka tarkoituksena on tarjota tehokas ratkaisu datan hallintaan organisaatioissa. Kirjoittajien mukaan viitekehystä tarkastellessa on huomionarvoista, että se ei erottele dataa ja informaatiota. Data governancea suunniteltaessa keskitytään päivittäisen päätöksenteon taustalla vaikuttaviin kysymyksiin, kuten mitkä ovat tärkeitä päätöksiä ja kenen pitäisi tehdä ne. Viitekehysten perustana käytetään Weillin ja Rossin (2010) IT governance- viitekehystä, joka linkittää yrityksellisen ja informaatioteknologisen hallinnoinnin toisiinsa.

Khatri ja Brown erottelevat data governancen (datan hallinnointi) ja data managementin (datan hallinta) Weillin ja Rossin (2004) perusteella seuraavasti: governance käsittää sen, mitä päätöksiä pitää tehdä, jotta voidaan varmistaa IT:n tehokas käyttö ja hallinta. Management taas sisältää päätösten tekemisen ja niiden toteuttamisen. (Khatri & Brown 2010, s. 148.)

Viitekehyksessä esitellään viisi data- päätöksenteon osa-alueita (eng. data decision domains), Kuva 7, ja käsitellään miksi nämä alueet ovat tärkeitä ja mitä päätöksiä niistä kuhunkin liittyy. Operationalisoimalla vastuut päätöksenteon keskiössä (eng. locus of accountability of decision making) jokaisessa alueessa, saadaan data governance- matriisi, Taulukko 2. (Khatri & Brown 2010, s. 149.)



Kuva 7. Päätöksenteon osa-alueet data governancessa (Khatri & Brown 2010, s. 149.)

Taulukko 2. Data governance- matriisi (Khatri & Brown 2010, s. 149.)

Data governance- alue	Alueen päätökset	Potentiaaliset roolit tai vastuiden keskittymät
<p>Data- käytänteet</p> <p>Datan roolin selvennys voimavarana/ resurssina (eng. Asset)</p>	<p>Mitä käyttöä liiketoiminnalla on datalle?</p> <p>Mitkä ovat datan liiketoiminnallisen käytön kommunikoinnin mekanismit?</p> <p>Mitkä ovat halutut tavat datan hyödyntämiseen voimavarana? Kuinka tilaisuudet uudelleenkäyttöä ja jakaa dataa tunnustetaan?</p> <p>Kuinka säädökset vaikuttavat datan liiketoiminnalliseen käyttöön?</p>	<p>Datan omistaja</p> <p>Datan huoltaja</p> <p>Data-vastaava</p> <p>Datan tuottaja/ datan toimittaja</p> <p>Datan käyttäjä</p> <p>Yrityksen data- ohjausryhmä</p>
<p>Datan laatu</p> <p>Vaitimusten luominen datan aiottua käyttöä varten</p>	<p>Mitkä ovat standardit datan laadulle tarkkuuden, ajantasaisuuden, eheyden ja luotettavuuden osalta?</p> <p>Mikä on suunnitelma datan laadun ja siitä kommunikoinnin vaikiinnuttamiseen?</p> <p>Kuinka datan laatua ja siihen liittyvää suunnitelmaa arvioidaan?</p>	<p>Datan omistaja</p> <p>Aihealueen asiantuntija</p> <p>Datan laatupäällikkö</p> <p>Datan laatuanalyttikko</p>
<p>Metadata</p> <p>Datan merkityksellistämisen ja sisällyksellistämisen niin, että se on käyttäjien tulkittavissa</p>	<p>Mikä on suunnitelma datan merkityksellisyyksien dokumentoinnille?</p> <p>Kuinka data määritellään ja mallinnetaan yhteneväisesti niin, että se on tulkittavissa?</p> <p>Mikä on suunnitelma erityyppisen metadatan ajantasaisena pitämiseen?</p>	<p>Yrityksen data- arkkitehti</p> <p>Yrityksen datan mallintaja</p> <p>Datan mallinnusinsinööri</p> <p>Data-arkkitehti</p>

		Kokonaisarkkitehtuuriohjausryhmä
<p>Datan saavutettavuus</p> <p>Datan saavutettavuusvaatimusten määrittely</p>	<p>Mikä on datan liiketoiminnallinen arvo?</p> <p>Kuinka jatkuvaa riskien arviointia tehdään?</p> <p>Kuinka arvioinnin tulokset integroidaan yleisen sisäisen valvonnan kanssa?</p> <p>Mitkä ovat datan saavutettavuuden standardit ja toimenpiteet?</p> <p>Mikä on suunnitelma kausittaiseen seurantaan ja ohjeiden mukaisuuden auditointiin?</p> <p>Kuinka levitetään tietoisuutta ja osaamista tietoturvallisuudesta?</p> <p>Mikä on suunnitelma varmuuskopiointiin ja palautumiseen?</p>	<p>Datan omistaja</p> <p>Datasta hyötyjä</p> <p>Tietoturvaohjaaja</p> <p>Tietosuojavaltuutettu</p> <p>Tietoturvatekniikka-analyytikko</p> <p>Kokonaisarkkitehtuuriohjausryhmä</p>
<p>Datan elinkaari</p> <p>Datan tuottaminen, säilytys ja poistuminen</p>	<p>Miten data inventoidaan?</p> <p>Mikä on suunnitelma datan määrittelyyn, tuottamiseen, säilyttämiseen ja käytöstä poistamiseen?</p> <p>Miten lainsäädäntöön liittyvät määräysten mukaisuuskysymykset vaikuttavat datan säilyttämiseen ja arkistointiin?</p>	<p>Data arkkitehti</p> <p>Informaatioketju- vastaava</p>

IT governancen liittyvät asiat ovat vahvasti relevantteja myös data governancen suhteen ja näiden kahden yhteensovittaminen, sekä yhtenevien menetelmien määrittely, lisää tehokkuutta. Sama taho, joka vastaa IT:n roolista ja siihen liittyvistä määrityksistä voisi myös vastata datasta resursina tai omaisuutena. Vastuiden keskittymisellä jokaisen edellä mainittujen alueen suhteen, on data governancen suunnittelussa liukuva rooli, keskittymisen ja hajautumisen välillä. Kuva 8 havainnollistaa data governance- matriisia ja dataan liittyvän päätöksenteon keskittymistä suhteessa päätöksenteon eri osa-alueisiin. Kuvan 8 esimerkissä dataan liittyvien toimintatapojen määrittely on erittäin keskittynyt yrityksen johtajille, jotka toimivat datavaltuutettuina. (Khatri & Brown 2010, s. 151.)

Päätöksenteon osa-alueet Vastuiden keskittymät	Data- käytänteet	Datan laatu	Metadata	Datan saavutettavuus	Datan elinkaari
Keskittynyt	X				
↑ ↓				X	X
			X		
		X			
Hajautettu					

Kuva 8. Esimerkki data governance- matriisista (Khatri & Brown 2010, s. 151).

5 TUTKIMUKSEN PERUSTA

Tämä luku käsittelee tässä opinnäytetyössä tehdyn vertailevan tutkimuksen perustaa. Tutkimusmenetelmäksi on valittu vertaileva kvalitatiivinen analyysi ja tehty tutkimus pohjaa toimeksiantajan järjestelmässä liikkuvan datan pohjalta laadittuihin vaatimuksiin.

