



Microsoft Power BI:n hyödyntäminen Tampereen kaupungin sote-tiedon analysoinnissa

Reija Tuominen

OPINNÄYTETYÖ
Huhtikuu 2019

Tietojärjestelmäosaaminen, ylempi AMK

TIIVISTELMÄ

Tampereen ammattikorkeakoulu
Tietojärjestelmäosaaminen, ylempi AMK

TUOMINEN, REIJA:

Microsoft Power BI:n hyödyntäminen Tampereen kaupungin sote-tiedon analysoinnissa

Opinnäytetyö 90 sivua, joista liitteitä 4 sivua
Huhtikuu 2019

Tämän opinnäytetyön toimeksiantajana toimi Tampereen kaupungin Tietohallinnon Konserniyksikkö. Tampereen kaupungilla on käytössä tietovarastointi- ja raportointijärjestelmänä SAP Business Warehouse. Kaupungilla on valittu ICT-palveluiden tarjoamaksi tiedon analysointivälineeksi Microsoft Power BI. Tampereen kaupungin edistämä digitalisaatio ja kokeileva pilotointimalli aiheuttavat haasteita nykyisille tietojärjestelmille ja toimintamalleille saada tuotettua data-aineistoja nykyisillä resursseilla.

Opinnäytetyö oli luonteeltaan laadullinen tutkimus, ja lähestymistapana käytettiin tapaustutkimusta. Opinnäytetyön tarkoituksena oli selvittää, mitä tietovarastointitekniikoita on olemassa, Power BI:n käyttömahdollisuudet analysointityökaluna sekä analysoida, minkälainen nykyinen Tampereen kaupungin tietovarastointi- ja raportointiympäristö on teoriaan verrattuna.

Tutkimustyön tavoitteena oli tietovarastoinnin ja raportoinnin toteutusratkaisujen kartoittaminen teoriaan perustuen, yleisesti käytettyjen ratkaisujen selvittäminen sekä analysointi, kuinka Tampereen kaupungin nykytilaa tulisi kehittää, jotta Power BI:tä voitaisiin hyödyntää paremmin sote-tiedon analysoinnissa.

Työssä perehdyttiin aihetta käsittelevään kirjallisuuteen ja muuhun kirjalliseen materiaaliin, tehtiin sähköinen kysely Power BI:n käyttötavoista sekä kartoitettiin Tampereen kaupungin tietovarastointi- ja raportointiympäristön nykytila. Tiedonkeruumenetelminä käytettiin kirjallisuuskatsausta, kvantitatiivista tutkimuskyselyä, havainnointia, sisäiseen dokumentaatioon perehtymistä sekä omakohtaista työkokemusta.

Tutkimus päättyi johtopäätökseen, että nykytilaa tulisi kehittää, jotta Microsoft Power BI:tä voitaisiin hyödyntää täysimääräisesti sote-tiedon analysoinnissa. Tutkimuksessa löytyi useita kehityskohteita Tampereen kaupungin kokonaisarkkitehtuurista. Tutkimuksessa tunnistettiin 15 kehittämiskohdetta, joiden avulla nykytilaa voisi kehittää.

Asiasanat: Microsoft Power BI, sote, tietovarastointi, liiketoimintatiedon hallinta, tietojohdaminen, tiedonhallinta

ABSTRACT

Tampereen ammattikorkeakoulu
Tampere University of Applied Sciences
Master's Degree Programme in Information Systems Competence

TUOMINEN, REIJA:

Utilizing Microsoft Power BI in Analyzing Social and Health Care Data of the City of Tampere

Master's thesis 90 pages, appendices 4 pages
April 2019

This thesis was commissioned by the city of Tampere Information Management Unit. The objective of this study was to investigate what data warehousing techniques exist, and to investigate the possibilities of using Microsoft Power BI as an analyzing tool. The aim of the study was to analyze the current state of data warehousing and business intelligence environment in the city of Tampere based on the results of the theory study.

The purpose of this study was to analyze how the current state of data warehousing and business intelligence environment in the city of Tampere should be developed in order to improve Microsoft Power BI utilization as an analyzing tool of social and health care data.

This master's thesis was implemented as a qualitative case study. Various research methodology was used in collecting information for this study. The theoretical base of the study was collected by a literary survey and a survey of internal documentation. The current state analysis was conducted through observations, researcher's professional skills and a survey of internal documentation. In addition to other methods, an electronic quantitative survey was carried out to collect data of the use of Microsoft Power BI.

The results of the research showed that the current state of enterprise architecture in the city of Tampere should be developed in the fields of business, data and applications. The research indicated 15 different targets for development in order to improve Microsoft Power BI utilization as an analyzing tool of social and health care data.

Key words: Microsoft Power BI, social and health care, data warehousing, business intelligence, knowledge management, data management

SISÄLLYS

1	JOHDANTO.....	7
1.1	Työn tausta	7
1.2	Työn tavoite, tarkoitus ja rajaukset	8
1.3	Työn toteutus	9
2	KESKEISET KÄSITTEET	11
2.1	Tiedonhallinta (Data management)	11
2.1.1	Mitä tieto on?	11
2.1.2	Sote-tiedon erityispiirteet	13
2.1.3	Tietojohtaminen	15
2.1.4	Sosiaali- ja terveystietojen tietojohdaminen.....	19
2.2	Tietovarastointi (Data Warehousing)	22
2.2.1	Tietovarastotyypit	25
2.2.2	Tietovarastojen arkkitehtuurit.....	28
2.2.3	Tietovarastojen suunnittelu	32
2.2.4	Tietovarastojen työkalut.....	35
2.3	Liiketoimintatiedon hallinta (Business Intelligence).....	38
2.3.1	Perusarkkitehtuuri.....	40
2.3.2	Keskeiset menetelmät	41
2.3.3	Tiedon laatu.....	46
3	MICROSOFT POWER BI ANALYSOINTITYÖKALUNA.....	49
3.1	Power BI	49
3.2	Power BI toimintamallit.....	54
3.3	Power BI tietolähteet ja tiedon muokkaus.....	57
4	NYKYTILA TAMPEREEN KAUPUNGILLA	60
4.1	Yleiskuva tietovarasto- ja raportointiympäristöstä.....	60
4.2	Tietovarastointi.....	63
4.3	Liiketoimintatiedon hallinta	66
4.4	Resurssien organisointi	70
5	TUTKIMUSTYÖN TULOKSET	73
5.1	Tietovarastointi ja raportointiympäristö	73
5.2	Tietojohtaminen	74
5.3	Microsoft Power BI:n hyödyntäminen tiedon analysoinnissa	76
5.4	Kehitysehdotukset.....	78
6	JOHTOPÄÄTÖKSET JA POHDINTA.....	83
	LÄHTEET	85
	LIITTEET	87

Liite 1. Power BI kyselylomakkeen kysymykset	87
--	----

LYHENTEET JA TERMIT

BI	Business Intelligence, liiketoimintatiedon hallinta
DW	Data Warehouse, tietovarasto
EDM	Enterprise Data Management, konseptinimi Tampe- reella käytössä olevalle tiedon välivarastolle
EDW	Enterprise Data Warehouse
ETL	Extract, Transform and Load – tiedon lataus taustajär- jestelmästä ja muokkausta esim. raportoinnin tarpeisiin
ODS	Operational Data Store
OLAP	OnLine Analytical Processing
PBI	Power BI
SAP BW	SAP Business Warehouse
SOTE	Sosiaali- ja terveystalvet

1 JOHDANTO

Vuosia meneillään ollut maakunta- ja sote-uudistus, jonka oli suunniteltu tulevan voimaan 1.1.2021, tulee asettamaan entistä enemmän vaatimuksia päätöksenteon tueksi tarvittavalle tiedolle. Väestön ikääntyminen ja huoltosuhteen kehittyminen epäedulliseen suuntaan tulevana vuosina edellyttävät toimia palvelurakenteen ja palveluverkoston kehittämiseksi, tuottavuuden parantamiseksi sekä panostusta ennaltaehkäisevään toimintaan. Tietoa tarvitaan nykyisten toimintojen tehokkuudesta, laadusta ja vaikuttavuudesta, jotta näihin toimiin pystyttäisiin vastaamaan. Tiedon avulla johtajat pystyvät suuntaamaan palvelut ja resurssit niihin kohteisiin, missä niillä pystytään tuottamaan mahdollisimman tehokkaasti hyvinvointia kansalaisille. (Klemola, Uusi-Ilkainen & Askola 2014, 9,11.) Klemolan, Uusi-Ilkaisen ja Askolan (2014, 11) mukaan Sosiaali- ja terveystieteiden johdon yksi suurimmista haasteista on lisääntyvän kysynnän ja käytettävissä olevan rahoituksen tasapainottaminen.

1.1 Työn tausta

Tampereen kaupungilla ollaan vuosien mittaan kehitetty sosiaali- ja terveystieteiden operatiivisen tason toimintaraportointia. Tampereen kaupungilla on käytössä tietovarastointi- ja raportointijärjestelmänä SAP Business Warehouse ja sitä käytetään lähinnä Hyvinvointipalvelujen toimintatietojen, Henkilöstöhallinnan ja Talouden tietojen operatiiviseen raportointiin. SAP BW:n käyttö tietovarastointi- ja raportointivälineenä on aloitettu noin viisitoista vuotta sitten ja sitä on laajennettu vuosien mittaan laajempaan käyttöön. Operatiivisessa raportoinnissa SAP BW täyttää pääsääntöisesti perusraportoinnin tarpeet ja sitä pystytään kehittämään pienimuotoisesti edelleen. Tietotarpeiden muuttuessa yhä enenevässä määrin perusraportoinnista ennakoivaan ja analysoivaan suuntaan tarvitaan nopeampia tapoja tuottaa laajoja data-aineistoja analyytikoille ja johtajille päätöksenteon tueksi.

Tämä opinnäytetyö on aiheeltaan ajankohtainen, sillä Tampereen kaupungin edistämä digitalisaatio ja kokeileva pilotointimalli tuovat paineita tuottaa moninaisia data-aineistoja kertaluonteisesti, nopealla aikataululla ja vaihtelevista tietolähteistä. Tämä tarve aiheuttaa haasteita nykyisille tietojärjestelmille ja toimintamalleille saada tuotettua data-aineistoja nykyisillä resursseilla. Nykyisen operatiivisessa raportoinnissa käytössä olevan SAP BW:n tarjoamat välineet eivät ole optimaalisin tapa tuottaa kertaluonteista, monesta eri toiminta-alueesta yhdisteltyä tietoa. Kaupungilla onkin valittu ICT-palveluiden tarjoamaksi tiedon analysointivälineeksi Microsoft Power BI.

1.2 Työn tavoite, tarkoitus ja rajaukset

Opinnäytetyön keskeisenä tavoitteena on tietovarastoinnin ja raportoinnin toteutusratkaisujen kartoittaminen teoriaan perustuen, yleisesti käytettyjen Power BI -ratkaisujen selvittäminen sekä nykyisen Tampereen kaupungin tietovarastointi- ja raportointiympäristön kartoittaminen. Opinnäytetyön tavoitteena oli myös selvittää soite-tiedon raportoinnin erityispiirteet, mm. lainsäädännöllisestä näkökulmasta ja kuinka soite-tiedon raportointi poikkeaa muista johdon raportoinnissa hyödynnetyistä tiedoista, esimerkiksi taloustiedon raportoinnista.

Opinnäytetyön tarkoituksena on tunnistaa Microsoft Power BI:n tarjoamia tiedon analysointimahdollisuuksia ja selvittää, kuinka Tampereen kaupungin nykytilaa tulisi kehittää, jotta Power BI:tä voitaisiin hyödyntää paremmin soite-tiedon analysoinnissa.

Opinnäytetyön toimeksiantajana toimi Tampereen kaupungin Tietohallintoyksikkö. Opinnäytetyöstä on hyötyä Tampereen kaupungin tietohallinnolle data- ja analytiikkapalveluiden toiminnan kehittämisessä ja tulevaisuuden tiekartan laati- misessa. Tavoitteena on myös lisätä tietoisuutta ja ymmärrystä tietojoh- tamisen ja liiketoimintatiedon hallinnan teknisten ratkaisujen merkityksestä kokonaisark- kitehtuuriin ja tietojen hyödynnettävyyteen. Suurin hyöty opinnäytetyön tekemi- sestä kuitenkin lienee opinnäytetyön tekijälle, sillä se kasvattaa omaa ammatil- lista osaamista ja antaa kattavan kuvan nykypäivän liiketoimintatiedon hallinnan eri ratkaisuista muuallakin kuin omassa työpaikassa.

Opinnäytetyön ulkopuolelle rajataan opinnäytetyössä havaittujen kehitysehdotusten jatkosuunnittelu ja toteuttaminen käytännössä. Tämä kokonaisuus rajataan opinnäytetyön ulkopuolelle aikataulullisista syistä sekä siksi, että tämän kokonaisuuden toteuttaminen opinnäytetyöntekijän nykyisessä työtehtävässä sovellusasiantuntijana ei ole mahdollista. Työn ulkopuolelle rajataan myös tutkiminen, kuinka sote-tiedon raportointi on nykytilassa toteutettu Tampereen kaupungilla verrattuna tulevaisuuden maakuntien käyttämiin Sote-tietopakettien mukaisiin tietorakenteisiin (Sote-tietopakettien käsikirja 2.1. 2018). Vaikka tämä kokonaisuus olisi hyvin mielenkiintoinen tutkia, on sen rajaaminen opinnäytetyön ulkopuolelle tässä vaiheessa perusteltua. Aihe on niin laaja, että siitä saisi tehtyä erillisen opinnäytetyön ja toiseksi valtakunnan tasolla sote-uudistuksen vaikutuksia ei vielä tiedä kukaan.

1.3 Työn toteutus

Opinnäytetyö on luonteeltaan laadullinen tutkimus ja lähestymistapana käytetään tapaustutkimusta, jossa pyritään syvälliseen ja laajaan ymmärrykseen tutkittavasta ilmiöstä ja tietoa tutkimukseen kerätään useista eri lähteistä eri menetelmin. Opinnäytetyössä selvitetään syvällisesti, mitä tietovarastointitekniikoita on olemassa ja selvitetään Power BI:n käyttömahdollisuudet analysointityökaluna. Opinnäytetyössä analysoidaan, minkälainen kaupungin teknologinen nykytila on teoriaosuuteen verrattuna sekä tunnistetaan kehittämisehdotuksia nykytilan parantamiseksi.

Opinnäytetyössä käytetään tiedon hakemiseen useita menetelmiä, riippuen kerättävästä tiedosta. Käyttämällä useita eri menetelmiä, tietoa saadaan kerättyä eri lähteistä ja eri näkökulmista. Eri näkökulmien ottaminen mukaan tiedonkeruumenetelmiin luo kattavan ymmärryksen alan käsitteistöstä, parhaista käytännöistä ja nykytilasta sekä auttaa analysoimaan olennaiset kehittämisen kohteet. Kirjallisuuskatsauksella luodaan tietoperusta, johon muilla menetelmillä muista alueista kerätty tieto tukeutuu ja pohjautuu.

Power BI:n käytöstä kerättiin tietoa tehdyllä sähköisellä kyselyllä, joka on määrällisen tutkimuksen tiedonkeruumenetelmä. Kyselyn tuloksia suunniteltiin hyödynnettävän analysoitaessa käyttökohteita, kuinka Power BI:n käyttö on jakautunut minkäkin tietovarastointitekniikan yhteyteen. Kyselyssä kysyttävät kysymykset määrittivät sitä mukaa, kun tietoperustaa saatiin kerättyä ja ymmärrys tietovarastointitekniikoista ja -ympäristöistä sekä Power BI:stä kasvoi.

Kysely toteutettiin sähköisellä kyselylomakkeella ja vastausaikaa kyselyyn oli 9.3. - 20.3.2019. Kyselylomakkeen kysymykset vastausvaihtoehtoineen on kuvattu liitteessä 1. Kysely halukkuudesta osallistua Power BI:tä koskevaan kyselyyn lähetettiin sähköpostitse yhdeksän eri yrityksen edustajalle. Yrityksistä kuusi edusti Power BI konsultointia harjoittavaa toimintaa ja kolme potentiaalista asiakaskuntaa. Kysely lähetettiin ennakkotiedustelun jälkeen kuuden eri yrityksen edustajalle, joista kaksi edusti potentiaalista käyttäjäkuntaa ja neljä konsultointiyritystä. Kyselyn saatteessa mainittiin, että kyselyä voi jakaa edelleen organisaatiossa henkilöille, jotka tekevät Power BI kehitystä. Tämän takia lopullista vastaanottajamäärää ei ole tiedossa.

Nykytilan kartoituksessa käytettiin menetelmänä havainnointia, jolla pystytään luotettavasti keräämään tietoa aihealueesta. Havainnoinnista saatua tietoa käytettiin pohjana tavoitetilan kuvaamiseen, sillä sen avulla voitiin määrittää vaatimuksia, mitä Power BI:stä analysointivälineenä vaaditaan. Nykytilan kartoituksessa käytettiin myös sisäiseen dokumentaatioon perehtymistä, jolla saatiin hyvä kuva olemassa olevasta dokumentaatiosta ja pystyttiin muodostamaan kuva nykytilasta, millaisena dokumentoitu materiaali sen esittää. Nykytilan kartoituksessa käytettiin havaintoja myös eri kehityshankkeisiin liittyneistä kokouksista ja palavereista, joissa käsiteltiin opinnäytetyöhön liittyviä aihealueita.

2 KESKEISET KÄSITTEET

2.1 Tiedonhallinta (Data management)

Tiedonhallinta sisältää tiedon tallentamiseen liittyvän teknisen näkökulman lisäksi myös organisaatioiden toiminnan ja tiedonkulun yhdistämistä. Teknisen näkökulman lisäksi tiedonhallinnassa on pohdittava tiedon omistajuuteen ja tiedon käytön vastuisiin kuuluvia kysymyksiä. Tiedonhallintaan kuuluu myös tietojen saatavuus ja luovuttaminen sekä kysymykset tietoturvasta ja -suojusta. Tämän opinnäytetyön aihealueeseen kuuluu tiedonhallinnan alueelta tiedon tekninen olemus ja tallentaminen, omistajuus ja tiedon hyödyntäminen. Opinnäytetyössä ei käsitellä tarkemmin tietoturvaa ja -suojaa kuin niiltä osin, jotka ovat sote-tiedon analysoinnin kannalta olennaisia.