5.1 Laadullisesta tutkimuksesta

Tutkimuksellisia lähestymistapoja voivat olla kvalitatiivinen, eli laadullinen-, ja kvantitatiivinen, eli määrällinen- suuntaus. Tutkimusmenetelminä laadullista ja määrällistä menetelmää ei voi täysin ehdottomasti erottaa toinen toisestaan, vaan ne voidaan nähdä ennemminkin toisiaan täydentävinä, usein rinnakkain käytettävinä, menetelminä. (Eskola & Suoranta 1998, 10; Hirsjärvi, Remes, Salovaara 1997, s. 131-132.)

Yksinkertaistetusti voidaan sanoa, että kvantitatiivinen tutkimussuuntaus keskittyy numeerisiin arvoihin, kun taas kvalitatiivinen kuvaa merkityksiä ja aineiston muotoa. On kuitenkin otettava huomioon, että kvalitatiivisten, ja kvantitatiivisten, aineistojen välillä voi olla riippuvuuksia. Eli näiden tyyppiset aineistot voivat selittää toinen toisiaan, kun mitattaessa käsitellään sekä kvalitatiivisia, että kvantitatiivisia ulottuvuuksia. (Eskola & Suoranta 1998, s. 13; Hirsjärvi ym. 1997, s.133.)

Tutkimuksen suuntausta, eli menettelytapaa, valitessa hyvä käytäntö on lähestyä asiaa ongelmalähtöisesti, pohjaten siihen, mikä lähestymistapa tarjoaa parhaan ratkaisun tai selvyuden käsillä olevaan ongelmaan. Kaikella tutkimuksella on myös aina oma päämääränsä ja tarkoituksensa ja tätä voidaan määritellä neljän erilaisen näkökulman kautta. Nämä näkökulmat, tai piirteet, ovat kartoittava, selittävä, kuvaileva tai ennustava. Tutkimusta voi määrittää myös useampi kuin yksi näkökulma. (Hirsjärvi ym. 1997, s. 133-134.)

Tutkimuksen toteuttamisen kannalta on myös olennaista määritellä käytettävät käsitteet. Käsitteet ovat työkaluja, joiden varaan koko tutkimustyö rakennetaan ja joiden avulla merkityksiä tulkitaan. Maailmaa jäsennetään erilaisten käsitteiden tai merkitysten avulla. Tieteellisessä tutkimuksessa esiintyy kahdenlaisia käsitteitä: konkreetteja ja abstrakteja. Abstrakteja käsitteitä voidaan kutsua myös teoreettisiksi. Toisin kuin abstraktit, konkreetit käsitteet ovat sidottuja tiettyyn aikaan ja paikkaan. (Eskola & Suoranta 1998, s. 45; Hirsjärvi ym. 1997, s. 142-143.)

Käsitteiden ongelmaksi muodostuu niiden määrittelyn vaikeus ja puutteellisesti määritellyn tai ymmärretyn käsitteen rajoittava vaikutus tutkimustyön tekemiseen. Tämän takia, käytettävien käsitteiden tulisikin olla vaikiintuneita, eikä mielivaltaisesti keksittyjä. Joidenkin käsitteiden merkitys

saattaa vaihdella kontekstien, kuten vaikkapa eri tutkijoiden tai kulttuurien, välillä tai jossain tapauksissa myös saman kulttuuripiirin sisällä. Käsitteiden monitulkintaisuuden yhteydessä nousee esiin myös aiheen rajaamisen tärkeys, koska tällöin minimoidaan käsitteistöstä johtuvat väärinymmärrykset. (Eskola & Suoranta 1998, s. 45; Hirsjärvi ym. 1997, s. 144-145.)

Aineistoa analysoitaessa lähestymistapa tulee valita sen mukaan, mikä palvelee parhaiten tutkimuksen tavoitteita. Lähestymistavat voidaan jakaa selittäviin ja ymmärtäviin. Selittämistä vaativien aineistojen tapauksessa tehdään usein tilastollista analyysiä ja päätelmiä. Ymmärtämistarkoituksessa tehtävässä tutkimuksessa taas valitaan laadullinen analyysi ja tehdään päätelmiä. (Hirsjärvi ym. 1997, s. 219.)

Tähän tutkimukseen on valittu kvalitatiivinen lähestymistapa, koska aineiston numeerisella analysoinnilla ei ole suurta merkitystä tutkimuksen tavoitteiden kannalta ja aineisto vaatii ymmärtämistä, ei selittämistä. Aineiston suhteen on olennaista tutkia erilaisia datan ominaisuuksia ja tehdä selittäviä ja kartoittavia havaintoja, joiden perusteella on mahdollista määrittellä se, mikä master data management- malli sopii parhaiten tutkimuksen tarkoitukseen, eli master datan omistajuuden määrittämiseen.

Toimeksiantajan tietojärjestelmissä liikkuvaan informaatioon on liitetty erilaisia, pääosin kuvailevia, metatietoja. Tältä kannalta katsottuna on mielekkäämpää keskittyä aineiston semanttiseen ulottuvuuteen kuin numeerisiin määriin.

Kvantitatiivisen tutkimusmenetelmän ongelmaksi tässä tapauksessa nousee tutkittavien master datan hallinnan mallien abstraktisuus ja vaikea yleistettävyyden. Tietojärjestelmissä liikkuva data sen sijaan on hyvin määrämuotoista ja siltä osin käytettävissä olevasta aineistosta on mahdollista tehdä yleistyksiä sen määrittämien vaatimusten osalta, esimerkiksi datan laatuvaatimusten, kuten ajantasaisuuden ja eheyden, osalta.

Mallien vertailu havainnollistetaan kvalitatiivisen vertailumatriisin avulla. Mallien ominaisuudet, suhteessa toimeksiantajan järjestelmässä liikkuvan datan vaatimukseen, pisteytetään kolmiportaisesti.

5.2 Pohdintaa datasta ja omistajuudesta aineiston pohjalta

Kuten luvussa 3.2 todettiin Pekkolan ja Vilminko-Heikkisen (2019) mukaan, datan omistajuus voidaan määritellä siten, että sillä tarkoitetaan tiettyä tahoja, joka voi tehdä päätöksiä datan suhteen. Nämä päätökset voivat käsittää esimerkiksi laadullisia ominaisuuksia tai datan saavutettavuuskysymyksiä. Omistajuudella ei tarkoiteta sitä, että data olisi niinkään sananmukaisesti merkityksessä konkreettisesti omistettu, vaan ennemminkin hallinnoitu. Data governance tarjoaa erilaisia lähestymistapoja ja käytettäviä rooleja tähän kysymykseen.

Tämän opinnäytetyön puitteissa on tarkoitus tarkastella datan omistajuutta perustuen dataan itseensä, eikä tietojärjestelmälähtöisesti. Tietojärjestelmälähtöinen lähestymistapa tarkoittaisi sitä, että omistajuuskysymys voitaisiin ratkaista sillä, mikä taho vastaa datan varsinaisesta säilytyspaikasta. Datalähtöisesti ajateltaessa omistajuuskysymykseen voidaan pyrkiä vastaamaan tarkastelemalla dataa itsessään, sekä käyttötapoja ja käytön asettamia laadullisia vaatimuksia.

Omistajuus ei ole synonyymi datan käytölle, mutta on perusteltua pohtia sitä, millä taholla on suurimmat intressit datan laadun suhteen. Luvussa 4.2 käsitellyssä Data quality: Setting organizational policies- tutkimuksessa Storey ym. (2012) pohtivat yksilöiden vaikutusta organisaatiossa käytettävään laatuun. Tutkimuksessa esitetään näkemys siitä, että datan kustannukset nousevat epäsuorassa suhteessa datan laatuun. On eri asia, onko tämä näkemys kuitenkin missä määrin yleistettävissä kaikenlaiseen datan käyttöön ja erilaisiin tilanteisiin.

Samassa tutkimuksessa todetaan myös, että työntekijät luontaisesti tavoittelevat alemmaa laadullista tasoa kuin olisi optimaalista organisaatiolle ja esitetään ratkaisuksi sitä, että datan omistajuus määritellään sen kautta, mikä taho hyödyntää eniten dataa. Sellaisessa tapauksessa, jossa ei ole yhtä selkeästi määriteltävää tahoja omistajaksi, koska dataa samoilla tavoilla käytäviä tahoja on useita, esitetään käytettäväksi erilaisia kannustimia tai omistajuuskäytänteitä.

Voidaan siis kääntäen sanoa, että asia voidaan ajatella myös arvontuottamisen ja hyötymisen näkökulmasta, eikä pelkästään keskittyen kustannuksiin. Tilanteessa, missä esimerkiksi teknisillä ratkaisuilla taataan datan korkea laatu, ei voida todeta, että kustannukset nousisivat suhteessa datan laatuun, vaan tällaisessa tilanteessa datan laatua voidaan pitää ennemminkin itseisarvona.

Vilminko-Heikkisen ja Pekkolan (2019) tutkimuksessa Changes in roles, responsibilities and ownership in organizing master data management, todettiin, että master data management- projektien omistajuuskysymykset osoittautuivat vaikeasti ymmärrettäviksi ja monimutkaisiksi. Datan omistajat eivät ymmärtäneet omaa rooliaan tai vastuitaan ja tämä aiheutti epäselvyyttä ja ongelmia. Tässä tutkimuksessa datan omistajuutta ei määritely hyötynäkökulman kautta, koska data oli tallennettuna eri järjestelmiin ja sitä käytti useat eri tahot, jolloin kysymykseen ei saatu selkeää ja toteuttamiskelpoista vastausta tällä lähestymistavalla, koska kukaan ei halunnut ottaa vastuuta datasta.

Tältä pohjalta voidaan siis todeta, että datan omistajuuden määrittely, niin käsitteenä itsessään, kuin siihen sisältyvien vastuiden osalta, on tärkeää. Samalla voidaan pohtia, olisiko Vilminko-Heikkisen ja Pekkolan tutkimuksessa tutkitun organisaation saamat tulokset olleet erilaisia, jos datan omistajuutta olisi lähestytty eri tavalla, eikä suhtautuen dataan rasitteena

ja kustannuksena esimerkiksi omistajuuden aiheuttaman lisävaivan merkeissä.

Kummankin ajattelutavan kannalta on kuitenkin yhtä lailla olennaista, että epäselvyydet minimoidaan ja data, sekä datan laadun ylläpito, tuodaan luonteviksi osiksi organisaation prosesseja, eikä ylimääräiseksi ja erilliseksi osakokonaisuudeksi, josta kukaan ei halua ottaa vastuuta. Storeyn ym. (2012) esittämien luontaisesti pienemmän vaivan näkemyksen kääntämällä voidaan esittää myös kysymys siitä, tavoittelisivatko yksittäiset työntekijät luontaisesti korkeampaa datan laatua, mikäli se tuottaisi konkreettisia hyötyjä yksittäisen työntekijän tasolla. Erityisen merkityksellistä tämä olisi esitellyssä ei-synergisessä keskinäisessä riippuvuussuhteessa, jolloin arvo organisaatiolle on pienempi kuin yksittäiselle työntekijälle.

Cheong ja Chang (2007) käsittelevät tapaustutkimuksessaan *The Need for Data Governance: A Case Study*, sitä, miten Yritys A:ssa ratkaistaan datan hallinnointiin liittyviä ongelmia ottamalla käyttöön data governance- viitekehyksen. Myös tässä tapauksessa on tunnistettu ongelmia datan omistajuudessa ja yleisiä epäselvyyksiä liittyen datan omistajuuteen ja dataa koskeviin kysymyksiin. Tutkimuksessa todetaan, että Yritys A:n tapauksessa datan tehokas hallinnointi vaatii yhteistyötä yrityksen johdolta, IT:ltä ja datasta vastaavalta taholta.

Yritys A:ssa datan käyttöä ja laatua lähestyttiin käyttäjäryhmien kautta, joihin kuului avainhenkilöitä, jotka olivat tekemissä datan kanssa. Avainhenkilöt koottiin käyttäjäryhmään eri liiketoiminnallisista yksiköistä ja sidosryhmistä. Käyttäjäryhmien tapaamisiin osallistui myös IT- henkilöstöä, jolla oli tarjota teknistä ymmärrystä ja tukea avainhenkilöille.

Tämänkaltainen toimintatapa lähestyy dataa ja datan laadullista ulottuvuutta hyvin kokonaisvaltaisella tavalla. Data governancen jakaminen strategiselle, taktiselle ja operatiiviselle tasolle sitouttaa sen kiinteäksi osaksi yrityksen tai organisaation toimintaa ja rakenteita, mutta voidaan esittää kysymys, näyttäytykö data tässä tapauksessa minkälaisena resurssina yksittäiselle työntekijälle.

Toinen huomion arvoinen kysymys on se, miten muut kuin avainhenkilöt suhtautuvat dataan ja otetaanko näiden käyttäjien tarpeet, huomiot ja toimet huomioon millä tavalla suhteessa avainhenkilöihin. Erilaista dataa liikkuu kuitenkin käytännössä kaikkialla, niin arkielämässä kuin työympäristössä ja on kohtuullisen turvallista väittää, että jokainen työntekijä on tavalla tai toisella jonkinlaisen yrityksellisen datan käyttäjiä tai tuottajia. Mikäli käyttäjätapaamisiin valitut henkilöt eivät edusta yrityksen tai organisaation datan käyttäjiä riittävän heterogeenisellä otoksella on mahdollista, että prosessit ja käytänteet muovautuvat palvelemaan vain osaa henkilöstöstä.

Mitä tulee näihin omistajuus- ja laatuksymyksiin tämän opinnäytetyön toimeksiantajan osalta, kysymyksiä tulee tarkastella myös kokonaisarkkitehtuuri ja tietoturva huomioon ottaen. Henkilötietoja sisältävän datan osalta vaatimuksia määrittää GDPR ja toiminnan strategisella tasolla Arter Framework sanelee myös datalle vaatimuksia.