2.1.1 Mitä tieto on?

Tiedolla on monta määritelmää, jotka ovat riippuvaisia tulkitsijan lähtökohdista. Yhtä totuutta ei ole olemassa, mutta tietoa voidaan jäsentellä eri tasoihin, jotta aihetta olisi helpompi käsitellä. Yleinen tietojohdamisen alalla vallitseva jäsentelytapa on jakaa tieto kolmeen eri tasoon: dataan, informaatioon ja tietämykseen. (Laihonen ym. 2013, 17-18.) Tällä perusjaottelun tasolla oleva tieto palvelee kuvassa 2 esiteltyä tietojohdamisen käytäntöjen ja johtamistyökalujen näkökulmaa. Datan, informaation ja tietämyksen yläpuolelle kerääntyy korkeampaa tietoa: ymmärrystä ja viisautta. Akateemisessa kirjallisuudessa esiintyy vielä näiden äsken kuvattujen käsitteiden lisäksi myös korkeamman tiedon jaottelut älykkyyteen ja totuuteen. (Laihonen ym. 2013, 18.) Tässä yhteydessä pitäydytään käsittelemään tietoa datan, informaation, tietämyksen, ymmärryksen ja viisauden tasoilla, koska niiden avulla kyetään selvittämään tietojohdamisen määritelmää. Älykkyys ja totuus kuuluvat enemmän filosofian tieteenalaan ja tässä opinnäytetyössä on tarkoituksena tutkia tietojärjestelmäosaamisen koulutuksen alan tutkimuskysymyksiä.

Taulukossa 1 on havainnollistettu tietoa viisiportaisesti esittäen tasoihin liittyvä esimerkki ja tulkinta jäsennystasosta. Alimmalla, datan tasolla tieto on pelkkää jäsentymätöntä datamassaa, josta ei pysty vielä päättelemään mitään. Data voi olla esimerkiksi mittaustuloksia tai laitteiden anturidataa, mutta siitä ei välttämättä vielä selviä, mistä mittauksista on kyse tai että mitä anturidataa on tallennettu. Kun dataan lisätään asiayhteys, voidaan datasta päätellä jo esimerkiksi, että kyseessä on BMI (kehon painoindeksi) mittaukseen liittyvät painon ja pituuden, sekä ikään ja sukupuoleen liittyvät mittaustiedot. Tällöin puhutaan jo informaatiosta. Yksittäisillä informaation palasillakaan ei vielä juuri voida tehdä johtopäätöksiä yleisemmin, vaan informaatiota pitää tulkita asiayhteydessä, jotta siitä syntyi tietämystä. Monesti vielä tietämyskään ei riitä päätöksenteon tueksi, vaan tietämystä pitää vielä tulkita inhimillisistä lähtökohdista aiemmin koetun tiedon, taidon ja kokemuksen kautta.

Nämä ylempien tiedon tasot – ymmärrys ja viisaus – auttavat tietojohdamisen näkökulmassa ilmiön ymmärtäminen, sillä ymmärryksen ja viisauden tasolla tiedosta muodostuu analysoitua tietämystä päätöksenteon tueksi. Mitä lähemmäksi tiedon portaita mennään viisautta, sitä voimakkaammin korostuu tiedon tulkitsijan merkitys, sillä tiedon tulkintaan tarvitaan myös hiljaista tietoa ja kokemusta, joita tulkitsijalla on.

TAULUKKO 1. Datan jalostuminen tietämykseksi ja viisaudeksi. (Laihonen 2009, 112, muokattu)

	Tiedon tasot	Esimerkki	Tulkinta																								
Tiedon konteksti	Viisaus	Laitoshoidossa olevien yli 85 vuotiaiden ravitsemukseen pitää kiinnittää huomiota ja puuttua aliravitsemuksen syihin.	Ymmärrys muuttuu viisaudeksi, kun ymmärrykseen lisätään tietoa, taitoa ja kokemusta.																								
	Ymmärrys	Laitoshoidossa olevien yli 85 vuotiaiden henkilöiden BMI sijoittuu keskimääräistä useammin luokkaan alipainoinen.	Tietämyksen tulkintannassa korostuu inhimillinen toiminta.																								
	Tietämys	Tietokantaan on tallennettu alueellamme 350 000 BMI-mittausta. Naisten osuus 65 – 75 vuotiaiden mittauksissa on 62 %.	Ihminen tulkitsee informaatiota asiayhteydessä.																								
	Informaatio	<table border="1"> <thead> <tr> <th>ID</th> <th>Sukupu.</th> <th>Ikä</th> <th>Pituus</th> <th>Paino</th> <th>BMI</th> </tr> </thead> <tbody> <tr> <td>8821</td> <td>1</td> <td>87</td> <td>160</td> <td>48</td> <td>18,8</td> </tr> <tr> <td>8822</td> <td>2</td> <td>56</td> <td>192</td> <td>90</td> <td>24,4</td> </tr> <tr> <td>8823</td> <td>2</td> <td>25</td> <td>178</td> <td>97</td> <td>30,6</td> </tr> </tbody> </table>	ID	Sukupu.	Ikä	Pituus	Paino	BMI	8821	1	87	160	48	18,8	8822	2	56	192	90	24,4	8823	2	25	178	97	30,6	Rakenteista dataa, johon yhdistyy asiayhteys.
	ID	Sukupu.	Ikä	Pituus	Paino	BMI																					
8821	1	87	160	48	18,8																						
8822	2	56	192	90	24,4																						
8823	2	25	178	97	30,6																						
Data	<table border="1"> <tbody> <tr> <td>8821</td> <td>1</td> <td>87</td> <td>160</td> <td>48</td> <td>18,8</td> </tr> <tr> <td>8822</td> <td>2</td> <td>56</td> <td>192</td> <td>90</td> <td>24,4</td> </tr> <tr> <td>8823</td> <td>2</td> <td>25</td> <td>178</td> <td>97</td> <td>30,6</td> </tr> </tbody> </table>	8821	1	87	160	48	18,8	8822	2	56	192	90	24,4	8823	2	25	178	97	30,6	Yksittäisiä tiedon jyväsiä, faktoja ja lukuja ilman asiayhteyttä.							
8821	1	87	160	48	18,8																						
8822	2	56	192	90	24,4																						
8823	2	25	178	97	30,6																						
			Tiedon sisäistäminen																								

Kaikkea tietoa ei voida tallentaa tietojärjestelmiin ja tietokantoihin, vaan organisaatioissa ja yksilöillä on paljon hiljaista tietoa, minkä ansiosta tiedon tulkintaa ja analysointia voidaan tehdä. Tämän tiedon jalostusprosessin ansiosta yksittäisistä tiedon jyvistä saadaan rikastettua ymmärrystä ja viisautta, jonka avulla pystytään tekemään organisaation edun mukaisia tietoon pohjautuvia päätöksiä. Myös organisaatioiden toimintatavoissa ja viestintäkulttuurissa on tallentuneena hiljaista tietoa ja osaamista, mikä muokkaa tiedon jalostumista datasta viisaudeksi. Tietojohtamisessa tarvitaan johtamiskyvykkyyttä tiedon kaikilla tasoilla ja varsinkin ymmärrystä siitä, että tietojohtaminen ei saa keskittyä pelkästään tiedon alimmilla tasoilla olevan tietojärjestelmiin ja tietokantoihin tallentuneen tiedon informaation hallintaan. (Kianto n.d., 5)

2.1.2 Sote-tiedon erityispiirteet

Sosiaali- ja terveystalvissa käytettäviin tietojärjestelmiin tallennetaan valtava määrä tietoa kansalaisten sote-palveluiden kontakteista päivittäin. Kansainvälistikin vertailtuna järjestelmät ovat erittäin kattavia. (Sund, Nylander & Palonen 2004) Digitaalisen tiedon kattavuudesta kertoo myös se, että sähköisen potilaskertomuksen levinneisyys julkisessa terveydenhuollossa on ollut vuodesta 2007 lähtien 100% ja käyttöaste hyvin korkea. Samassa tutkimuksessa kerrottiin sähköisen potilaskertomusjärjestelmän levinneisyyden olevan 100% myös erikoissairaanhoidossa. (Reponen, J., Kangas, M., Hämäläinen, P., Keränen, N. & Haverinen 2018, 40-41.) Julkisissa sosiaalipalveluorganisaatioissa sähköiseen muotoon asiakastietojärjestelmiin tallennettavan dokumentaation osuus ei ole Kuusisto-Niemen, Ryhäsen ja Hyppönen (2018) tekemän raportin mukaan ihan niin kattava kuin julkisessa terveydenhuollossa, mutta olevan kuitenkin valtaosin yli 80 %.

Vaikka tietoa kertyykin määrällisesti paljon, se hajautuu useaan eri tietojärjestelmään. Valtakunnalliset tietojärjestelmät perustuvat omaan toimintatapaansa ja paikalliset omaansa. Informaatiojärjestelmien monijakoisuudesta kertoo sekin, että myös paikallisella tasolla olevat järjestelmät jakaantuvat toiminnoittain. Niin sosiaalihuollossa, erikoissairaanhoidossa kuin perusterveydenhuollossa on mo-

nosti omat järjestelmänsä. (Sund ym. 2004.) Näihin eriytyneisiin järjestelmiin kerätään yhden perustoiminnon dataa ja tietojen yhdistäminen on haastavaa. Sundin ym. (2004) mukaan tietojärjestelmien rakentaminen palvelemaan yhtä perustoimintoa varten on aiheuttanut järjestelmien pirstaloitumista. Järjestelmistä löytyy pääosin palvelujen käyttötietoja Sundin ym. (2004) mukaan siksi, että ne ovat monesti alun perin kehitetty palvelemaan lähinnä hallinnollisia tarpeita.

Tietojen runsaudesta ja hajanaisuudesta johtuen datasta on vaikea tulkita ja yhdistellä olennaista informaatiota päätöksenteon tueksi. Sundin ym. (2004) tutkimuksen teon jälkeen sote-alan tietojärjestelmät ovat kehittyneet avoimempaan ja toiminnot yhdistävään suuntaan, vaikka paikallisella tasolla sote-alan toiminta perustuukin vielä erillisiin tietojärjestelmiin. Esimerkiksi valtakunnallinen Kanta -palvelu tuo kansalaisille yhtenäisen ja digitaalisen käyttöliittymän omiin sosiaali- ja terveydenhoidon palveluihin asuinpaikasta ja palveluntarjoajasta riippumatta. Vielä ei kuitenkaan voida luopua paikallisista tietojärjestelmistä, sillä Kanta -palvelu kattaa vain osan lukuisiin tietojärjestelmiin kirjattavasta tiedosta.

Sundin, Nylanderin ja Palosen (2004) mukaan suomalaisen sote-käytännön erityispiirteitä ovat 1) yhtenäinen henkilötunnuksen käyttö kaikessa valtakunnallisessa ja paikallisessa toiminnassa, 2) pitkät perinteet henkilötunnisteellisen materiaalin keräämisessä perustietojärjestelmätasolla (väestötiedoista lähtien) ja 3) atk-toimintojen tulo jo varhaisessa vaiheessa osaksi informaatiojärjestelmää. Sote-tiedon erityispiirteisiin kuuluu tiedon tallennuksen epävarmuus ja alttius muutoksille. Toisin kuin talous- tai myyntitietojen käsittely, sote-tieto on moniulotteisempaa ja muuttuvampaa kuin pääsääntöisesti talous- ja myyntitietokirjaukset ovat. Kirjanpitoon kirjataan talouslukuja ja ne eivät juuri muutu, paitsi vastakirjauksilla tai laskut hyvityslaskuilla. Näissä toiminnoissa syötetyt tiedot yleensä kirjataan tapahtumahetkellä, niitä ei jälkikäteen muuteta (paitsi vastakirjauksilla ja hyvityslaskuilla) tai poisteta.

Toisin on sote-tiedon kanssa: tietoja ei välttämättä syötetä tapahtumahetkellä, vaan ne voidaan kirjata pitkänkin ajan kuluttua, kirjaaja ei välttämättä ole tapahtuman suorittaja, tietoja poistetaan takautuvasti ja tietoja muutetaan jälkikäteen yleisesti. (Kimball & Ross 2013, 352) Tämä aiheuttaa analysoinnille haasteita,

sillä tietojärjestelmiin tallentunut tieto on jatkuvassa muutoksessa, eikä stabiilia tilannetta koskaan saavuteta.

Toinen erityispiirre sote-tiedossa on tietojen aikariippuvaisuus ja sidonnaisuus laissa määrättyihin prosesseihin, varsinkin sosiaalihuollon alalla. Tietojärjestelmiin kirjataan esimerkiksi hoitajaksoja, jolloin jakso alkaa jonain tiettyinä päivinä ja päättyy tulevaisuudessa. Tämä aiheuttaa tiedon analysoinnin osalta tulkintakysymyksiä, kun tapahtuma (eli hoitajakso) ei ole sidottu yhteen ajanhetkeen, vaan on voimassa jonkun ennalta määrittämättömän ajanjakson ajan.

Kolmas sote-tiedon erityispiirre on tiedon arkaluonteisuus ja sidonnaisuus henkilötunnukseen. Lähes kaikki sote-tieto sisältää asiakkaan henkilötunnuksen, jolla tieto pystytään identifioimaan tiettyyn henkilöön ja henkilö tunnistamaan henkilötunnuksen avulla. Vain vastasyntyneillä lapsilla, ulkomaalaisilla henkilöillä ja turvapaikanhakijoilla ennen virallisen henkilötunnuksen myöntämistä ei ole käytössään suomalaista henkilötunnusta sote-tietojärjestelmissä. Poikkeuksen tekee myös lastensuojelun vireillepanoilmoitus, jossa ei aina pystytä kohdentamaan henkilöä, jota ilmoitus koskee. Henkilötunnusriippuvaisuus tekee sote-tiedon käsittelystä luvanvaraista ja käsittelijällä pitää olla aina oikeus ja lupa käsitellä arkaluonteista tietoa. EU:n Yleisen tietosuojasetuksen (GDPR) soveltamisen voimaantulon 25.5.2018 jälkeen sote-tietojen käsittelyyn kohdistuu entistä enemmän sääntelyä ja rajoitteita sen käytöstä.

2.1.3 Tietojohtaminen

Yrityksien käytössä oleva tieto lisääntyy hurjaa vauhtia ja tieto- ja viestintäteknologian nopea kehitys sekä muutosvauhdin kiihtyminen asettavat organisaatioiden johtamiselle uusia haasteita. Osaaminen ja tieto ovat nousseet nykyaikaisessa yrityksessä kilpailukykyyn ja menestyksen tärkeimmiksi tekijöiksi. Kilpailukykyisellä yrityksellä on kykyä johtaa ja luoda tietoa, joten avainasemassa on se, kuinka hyvin yritykset pystyvät muuttamaan tiedon ja osaamisen menestyväksi liiketoiminnaksi. (Laihonen ym. 2013, 7; Kianto n.d., 1) Yhteiskunnan muuttuessa teollisesta valmistuskulttuurista jälkiteolliseen tieto- ja palveluyhteiskuntaan johtamisen mallit ja käytännöt kehittyvät tiedon ja osaamisen merkityksen korostuessa.

Tietojohtaminen johtamisen osa-alueena alkoi vakiinnuttaa jalansijaa Suomessa 1990-luvulla, jolloin nopea tieto- ja viestintäteknologian kehittyminen tarjosi mahdollisuuden tiedon ja datan tallennukseen, analysointiin ja välittämiseen. (Laihonen ym. 2013, 6.) Toimintaympäristön muuttuessa entistä nopeammalla tahdilla yritysten on kyettävä uudistamaan tietopääomaansa jatkuvasti. Tietopääomaa kasvattavat yrityksessä oleva osaaminen ja innovaatiot, jolloin uudistumiskykyinen organisaatio menestyy muuttuvissakin oloissa. (Kianto n.d., 1)

Johtamisen osa-alueena tietojohtaminen on uudehko, vasta viime vuosikymmeninä organisaatioiden johtamiseen kehittynyt näkökulma. Johtuen ilmiön pääsääntöisesti aineettomasta luonteesta tietojohtamisen ymmärtämiseen tarvitaan selkeitä ja yksiselitteisiä käsitteitä. Tietojohtamisessa käytetään erilaisia ja osittain myös ristiriitaisia käsitteitä, joka johtuu tietojohtamisen käsitteiden vakiintumattomuudesta. (Laihonen ym. 2013, 6.)

Käsitteiden vakiintumattomuudesta huolimatta tietojohtamiseen liittyvää toimintaa harjoitetaan yleisesti yrityksen koosta riippumatta. Tietojohtamisesta puhutaan yrityksissä eri termeillä, organisaation kehittäminen, tiimityöskentely ja jatkuva parantaminen ovat tietojohtamisen muotoja. (Kianto n.d., 2.) Yrityksissä organisaatioiden monissa eri työrooleissa tarvitaan kykyä hallita tietämystä, hakea, luoda ja hyödyntää tietoa, joten organisaatioissa tiedon johtamisesta on tullut entistä tärkeämpää. Tietoa luodaan ja hyödynnetään organisaation joka tasolla, mutta onnistuakseen tietojohtaminen tarvitsee välttämättä organisaatiossa ylimmän johdon tuen. (Kianto n.d., 2.)

Laihonen ym. (2013) käsittelee kirjassa tietojohtamisen ilmiötä ottamalla lähtökohdaksi johtamisen haasteet tieto- ja palveluyhteiskunnassa. Tähän haasteeseen voidaan vastata kolmella eri tarkastelutasolla: ilmiön ymmärtäminen, johtamisen käytännöt ja johtamisen työkalut. Kuvassa 1 Laihonen ym. (2013) havainnollistaa tietojohtamisen lähtökohdan ja näkökulmat.