Hankkeilla ja luonnollisesti hankkeisiin sisältyvällä datalla on siis konkreettisen liiketoiminnallisen arvon lisäksi merkitystä raportoinnin ja edelleen raportointiin pohjautuen strategisessa mielessä. Käytännön tasolla tämä merkitsee siis sitä, että operatiivisella tasolla kerättävästä tai kasaantuvasta tiedosta muodostuu loogisella tasolla informaatioresurssi. Data, informaation rakennuspalikkana, ei sinällään merkitse mitään, joten dataan liittyvät erilaiset metatiedot ovat elintärkeitä merkityksellistäjää, jotta dataa on mahdollista tulkita, eli toisin sanoen, sitä on mahdollista hyödyntää.

Kuten luvussa 2.3 käytiin läpi, HAMKin TKI-hankesalkun hallintajärjestelmässä liikkuvaa dataa käytetään eri intervalleilla. Osa datasta on käytännössä reaaliaikaista ja osa päivittyy harvemmin. Tämä asettaa ajantasaisuuteen liittyviä laatuvaatimuksia datalle ja sen käytettävyydelle. Eli tiheästi päivittyvä data on käytettävää ja laadukasta kun se on mahdollisimman ajantasaista.

HAMKissa liikkuvasta datasta koostetaan erilaisia raportteja ja yhteenve-toja, eli datan on oltava myös keskenään yhteensopivaa ja yhtenäistä, jotta raportit ovat todenmukaisia. Raportoinnin kannalta datan tulee olla myös ajantasaista ja oikeellista.

Mitä tulee erilaisiin projektikohtaisiin tietoihin, on huomionarvoista todeta, että erilaisia kuvaavia ja määrittäviä tietoja on suuri määrä. Tällaisessa tapauksessa datan laatua määrittää myös se, onko data asianmukaisesti tuotettua tai onko sitä tuotettu ylipäätään ja mikä data on tärkeämpää kuin jokin muu, eli mikä on varsinaista ydintietoa. Tässä suhteessa henkilöstön sitoutuminen datan asianmukaiseen tuottamiseen on tärkeää, koska sitoutumaton henkilöstö tuottaa huonompilaatuista dataa tai tässä tapauksessa, saattaa jättää jotain olennaista dataa tuottamatta kokonaan.

Datan tuottamatta jättäminen on huomionarvoinen ongelma nimenomaan HAMKin TKI-hankesalkun hallintajärjestelmän tapauksessa. Koska projektikohtaisia tietoja on paljon, yksittäiselle työntekijälle merkityksetön tieto, joka saattaisi kuitenkin olla arvokasta esimerkiksi raportoinnin osalta, saattaa jäädä täyttämättä.

Datan sisällöllisen ja määrällisen runsauden suhteen saattaa olla myös perusteltua pohtia sitä, mikä on tarkalleen ottaen master dataa. Master dataksi määritellään yrityksessä tai organisaatiossa liikkuva kriittinen tai elintärkeä ydindata, mutta kriittisyys ja tärkeys on samaan tapaan subjektiivinen käsite kuin datan arvokin on suhteessa yksittäiseen työntekijään ja koko organisaatioon. Tämä ei ole tämän opinnäytetyön rajauksen kannalta

olennainen kysymys, joten datan syvällisempi tutkiminen ei ole perusteltua, vaan tässä tapauksessa voidaan puhua vain datasta yhtenä kokonaisuutena, osiensa summana. Tutkimuksen osalta ei siis käsitellä sitä, mikä on ja mikä ei ole master dataa, vaikka jossain tapauksessa omistajuuskysymyksen osalta sillä voisikin olla merkitystä.

HAMKissa liikkuvaa, henkilötietoja sisältävää, dataa käsitellään GDPR:n mukaisesti. Eli datan tulee muun muassa olla saavutettavissa, sitä tulee käyttää vain alkuperäiseen ja perusteltuun käyttötarkoitukseensa, sen tulee olla muokattavissa ja poistettavissa tarpeen vaatiessa ja sitä tulee säilyttää niin turvallisesti, ettei se päädy ulkopuolisten käsiin. Eli henkilötietoja sisältävän datan vaatimukset huomioon ottaen voidaan todeta, että tämänkaltainen data vaatii ylläpitoa, eli sen, että on selkeästi määritelty taho, joka on vastuussa edellä mainituista datan huolto- ja ylläpitotoimenpiteistä.

Arter ARC- ohjelmistossa voidaan tarkastella erilaisia riippuvuussuhteita, sekä suodattaa ja tutkia dataa muilla tavoin. Myös tämä, kuten raportointikin, asettaa datan laadulle vaatimuksia eheydestä, ajantasaisuudesta, saavutettavuudesta ja oikeellisuudesta.

6 VERTAILUTUTKIMUS

Tämän opinnäytetyön tutkimuksellisessa osassa vertaillaan master data management- malleja toisiinsa. Vertailun määrittäjänä toimii toimeksiantajan järjestelmissä liikkuvan datan vaatimukset. Vertailua havainnollistaa vertailumatriisi.

6.1 Vertailu

Master data management- mallien vertailun lähtökohtana on HAMKin TKI-hankesalkun hallintajärjestelmässä liikkuva data, sekä sen käyttöä määrittävien Arter Frameworkin, GDPR:n ja raportoinnin asettamat vaatimukset. Datan omistajuuskysymyksen kannalta vertailussa keskitytään siihen millä tavalla ja millä tasolla mallit ottavat kantaa omistajuuteen. Tavoitteena on tunnistaa toimeksiantajan kannalta hyödyllisimpiä ja käyttökelpoisimpia toimintamalleja ja -tapoja.

Lopulliseen vertailuun valitut tutkimukset ja mallit valikoitiin toimeksiantajan antamasta aineistosta tähän tutkimukseen mukaan niiden laadullisten ja sosiaalisten, eli omistajuuteen kantaa ottavien ulottuvuuksien takia.

Ennen varsinaista lopullista vertailua MIT Sloan Management Review- julkaisussa esitelty Tiered Data Storage- malli karsiutui pois, koska se painotti väärä asioita, kuten teknologista arkkitehtuuria hyvin yleisellä tasolla, eikä ottanut sinällään kantaa laatuun tai sosiaalisiin aspekteihin, vaikka tämä nimenomainen malli muuten olikin relevantti HAMKissa käytettyjen datansäilytysratkaisujen osalta. Tämän nimenomaisen mallin etuna olisi tosin ollut kustannustehokkuuden huomiointi ja datan järjestäminen varsinaisen käytön perusteella, mutta se ei olisi vastannut tämän opinnäytetyön toimeksiantoa ja aiheen rajausta.

Mallien vertailu on havainnollistettu taulukossa 3. Vertailtavat ominaisuudet on jaoteltu otsikoiden laatu, tietoturva, omistajuus ja käytettävyys, alle.

Vertailu pisteytetään kolmiportaisesti asteikolla:

0 = malli ei ota kantaa kyseiseen asiaan

1 = malli ottaa jossain määrin kantaa

2 = malli ottaa seikkaperäisesti kantaa asiaan

Koska mallit eivät ole selkeitä ja helposti jäseneltäviä malleja, vertailu tapahtuu korkealla abstraktitasolla. Pisteytyksessä käytettävä termi ”ottaa kantaa” merkitsee siis tässä yhteydessä sitä, että asia on tavalla tai toisella otettu huomioon mallissa. Seikkaperäisellä kannan ottamiselle tarkoitetaan sitä, että kyseinen asia on selitetty selkeästi, mahdollisesti esimerkkejä käyttäen. Vertailua ei ole mielekäästä tehdä laajemmalla skaalalla,

koska mallien välisiä ominaisuuksia ei ole tämän tutkimuksen puitteissa mielekästä vertailla tätä tarkemmalla tasolla. Vertailussa ei myöskään käytetä painokertoimia, vaan ominaisuuksia tarkastellaan saman arvoisina ja tämä otetaan huomioon tulosten tulkinnessa.

Laadullisista kysymyksistä eheyden, ajantasaisuuden ja oikeellisuuden huomiointi tarkoittavat sitä, ottaako kyseinen malli kantaa näihin datan laadullisiin ominaispiirteisiin. Datan laatuvaatimuksilla tarkoitetaan sitä, onko mallissa määritelty tapoja tai toimenpiteitä, prosesseja tai muita toimia koskien organisaation tarpeita datan laadulle. Esimerkiksi tulisiko datan vähimmäislaatu määrittellä. Laadunvalvonnan suunnitelmalla tarkoitetaan sitä, onko mallissa otettu kantaa tai esitelty tapoja datan laadunvalvonnan sitouttamiseksi osaksi organisaation prosesseja tai onko sille annettu jollain muulla tavoin painoarvoa. Datan keräämisen taustaprosesseilla tarkoitetaan työnkulkua prosesseja ja toimia, joiden seurauksena dataa kerääntyy tai kerätään ja vertailussa tutkitaan sitä, ottaako malli tähän asiaan kantaa ja millä tasolla.

Tietoturvan osalta malleja peilataan GDPR:n henkilötietoja koskeviin vaatimuksiin. Datan elinkaaren huomiointilla tarkoitetaan sitä, ottaako malli huomioon datan arkistointia, poistumista tai uudelleen käyttöä koskevia asioita. Elinkaari on vahvasti sidoksissa GDPR:n, joka vaatii muun muassa, ettei tietoja säilytettäisi tarpeettomasti. On myös perusteltua olettaa, että turhaan säilytettävä tieto, esimerkiksi henkilötietojen osalta, on aina potentiaalinen riski, joka on mahdollista minimoida kontrolloimalla datan elinkaarta, johon myös säilytysratkaisut liittyvät.

Omistajuuden kannalta vertailussa tutkitaan sitä, onko roolit nimetty, eli onko datan omistavat tahot nimetty tai muuten yksilöitävissä. Rooleihin sisältyvät vastuut tarkoittavat, onko mallissa määritelty niitä tehtäviä, vastuita tai työnkuvia, joita datan omistavalla taholla on hoidettavanaan. Roolien määrittely osaksi organisaation toimintaa taas käsittää sen, onko datan omistavan tahon, mikäli sellainen on määritelty, vastuut omistajana sidottu organisaation sisäisiin prosesseihin, työnkulkuihin tai toimintatapoihin ja millä tasolla; hyvin yleisellä tasolla määriteltynä, vai seikkaperäisemmin kuvattuna.

Omistajuuden määrittely tarkoittaa, että mallissa on määritelty se, mitä omistajuudella tarkoitetaan ja mitä siihen sisältyy, sekä omistajuuskäsitteen tiettäväksi tekeminen työntekijätasolla on otettu huomioon. Osaston määrittely omistajaksi tarkoittaa sitä, määritelläänkö mallissa jokin organisaation osasto, kuten talous tai tietohallinto, omistajaksi. Tämä on relevantti kysymys siksi, koska omistajien tunnistaminen työntekijätasolla ei ole kaikissa tapauksissa tarkoituksenmukaista ja osastokohtainen ajattelu luo omalta osaltaan joustavuutta mallin soveltamiseen.

Käytettävyys-kategoria tutkii malleista sitä, ovatko ne joustavia ja käyttöön otettavia erilaisissa tapauksissa. Tässä tutkimuksessa tutkitaan HAMKin

TKI-hankesalkun hallintajärjestelmää, mutta on huomionarvoista myös pohtia sitä, olisiko malli sovellettavissa myös muuhun tarkoitukseen. Joustavuus ja sovellettavuus tarkoittaa sitä, onko malli hyödynnettävissä erilaiseen käyttöön, eli palvelisiko se tarkoitusta myös jossain muussa käytössä ja onko se tarpeeksi joustava otettavaksi käyttöön HAMKissa. Tutkimustulosten ja käyttökokemusten osalta tutkitaan sitä, onko malli kokeiltu käytännössä ja perustuuko se vahvaan tieteelliseen pohjaan.

Taulukko 3. Kvalitatiivinen vertailumatriisi tutkituista malleista

	Malli 1	Malli 2	Malli 3	Malli 4
Laatu				
Datan eheys, ajantasaisuus ja oikeellisuus huomioitu	0	0	1	2
Datan laatuvaatimukset määritelty	1	0	1	2
Laadunvalvonnan suunnitelma huomioitu	0	0	1	2
Datan keräämisen taustaprosessit huomioitu	1	0	1	2
Yhteensä:	2	0	4	2
Tietoturva				
Datan elinkaari huomioitu	0	0	1	2
Tietoturvanäkökulma (GDPR) huomioitu	0	1	0	2
Yhteensä:	0	1	1	4
Omistajuus				
Roolit nimetty	0	1	2	2
Rooleihin sisältyvät vastuut määritelty	0	0	2	2
Roolit määritelty osaksi organisaation toimintaa	0	1	2	2
Omistajuus määritelty	0	0	2	2
Omistajuus määritelty osastotasolla	2	1	2	0
Yhteensä:	2	3	10	8
Käytettävyys				
Malli on joustava ja sovellettavissa	2	1	2	2
Malli on selkeä	1	0	2	2
Mallista on tutkimustuloksia ja käyttäkokemuksia	1	2	2	2
Yhteensä:	4	3	6	6
Kaikki yhteensä:	8	7	21	20

Malli 1, eli Datan laatu: organisationaalisten toimintatapojen määrittely, on hyvin abstraktitasoinen. Malli 1 painottuu yleiseen datan laatuun ja laadun kustannuksiin, eikä määrittele sen tarkemmin datan laadullisia erikoisvaatimuksia tai ominaispiirteitä. Datan laatuvaatimusten määrittelystä malli saa yhden pisteen, koska datan laatutasoa käsitellään yleisellä tasolla. Yksi piste datan keräämisen taustaprosessien huomioinnista tulee

Malli 1 = Datan laatu: organisationaalisten toimintatapojen määrittely
Malli 2= Muutokset rooleissa, vastuissa ja omistajuudessa master data managementia organisoidessa
Malli 3 = Tarve data governancelle: tapaustutkimus
Malli 4= Data governancen suunnittelu

0 = malli ei ota kantaa kyseiseen asiaan

1 = malli ottaa jossain määrin kantaa kyseiseen asiaan

2 = malli ottaa seikkaperäisesti kantaa kyseiseen asiaan

siitä, että mallissa datan kaappaus ja prosessointi on yhdistetty päätöksentekoon. Tämän mallin kannalta on huomionarvoista, että se tarjoaa näkökulmia ja ratkaisuja datan laadullisiin ongelmiin organisaatiossa. Malli ei myöskään esitä minkäänlaisia konkreettisia tai abstraktitasoisia suunnitelmia datan laadullisen valvonnan käytännön toteutukselle. Malli 1 ei ota kantaa myöskään datan tietoturvasuhteisiin ominaispiirteisiin.