KUVA 1. Tietojohdamisen näkökulmat (Laihonen ym. 2013, 7)

Johtamisessa, kuten myös missä tahansa muussa asiassa on olennaista ymmärtää aihealue, käsitteet ja viitekehys, mistä puhutaan. Ilman yhtenäisiä käsitteitä ja taustatietoa aiheesta on vaikea keskustella samasta asiasta. Tietojohdamisen näkökulma ilmiön ymmärtäminen tuo ymmärrystä tiedon arvonluonnista, kuinka tiedon avulla voidaan luoda lisäarvoa erilaisissa liiketoimintaprosesseissa- ja ympäristöissä. (Laihonen ym. 2013, 7) Tietojohdamisessa on ymmärrettävä tiedon merkitys kulloisessakin yhteydessä ja sen vaikutukset päätöksentekoon. Kun perusymmärrys organisaation tiedosta ja osaamisesta on olemassa, on valmiudet miettiä tiedon hyödyntämistä ja siitä saatavaa lisäarvoa liiketoimintaan.

Aivan yhtä tärkeää on ymmärtää, mitä osaamista ja tietoa organisaatiossa on olemassa ja mistä se on saatavilla. Jos esimerkiksi organisaatiossa tarvittaisiin tietoa väitteen 'psykiatrisen hoidon tarve on lisääntynyt lasten keskuudessa' tueksi, olisi hyvä tietää, mitä taustatietoa lähdejärjestelmissä on tämän väitteen tueksi olemassa. Tässä esimerkissä voitaisiin lähteä hakemaan tietoa pakollisten viidesluokkalaisten terveystarkastuksista (Valtioneuvoston asetus neuvolatoiminnasta, koulu- ja opiskeluterveydenhuollosta sekä lasten ja nuorten ehkäisevästä

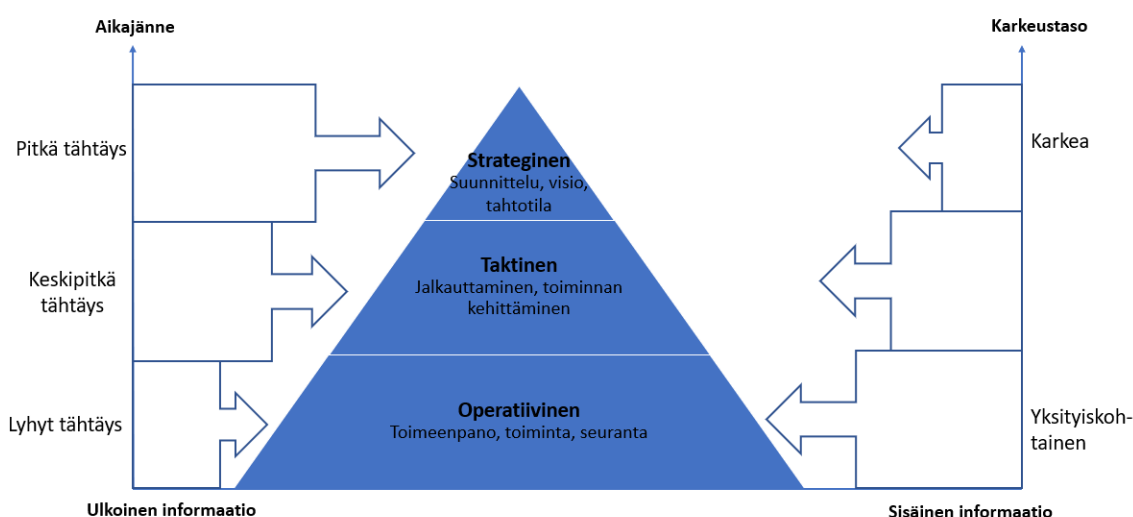
suun terveydenhuollosta VNA 338/2011) ja törmätä siihen, että kyseisistä terveystarkastuksista järjestelmiin ei välttämättä kirjata kattavasti ja järjestelmällisesti tietoa tarkastuksessa havaituista diagnooseista tai muuta lisäinformaatiota. Aina ei ole olemassa tietoa päätöksenteon tueksi, mutta tietojohdamisen tavoitteena on myös kyvykyys ymmärtää organisaation tiedon hallinnan taso niin osaamisen kuin teknologiankin näkökulmasta.

Johtamisen käytännöt -näkökulma tuo käytäntöjä ja näkökulmia siihen, kuinka organisaatiossa johdetaan tietoresursseja ja kuinka organisaatiota johdetaan tiedon avulla. (Laihonen ym. 2013, 7) Johtamisen käytäntöjen näkökulmassa painottuu eritoten henkilöstön inhimillisen ja sosiaalisen pääoman johtaminen, sillä ilman osaavaa ja motivoitunutta henkilöstöä organisaatiokaan ei pysty toimimaan tuottavasti. Henkilöstöä ei voida pakottaa jakamaan tietoa, kasvattamaan osaamistaan eikä osallistua tiedon tuottamiseen. Luomalla johtamisen käytännöillä hyvät edellytykset luottamuksen, sosiaalisten verkoston ja jaetun kulttuurin keinoin voidaan edistää henkilöstön kykyä luoda, jakaa ja hyödyntää tietoa. (Kianto n.d., 5-6.)

Kolmas näkökulma, johtamistyökalut, tuo tietojohdamiseen käytännön johtamistyökalut ja teknologiaa, minkälaisilla välineillä johtamista tuetaan ja miten välineitä hyödynnetään. (Laihonen ym. 2013, 7) Tietojohdamisen kehityksen alkuaikoina 1990-luvulla tietojohdaminen painottui vahvasti teknologianäkökulmaan ja keskittyi enemmän tiedon tuottamiseen ja jakamiseen. Kehityksen ja ymmärryksen myötä johtamistyökalujen painopiste onkin siirtynyt tiedon tuottamisesta ja hallinnasta enemmän tiedon hyödyntämisen suuntaan. (Laihonen ym. 2013, 10.; Kianto n.d., 13-14.) Teknologian kehittymisen myötä tietojohdamisen välineet monipuolistuvat ja se mahdollistaa myös uudenlaisen tiedon hyödyntämistä ja yhdistämistä perinteiseen tietoon. Enää ei niinkään keskitytä tiedon tallentamiseen tietokantaan ja peruseräraportointiin, vaan tallennettua tietoa voidaan hyödyntää erilaisilla analyysityökaluilla ja yhdistää sitä myös organisaation ulkopuolelta tulevaan tietoon.

Johtaminen jakaantuu kolmeen tasoon: strategiseen, taktiseen ja operatiiviseen tasoon. Ylimmän johdon tietojohdamisen tarpeet ovat yleensä strategisen tason johtamistarpeita varten, jolloin tietojohdamista tarkastellaan pitkällä aikavälillä,

noin 5-10 vuotta eteenpäin. Tietojohtamisen taktinen taso keskittyy keskipitkälle aikavälille ja operatiivisella tasolla johdetaan päivittäistä toimintaa. (Klemola ym. 2014, 11.) Mitä ylemmälle tasolle johtamisen tarpeissa mennään, sitä summattumpaa ja karkeammalla tasolla tietoa tarvitaan. Samalla ulkoisen informaation hyödyntämisen tarve kasvaa, mitä strategisemmalla tasolla tietoja tarkastellaan. Kuva 2 havainnollistaa tietojohtamisen tarpeita ja tiedon yksityiskohtaisuutta eri tasoilla. Operatiivisessa johtamisessa tavoitteena on toiminnan seuranta ja kehittäminen, jolloin tarvitaan yksityiskohtaisempaa tietoa, eikä ulkoiselle informaatiolle ole juuri tarvetta.



KUVA 2. Tietojohtamisen tarpeet eri tasoilla (Aho 2011, 23, muokattu)

2.1.4 Sosiaali- ja terveyspalveluiden tietojohtaminen

Sosiaali- ja terveyspalveluiden strategista tietojohtamista ohjaavat organisaation pitkän aikavälin tavoitteet, jotka määrittävät johdon tietotarpeita toiminnan kehittämiseksi tavoitteenmukaiseen suuntaan. Suurimpia haasteita julkisissa sote-palveluissa on tällä hetkellä voimakkaasti lisääntyvä kysyntä ja palveluihin käytettävissä olevan rahamäärän riittämättömyys. Tulevaisuuden tietotarpeet sote-palveluiden tietojohtamisessa painottuvat palveluiden palvelutarpeesta sekä nykyisten palveluiden vaikuttavuudesta, tehokkuudesta ja laadusta. (Klemola ym. 2014, 11.)

Markkinataloudessa kysyntä ja tarjonta määräävät taloustieteen teorian mukaan hyödykkeen hinnan ja tuotannon määrän markkinoilla. Vapaassa markkinataloudessa kysynnän ja tarjonnan tasapaino muodostuu hintatasolla, jolla ostaja on valmis ostamaan hyödykettä ja valmistaja sitä valmistamaan. Julkisella sektorilla on otettava huomioon myös säädetyt lait ja asetukset, jotka luovat tarpeita erilaisten sosiaali- ja terveystalouden järjestämistä varten. Toisen erottavan seikan vapaaseen markkinatalouteen tekee pääsääntöisesti julkisin verovaroin järjestettävät palvelut, joten kysynnän ja tarjonnan tasapaino ei kaikilta osin päde julkiseen sosiaali- ja terveydenhuoltoon. (Klemola ym. 2014, 11.)

Sote-palveluiden tietojohdamisen erityispiirteisiin liittyy myös ulkoinen informaatio-ohjaus, jota ei juurikaan yksityisellä, kaupallisella sektorilla esiinny. Laihonen (2009, 172) on väitöskirjassaan jakanut ulkoisen informaatio-ohjauksen kolmeen tasoon: valtionhallinnon informaatio-ohjaukseen, hallinnolliseen informaatio-ohjaukseen ja asiantuntijaohjaukseen. Tietojohdamisen näkökulmista näistä valtionhallinnon informaatio-ohjausta voidaan pitää strategisena johtamisena, hallinnollista informaatio-ohjausta taktisena johtamisena ja asiantuntijaohjausta operatiivisena johtamisena. Taulukossa 2 Laihonen (2009, 173) on käsitellyt informaatio-ohjausta kysymysten kuka (ohjaaja), mitä (sisältö) ja kenelle (kohde) näkökulmista.

TAULUKKO 2. Ulkoisen informaatio-ohjauksen luokittelu ohjaajan (kuka), sisällön (mitä) ja kohteen (kenelle) perusteella. (Laihonen 2009, 173)

	Valtionhallinnon informaatio-ohjaus	Hallinnollinen informaatio-ohjaus	Asiantuntijaohjaus
Kuka?	Valtionhallinto	Hallinnon asiantuntijat	Lääke-/hoitotieteen asiantuntijatahot
Mitä?	Terveyspolitiikka, strateginen ohjaus	Taktinen/operatiivinen ohjaus	Toiminnanohjaus, kehittäminen, tiedon tulkinta, osaamisen jakaminen
Kenelle?	Omistajat (kunnat), yhtymävaltuusto, yhtymähallitus, johtoryhmä	Hallinnon asiantuntijat, koko organisaatio	Lääke-/hoitotieteen asiantuntijatahot

Palveluiden moninaisuudesta johtuen sosiaali- ja terveystaloudellisia tuottavissa organisaatioissa tietojärjestelmien ja raportointiympäristöjen lukumäärä on iso ja järjestelmien kirjo on suuri. Tietojärjestelmät on suunniteltu operatiivista käyttöä varten ja niihin syötettävän tiedon tarkkuus on yksityiskohtainen. Raportointi

näistä operatiivisista järjestelmistä on suunnattu etupäässä palvelemaan toiminnan tarpeita, mutta niissä oleva tieto on hajautunut eri järjestelmiin strategisen johtamisen näkökulmasta. Strategisessa johtamisessa tarvittavan tiedon tulisi olla useasta lähteestä yhdisteltyä, yhteismitallista ja sisältää myös ulkoisia tietolähteitä, jotta sitä pystyttäisiin hyödyntämään ylimmän tason päätöksenteossa.

Eri tietolähteiden moninaisuus aiheuttaa tietojen yhdistämisen vaikeuden, kun tietolähteiden tiedot eivät ole yhteismitallisia eikä tietojen käsitteet ole välttämättä yhtenäisiä. Sosiaali- ja terveystietojärjestelmien tietojärjestelmiin syötetty tieto ei usein myöskään ole helposti saatavilla puutteellisten rajapintojen takia. (Klemola ym. 2014, 12.) Sosiaali- ja terveysministeriö ja Sitra ovat kehittäneet Sote-tietopaketteja, joiden avulla pystytään keräämään luotettavaa, läpinäkyvää ja vertailukelpoista tietoa valtakunnallisen ja maakunnallisen sosiaali- ja terveystietojärjestelmien päätöksenteon avuksi. Sote-tietopakettien tavoitteena on tuottaa vertailukelpoista tietoa maakuntien käyttämien sosiaali- ja terveystietojärjestelmien kokonaisuudesta. (Sote-tietopakettien käsikirja 2.1. 2018, 3.)

Yhtenäisellä ja valtakunnallisella tiedon määrittelyllä ja laskentaperiaatteilla edistetään tietojen yhdistämistä ja vertailtavuutta, kun käsitteet määritellään yhtenäisiksi ja laskentaperiaatteet yhdenmukaisesti. Määrittelyjen ja laskentaperiaatteiden luonti valtakunnan tasolla ei ole helppoa ja yksiselitteistä, sillä valmiita malleja tai kokemuksia näin laajaan sote-toimintokokonaisuuden ohjaamiseen ei löydy edes kansainvälisesti. (Sote-tietopakettien käsikirja 2.1. 2018, 3.)

Strategista päätöksentekoa varten tietoja joudutaan usein keräämään manuaalisesti kertaluontoisina poimintoina ja yhdistelemään ulkoisiin tietolähteisiin analysointia varten, koska tieto on hajallaan eri järjestelmissä ja rajapinnat niiden poimimiseksi ovat puutteellisia. Manuaaliseen kertaluonteiseen poimintaan ja tietojen jalostamiseen analysointia varten kuluu organisaatiolta paljon aikaa ja työtä, eikä tietojen poiminta ole useinkaan ole uusittavissa täysin samanlaisena kuin ensimmäisellä kerralla. (Klemola ym. 2014, 12.)

2.2 Tietovarastointi (Data Warehousing)

Tietovarastointitermin määritelmä julkaistiin ensimmäisen kerran helmikuussa 1988 IBM Systems Journal -julkaisussa Barry Devlin ja Paul Murphyn julkaisemassa artikkelissa. Tietovarastoinnin kehitys on pysynyt melko vakaana kolmen vuosikymmenen aikana, mutta aivan viime vuosina perinteisen tietovarastoinnin on haastanut modernimman tietovaraston käsite, tietojärvi (data lake). (Devlin 2018.) Modernimman tietovarastoinnin ratkaistut vastaavat nykypäivän haasteeseen valtavasti kasvavan ja laajenevan datamäärän, kuten internet-tilastodata, IoT (IoT, internet of things) -data ja sosiaalisen median tuottama datan, analysointiin ja käsittelyyn.

Tietojärvi (Data Lake) -termin juuret juontavat vuoteen 2010, jolloin Pentaho -yrityksen teknologiajohtaja James Dixon lanseerasi kyseisen termin Pentahon Hadoop -julkistuksen yhteydessä. Dixon käytti tietovaraston ja tietojärven eroista vertausta: tietovarasto on kuin pullollinen puhdistettua vettä, kun taas Data Lake on järvi luonnollisemmassa tilassaan. (Dixon 2010.) Tietojärvi -termillä on erilaisia tulkintoja riippuen siitä, mitä tietoa tietojärvestä on. Devlinin (2018) tulkinnan mukaan termin alkuperästä johtuen se sisältää vain ja ainoastaan internetiin liittyvää dataa. Termin määritelmä ei ole kovin vakiintunut ja siitä onkin johdettu monenlaisia uusia termejä kuvaamaan tietojärveä sen sisällön perusteella. Devlin (2018) on luonut termin data lough kuvaamaan tietoallasta, johon voidaan laajemmin ymmärtää kuuluvaksi muutakin kuin internetiin liittyvää dataa. Muita termejä ovat mm. data pond, tietoallas ja data-allas.

Tietovarastoista tietoa hakiessa ei voi olla törmäämättä kahteen nimeen: Bill Inmon ja Ralph Kimball. Näitä kahta herraa voidaan pitää tietovarastoinnin uranuurtajina ja perustan luojina. Inmonin (2005, 29) määritelmän mukaan tietovarasto on kokoelma pysyvää, asiakokonaisuuksittain integroitua tietokantoja, jotka on suunniteltu tukemaan päätöksentekoa ja joiden tieto on sidottu johonkin ajanhetkeen. Inmonin koulukunta näkee tietovaraston enemmänkin tietokantana, kun taas Kimballin koulukunta sisällyttää tietovaraston määritelmään myös tavan, millä tietovarasto on mallinnettu sekä tietojen latausprosessin. Kimballin ja Casertan (2004, 23) määritelmän mukaan tietovarasto on järjestelmä, joka poimii,

puhdistaa ja yhdenmukaistaa tietoa lähdejärjestelmistä dimensionaalisen tietovarastoon sekä tukee ja mahdollistaa kyselyt sekä analyysit päätöksenteon tueksi.

Sekä Inmonin että Kimballin koulukuntien perusnäkemys tietovarastosta on samanlainen, molemmat näkevät tietovaraston olevan yrityksen keskitetty tietovarasto, johon ladataan tietoa ETL:n (Extract – Transform – Load) avulla ja joka ensisijaisesti palvelee yrityksen raportointitarpeita. Erot koulukuntien välillä tulevat esiin siinä, kuinka tietovarasto on mallinnettu, kuinka tieto tietovarastoon ladataan ja kuinka se sinne tallennetaan. (Rangarajan 2016.)