Omistajuuden malli 1 esittää osastotasolla tapahtuvaksi, jolloin dataa hallinnoitaisiin osastotasolla, sen sijaan, että käytettäisiin keskitettyä tietokantaa. Tällöin datan omistajuuden lähtökohtana on järjestelmät, jotka osasto omistaa. Rooleja tai niihin sisältyviä vastuita ei ole kuvattu.

Käytettävyyden osalta malli 1 saa täydet kaksi pistettä, sekä yhden pisteen selkeydestä ja tutkimustuloksiin ja käyttökokemuksiin perustumisesta. Malli on niin abstrakti, että se on joustavasti sovellettavissa muuallekin, mutta toisaalta on hyvin yleisluontoinen. Malli on kohtuullisen selkeä, mutta se ei tarjoa konkreettista piirrettyä mallia tai seikkaperäisiä kuvauksia tai toimintaohjeita. Malli 1 perustuu vahvasti aiemmalle tutkimustyölle ja esitetyt havainnot, väitteet ja ratkaisuehdotukset pohjaavat vahvasti matemaattisiin malleihin ja kaavoihin.

Malli 2, eli Muutokset rooleissa, vastuissa ja omistajuudessa master data managementia organisoidessa, on tutkimus siitä, miten omistajuus, vastuut ja roolit muuttuvat master data management- projektin edetessä. Tutkimus sivuaa laadullisia kysymyksiä, mutta ei sinällään ota niihin sen tarkemmin kantaa, siksi malli 2 ei saa yhtään pistettä laadun osalta. On myös huomionarvoista, että mallissa 2 datan ylläpito on kuvattu tehdyn tietojärjestelmä, ei datalähtöisesti.

Tietoturvan osalta malli 2 saa yhden pisteen tietoturvanäkökulman huomiointista. Yksityisyys, tietoturva, laadunhallinta ja informaatiojärjestelmien integraatiot kuvattiin tunnistetuiksi datan tukitoimitasolla, osana datan kolmiportaista toimialuetasojärjestelmää.

Omistajuuden osalta tutkimuksessa kuvatuissa master datan hallinnan käyttöönoton projekteissa kerrottiin olleen suuria ongelmia roolien ja niihin kuuluvien vastuiden osalta. Roolit kuvattiin määritellyiksi, siitä yksi piste, mutta sisällön ja vastuiden kerrottiin jääneen epäselviksi ja abstrakteiksi henkilöstölle. Sama omistajuuden määrittelyn tapauksessa, omistajuutta ei ole selostettu tarkemmin, mutta sitä on kuvattu osastotasolla tapahtuvana, esimerkkinä tästä on IT-osasto, jota ehdotettiin omistajaksi datalle, koska se vastasi informaatiojärjestelmistä, joihin dataa tallennettiin.

Malli 2 saa joustavuudesta ja sovellettavuudesta yhden pisteen, koska se on käyttöönottokokemusten kuvaus, eikä varsinainen toimintaohje tai malli. Samasta syystä malli 2 saa selkeydestä nolla pistettä, koska tutkimus liikkuu hyvin abstraktilla tasolla, eikä se tarjoa konkreettisia tai helposti havainnollistettuja kuvauksia. Tutkimustulosten ja käyttökokemusten osalta

malli 2 saa kaksi pistettä, koska kyse on tarkasti dokumentoidusta tapaus-
tutkimuksesta.

Malli 3 on tapaus tutkimus, Tarve data governancelle: tapaus tutkimus.
Malli 3 saa vertailumatriisin laatukategoriassa jokaisesta laatu kysymyk-
sestä yhden pisteen. Malli ei kuvaa datan laadullisia vaatimuksia seikkape-
räisesti eri konteksteissa, mutta se kuvaa yksityiskohtaisesti prosesseja ja
rakenteita datan laadulliseen seurantaan ja määrittelyyn, sekä toimia,
joilla työntekijöiden laatu työtä voidaan tukea. Esimerkkinä tällaisesta laa-
tutyön tukemisesta on IT-henkilöstön mukaan tuominen avainhenkilöiden
palavereihin.

Datan elinkaaren liittyviä kysymyksiä sivutaan datan hallinnointitekno-
logioiden osalta, jossa on esitelty datan profilointi- ja puhdistustyökalut,
sekä metadata-repositorio. Malli 3 ei ota kantaa muihin tietoturvakysy-
myksiin.

Omistajuuden osalta malli 3 käsittelee data governanca strategisella,
operatiivisella ja taktisella tasolla, joille kaikille on nimetty oma tahonsa,
jolle on kuvattu omat tehtävänsä. Mallissa on myös kuvattu eri tahojen vä-
liset suhteet ja vastuut. Omistajuus on myös kuvattu selkeästi. Omistajuus
on määritelty osastotasolla osaston edustajiston kautta. Myös omistajuu-
den osalta malli 3 saa vertailumatriisin laatukategorian jokaisesta kohdasta
kaksi pistettä.

Käytettävyysskysymysten osalta malli 3 saa myös täydet pisteet, koska malli
on selkeä, selkeästi kuvattuna se on joustava ja sovellettavissa, sekä se on
riittävän abstraktitasoinen, mutta myös seikkaperäisesti avattu. Malli
myös perustuu käytännön kokemukseen tapaus tutkimuksen osalta.

Malli 4, Data governancen suunnittelu, kuvaa laadullisia kysymyksiä tar-
kasti ja monesta eri näkökulmasta data governance-matriisissa, joka on
koostettu viiden päätöksenteon osa-alueen mukaan. Tämä malli kysyy
seikkaperäisiä kysymyksiä laatu standardien määrittelystä ja esittelee
näistä kysymyksistä vastuun kantavia rooleja. Tästä syystä malli 4 saa täy-
det pisteet kaikista laatukategorian kohdista. Malli ottaa huomioon kaikki
laatu työssä osallisina olevat tahot ja erilaiset taustaprosessit ja jatkosuun-
nitelmat.

Tietoturvan osalta malli 4 saa kaksi pistettä elinkaaren huomioinnista ja
tietoturvan näkökulmasta, koska data governance -matriisissa on määritelty
yksityiskohtaisesti kaikki roolit ja vastuut, jotka linkittyvät tietoturvakysy-
myksiin ja datan elinkaaren. Roolien määrittelyn kautta mahdollistetaan
myös parempi tietoturva, koska prosessit ja työnkuvat ovat jo määriteltyjä.

Samoista syistä malli 4 saa täydet pisteet muista omistajuuskategorian
kohdista kuin omistajuuden määrittelystä osastotasolla. Tämä malli kuvaa
rooleja yksittäin ennemmin kuin rooleja osastokohtaisesti.

Käytettävyyden osalta malli 4 saa myös täydet pisteet. Malli on erittäin helpposti tulkittavissa, se tarjoaa yksityiskohtaisia roolikuvauksia ja esittää erilaisia kysymyksiä ja lähestymistapoja datan hallintaan. Malli perustuu myös käytännön kokemuksiin ja on testattu tapaustutkimuksen osalta.