Hovi, Koistinen ja Ylinen (2001, 51) tiivistävät kirjassaan tietovaraston määrittelyn seuraavasti:

- sisältää usean operatiivisen järjestelmän tietoja integroituna
- on vain lukukäytössä ja on pohjana käyttäjien päätöksenteolle
- on monien käyttäjäryhmien yhteiskäytössä
- on suunniteltu raportoinnin ja kyselyiden kannalta tehokkaaksi
- ladataan ajoittain operatiivisista järjestelmistä, ei ajantasaista päivitystä
- tiedot siirretään uusina sukupolvina; vanhat tiedot jäävät (historia)

Tietovarasto tarkoittaa yleensä liiketoimintatiedon hallintaan tarkoitettua analyysitietokantaa, johon ladataan säännöllisin väliajoin yrityksen operatiivisista järjestelmistä tietoa. Tietovarastointi tarkoittaa tietojen lataamisen ja jalostamisen ketjua, kuinka tiedon saadaan hallitusti ladattua operatiivisista järjestelmistä tietovarastoon säännöllisin väliajoin. Hovin ym. (2001) mukaan tietovarastointi tarkoittaa:

Joukkoa ajatusmalleja, suunnittelumenetelmiä, organisointia ja työkaluja, joiden avulla erillisistä operatiivisista tietokannoista saadaan tiedot yhdistettyä yhtenäiseksi, integroiduksi tietovarastoksi, jota voidaan käyttää monenlaisilla helppokäyttöisillä kysely- ja raportointivälineillä.” (Hovi ym. 2001, 32)

Yrityksillä on käytössään useita eri tietojärjestelmiä operatiivisessa käytössä ja niihin on tallennettu paljon tietoa. Järjestelmät on optimoitu operatiivista käyttöä varten ja niiden tietorakenteet palvelevat jatkuvaa, nopeaa tiedon tallennusta ja

päivitystä. Operatiivisten järjestelmien tavoitteena on tehostaa organisaatioiden toimintaa ja parantaa prosessien läpivientiä. Operatiiviset järjestelmät eivät ole suunniteltu niihin tallennettavan tiedon raportointiin ja analysointiin, koska tieto on niissä hajallaan, tietorakenteet hankalia ja tietoa ei ole kuvattu riittävästi. (Hovi, Hervonen & Koistinen 2009, XI.)

Operatiiviset tietojärjestelmät ovat hajaantuneet yrityksen sisällä eri organisaatio-osien käyttämiin sovelluksiin ja niihin tallennettu tieto on käytettävissä lähinnä organisaatio-osien sisällä. Tämä aiheuttaa tiedon siiloutumista, jolloin yrityksen kokonaisnäkyminen olemassa olevasta tiedosta on puutteellinen. (Hovi ym. 2009, XI.) Tieto on kuitenkin tärkeä yrityksen pääoma, jonka avulla liiketoiminnan pitäisi saada ajantasaista ja luotettavaa informaatiota päätöksenteon tueksi. Eri järjestelmiin siiloutuneen tiedon hyödyntämistä varten tarvitaan integraatioita ja tietoa yhdistävä kerros, jonka avulla tietoa saadaan yhdistettyä, jalostettua ja rikastettua palvelemaan paremmin ja nopeammin organisaation päätöksentekijöitä.

Yleisin syy tietovaraston rakentamiseen yrityksissä on eri lähdejärjestelmien tietojen kokoaminen yhteen paikkaan raportointia ja analysointia varten. Operatiiviset järjestelmät eivät ole suunniteltu raportoinnin ja tietojen analysoinnin tarpeisiin, eikä niiden sisältämä transaktiodata ole useinkaan synkronisoitua, organisoitua tai integroitua. Tietovarastossa tietoa yhdistellään tarkoituksenmukaisiksi kokonaisuuksiksi ja tietovarastoon ladattavasta tiedosta on suodatettu pois operatiivisten lähdejärjestelmien tietokannoissa oleva epäoleellinen, tarpeeton tieto. Tietovarastot on suunniteltu ja optimoitu raportointia ja analysointia varten, joten tietovarastoon voidaan tehdä raskaitakin kyselyitä ja analyysyjä. Hyvin suunniteltu ja rakennettu tietovarasto tarjoaa käyttäjilleen luotettavaa, johdonmukaista, oikea-aikaista sekä saatavissa olevaa dataa ilman, että operatiivisten järjestelmien suorituskyky häiriintyy.

Yksi tietovarastoiden merkittävä hyöty raportoinnissa ja analysoinnissa on se, että tietovarastoon tallennetaan operatiivisista järjestelmistä historiatietoa. Operatiiviset järjestelmät ovat usein päivittyviä järjestelmiä tietojen suhteen, jolloin niistä ei saada yhteneväistä ja useat vuodet kattavaa tilastotietoa helposti. Tietovarastoissa olevan historiatiedon avulla voidaan tietovarastoista tehdä pitkänkin

aikavälin trendianalyseja. Tietovarastoon ei kuitenkaan ladata kaikkea operatiivisissa järjestelmissä tallennettua tietoa, vaan vain tietoa ladataan jotain tarvetta varten, yleensä päätöksenteon tueksi.

2.2.1 Tietovarastotyypit

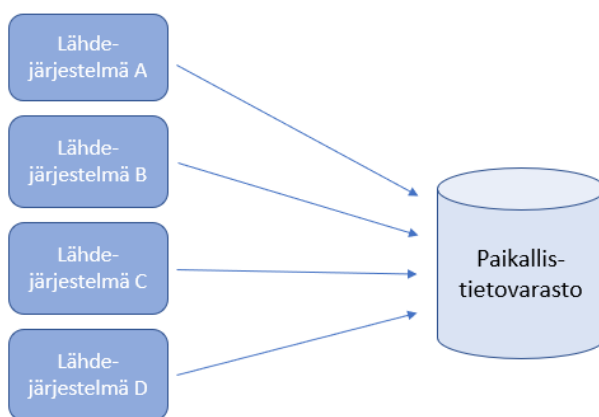
Tietovarasto (Data Warehouse)

Edellisessä aluvuossa käsiteltiin tietovaraston määritelmää. Tietovarasto -termi voidaan käsittää myös yhtenä tietovarastotyyppinä, joskin sitä käytetään yleisterminä koskien kaikenlaisia tietovarastotyyppisiä. Tietovarastoa voidaan kutsua myös keskustietovarastoksi, jonka avulla se voidaan erottaa paikallistietovarastosta. Tietovarastotyyppinä termi tietovarasto käsittää monien käyttäjäryhmien yleiskäyttöisen, yritystason tietokannan, johon kootaan tietoa useista eri lähdejärjestelmistä ja joka sisältää historiatietoa. (Hovi ym. 2001, 50; Hovi ym. 2009, 23)

Tietovarastojen ETL-prosessit vaativat toimiakseen lastausalueen tai työalueen (staging area), johon tiedot tuodaan operatiivisista järjestelmistä joko suoraan tietokantaan tai tiedostoina. Lastausalueelle tuodun tiedon rakenne on yleensä samalainen kuin operatiivisissa järjestelmissä, koska tietojen siirto toteutetaan monesti tietokantakopiona tai materialisoituina näkyminä, jotka kuormittavat vähiten operatiivisen järjestelmän suorituskykyä latausvaiheessa. Latausalueella tiedot odottavat ETL-prosessin käsittelyä ja tietovarastoon vientiä. (Hovi ym. 2009, 25) Kimball ja Caserta (2004, 16-17.) jakavat tietovarastoinnin kahteen alueeseen, tiedonhallintaan ja tiedon saatavuuteen. Hovin ym. (2009) kuvaamat tiedon lastausalue ja ETL-prosessi kuuluvat tiedonhallinnan alueeseen, jossa tietoa hankitaan lähdejärjestelmistä, muokataan ja ladataan tietokantaan. Tiedonhallinnan alueella tietoa ei jaeta käyttäjille kyselyjä varten, vaan tätä tarkoitusta varten on olemassa tiedon esityskerros, jonka kautta tieto on tarjolla käyttäjille raportointia ja analysointia varten optimoidussa dimensionaalisessa tietokannassa. (Kimball & Caserta 2004, 16)

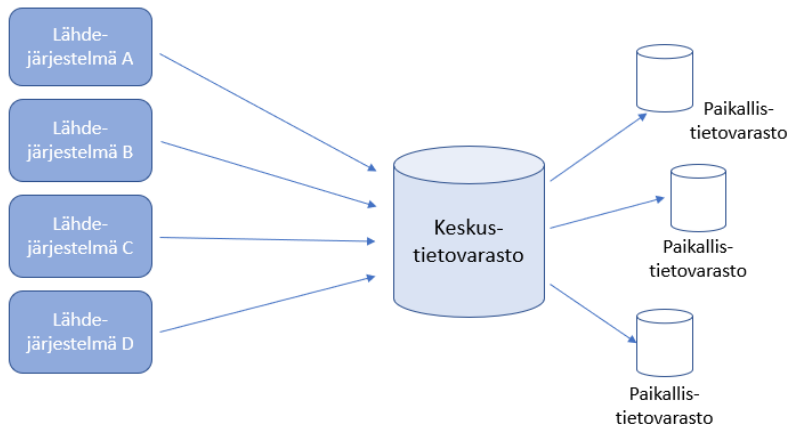
Paikallistietovarasto (alitetovarasto, otoskanta, Data mart)

Paikallistietovarasto eroaa keskustietovarastosta siinä, että paikallistietovarastoon ladataan usein aihe- tai organisaatiokohtaista tietoa ja ne ovat suunniteltu tukemaan nimenomaan helppoja kyselyitä ja raportointia. (Hovi ym. 2001, 52; Hovi ym. 2009, 24) Paikallistietovaraston etuja ovat keskustietovarastoa pienempi koko ja kun ne ovat usein suunniteltu jonkun tietyn käyttäjäryhmän tarpeisiin, ovat ne paremmin hallittavissa ylläpitomielessä. Kuvassa 3 kuvataan Inmonin (2005, 370) määritelmän mukaan itsenäinen paikallistietovarasto on tietorakenne, joka on omistettu yhden käyttäjäjoukon analyttisiä tietotarpeita varten.



KUVA 3. Itsenäinen paikallistietovarasto (Inmon 2005, 370, muokattu)

Kimballin koulukunnan oppien mukaan paikallistietovarasto voi olla yhden lähde-tietojärjestelmän tiedoista koostuva tähtimallin mukaan mallinnettu tietovarasto, jolloin tiedot ladataan suoraan lähdejärjestelmästä ETL-prosessin kautta paikallistietovarastoon. Tällöin ei voida puhua vielä Inmonilaisesta keskustietovarastosta, sillä se sisältää aina tietoa useammasta lähteestä. Inmonin koulukunnassa keskustietovaraston tehtävä on tarjota tietoa johdettuun paikallistietovarastoon (Inmon 2005, 157), jolloin keskustietovarasto on johdetun paikallistietovaraston lähdejärjestelmänä, kuten kuvassa 4 on havainnollistettu.



KUVA 4. Johdettu paikkaitietovarasto (Inmon 2005, 371, muokattu)

Monidimensionaalinen tietovarasto (OLAP, OnLine Analytical Processing)

OLAP eli OnLine Analytical Processing -termillä tarkoitetaan moniulotteisen tietorakenteiden analysointiprosessia sekä -välineitä eli moniulotteisen tiedon tallennukseen ja analysointiin tarkoitettua ohjelmistoa. Usein OLAP -termin yhteydessä esiintyy myös käsite kuutio. Ulottuvuuksia moniulotteisessa tietomallissa on usein enemmän kuin kuutiolla sivuja, mutta ajatuksena kuutio kuvaa hyvin moniulotteista tietoa. (Hovi ym. 2001, 52-54.) Moniulotteisen tietomallin erityispiirre verrattuna esimerkiksi relaatiotietokantaan on mahdollisuus tarkastella ja analysoida tietoja monesta eri näkökulmasta. Monidimensionaalinen tietomalli mahdollistaa tiedon tarkastelun helposti summatasolla, mutta tietoon voidaan myös porautua (drill down) yksityiskohtaisemmalle tasolle tuoden tietoon lisää ulottuvuuksia. Porautumisen vastakohta on tiedon karkeistaminen (roll up), jolloin yksityiskohtaisesta tiedosta pääsee siirtymään ylemmälle tasolle. Moniulotteisen tietomallin etuja ovat porautumisen lisäksi mahdollisuus tarkastella tietoa erilaisien hierarkioiden avulla sekä erilaisten luokittelujen tasolla. (Hovi ym. 2001, 56-61.)

Moniulotteinen OLAP -ratkaisu on parhaimmillaan paikallistietovarastona, johon tieto on ladattu keskustietovarastosta. Tällöin tiedon esikäsittely ja valmistelu on tehty keskitetysti keskustietovarastossa ja moniulotteinen tietomalli tarjoaa tehokkaan analyttisen alustan tiedon monipuoliseen analysointiin. Useimmin OLAP -tuotteista puhutaan kuitenkin liiketoimintatiedon hallinnan yhteydessä, joka kuvastaa OLAP:in luonnetta enemmän analysoinnin ja raportoinnin välineenä kuin varsinaisena tietovarastoratkaisuna. Kimball ja Caserta (2004, 45) pitävät OLAP-

kuutioita tiedon esityskerrokseen kuuluvana osana, johon tieto ladataan dimensionaalista tietovarastosta. Kimball ja Caserta (2004, 45) mukaan OLAP -kuutiot ovat huipputehokkaita dimensionaalisia tietomalleja, jotka ovat toteutettu erityisohjelmistoilla.

Tilannekanta (Operational Data Store, ODS)

Usein tietovarastoinnin yhteydessä puhutaan myös tilannekannoista, ODS-kannoista. Niiden kuuluminen tietovarasto -kattotermin alle on kyseenalainen, sillä tilannekannat eivät ole varsinaisia tietovarastoja. Tilannekanta voidaan ymmärtää tiedon välitystietokantana eri operatiivisten tietojärjestelmien välillä, tai vaihtoehtoisesti mahdollistamaan helpompaa päivätasoisia raportointia operatiivisista järjestelmistä. Tilannekanta voidaan pitää myös varsinaisen tietovaraston tietolähteenä tai työalueena, jossa tietoja on jo yhdistelty eri operatiivisista järjestelmistä. (Hovi ym. 2001, 62-65.; Hovi ym. 2009, 25.; Inmon 2005, 430-442.)

Tilannekannassa ei yleensä ole tallennettu historiatietoa, tai sitä on tarjolla vain lyhyeltä aikaväliltä toisin kuin tietovarastossa. (Inmon 2005, 441) Tilannekanta voidaan hyödyntää tietojen yhdenmukaistamiseen, kuten esimerkiksi asiakastiedon yhtenäistämiseen eri lähetietojärjestelmien välillä. Tällöin usein puhutaan Master Datan hallinnasta (Master Data Management). (Hovi ym. 2009, 25.) Tilannekannassa yhtenäistettyä masterdataa voidaan ladata tietovarastoon, jolloin tietovarastossa pystytään yhdistelemään eri tietolähteiden erilaisilla avaintiedoilla tuottamien tietoja yhdeksi kokonaisuudeksi.

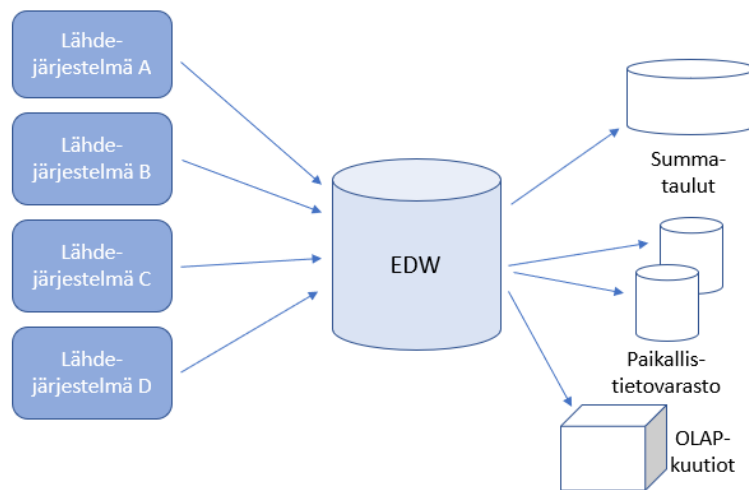
2.2.2 Tietovarastojen arkkitehtuurit

Tietovarastojen arkkitehtuureina esitetään yleisesti kolme vaihtoehtoa niiden yleisyyden takia. Viitearkkitehtuurit tietovarastojen rakentamiseen ovat yhdenmukaistetut paikallistietovarastot, yritystasoinen keskitetty tietovarasto ja itsenäiset paikallistietovarastot. (Hovi ym. 2001, 66-70.; 26-29.) Kaikki näistä viitearkkitehtuureista puolustavat paikkansa tietentyypisissä ympäristöissä, joten mitään näistä ei voida nostaa muiden yläpuolelle. Tietovarastojen viitearkkitehtuurit tarjoavat yleiskäyttöisiä suuntaviivoja tietovarastojen toteuttamiseen, mutta ne eivät ota kantaa siihen, mitä yksittäisiä teknologioita pitäisi käyttää, miten esimerkiksi

metadatan hallinta tulisi toteuttaa, kuinka ETL-järjestelmä pitäisi tehdä tai että kuinka eri ikäisten tietojen hallinta tulisi toteuttaa. Perinteisen tietovaraston arkkitehtuureissa on tyypillisesti ollut varsinaisen tietojen tallennukseen käytetyn alueen lisäksi työalue ja tiedon hyödyntämisrajapinta.

Yritystasoinen keskitetty tietovarasto (EDW, Enterprise Data Warehouse)

Yritystasoisesta keskitettyä tietovarastoa kutsutaan monesti inmonilaiseksi tietovarastoksi, jossa yrityksen tuottamaa olennaista tietoa kerätään yhteen tai muutamaaan isoon keskitettyyn tietovarastoon yhdenmukaisella tavalla tallennetuksi kokonaisuudeksi. Keskitettyyn tietovarastoon ladataan tietoa yritystason näkökulmasta, jolloin eri järjestelmiin kertynyt tieto on saatavissa helposti yhden keskitetyn tietovaraston kautta yhdenmukaistettuna. (Hovi ym. 2009, 27.) Kuvassa 5 havainnollistetaan keskitetyn tietovaraston perusajatus, toimia kokoavana paikkana yhtenäiselle tiedolle, josta voidaan tehdä pienempiä paikallistietovarastoja, multidimensionaalisia OLAP-kuutioita tai muodostaa summatauluja.