6.2 Vertailun tulokset

Mallien kokonaispistemäärät:

Malli 1: Datan laatu: organisationaalisten toimintatapojen määrittely: 8 pistettä

Malli 2: Muutokset rooleissa, vastuissa ja omistajuudessa master data managementia organisoidessa: 7 pistettä

Malli 3: Tarve data governancelle: tapaustutkimus: 21 pistettä

Malli 4: Data governancen suunnittelu: 20 pistettä

Vertailun tuloksia analysoitaessa on otettava huomioon, että mallit eivät sinällään ole suoraan keskenään verrannollisia. Tästä syystä pisteet eivät tarjoa suoraa vastausta siitä, mikä malleista olisi paras tai käytettävien kokonaispistemäärän perusteella. Kokonaispistemäärää tarkastellessa voidaan kuitenkin todeta, että tässä tutkimuksessa määriteltyjen ehtojen puitteissa malleja toimeksiantajan datan asettamia vaatimuksia vasten tulkittaessa, malli 3 vastaa parhaiten tarpeita.

Korkea pistemäärä selittyy mallin 3 osalta sillä, että malli on kuvattu ja selitetty seikkaperäisesti konkreettisella tasolla. Ero toiseksi tulleeeseen malliin 4 on niin pieni, ettei tulkinnasta johtuvan virhemarginaalin nimissä voida tehdä eroa näiden kahden mallin paremmuuden välillä. Malli 4 tarjoaa myös hyvin käytännönläheisen ja tarkasti havainnollistetun ratkaisun datan hallintaan.

7 JOHTOPÄÄTÖKSET

Master data management- mallien vertailun perusteella ei voida antaa yksiselitteistä vastausta siitä, millä tavoin TKI- hankesalkun hallintajärjestelmässä liikkuvaa dataa olisi perustelluinta hallinnoida. Mallien saamien pisteiden perusteella on kuitenkin mahdollista esittää johtopäätöksiä ja tehdä havaintoja siitä, miltä osin mikäkin malli esittää oikeita kysymyksiä, tarjoaa vastauksia tai ratkaisuja.

7.1 Tutkimuksen tuloksista

Mallien vertailun pohjalta voidaan todeta, että mallit 3 ja 4, Tarve data governancelle: tapaustutkimus ja Data governancen suunnittelu, tarjoavat konkreettisia ja helposti hahmotettavia ratkaisuja organisaation datan hallintaan. Mallit 1 ja 2, Datan laatu: organisaation toimintatapojen määrittely ja Muutokset rooleissa, vastuissa ja omistajuudessa master data managementia organisoidessa, tarjoavat kuitenkin myös erittäin hyödyllisiä ja käyttökelpoisia lähestymistapoja ja teorioita toimeksiantajan käyttöön.

Erityisesti malli 2 tarjoaa arvokasta tutkimustietoa master data management- mallien käyttöönotosta ja siitä, miten roolit ja vastuut muuttuvat projektin edetessä. Mallissa 2 esitetyt ongelmat, etenkin käsitteiden määrittelyiden ja vastuiden ymmärtämisen osalta ovat luonteeltaan hyvin yleismaallisia ja oletettavasti hyvin todennäköisiä lieveilmiöitä vastaavan suuruusluokan projekteissa.

Tämän opinnäytetyön ensimmäinen tutkimuskysymys on: miltä osin tai mikä valituista master data management- malleista olisi sopivin TKI- hankesalkun hallintajärjestelmässä liikkuvan datan laadulliseen parantamiseen? Vertailun ja tehtyjen olettamusten perusteella tähän voidaan vastata, perustuen vertailumatriisin laadukategorian pisteisiin, että kokonaisuutena mallit 1, 3 ja 4, mutta kuten aiemmin on todettu, tämä vastaus ei ole täysin yksiselitteinen.

Tämän opinnäytetyön toinen tutkimuskysymys on: miltä osin tai mikä valituista master data management- malleista olisi sopivin TKI- hankesalkun hallintajärjestelmässä liikkuvan datan omistajuuden määrittämiseen perustuen dataan itseensä, ei järjestelmään? Vertailun perusteella omistajuuskysymykseen otti parhaiten kokonaisuutena kantaa mallit 3 ja 4.

Mallien tarkastelu omina kokonaisuuksinaan ei ole kuitenkaan tarkoituksenmukaista, eikä toimivaa. Tutkimuskysymysten puitteissa onkin esitetty kysymykset myös siitä, mitkä osat tutkituista malleista vastaisivat parhaiten laadullisiin tai omistajuutta määrittäviin tarpeisiin.

Abstraktilla tasolla liikuttaessa mallit 1 ja 2 tarjoavat näkökulmia muun muassa laadun kustannuksiin ja inhimilliseen näkökulmaan organisaatioiden datanhallinnassa. Mallin 1 esittämät havainnot yksittäisten työntekijöiden suhtautumisesta laatutyöhön ja vaadittuun laatu tasoon ovat relevantteja kysymyksiä organisaatiosta riippumatta. HAMKin TKI-hankesalkun hallintajärjestelmän osalta mallia voidaan tulkita esimerkiksi siten, että määrittelemällä toimintatapa siihen, mitä tietoja järjestelmään tulisi minimitasolla syöttää ja tuomalla tiettäväksi sen, mitä konkreettisia hyötyjä tästä seuraa. Datan eheyden laatuvaikutuksen tiettäväksi tuominen voisi olla joko data-analytiikan käytänteiden ja vaatimuksien parempaa esiin tuomista projektien parissa työskenteleville tai datan eheydestä seuraavien konkreettisten hyötyjen esiin tuomista.

Mallin 2 pohjalta voidaan myös todeta, että on erittäin tärkeää määritellä datan parissa työskentelevien tahojen, niin työntekijätasolla kuin osastotasolla liikuttaessa, roolien kuvaukset ja vastuut. Datan omistajuus tulee määritellä selkeästi ja ymmärrettävästi, sekä vastuiden tulee olla ymmärrettäviä. Mallin 2 pohjalta voidaan todeta myös, että kokonaiskuvan kannalta selkeintä on välttää liian monimutkaisia rakenteita, etenkin kun kyseessä on kokoluokaltaan sen suuruinen organisaatio, että inhimillisen erheen tai väärinymmärryksen riski moninkertaistuu henkilöstömäärän kasvaessa.

Käytännössä mikään malli ei ottanut tietoturvakysymyksiin kantaa riittäväällä selkeydellä. Malli 4:n osalta määritellyt selkeät tietoturvaroolit ja esitetyt kysymykset näkyivät pisteytyksessä, mutta sellaisenaan malli 4 ei tarjoa mitään ehdottoman käyttökelpoista HAMKin henkilötietoja sisältävän datan käsittelyn tarpeisiin, jos asiaa katsotaan GDPR:n näkökulmasta. Roolien määrittely luo pohjan henkilötietoja sisältävän datan hallitsemiselle ja huoltamiselle, mutta tarvittavat prosessit ja työkulut tulisi luonnollisesti vielä määritellä tarkasti ja sitoa selkeästi roolituksiin ja vastuisiin.

7.2 Suosituksia datan hallinointiin

Master data management- mallien vertailun pohjalta voidaan esittää yleisluontoisia suosituksia HAMKille.

- Mahdollisten roolien kuvausten tulisi olla selkeitä ja mahdollisimman helposti ymmärrettäviä
- Rooleihin sisältyvien vastuiden tulisi olla selkeitä ja mahdollisimman helposti ymmärrettäviä
- Datan omistajuus voitaisiin määrittää perustuen siihen, mikä taho hyötyy eniten käytettävästä datasta, eli kenelle datalla on suurin arvo
- Datan laatustandardit tulisi olla selkeästi määritellyt ja dataa tuottavien ja käyttävien henkilöiden tiedossa
- Datan laadun vaikutukset organisaation toimintaan tulisi olla dataa tuottavien ja käyttävien henkilöiden tiedossa

- Datan laatuun ja muuhun datan ylläpitoon liittyvä työ olisi perusteltua sitouttaa kiinteästi ja selkeästi hahmotettavalla tavalla organisaation prosesseihin
- Dataan liittyvät työkulut ja vastuunjaot tulisivat olla riittävän selkeästi hahmotettavat ja henkilöstön saavutettavissa. Esimerkiksi kenen tai minkä tahon puoleen on mahdollista kääntyä dataan liittyvässä ongelmatilanteessa tai miten datan laadullisiin ominaisuuksiin on mahdollista vaikuttaa
- Datan laatutyön merkitys tulisi tehdä selkeäksi ja ymmärrettäväksi

7.3 Pohdintaa lopputuloksesta

Tämä työ ei lopullisessa muodossaan vastaa toimeksiantajan alkuperäiseen tarpeeseen, joka oli datanhallinnan mallin laatiminen annetun aineiston pohjalta, mutta olen pyrkinyt ottamaan tämän parhaani mukaan huomioon. Opinnäytetyöni toissijaisena tavoitteena, tässä työssä määriteltyihin tutkimuskysymyksiin vastaamisen lisäksi, on tarjota lisätietoa ja tehdä havaintoja, jotka tukevat tätä alkuperäistä tavoitetta.

Aiheen monimutkaisuus ja vaikeus olisivat kuitenkin vaatineet enemmän aikaa, jolloin olisi ollut mahdollista valita paremmin tarkoitusta palveleva ja perusteellisempi lähestymistapa. Tältä pohjalta kvalitatiivisen vertailun tekeminen ei palvellut lopputulosta parhaalla mahdollisella tavalla, koska lopullinen lähestymistapa muovautui huomattavan käytännönläheiseksi ja pakotti tekemään paljon olettamuksia ja yleistyksiä tutkittavan järjestelmän suhteen. Toisaalta teoreettisempi ote olisi vaatinut myös huomattavan syvällisempää perehtymistä HAMKin järjestelmissä liikkuvaan dataan ja data-analytiikkaan, mikä ei olisi ollut mahdollista salassa pidettävien aineistojen ja materiaalien osalta, eikä suhteessa käytettävissä olevaan aikaikkunaan.

Tämä työ tarjosi syvempää ymmärrystä datasta, master datasta ja master data managementista, sekä siitä, miten edellä mainittuja sovelletaan yritysmaailmassa. Tavoitteena on, että tätä opinnäytetyötä voidaan käyttää taustamateriaalina HAMKin datanhallinnan suunnittelussa.

LÄHTEET

- Akeroyd, J, Dmitriev, D & Martin, A. 2009. Resurgence of interest in Information Architecture. *International Journal of Information Management* 2010 (30), 6-12.
- Arter n.d. Arter Framework – Kokonaisarkkitehtuuri – Muutosta laadun ehdoilla. Haettu 2.3.2019 osoitteesta <https://www.arter.fi/arter-framework-kokonaisarkkitehtuuri/>
- Aunola, S. (2019). Uusi vaihde kehittämiseen kyvykkyyksien hallinnalla. Blogijulkaisu 16.1.2019. Haettu 2.3.2019 osoitteesta <https://www.arter.fi/uusi-vaihde-kehittamiseen-kyvykkyyksien-hallinnalla/>
- Beath, C., Becerra-Fernandez, I., Ross, J. & Short, J. 2012. Finding Value in the Information Explosion. *MIT Sloan Management Review* 53 (4), 17-20.
- Cheong, L., K. & Chang, V. 2007. The Need For Data Governance: A Case Study. *ACIS 2007 Proceedings*, 2007, 999-1008.
- Eskola, J. & Suoranta, J. 1998. *Johdatus laadulliseen tutkimukseen*. Jyväskylä, Gummerus Kirjapaino Oy.
- EU:n yleinen tietosuoja-asetus 2016/679. Haettu 2.3.2019 osoitteesta <https://eur-lex.europa.eu/legal-content/FI/TXT/HTML/?uri=CELEX:32016R0679&from=FI>
- Henttonen, P. & Valtonen, R. 2010. Asiakirjatiedon hallinta. Teoksessa Serola S. (toim.) *Ote Informaatiosta; Johdatus informaatiotutkimukseen ja interaktiiviseen mediaan*. Helsinki, BTJ Finland Oy, 211-239.
- Hirsjärvi, S., Remes, P., Sajavaara, P. 1997. *Tutki ja kirjoita*. Keuruu, Otavan Kirjapaino Oy.
- Kari, J. 2010. Mitä on tiedonkäyttö? Teoksessa Serola S. (toim.) *Ote Informaatiosta; Johdatus informaatiotutkimukseen ja interaktiiviseen mediaan*. Helsinki, BTJ Finland Oy, 115-154.
- Kartturi – korkeakoulujen kokonaisarkkitehtuurin menetelmäopas. 2013. Korkeakoulujen KA-Pilotti ja KA-SIG ryhmä. CSC Tieteen tietotekniikan keskus Oy.
- Khatri, V., & Brown C., V., 2010. Designing Data Governance. *Communications of the ACM* 2010, 58 (1), 148-152.
- Savolainen, R. 2009. Epistemic work and knowing in practice as conceptualizations of information use. *Information Research* 14 (1). Viitattu

22.1.2019. Saatavissa Information Research, an international electronic journal- tietokannassa:

<http://www.informationr.net/ir/14-1/infres141.html>

Pekkola, S. & Vilminko-Heikkinen, R. 2019. Changes in roles, responsibilities and ownership in master data management. *International Journal of Information Management* 2019 (47), 76-87.

Pekkola, S. & Vilminko-Heikkinen, R. 2013. Establishing an Organization's Master Data Management Function: A Stepwise Approach. *Hawaii International Conference on System Sciences*, 2013 (46).

Spruit, M. & Pietzka, K. 2015. MD3M: The Master Data Management Maturity Model. *Computers in Human Behavior* 2015 (51), 1068-1076.

Storey, V. C., Devan, R. M. & Freimer, M., 2012. Data quality: Setting organizational policies. *Decision Support Systems* (54) 2012, 434-442.

Van Der Lans, R. 2012. *Data Virtualization for Business Intelligence Systems : Revolutionizing Data Integration for Data Warehouses*. Elsevier Science & Technology.

Zins, Chalms, 2007. Conceptual approaches for defining data, information and knowledge. *Journal of the Association for Information Science and Technology* 2007, 58 (4), 479-493. Viitattu 22.1.2019. Saatavissa Wiley Online Direct tietokannassa:
<https://onlinelibrary-wiley-com.helios.uta.fi/doi/full/10.1002/asi.20508>

Arter Framework metamalli (Arter n.d.)