KUVA 5. Yritystasoinen keskitetty tietovarasto (Hovi ym. 2009, 27, muokattu)

Yritystasoisesta keskitetyn tietovaraston etuina ovat tietojen integrointi operatiivisista järjestelmistä suoraan yhteen keskustietovarastoon, jolloin se sopii pohjaksi suuriin yrityksiin, jossa tarvitaan tietoa useasta organisaatiosta. EDW on kuitenkin työläs ja aikaavievä rakentaa ja ylläpitää, jolloin sitä ei monesti pienissä yrityksissä lähdetä toteuttamaan. EDW:n rakentaminen ja ylläpito vaativat yritykseltä merkittävää rahallista ja resurssien panostusta, jolloin sen rakentamisesta saavutettavat hyödyt realisoituvat tietoon perustuvien päätösten muodostamasta

arvonluonnista. Yritystasoinen keskitetty tietovarasto on laajalti käytössä sen pitkällä aikavälillä tarjoamien hyötyjen takia ja sen käyttö on vakiintunut toiseksi yleisesti käytössä olevaksi tietovarastoarkkitehtuuriksi. Yritystason tietovarasto perustuu yleisesti normalisoituun relaatiotietokantatoteutukseen, jonka takia se ei ole useinkaan optimoitu raportointia ja analysointia varten. Viime aikoina myös data vault -menetelmää on käytetty yritystason keskitettyjen tietovarastojen toteutusmenetelmänä. Data vault -tietovarasto vaatii aina dimensionaalisen tai vastaavan julkaisukerroksen, jotta tietoja pystyttäisiin helposti raportoimaan ja analysoimaan.

Yhdenmukaistetut paikallistietovarastot (Data Warehouse Bus Architecture)

Toinen yleisesti yrityksissä käytössä oleva tietovarastoinnin arkkitehtuuri on kimballilainen yhdenmukaistettu paikallistietovarasto. Kimballin suositusten mukaan paikallistietovarastot tulisi suunnitella tähtimallinmenetelmällä siten, että osa dimensioista ovat yhteneväisiä. (Kimball & Caserta 2004, 26.) Toisin kuin keskitetyssä tietovarastossa, yhdenmukaistetussa paikallistietovarastoissa tieto mallinetaan ja tallennetaan useisiin paikallistietovarastoihin käyttäen päällekkäisiltä osiltaan yhteisiä dimensioita. (Hovi ym. 2001, 68.; Hovi ym. 2009, 28.) Yhdenmukaistettuja paikallistietovarastoja on helpompi ja nopeampi toteuttaa, koska niiden sisältämä tietosisältö rajoittuu yhteen aihealueeseen eikä koko yrityksen kaikkia tietoja tarvitse tuoda tietovarastoon kerralla. Tähtimallinmenetelmä myös selkeyttää tietojen jäsenystä tietovarastossa ja tähtimallin paikallistietovarastoa voidaan hyvin pitää raportoinnissa ja analysoinnissa tietolähteenä toisin kuin EDW tietovarastoa, joka usein vaatii erillisen tähtimallin tietomallikerroksen raportoinnin ja analysoinnin tarpeisiin.

Itsenäiset paikallistietovarastot

Kolmas yrityksissä käytetty tietovarastoarkkitehtuuri on itsenäinen paikallistietovarasto. Tätä tietovarastoarkkitehtuuria käytetään erillisissä paikallistietovarastoratkaisuissa, joissa tietoa ladataan sinne joko yhdestä organisaatiosta tai operatiivisesta järjestelmästä. Itsenäisessä paikallistietovarastossa tiedot ovat irrallaan toisistaan, vaikka itsenäisiä paikallistietovarastoja voi yrityksessä olla useitakin. Tietojen yhdistäminen raportointia ja analysointia varten erillisistä paikallistietovarastoista on usein hankalaa, sillä esimerkiksi koodistot tai asiakastunnukset ovat

usein toisista poikkeavia eri lähdejärjestelmissä. (Hovi ym. 2001, 68-69.; Hovi ym. 2009, 26-27.) Kimballin ja Rossin (2013, 27) mukaan itsenäiset paikallistietovarastot ovat etenkin isoissa yrityksissä varsin yleisiä, vaikka mikään johtava tietovarastoratkaisuja tarjoava taho ei tätä ratkaisua suosittelen käytettäväksi.

Itsenäisten paikallistietovarastojen rakentamiseen on usein syynä se, että halutaan rakentaa nopeasti ja pienelle käyttäjäkunnalle käyttöön raportointiympäristö, jonka avulla he saavat omaan käyttöönsä oman alueensa tietoja. Tällaisessa käytössä itsenäinen paikallistietovarasto puolustaa paikkansa, samoin yrityksissä, jotka ovat pieniä, eivätkä tarvitse massiivista tietovarastohanketta saadakseen riittävät tiedot päätöksentekonsa tueksi. Erona kimballilaiseen yhdenmuikaistettuun paikallistietovarastoon on se, että itsenäisissä paikallistietovarastoissa ei huomioida yritystason tietotarpeita, vaan ratkaisut rakennetaan erillisinä toteutuksina.

Usein kuitenkin yrityksissä ajaudutaan rakentamaan itsenäisiä paikallistietovarastoja vain sen takia, ettei yrityksessä ole ollut kunnollista tietovarastostrategiaa, jossa suunnitellaan yrityksen strategian mukaisia tietovarastojärjestelmiä suunnitelmallisesti kokonaisarkkitehtuuria noudattaen. Kunnollisen tietovarastostrategian puuttuessa yrityksellä ei ole useinkaan suunnitelmaa, kuinka pitkällä aikavälillä pystytään vastaamaan yrityksen tietojohdamisen asettamiin tietotarpeisiin. Kimballin ja Rossin (2013, 27) esittämän näkemyksen mukaan isojen yritysten itsenäiset paikallistietovarastot kuvastavat sitä, kuinka IT -projekteja organisaatioissa yleisesti rahoitetaan. Liiketoimintarahoitteisessa mallissa tietovarastokehitys keskittyy saamaan hyötyä liiketoiminnan omiin tarpeisiin ja tietovarastoon tehdyt arkkitehtuuriset ratkaisut palvelevat yhden liiketoiminta-alueen tarpeita, mutta yrityksen kokonaisuuden kannalta yhtenäinen ratkaisu jää ilman rahoitusta.

Yrityksen tietovarastojen arkkitehtuuriin liittyvät päätökset eivät ole monesti kovin yksikertaisia ja puhtaasti yhden arkkitehtuurin mukaisia tietovarastoja yritykseen harvoin pystytään toteuttamaan. Viime aikoina eri tietovarastoarkkitehtuurien väliset eroavaisuudet ovatkin kaventuneet ja eri arkkitehtuurien väliset rajat ovatkin hämärtyneet. (Kimball & Ross 2013, 26-30.) Yrityksissä onkin monesti päädytty käyttämään eri tietovarastoarkkitehtuureista soveltuvia osia tietovarastossa ja siten hybridimallit ovat yleisiä.

Tietovarastolla on merkittävä rooli tietojen yhtenäistämässä ja yhden totuuden tarjoajana, oli tietovaraston arkkitehtuuri mikä tahansa näistä yllä esitellyistä tai niiden yhdistelmä, hybriditietovarasto. Johdettujen tietojen käsittelysäännöt ja kaavat tulisi tallettaa tietovaraston metatietoihin ja tietovarastoon tulisi laskea johdetut tulokset valmiiksi oikeaoppisen tietovarastoarkkitehtuurin mukaan. (Hovi ym. 2001, 117) Tällöin tietovarastosta saa samoja tuloksia huolimatta siitä, mitä raportointi- tai analysointivälinettä käyttäjät käyttävät. Tietojen laskentasääntöjä ja kaavoja ei tulisi sisällyttää raportointivälineisiin eikä OLAP-kuutioihin, sillä se aiheuttaa erilaisia tuloksia tiedosta eri välineillä raportoituina.

2.2.3 Tietovarastojen suunnittelu

Tietovarastojen rakentaminen alkaa aina jonkin tarpeen seurauksena, sillä niillä pyritään tuottamaan yhdenmukaistettua, luotettavaa ja helposti saatavilla olevaa informaatiota päätöksenteon tueksi. Tietovarastot pitäisi olla suunniteltu täsmälleen tarkoin määriteltyä liiketoimintatarvetta varten. Liiketoimintatarpeet puolestaan määräytyvät monesti liiketoiminnan johdon tarvitseman informaation mukaan. (Surma, Górniakowska & Gee 2011, 25.) Ilman huolella tehtyä tietovaraston suunnittelua tietovaraston hyödynnettävyys kärsii. Suunnittelu on oleellisen tärkeää myös siksi, että tätä vaihetta eivät mitkään työkalut tai tuotteet voi korvata.

Tietovaraston arkkitehtuurin suunnittelussa on tyypillisesti kaksi lähestymistapaa. Lähestymistavat ovat alhaalta ylös, jota käytetään yleensä paikallistietovarastojen arkkitehtuurissa ja ylhäältä alas, jota käytetään yleensä keskitetyn tietovaraston arkkitehtuurissa. Ylhäältä alas -lähestymistavassa tieto virtaa lähdetietojärjestelmistä keskitettyyn tietovarastoon ja sieltä eteenpäin paikallistietovarastoihin raportointia ja analysointia varten. Alhaalta ylös -lähestymistavassa arkkitehtuuri rakentuu paikallistietovarasto kerrallaan liiketoiminnan tarpeisiin ja tietoa ladataan paikallistietovarastosta keskitettyyn tietovarastoon. (Maheshwari 2015, 39.) Taulukossa 3 on vertailtu toimintokohtaisen paikallistietovaraston ja keskitetyn tietovaraston eroja.

TAULUKKO 3. Toimintokohtaisen paikallistietovaraston ja keskitetyn tietovaraston vertailutaulukko (Maheshwari 2015, 39, muokattu)

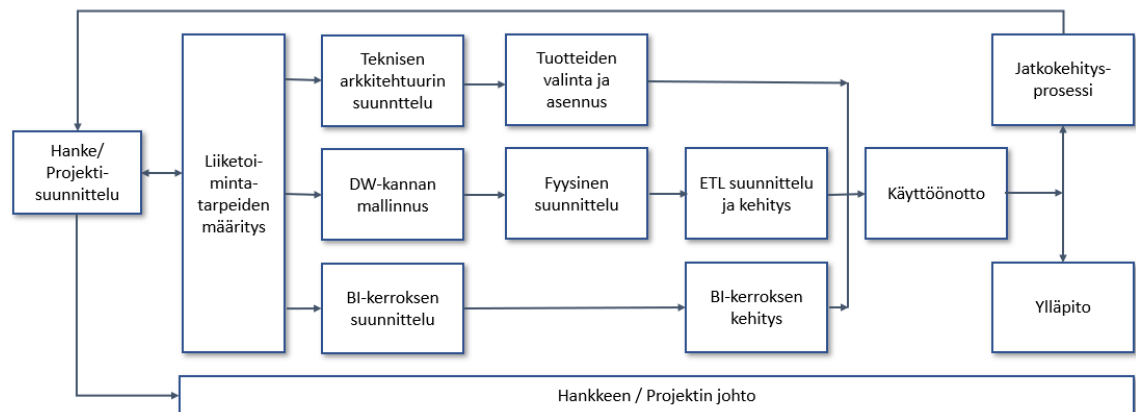
	Toimintokohtainen paikallistietovarasto	Keskitetty tietovarasto
Laajuus (Scope)	Yksi toiminta-alue tai aihe	Kaikki yrityksen tietotarpeet
Arvo (Value)	Toimintokohtainen raportointi ja ymmärrys	Syvempi, yhtenäinen ymmärrys useasta eri toiminnon alueesta
Kohdeorganisaatio (Target organization)	Hajautettu johto	Keskitetty johto
Aika (Time)	Lyhyestä keskipitkään	Pitkä
Kustannukset (Cost)	Matalat	Korkeat
Koko (Size)	Pieni ja keskisuuri	Keskisuuri ja iso
Kehitys (Approach)	Alhaalta ylös	Ylhäältä alas
Monimutkaisuus (Complexity)	Pieni (vähemmän datan muunnoksia)	Iso (datan standardointi)
Teknologia (Technology)	Pienen kokoluokan serverit ja tietokannat	Teollisuusmittaluokan serverit ja tietokannat

Karkealla tasolla tietovaraston suunnittelussa on kolme tärkeää lähtökohtaa: käyttäjien tarpeet, tietovaraston laajuus ja saatavissa olevat tiedot. (Hovi ym. 2009, 31.) Liiketoiminta sanelee tarpeita, mitä informaatiota sen pitäisi saada pystyäkseen tekemään tietoon pohjautuvia päätöksiä. Liiketoimintatarpeiden lisäksi selvitetään kokonaisuuden kannalta ne kokonaisuudet, mitä tietoa tietovarastoon lähdetään lataamaan. Samalla selvitetään, mitä tietoa on saatavissa yrityksen sisällä ja mitä puuttuu. Usein törmätään tilanteeseen, että tarvittavaa tietoa ei ole käytettävissä operatiivisissa lähdejärjestelmissä tai tieto on puutteellista. Silloin tietoa on hankittava ulkoisista tietolähteistä, kuten esimerkiksi väestötietojärjestelmästä, paikkatietojärjestelmästä, säätilastoista tai muista avoimen datan tietolähteistä.

Hovi ym. (2009) suosittelee tietovarastojen suunnitteluun mallia, jossa suunnittelussa ensin laaditaan ylätason kokonaisuusmalli, joka kattaa tietovarastoinnin koko alueen. Ylätason mallin suunnittelun jälkeen suunnitellaan ensimmäinen toteutettava osa-alue perustuen liiketoiminnan tarpeisiin ja käytettävissä oleviin tietoihin. Ensimmäisen osa-alueen suunnittelun valmistuttua aletaan toteuttaa ensimmäisen vaiheen tietovarastoa. Tämän jälkeen siirrytään seuraavan osa-alueen suunnitteluun ja toteutukseen. Mallin ansiosta käyttäjille saadaan tuloksia nopeasti, mutta osatoteutukset noudattavat yhdenmukaisen tietovaraston kokonaiskuvaa ja tavoitemallia.

Ralph Kimball on kehittänyt oman tietovaraston elinkaaren viitekehysten tietovarastojen suunnitteluun, toteutukseen ja hallintaan. Kimballin (2008) oppien mukaan menestyvän tietovarastoinnin perustana ovat kolme olennaista ajatusta: keskity liiketoimintaan, dimensionaalisesti mallinnettu data, joka on liiketoiminnan käytettävissä ad-hoc -kyselyillä tai raporteilla ja kolmanneksi tietovaraston inkrementaalinen kehitys. Kimball painottaa tietovaraston suunnittelussa liiketoimintatarpeiden tärkeyttä, sillä tietovarastot rakennetaan nimenomaan liiketoiminnan tarpeisiin.

Tietovaraston elinkaaren hallintamallissa tämä näkyy järjestyksessä, miten tietovaraston suunnittelu etenee. Kuvassa 6 olevan kaavion mukaan hanke/projektisuunnittelun jälkeen alkaa liiketoimintatarpeiden kartoitus, jonka jälkeen vasta aloitetaan varsinaiset tietovaraston ja liiketoimintatietokerroksen suunnittelu sekä teknisen arkkitehtuurin suunnittelu. Kimballin tietovaraston elinkaaren viitekehys käsittää tietovarastointi -termin sisään laajemman kokonaisuuden kuin aliluvussa 2.2.2 Tietovarastoarkkitehtuurit käsitellyt tietovarastokäsitteet. Tietovaraston elinkaaren viitekehyksessä on kuvattu sekä tietovaraston (DW, Data Warehouse) että liiketoimintatiedon hallinnan (BI, Business Intelligence) suunnittelun, kehityksen ja hallinnan toiminnot. Liiketoimintatiedon hallintaa käsitellään tarkemmin luvussa 2.3 Liiketoimintatiedon hallinta.



KUVA 6. Kimballin tietovaraston elinkaaren viitekehys (Kimball & Ross 2013, 404, muokattu)

Tietovarastojen suunnittelu ei ole pelkästään IT -osaston työtä, vaan suunnitteluun tarvitaan sitoutunut johto liiketoiminnasta. Liiketoiminnalla tulee olla kokonaisnäkemys kohdealueen strategiasta ja tietovarastojen suunnittelu tulee kytkeä

yrittäjästrategiaan tavoitteisiin. Myös IT -osasto tulisi olla strategiatyössä mukana, jolloin pystyttäisiin muodostamaan koko yritystä koskeva tietovarastointi-strategia tukemaan liiketoiminnan tarpeita pidemmällä aikavälillä. Jos selkeää tietovarastostrategiaa ja tiekarttaa ei yrityksessä ole tehty, on vaarana tietovarasto-suunnittelun pistemäisyys ja erillisyys. Hovi ym. (2009, 33) että Kimball (2008) suosittelevat suunnittelemaan tietovarastoja laajasti, mutta toteuttamaan ne pieninä kokonaisuuksina.

2.2.4 Tietovarastojen työkalut

Tietovarastoinnissa käytetään erilaisia työkaluja pääsääntöisesti neljään eri tarkoitukseen. Nämä ovat ETL-välineet, tietokantaohjelmistot, mallinnusvälineet ja raportointivälineet. Kaikkiin osa-alueisiin on olemassa tarkoitukseen suunniteltuja työkaluja, joita löytyy markkinoilta useita eri vaihtoehtoja. Työkalujen valinnassa pätee sama periaate kuin koko tietovaraston suunnittelussa, työkalut tulee valita huolella pohtien tulevaa käyttöympäristöä sekä organisaatiossa käytettävissä olevia resursseja. Raportointivälineitä ja -menetelmiä käsitellään tarkemmin liiketoimintatiedon hallintaa koskevassa luvussa 2.3.2 Keskeiset menetelmät.

Mallinnustyökalut

Mallinnustyökalulla luodaan looginen tietomalli toteutettavasta tietovarastosta, jossa kuvataan tiedon kulku tietovirtana lähdejärjestelmistä tietovarastoon. Loogisessa tietomallissa hallitaan myös tietovarastoon liittyvää metadatan ja kuvataan tietovirtaan tiedon muunnokseen liittyvät kaavat ja laskennat. Mallinnustyökalua voidaan pitää myös tiedon metadatan koonti- ja ylläpitotyökaluna, josta pystytään jäljittämään tietoon liittyvät käsittelysäännöt ja muunnokset. Mallinnustyökalussa voidaan ylläpitää myös tiedon liiketoimintakuvauksia, joka on tiedon ymmärrettävyyden kannalta olennaista olla kuvattuna. (Kimball & Caserta 2004, 50.)

Mallinnustyökalussa olevasta loogisesta tietomallista luodaan fyysinen tietomalli, jossa kuvataan tiedon rakenteeseen liittyvää informaatiota. Mallinnustyökaluilla voidaan myös generoida tietokannan luontilauseet valitulle tietokannan hallinta-

järjestelmälle. (Hovi ym. 2009, 63) Paljon käytettyjä mallinnustyökaluja ovat esimerkiksi Embarcaderon kehittämä, sittemmin yrityskaupan jälkeen Ideralle siirtynyt ER/Studio Data Architect, Oraclen SQL Developer Data Modeler ja erwin Data Modeler.

Tietokantaohjelmistot

Tietovarastot tarvitsevat toimiakseen tietokannan, kuten luvussa 2.2 Tietovarastointi kuvattiin. Perinteisesti tietovarastojen tiedot tallennetaan SQL-relaatiotietokantoihin, joita operatiivisissakin järjestelmissä käytetään. Nämä tietokannat on suunniteltu alun perin operatiivisissa järjestelmissä tapahtuvaan tietojen käyttöön, kuten tietojen ajantasainen tapahtumankäsittely, jossa tietojen lukitukset, toipumisominaisuudet ja online -varmistukset ovat keskeisessä asemassa. Tietovarastojen tietokannan käyttötarve on erityyppistä, sillä tietoa ei pääsääntöisesti päivitetä, vaan tietoja kirjoitetaan tietokantaan massalatauksina ja tietoja luetaan erilaisilla suorituskykyä vaativilla kyselyillä. Tietokantoja on kehitetty entistä enemmän tietovarastointiin soveltuviksi ja tietovarastoja varten kehitettyjä erikoistuotteita onkin nykyään jo tarjolla. (Hovi ym. 2009, 62.) Perinteisempiä relaatiotietokantoja ovat esimerkiksi Oracle, Microsoftin SQL Server ja IBM:n Db2, avoimen lähdekoodin relaatiotietokantoja ovat esimerkiksi MySQL ja PostgreSQL. Tietovarastoihin erikoistuneita tietokantoja tarjoavat esimerkiksi Teradata ja SAP IQ.

Viime aikoina pilvipalveluina toimivat tietokannat (DBaaS, Database as a Service) ovat saaneet jalansijaa tietovarastojen tietokantoina teknologian kehityttyä kovaa vauhtia. Monia tietokantoja saa nykyään myös pilvipalveluina, kuten esimerkiksi MySQL -tietokantaa saa usealta pilvitoimittajalta. Microsoft tarjoaa SQL Serveriä myös Amazonin pilvipalveluna. Amazonin RDS -ratkaisussa voi valita Oracle, MySQL SQL Server tai PostgreSQL -relaatiokannan valmiiksi asennettuna. Tietovarastoratkaisuille tarkoitettuja pilvipohjaisia erikoistietokantoja ovat esimerkiksi Snowflake, Amazon Redshift ja Microsoftin Azure Data Warehouse. Pilvipalveluiden käyttöönottoa harkitessa pitää huolehtia yrityksen tietoturvapoliittikan mukaisesta toiminnasta, sillä julkiset pilvipalvelimet sijaitsevat lähes poikkeuksetta Suomen rajojen ulkopuolella. Tämä on otettava huomioon yrityksen tietojen sijoittamista koskevissa päätöksissä.

ETL -välineet

Yksi suurimmista ja tärkeimmistä tietovarastojen suunnittelun päätöksistä koskee tietovarastojen ETL-prosessia, eli kuka tämän toteuttaa ja ylläpitää. Organisaatiossa käytettävissä olevan henkilöstön osaaminen ja kyvykkyys vaikuttaa suuresti siihen, kannattaako ETL-prosessi tehdä itse ohjelmoimalla vai käyttää valmiita ETL-ohjelmistoja latausten toteutukseen. Jos henkilöstöllä ei ole riittävää ohjelmointikokemusta tai kyvykkyyttä, niin silloin on hyvä valita käytettäväksi jonkun tunnetun ETL-välinetoimittajan valmis ohjelmisto tähän tarkoitukseen.

Päätös, valitseeko ETL-prosessin toteutukseen valmiin ETL-ohjelmiston vai päättääkö ohjelmoida sen itse, on tietovarastoarkkitehtuurissa perustavanlaatuinen päätös ja se vaikuttaa olennaisesti kaikkiin arkkitehtuurin osa-alueisiin. (Kimball & Caserta 2004, 9.) Se, kumman tavan valitsee, riippuu täysin yrityksestä ja organisaatiosta, kummallakin tavalla on omat hyvät ja huonot puolensa. Päätöstä pitääkin siksi harkita tarkoin ja hankkia päätöksenteon tueksi tietoa kummankin tavan hyvistä ja huonoista puolista. Tähän päätökseen tarvitaan myös yrityksen sisäisen toimintatavan, osaamisen ja kyvykkyyksien tuntemista, jolloin tietojohdamisen näkökulmista johtamisen käytäntöjen menetelmät ovat tässä kohtaa avuksi. Päätökseen vaikuttaa merkittävästi myös se, kuinka tietovaraston ylläpito on suunniteltu toteutettavan projektien jälkeen. Tietovaraston toteutustyötä voidaan ostaa projektiluontoisesti ulkopuolisilta toimittajilta, mutta tietovaraston ylläpito on jatkuvaa toimintaa, johon yrityksessä pitää olla resursoituna riittävästi osaavaa henkilöstöä.

Useimmiten yritykset päätyvät toteuttamaan ETL-prosesseja jonkin tunnetun toimittajan valmiilla ETL-ohjelmistoilla, ellei sitten yrityksen ydinliiketoimintaa ole ohjelmistokehitys, jolloin yrityksen henkilöstön osaaminen ei ole ETL-ohjelmoinnin esteenä. Valmiita ETL-ohjelmistoja markkinoidaan myös integraatiotyökaluina kuvaten työkalun luonnetta tietojen integroimisen välineenä. Markkinoiden suurimpia ja laajimmin yrityksissä käytettyjä ETL-ohjelmistoja ovat Informatican PowerCenter, IBM:n InfoSphere DataStage ja Microsoftin SQL Server Integration Services.

2.3 Liiketoimintatiedon hallinta (Business Intelligence)

Jokaisen yrityksen ja organisaation pitää seurata ympäröivää liiketoimintakenttää sekä omaa suorituskykyään ja mukauttaa tulevaisuuden suunnitelmiaan nopeasti pysyäkseen kilpailukykyisenä ja elinvoimaisena. Oman suorituskyvyn ja ympäristön seurannan onnistumiseksi tarvitaan järjestelmä, jolla liiketoimintatietoa kerätään ja analysoidaan. Tietovarastointia ei oikein voida käsitellä ilman, että puhutaan myös liiketoimintatiedon hallinnasta (BI, Business Intelligence), sillä nämä liittyvät toisiinsa kiinteästi.

Liiketoimintatiedon hallinta -termi yleistyi 1990 -luvulla viitaten tietovarastoissa olevan datan raportointiin ja analysointiin. Termin määritelmä vaihtelee tulkitsijasta ja joillekin se tarkoittaa sateenvarjokäsitettä tietovaraston ja liiketoimintatiedon kokonaisuudelle, kun taas osa tulkitsee tietovarastointi -termin olevan sateenvarjokäsite sekä tietovarastoinnille että liiketoimintatiedon hallinnalle. (Kimball 2008.) Yhden määritelmän mukaan liiketoimintatiedon hallinnalla tarkoitetaan laajaa IT-sovellusten kokoelmaa, jolla kerätään, analysoidaan ja raportoidaan organisaation suorituskykyä ja ympäristöä koskevaa tietoa. (Maheshwari 2015, 3) Yleisemmin liiketoimintatiedon hallinta käsitetään tiedon analysointiin ja raportointiin liittyväksi päätöksenteon tukijärjestelmäksi tai raportointiratkaisuksi, jolla pyritään tuottamaan datasta informaatiota päätöksenteon tueksi.

Pelkillä hyvilläkään raportointiteknologioilla ei pystytä keräämään, jalostamaan ja raportoimaan tietoa, jos tietoa ei välillä jossain muodossa tallenneta. Polku datasta päätöksenteossa tarvittavaksi tietämykseksi on yleensä pitkä ja käsittää monenlaista tiedon käsittelyä ja muokkausta. Pelkillä raportointi- ja visualisointityökaluilla ei monestikaan pystytä ratkaisemaan kaikkea sitä jalostusprosessia, mitä raakadata tarvitsee muovautuakseen tuottamaan lisäarvoa liiketoiminnalle. Toisaalta pelkän tietovaraston rakentamisella ei välttämättä pystytä täyttämään liiketoiminnan vaatimuksia tiedon analysoinnista ja helposta saatavuudesta. Tämän takia tarvitaan sekä tietovarastointia että liiketoimintatiedon hallintaa, molemmilla on tärkeä osa toimivalle analyttiselle prosessille, joka tuottaa ehdotuksia, näkemyksiä ja suosituksia päätöksentekijöille.

Maheshwari (2015, 21) nostaa tiedon välittämisen sitä tarvitseville käyttäjille yhdeksi liiketoimintatiedon hallintaan kuuluvaksi alueeksi. Hän myös liittää tiedon louhinnan (Data Mining) termin liiketoimintatiedon hallinnan yhdeksi komponentiksi tiedon varastoinnin, analysoinnin ja raportoinnin rinnalle.

Liiketoimintatiedon hallintaa voidaan tulkita myös sisällön näkökulmasta, mistä lähteistä tietoa yrityksessä kerätään ja analysoidaan. Sisäisellä eli kvantitatiivisella näkemyksellä tarkoitetaan yrityksen sisäisesti keräämään liiketoimintatiedon analyyttistä hyödyntämistä ja ulkoisella eli kvalitatiivisella näkemyksellä tarkoitetaan yrityksen kilpailijoista ja markkinoista saatavan tiedon hyödyntämistä. Sisäinen tieto on luonteeltaan yleensä strukturoitua eli jäsenneiltyä ja rakenteista tietoa, kun taas ulkoinen tieto on luonteeltaan pääsääntöisesti strukturoimatonta eli esimerkiksi rakenteettomia dokumentteja, uutisvirtaa ja sosiaalisen median tietoja. Kuvassa 7 havainnollistetaan liiketoimintatiedon tulkinta tiedon syntyipaikan ja tiedon muodon näkökulmista. (Hovi ym. 2009, 78-79.)

Tiedon muoto		Tiedon syntypaikka	
	Ulkona	Sisällä	
Strukturoitu	esim. pörssikurssit	esim. myynti-, tuotanto-, talous- ja henkilöstötietoja	
Strukturoimaton	esim. markkina-, kilpailijatiedot/uutiset	esim. dokumentinhallinta, sähköinen laskutus	

KUVA 7. Liiketoimintatiedon hallinta -termin kaksi eri tulkintaa (Hovi ym. 2009, 79, muokattu)

Tässä opinnäytetyössä liiketoimintatiedon hallinta -käsite keskittyy kuvaamaan termin sisäistä tulkintaa, sillä sote-tieto on luonteeltaan organisaatiossa sisällä kirjattua ja tallennettua, pääsääntöisesti rakenteista ja jäsenneiltyä tietoa. Organisaation ulkopuolella syntynyttä tietoa, kuten esimerkiksi Sosiaali- ja terveydenhuollon ylläpitämän koodistopalvelimen julkaisemat tautiluokitus ICD-10 koodistot ja laboratoriotutkimusnimikkeistö, voidaan hyödyntää sisäisen tiedon täydentämiseksi, mutta tässä yhteydessä tieto tulkitaan sisäiseksi tiedoksi.

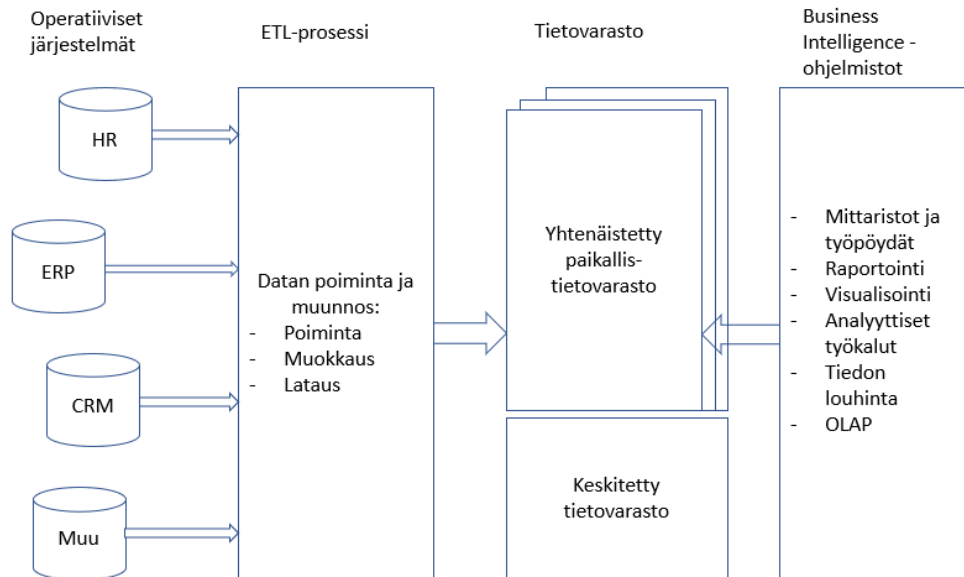
Liiketoimintatiedon hallinnalla voisi ajatella olevan kaksijakoinen tarkoitus: teknisemmästä näkökulmasta sitä voidaan pitää tiedon hallinnan välineenä ja liiketoiminnallisesta näkökulmasta sitä voidaan pitää tiedolla johtamisen välineenä. Tämän kaksoisroolin takia onkin olennaista pitää liiketoiminnan ja tietohallinnon keskusteluyhteys saumattomana ja molemminpuolisena. Muutoin voidaan ajautua tilanteeseen, että tiedon hallintaa kehitetään teknologiapainotteisesti, eikä suunnitella niitä liiketoiminnan tarpeita varten. Liiketoimintatarpeet pitää jaotella strategisen johdon, taktisen johdon ja operatiivisen tason tarpeisiin, sillä jokaisella tasolla on erilaiset tarpeet ja vaatimukset tiedon analysoinnille ja raportoinnille. Monesti jokaisella tasolla on erilaiset teknologiset tarpeet tiedon hyödyntämiselle, jolloin liiketoimintatiedon hallinnassa pitää olla teknologiset valmiudet kaikkien tasojen tiedon hyödyntämistä varten.

2.3.1 Perusarkkitehtuuri

Liiketoimintatiedon hallinnan perusarkkitehtuuri koostuu koko ketjusta tietojen lähteeltä aina sen hyödyntämiseen käyttäjille saakka. Perusarkkitehtuurissa on tyypillisesti neljä eri teknologista komponenttia, joilla jokaisella on oma tehtävänsä tiedon jalostamisen prosessissa. Komponentit koostuvat operatiivisista eli lähdejärjestelmistä, ETL-prosessista eli tiedon poiminta- ja muokkausprosessista, tietovarastosta sekä tiedon esitys ja analysointikerroksesta. Operatiiviset järjestelmät toimivat tiedon raaka-aineena, josta tietoa haetaan jalostettavaksi. ETL-prosessin avulla tietoa poimitaan lähdejärjestelmistä, tietoa muokataan ja yhtenäistetään sekä lopuksi prosessi lataa tiedot tietovarastoon. ETL-prosessi toteutetaan tiedon integrointiin tarkoitetuilla ohjelmistoilla ja prosessointi tapahtuu yleensä omassa palvelinympäristössä erillään tiedon tallennuksesta.

Tietovarastossa informaatio on tallennettuna jäsennellysti ja yhtenäistetysti tiedon raportointia ja analysointia varten. Tietovarastossa tieto on tallennettuna yksityiskohtaisella rivitasolla, mutta tietoa tallennetaan myös summatauluihin, jolloin summataulujen laskenta toteutetaan ETL-prosessissa. Business Intelligence ohjelmistot ovat liiketoimintatiedon hallinnan arkkitehtuurin viimeinen komponentti, joka on olennaisin ja tärkein käyttäjälle näkyvä komponentti. (Hovi ym. 2009, 86.) Kuva 8 havainnollistaa liiketoimintatiedon hallinnan eri komponenttien

väliset suhteet ja tiedon jalostuksen polun operatiivisista järjestelmistä käyttäjille näkyvään kerrokseen.



KUVA 8. Liiketoimintatiedon hallinnan tavanomainen perusrakitehtuuri (Kimball & Ross 2013, 19; Aho 2011, 30; Hovi ym. 2009, 86, muokattu)

Tietojen helppoa ja yksinkertaista raportointia varten tarvitaan vielä tietovaraston ja BI-ohjelmistojen väliin datan esityskerros eli metadatatamalli, jonka avulla tietovarastossa olevat ehkä teknisetkin termit käännetään käyttäjien ymmärtämään muotoon. Tämä esityskerros voi olla sisäänrakennettuna BI-ohjelmistossa, mutta joskus teknologia voi olla monikerroksisempi. Kuten luvussa 2.3 Liiketoimintatiedon hallinta (Business Intelligence) kerrottiin, liiketoiminnalla on erilaisia käyttötarpeita tiedon hyödyntämisen eri tasoilla. BI-ohjelmistoissa erilaiset tarpeet kullminoituvat erilaisiin raportointitekniikoihin, eikä mikään yksittäinen BI-ohjelmisto tai -ratkaisu taivu kaikkien tarpeiden täyttämiseen.

2.3.2 Keskeiset menetelmät

Liiketoiminnan erilaiset käyttötarpeet tiedon hyödyntämisen eri tasoilla aiheuttavat erilaisia vaatimuksia BI-ohjelmistojen teknologioille. Strategisen tason analysointia tekevien käyttäjien tiedon käyttötarve on huomattavan erilainen kuin suorittavan tason perusraportoijan käyttötarve. Käyttäjien teknologisessa osaami-

sessä on näiden ääripäiden välillä usein suurikin ero. Strategisen tason analysoinnissa käyttäjät, usein puhutaan tehokäyttäjistä tai analytikoista, käyttävät tiedon analysoinnissa erilaisia ad-hoc -kyselyjä, suorahakuja, OLAP-kuutioita ja tiedon louhintaa. Toisessa ääripäässä, usein operatiiviset käyttäjät, käyttävät valmiiksi rakennettuja raportteja, jotka ajamalla he saavat tarvitsemansa tiedon omaan tarpeeseensa. (Kimball 2008.) Operatiivisen raportoinnin tarpeisiin on tehty myös ajastettuja raportteja, jotka ajetaan automaattisesti säännöllisen aikataulun mukaan ja valmis raportti lähetetään käyttäjälle esimerkiksi sähköpostiin. Operatiivisten käyttäjien ja strategisen tason analytikoiden väliin ja käyttäjäkunta, jolle ei ihan riitä pelkät valmiiksi tehdyt standardiraportit, vaan he ajavat parametroituja raportteja, jotka ovat rakenteeltaan vakiomuotoisia, mutta käyttäjät voivat valita raportin ajamiseen haluamansa parametrit. (Hovi ym. 2001, 115)

Kimballin (2008) arvion mukaan valtaosa, noin 80-90 % käyttäjistä kuuluu alimpaan tai keskitason käyttäjäkuntaan tiedon hyödyntämisen tarpeiden näkökulmasta. Tällöin suurin osa raportoinnin tarpeista kohdistuu parametroituihin valmisraportteihin. Monet keskitason tiedon hyödyntäjien asiantuntijat kokevat parametroitujen valmisraporttien käyttämisen ad-hoc -raportointina, sillä he saavat itse valitsemillaan muuttujilla luoda raportin omaan tarpeeseensa. (Kimball 2008.) Siksi liiketoimintatiedon hallinnassa käyttäjien tarpeiden kuunteleminen onkin erityisen tärkeää, että ymmärtää mitä käyttäjät oikeasti tarkoittavat puhuessaan esimerkiksi ad-hoc -raportoinnista. Joillekin ad-hoc -raportiksi riittää, että tekee valmisraportin, jota ajettaessa käyttäjä voi valita esimerkiksi aikavälin, miltä ajalta raportti ajetaan, mutta muuten raportti on vakiomuotoinen. Toiset käyttäjät voivat tarkoittaa ad-hoc -raportoinnilla pääsyä tietovaraston tietokantoihin, josta he itse haluavat tehdä kyselyjä esimerkiksi SQL-kyselykielellä. Tällöin käyttäjät tarvitsevat aivan toisenlaiset tekniset välineet päästäkseen tekemään ad-hoc SQL-kyselyjä tietokannoista kuin käyttäjät, jotka mieltävät ad-hoc -raportoinnin parametroitujen valmisraporttien ajamiseksi.

Tyypillisessä organisaatiossa, jossa liiketoimintatiedon hallinta on jo kehittynyt, on usein kymmeniä, ellei satoja valmisraportteja eri tarpeisiin. Jotta raporttien jakelu olisi hallittua ja yhtenäistä, tarvitaan jonkinlainen yhteinen käyttöliittymä eli BI-portaali, josta raportteja voidaan ajaa ja käyttäjät pääsevät tarvitsemiinsa BI-

ohjelmistoihin. (Kimball 2008.) Portaaleita voidaan käyttää myös käyttöoikeuksien hallinnan välineenä, jolloin käyttäjillä on pääsy vain niihin raportteihin ja työkaluihin, mihin hänelle on myönnetty oikeudet. BI-portaaleita rakennetaan yleisesti sen takia, että se tarjoaa käyttäjille hyvin organisoidun, tarpeellisen ja helposti ymmärrettävän paikan hakea BI-työkaluja ja informaatiota. (Kimball 2008)

Muita keskeisiä menetelmiä liiketoimintatiedon hallintaan valmisraporttien, parametroitujen raporttien ja BI-portaalin lisäksi ovat OLAP-kuutioiden analysointiin tarkoitettut ohjelmistot, johdon työpöydät, mittaristot ja tuloskortit, tiedon louhintaohjelmistot ja erilaiset analyttiset ohjelmistot. (Kimball 2008) Kaikilla näillä eri liiketoimintatiedon hallintaohjelmistojen menetelmillä on oma tärkeä roolinsa tukea organisaatioiden käyttäjiä hyödyntämään informaatiota jokaisella tiedon käyttötarvetasolla. Kuten yllä perusteltiin, käyttötarpeita on monenlaisia riippuen käyttäjän työtehtävästä organisaatiossa. Tämän lisäksi käyttäjiä on erilaisia ja käyttäjien teknologiset valmiudet vaihtelevat huomattavasti. Tämä on myös yksi syy, minkä takia BI-ohjelmistoja on niin paljon erilaisia eri tarkoituksiin.

Luvussa 2.2.1 Tietovarastotyytit käsiteltiin monidimensionaalista tietovarastoa tietovarastoinnin näkökulmasta, mutta yleisemmin OLAP-kuutioita hyödynnetään raportoinnin ja analysoinnin työvälineenä. OLAP-kuutiot tuovat tiedon ad-hoc analysointiin enemmän mahdollisuuksia ja tehoa, kuin perinteisen tähtimallin relaatiotietokannan päälle rakennetut raportit. OLAP-kuutiot tarjoavat esimerkiksi controllereille ja analyttikoille valmiiksi paketoitua informaatiojoukon valitusta aihealueesta, jolloin he voivat analysoida dataa haluamistaan näkökulmista. Käyttäjät ovat tyypillisesti oman substanssinsa asiantuntijoita, mutta heillä on myös teknistä osaamista käyttää OLAP-kuutioiden raportointiin tarkoitettuja työkaluja. OLAP-kuutioiden päälle voidaan myös rakentaa erilaisia standardiraportteja sekä parametroituja raportteja, mutta näiden raporttien tekemiseen vaadittavat välineet ovat usein teknologialtaan vaativia ja raporttikehittäjien on oltava alan asiantuntijoita.

Johdon työpöydät, mittaristot ja tuloskortit ovat tarkoitettu etupäässä johdon tarpeisiin visualisoimaan liiketoiminnan suoriutumista asetettuihin tavoitteisiin nähden. Organisaation suorituskyvyn mittaamiseen on olemassa erilaisia menetel-

miä ja malleja, joiden teoriaan tässä ei syvällisemmin mennä. Yleinen organisaation suorituskyvyn mittaamiseen liittyvä suunnittelun ja johtamisen väline on tasapainotettu mittaristo (BSC, Balanced Scorecard), jonka visualisointiin voidaan käyttää johdolle rakennettuja työpöytiä.

Tasapainotettu mittaristo on tunnettu suomen kielessä myös termillä tuloskortti. Jokaisella yrityksellä ja organisaatiolla on omat mitattavat tavoitteensa ja tarpeensa suorituskyvyn mittaukselle, joten termistö ei ole vakiintunutta ja saattaa tarkoittaa eri yrityksissä eri asioita. Yhteistä näille kaikille mittaristoille kuitenkin on se, että jokaisen yrityksen tulee valita, määritellä ja päättää omat mittarinsa ja niihin liittyvät mitattavat suoritteet, joka usein on työläin vaihe suorituskymittariston rakentamisessa. Suorituskyvyn johtaminen vaatii visualisoidun mittariston lisäksi taustalle yleensä yhdenmukaistetun tietovaraston, jossa mittariston vaatima informaatio on saatavilla siistittynä, yhdenmukaisena, oikea-aikaisena ja joustavasti saatavilla. (Aho 2011, 46)

Liiketoimintatiedon hallinnan menetelmiksi voidaan käsittää myös tiedon louhinta ja erilaiset analytiikkavälineet. Tiedon louhinnassa yhdistyy tietämyksen etsintä, näkemys ja datassa toistuvien kaavojen tunnistaminen. Tiedon louhinta on monitieteellinen kenttä, jossa yhdistyy tietämys useasta eri tieteen ja tekniikan alasta. Siinä hyödynnetään tietokannoissa olevaa datan laadun ja organisoinnin ymmärrystä, tilastotieteen ja tietotekniikan tekniikoita sekä liiketoimintajohdon päätöksentekokykyä. (Maheshwari 2015, 45.)

Tiedon louhinnassa pyritään yleensä löytämään laajasta datamassasta datan toistuvia ilmiöitä, joiden perusteella pyritään tekemään ennustemalleja tulevaisuuden käyttäytymiseen. Tähän käytetään erilaisia tiedon louhinnan työkaluja, jotka perustuvat erilaisiin ohjelmointikieliin ja tilastollisiin malleihin. Tiedon louhinnan ammattilaiset ovat usein tietoteknisiltä kyvyiltään hyvin edistyneitä, mutta heillä on oltava osaamista myös tilastotieteen alueelta sekä aihealueen syvällistä tuntemusta. Vain harvoilla henkilöillä on tämä yhdistelmä hyvin erilaistuneista kyvyistä, jolloin tiedon louhintakaavoja ja -malleja kehitetäänkin yleensä moniammatillisissa tiimeissä. Juuri tämä usean alueen syvällisen osaamisen keskittymisen harvoille tekijöille aiheuttaa tiedon louhinnan vaikeuden, sillä löytääkseen

merkityksellisiä ennustemalleja pitää löytää ensin tiimi, joka pystyisi yhdessä nämä rakentamaan.

Kuten suorituskyvyn johtaminenkin, tiedon louhinta vaatii toimiakseen moniammatillisen tiimin lisäksi myös hyvin organisoidun ja yhtenäistetyn tietovaraston. Tiedon laatu on kriittinen menestystekijä arvoa tuottavalle tiedonlouhintahankkeelle. Lähes poikkeuksetta tiedon louhinnassa tarvitaan lähdedatan siivousta ja yhdenmukaistamista ennen datan hyödyntämistä, jota varten yhtenäistettyä tietovarastoa tarvitaan. (Maheshwari 2015, 48.) Maheshwarin (2015) mukaan tiedon puhdistaminen ja esivalmistelu on hyvin työllistävää ja vaikeasti automatisoitavaa ja tiedonlouhintahankkeista jopa 60-70% ajasta saattaa mennä tähän vaiheeseen.

Tässä opinnäytetyössä sote-tiedon analysoinnilla käsitetään olevan laaja tulkinta liiketoimintatiedon hallinnan tiedon analysoinnin menetelmistä. Tähän tulkintaan katsotaan kuuluvan raportoinnin eri menetelmät, OLAP-kuutioiden analysointi sekä eri ad-hoc -tekniikat. Sote-tiedon analysoinnilla ei kuitenkaan tarkoiteta tiedon louhintaa eikä siinä yhteydessä käytettäviä analysointityökaluja. Rajaus on tehty, koska opinnäytetyössä tutkittavaksi valittu työkalu, Power BI ei ole tiedon louhintaan tarkoitettu työkalu, vaikka siellä voikin suorittaa esimerkiksi R-kielellä toteutettuja komentosarjoja tai hyödyntää Python -kieltä kyselyeditorissa. (Microsoft n.d.)

Liiketoimintatiedon hallintaprojekteihin ja -hankkeisiin ei pidä lähteä teknologia-valinta edellä olettaen, että käyttäjät rakentaisivat omat raporttinsa annetuilla työkaluilla. (Kimball 2008) Raporttien ja esimerkiksi johdon työpöytien rakentaminen vaatii merkittävän paljon taustatyötä sekä tekniseltä että liiketoimintahenkilöstöltä. (Aho 2011, 46) Tällöin on ensiarvoisen tärkeää, että teknologiavalintoja tehdään vasta sen jälkeen, kun on tiedossa käyttötarpeet, miten käyttäjät tulevat tietoa hyödyntämään. Monesti myöskään ei ole tavatonta, että liiketoimintatarpeet täsmentyvät tai vaihtuvat kokonaan kesken BI-projektin. (Hovi ym. 2009, 123)

2.3.3 Tiedon laatu

Operatiiviset lähdejärjestelmät tuottavat tietovarastoille ja liiketoimintatiedon hallintaan informaation raaka-ainetta, kuten luvussa 2.3.1 Perusarkkitehtuurit kerrottiin. Operatiivisissa järjestelmissä oleva tieto ei useinkaan ole yhdenmukaista, täydellistä ja oikein syötettyä. Tietojen oikeellinen syöttö on pitkälti kiinni käyttäjistä, jotka tietoa järjestelmiin syöttävät. Kuten tiedon hyödyntäjissäkin, tiedon syöttäjissä on myös monen tasoista osaamista niin teknisiltä valmiuksiltaan kuin ammattitaidoiltaan.

Myös organisaatiokulttuuri vaikuttaa operatiivisten järjestelmien tiedon laatuun. Jos käyttäjiä ei opasteta riittävästi oikean toimintatavan mukaiseen toimintaan tai aikaa ei yksinkertaisesti ole tehdä riittäviä ja oikeita kirjauksia ajoissa, sillä on vaikutusta operatiivisten järjestelmien sisältämän tiedon oikeellisuudelle. Yksi syy operatiivisten järjestelmien tiedon heikkoon laatuun on myös se, etteivät tiedon syöttäjät useinkaan analysoi itse tekemiään kirjauksia, jolloin virheellinen tieto ei tule heille mitenkään näkyväksi sen jälkeen, kun tieto on tallennettu. (Hovi ym. 2001, 37)

Sosiaali- ja terveyspalveluiden alalla on käytössä monia mutkikkaita ja vaikeakäyttöisiäkin järjestelmiä, joihin sote-tietoa kirjataan. Tiedon syöttäjillä ei pääsääntöisesti ole koulutusta it-alalta, vaan he ovat ammattilaisia ihmisten hoitamisessa sekä terveyden että sosiaalihuollon alueilla. Tällöin heidän ensisijainen motivaationsa on auttaa potilasta tai asiakasta ja vasta sen jälkeen tulee motivaatio kirjata tiedot tarvittaviin järjestelmiin. Luvussa 2.1.2 Sote-tiedon erityispiirteet kuvattiin myös muita sote-tietojärjestelmien tiedon laatuun vaikuttavia seikkoja.

Poikkeus sote-tiedoissa tiedon laadussa ovat tiedot, joita käytetään esimerkiksi lääkärin palkkioiden maksun perusteena ja lain vaatimat pakolliset kirjaukset. Näiden tietojen laatu on lähes poikkeuksetta hyvää ja ajantasaista, sillä käyttäjillä on henkilökohtainen motivaatio saada tiedot kirjattua oikein, että saavat palkkion tekemästään toimenpiteestä. Myös lain vaatimia kirjauksia seurataan aktiivisesti, jolloin näiden tietojen oikeellisuus on yleensä hyvällä tasolla. Toisin on vain tilastointia varten tehtyjen kirjaustapojen noudattaminen, sillä ne koetaan monesti

olevan vain ylimääräistä työtä, johon menevä aika on pois asiakkaan hoitamisesta.

Operatiivisissa järjestelmissä olevan tiedon laatuongelmat paljastuvat usein vasta siinä vaiheessa, kun tietoja aletaan systemaattisesti ladata tietovarastoon. (Hovi ym. 2009, 68) 2.1.3 Tietojohdaminen -luvussa oli yksi esimerkki tapauksesta, jossa tietoa olisi tarvittu tukemaan esitetyn väitteen paikkaansa pitävyyttä, mutta johtuen puutteellisesta kirjaustavasta, tietoa ei ollut saatavilla. Tämä on yksi yleisimmistä tiedon laatua koskevista ongelmista. Jos tietoa ei kirjata operatiivisiin järjestelmiin, sitä ei ole käytettävissä myöskään tietovarastossa eikä raportoinnissa. Toinen tiedon laatua koskeva harhakäsitys on myös se, että virheellinen tieto muuttuisi tietovarastossa oikeaksi. Näin ei ole, vaan tietovarastoon ladataan sitä tietoa, mitä operatiivisissa lähdejärjestelmissä on kirjattuna. (Hovi ym. 2001, 37.)

Tiedon laatuun voidaan vaikuttaa ensisijaisesti panostamalla oikeisiin kirjaustapoihin ja tiedon oikeellisuuteen lähdejärjestelmissä. Tätä varten organisaation tulisi laatia laatustrategia, jossa määritellään tietojen laadulle asetettavat tavoitteet. Laatustrategiassa määritellään erityyppisille tiedoille omat laatutavoitteet ja niiden seurantaan tukevat mittarit. (Hovi ym. 2009, 69.) Kimball (2008) esittää perustettavaksi erillistä laadunvarmistajan roolia tai toimenkuvaa, jonka tehtävänä on valvoa ja varmistaa, että tietovarastoon ladattava tieto on oikeellista. Laadunvarmistajan tehtäviin kuuluu myös virheellisten tietojen selvittely ja tietovirheiden palauttaminen liiketoiminnan edustajille, jotka korjaavat liiketoiminnan toimintatapoja saadakseen tiedon oikein operatiivisiin järjestelmiin. (Kimball 2008.)

Monesti operatiivisten järjestelmien tiedon laatuongelmien syyt ovat syvemmillä organisaatorakenteissa, prosesseissa ja organisaatiokulttuurissa, kuin voisi ajatella. Organisaatioiden toimintoprosessit eivät välttämättä tue loogista ja systemaattista tietojen syöttämistä ja hallintaa, eikä organisaatorakenteet aina kannusta laadukkaan tiedon tallentamiseen järjestelmiin. Sellaisetkin ammattiryhmät, joiden päätehtävänä ei ole tietotekniset järjestelmät, joutuvat nykyään kirjaamaan itse omat työnsä myös tietokoneelle eri järjestelmiin, kun aiemmin niiden tekeminen saattoi kuulua sihteereille ja tiedon kirjaajille. Tämä kehityssuunta on ollut vallitseva viime vuosikymmenet ja tulee enenevässä määrin lisääntymään.

Tietojen laatua voidaan parantaa myös tekemällä tiedon profilointia, yhdistämistä ja monitorointia. Profiloinnilla tarkoitetaan tiedon laadun tarkkailemiseksi tehtyjä kyselyjä, joiden avulla esimerkiksi tutkitaan, puuttuuko jotain tietoa, onko tietojen arvot oikein tai onko kenttään syötetty virheellisiä merkkejä tai lukuja. (Hovi ym. 200, 69.) Datan profiloinnilla tarkoitetaan laajemmin myös systemaattista tutkimusta tiedon laadusta, laajuudesta ja asiayhteydestä, johon tietoa on tarkoitettu käytettävän. (Kimball & Caserta 2004, 5) Yhdistämisellä tarkoitetaan samaa tietoa tarkoittavien, eri lähteistä tulevien tietojen integrointia ja yhdistämistä samaan muotoon. Monitoroinnilla taas tarkoitetaan tiedon laadun jatkuvaa valvontaa, joka voi tapahtua joko automaattisesti profilointikyselyillä tai erilaisilla muilla valvontamekanismeilla. Laadun tarkkailuun ja valvontaan markkinoilta löytyy erilaisia ohjelmistoja.

Ennen jokaisen tietovarastohankkeen aloitusta olisi hyvä tehdä systemaattinen profilointi tiedon laadusta sillä alueella, mihin tietovarastointia suunnitellaan. Tiedon profilointi paljastaa lähdejärjestelmän tiedon laadun ja käyttökelpoisuuden tietovarastossa käytettäväksi. Profilointi antaa myös kuvan siitä, kuinka työläs ja monimutkainen ETL-prosessi tulee olemaan, sillä laadukkaalle, hyvin ylläpidetylle ja puhtaalle datalle ei juuri tarvitse tehdä ETL-prosessin aikana käsittelyjä. Toisessa ääripäässä oleva huonosti ylläpidetty ja puutteellinen lähtödata voi vaatia merkittävän työpanoksen ETL-prosessissa, jolla se saadaan ladattua yhtenäiseen muotoon tietovarastoon. Äärimmäisissä tapauksissa datan profilointi ennen tietovarastohanketta saattaa tuottaa tuloksen, jossa lähteenä oleva data ei vastaa suunniteltua tarvetta ja koko hankkeesta luovutaan tai vähintään perustetaan tiedon laatua parantavia toimenpiteitä ennen tietovarastohankkeen aloitusta. (Kimball & Caserta 2004, 5-6)

3 MICROSOFT POWER BI ANALYSOINTITYÖKALUNA

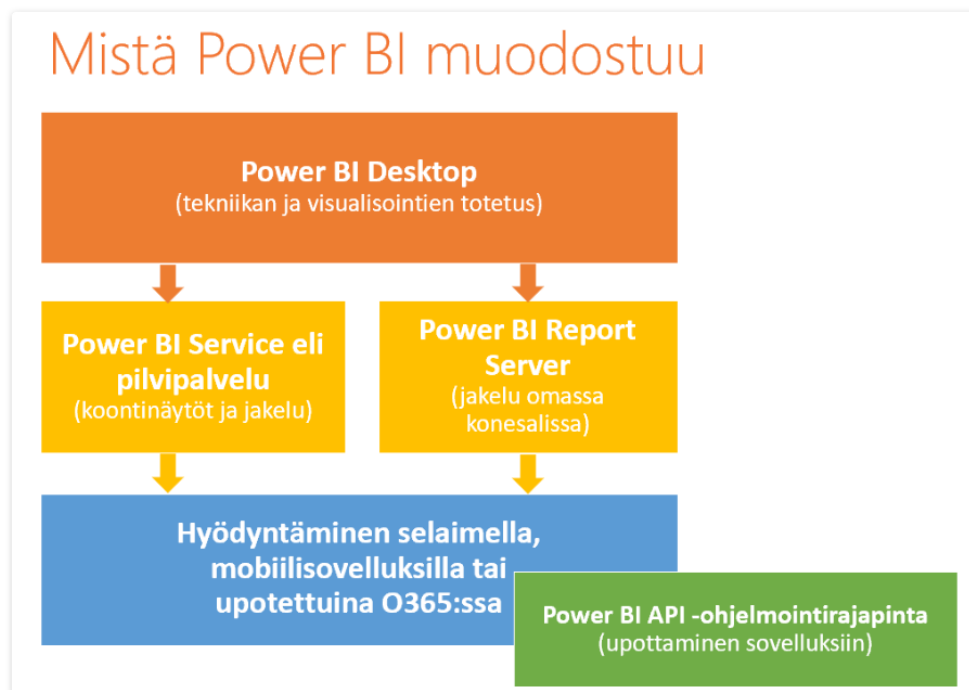
3.1 Power BI

Microsoft Power BI on vuonna 2015 julkistettu datan raportointi- ja analysointipalveluratkaisu, joka avulla dataa voidaan visualisoida ja yhdistellä yrityksen omista tietolähteistä, tiedostoista, internetistä, erilaisista pilvipalveluista ja avoimen datan lähteistä. (Enho 2018) Power BI:llä saadaan muunnettua toisiinsa liittymättömistä lähteistä peräisin olevaa tietoa johdonmukaisiksi, visuaalisesti vaikuttaviksi ja vuorovaikutteisiksi näkemyksiksi. (Microsoft n.d.) Kuvassa 9 havainnollistettu Power BI kokonaisuus koostuu neljästä eri osasta: Windows-työpöytäsovellus Power BI Desktopista, Power BI Service -pilvipalvelusta (Saas, Software as a Service), Power BI Report Server -raporttipalvelusta sekä tiedon hyödyntämiskerroksesta, kuten Power BI Mobile -mobiilisovelluksesta ja sovelluksiin, julkisille verkkosivuille tai intraan upotetuista raporteista. (Enho 2018)

Jokaisella näillä Power BI:n osalla on oma käyttötarkoituksensa ja monesti yrityksissä eri osille on omat tekijänsä ja käyttäjänsä. Karkeasti jaoteltuna Power BI käyttö voidaan jakaa neljään eri rooliin: suunnittelija, kehittäjä, kuluttaja eli loppukäyttäjä sekä järjestelmänvalvoja. Suunnittelijan rooliin kuuluu yleensä tietomallien ja raporttien tekeminen PBI Desktopilla ja tiedostojen julkaisu PBI Service -pilvipalveluun tai PBI Report Server -raporttipalvelimelle. Kehittäjän rooliin kuuluu yleensä Power BI sisällön upottaminen eri sovelluksiin, mukautettujen visualisointien kehittäminen sekä Power BI ohjelmointirajapintojen hyödyntäminen tiedon tuomiseksi Power BI:hin nähtäville. Kuluttajan eli loppukäyttäjän rooliin kuuluu koontinäyttöjen, raporttien ja sovellusten vastaanottaminen sekä Power BI:ssä olevan sisällön hyödyntäminen analysointiin, valvontaan, tarkasteluun ja päätöksentekoon. Järjestelmänvalvojan rooliin kuuluu Power BI -ympäristön hallinta, kuten hallintakäytäntöjen määrittäminen, käytön valvonta sekä käyttöoikeuksien, kapasiteettien ja organisaatioresurssien valmistelu. Järjestelmänvalvojan tehtävänä on myös parantaa eri rooleissa toimivien käyttäjien tuottavuutta sekä huolehtia ympäristön suojauksesta ja lakien ja asetusten noudattamisesta.

Isoissa organisaatioissa Power BI tietomallien ja raporttien kehitystä tehdään tiimeissä, joissa on henkilöitä eri rooleista. Kuluttajat eli loppukäyttäjät hyödyntävät kehitystiimin valmiiksi tekemiä koontinäyttöjä ja raportteja. Loppukäyttäjät voivat myös itse analysoida tietoa vuorovaikutteisilla visualisoinneilla, mutta eivät juuri itse kehitä tai koosta raportteja. Pienemmissä organisaatioissa yksittäiset henkilöt voivat toimia kaikissa näissä rooleissa ja huolehtia tietolähteestä saadun tiedon jalostamisesta valmiiksi koontiraportteiksi saakka.

Tapahtui Power BI -kehitys ja roolijaot miten tahansa, niin yrityksen on kuitenkin huolehdittava siitä, että Power BI -ratkaisun käyttöönoton yhteydessä sovitaan yhtenäiset raportointikäytännöt esimerkiksi koontinäyttöjen laatimisen ja jakelumenetelmien osalta. Muuten kokonaisuus voi näyttäytyä loppukäyttäjille sekavalta ja epäyhtenäiseltä. (Enho 2018.) Power BI:n käyttöönottoa suunnitellessa on hyvä sopia etukäteen eri roolien vastuut liiketoiminnan ja tietohallinnon organisaatioiden kesken, että kaikki Power BI:n vaatimat toiminnot tulisi resursoitua riittäväillä ja osaavilla henkilöillä.



KUVA 9. Power BI:n eri osat (Enho 2018)

Power BI Desktop

Ilmaiseksi paikalliseen tietokoneeseen asennettavalla Power BI Desktop -soveluksella toteutetaan kyselyt, tiedon mallinnus ja laskentalogiikka. Tietomallia käyttämällä PBI Desktopissa luodaan myös raporttitiedostoja ja -sivuja sekä koostetaan raportteja vuorovaikutteisiksi visualisoinneiksi. Yhdessä PBI Desktop -tiedostossa olevien visualisointien kokoelmaa kutsutaan raportiksi ja raportti voi koostua joko yhdestä tai useammasta sivusta. (Microsoft n.d.) Yhdellä sivulla on yleensä useita visualisointeja, kuten pylväitä, palkkeja, karttoja, kaavioita ym. Visualisoinnit ovat yleensä keskenään vuorovaikutteisia, jolloin yhden visualisoinnin suodattaminen vaikuttaa myös muihin sivulla oleviin visualisointeihin.

PBI Desktop kehitys kuuluu tavallisesti suunnittelijan rooliin, kuten myös valmiin tietomallin ja raporttien julkaisu PBI Service -pilvipalveluun tai PBI Report Server -raporttipalvelimelle. Suunnittelijan vastuulle voi kuulua myös raporttien ja koonnäytöjen laatiminen pilvipalvelussa. Tälle roolille on kuitenkin monesti tarve löytää organisaatiosta suunnittelijoista erilliset henkilöt, jotka tuntevat dataa riittävän hyvin ja jotka kykenevät laatimaan olemassa olevista malleista raportteja.

PBI Desktopin ilmaisuudesta johtuen myös yksittäiset kuluttajat voivat käyttää PBI Desktopia veloituksetta omiin käyttötarpeisiinsa ja rakentaa itselleen esimerkiksi omasta taloudestaan tai harrastuksistaan raportteja. PBI Desktopin ilmaisuus hyödyttää myös pieniä yrityksiä ja yksityisyrittäjiä, joiden raportointitarpeet ovat usein pelkistettyjä ja yksinkertaisia. PBI Desktopilla tehdyillä tietomalleilla ja sen kautta rakennetuilla raporteilla pystyy raportoimaan tietoa myös paikallisesti omalla tietokoneella, jolloin yrityksen ei välttämättä tarvitse ostaa maksullisia lisenssejä raporttien julkaisemiseen Power BI Service -pilvipalveluun tai Power BI Report Server -raportointipalvelimelle.

Henkilökohtaiseen käyttöön ja julkiseen jakamiseen on käytettävissä myös ilmainen pilvipalvelutili Power BI Free, joka sopii henkilökohtaiseen käyttöön ja julkiseen jakamiseen. Muuten PBI Service -pilvipalvelun käyttö ja PBI Report Server -raporttipalvelimen käyttö vaativat maksulliset lisenssit. Power BI:n hinnoittelua ja lisensointia ei tässä opinnäytetyössä käsitellä tämän syvällisemmin, sillä se ei ole olennainen asia opinnäytetyön aiheen kannalta. Microsoftin verkkosivuilta

löytyy ajankohtaista tietoa hinnoittelumallista ja lisenssihinnoista. Myös Heidi Enhon (2018) Power BI:tä koskevalla verkkosivustolla on kuvattu asiaa kattavasti.

Power BI Service ja Power BI Report Server

Power BI Desktopilla muodostetut tiedostot julkaistaan joko Power BI Service -pilvipalveluun tai sisäisesti Power BI Report Server -raportointipalvelimelle ja jaetaan loppukäyttäjille hyödynnettäväksi. (Enho 2018.) PBI Servicen työtiloihin julkaistujen tietomallien ja/tai raporttien tärkeimmistä mittareista ja visualisoinneista koostetaan PBI Service -pilvipalvelussa koontinäyttöjä. Koontinäyttö on PBI Service -palvelussa oleva yksi sivu, johon voidaan kiinnittää useasta eri raportista useita eri visualisointeja. Koontinäyttöjä hyödynnetään esimerkiksi johdolle tarkoitettussa raportoinnissa, jossa on kerätty yhteen näkymään keskeisimmät ja olennaisimmat liiketoiminnan suorituskykyä kuvaavat mittarit.

PBI Service -pilvipalvelu mahdollistaa yksittäisen Power BI -ratkaisun jakamisen kahteen eri osaan: tekniseen tietomalliin ja erillisiin visualisointitiedostoihin. Tämä ratkaisu erottaa tietomallin toteuttamisen ja kehittämisen ja tietomallin raporttikehityksen toisistaan, jolloin Power BI -ratkaisun kehittämistä voidaan jakaa organisaatiossa useammalle eri vastuutaholle. PBI Service -pilvipalvelussa voidaan tehdä myös pienimuotoista raporttikehitystä, mutta sen tarjoamat ominaisuudet ovat huomattavasti rajoitetummat PBI Desktopiin verrattuna.

PBI Service -pilvipalvelu on monipuolisempi, kattavampi ja yleisemmin käytössä oleva julkaisupalvelu kuin sen rinnalle kesällä 2017 julkistettu PBI Report Server -raportointipalvelin. PBI Report Server -raportointipalvelin on kehitetty tarpeeseen, jossa yritykset eivät voi tai halua julkaista tietojansa julkiseen pilvipalveluun, vaan haluavat pitää tietonsa paikallisesti yrityksen omassa sisäverkossa palomuurien takana. Microsoft kehittää Power BI:tä pilvipalveluvetoisesti, jolloin uusimmat ominaisuudet julkaistaan ensin pilvipalveluversiona kuukausittain. PBI Report Serverin julkaisu tiheys on neljä kertaa vuodessa, eikä kaikkia pilvipalvelussa julkaistuja ominaisuuksia ole PBI Report Serverille saatavissa. PBI Report Server -raporttipalvelimelta puuttuvat esimerkiksi koontinäytöt (Dashboard) kokonaan, eikä Power BI -ratkaisua voida jakaa kahteen osaan (tietomalli ja visualisoinnit), kuten PBI Servicessä pystytään tekemään. Raporttien kehitys PBI Report Serverille vaatii myös oman, raporttipalvelimelle optimoidun version PBI

Desktopista, jolle julkaistaan neljä kertaa vuodessa päivityspaketti, kun pilvipalveluun tarkoitettuun PBI Desktopiin päivityksiä julkaistaan kuukausittain. Raporttipalvelimelle optimoidussa PBI Desktopissa ei myöskään ole kaikkia ominaisuuksia saatavilla, mitä pilvipalveluun tarkoitettussa PBI Desktopissa on julkaistu.

Kattava ja ajantasainen lista PBI Servicen ja PBI Report Serverin samankaltaisuuksista ja olennaisista eroista löytyy Microsoftin verkkosivuilta. Eroja ja yhtäläisyyksiä ei tässä opinnäytetyössä käydä läpi tämän tarkemmin, sillä asia ei ole olennainen opinnäytetyön aiheen kannalta. Tampereen kaupungilla teknologia-valintoihin tehtyjen päätösten kannalta asiaan liittyvät olennaiset ja merkitykselliset erot käsitellään tässä opinnäytetyössä nykytilan kuvausta, johtopäätöksiä ja pohdintaa koskevissa luvuissa niiltä osin, kuin on tarpeen.

Power BI sisällön hyödyntäminen

Kun tietomalli, visualisoinnit ja raportit on tehty PBI Desktopilla ja julkaistu joko PBI Services -pilvipalveluun tai PBI Report Server -raporttipalvelimelle, julkaistua sisältöä voidaan alkaa hyödyntää eri käyttötarkoituksiin. Suunnittelija tai muu dataa riittävästi tunteva henkilö koostaa loppukäyttäjiä varten koontinäyttöjä, sovel-luspaketteja ja raportteja pilvipalveluun, joita loppukäyttäjät hyödyntävät datan analysoinnissa ja päätöksenteossa. Loppukäyttäjät kirjautuvat sekä PBI Service -pilvipalveluun että PBI Report Server -raporttipalvelimelle selaimella verkkoportaalin kautta. Molempiin palveluihin julkaistuja raportteja voidaan käyttää myös Power BI Mobiilisovelluksen kautta. Olennainen ero PBI Services -pilvipalvelun ja PBI Report Server -raporttipalvelimen sisällön hyödyntämisessä on se, ettei PBI Report Server -raporttipalvelimeen voida tehdä koontinäyttöjä ja sen tarjoamat raporttien muokkausmahdollisuudet ovat vähäisemmät. PBI Report Server -raporttipalvelimella toimiva verkkoportaali toimii yrityksen sisäverkossa, joten sinne voidaan julkaista myös esimerkiksi SQL Server Reporting Services -raportteja.

PBI Report Server -raporttipalvelimen verkkoportaalin kautta käyttäjille jaettu sisältö poikkeaa PBI Service -pilvipalvelun portaalista mm. siten, että PBI Report Server verkkoportaalissa sisältö on järjestetty perinteiseen kansiohierarkiaan. Kansiot on ryhmitelty sisällön tyyppin mukaan, Power BI -raportit, mobiiliraportit,

sivutetut raportit, suorituskykyilmaisimet ja Excel -tiedostot löytyvät omista kansioistaan. Myös jaetuille tietolähteille ja tietojoukoille löytyy omat kansionsa. PBI Report Server -verkkoportaalin käyttö on enemmän tiedon kuluttamista varten kuin PBI Service -pilvipalvelussa, jossa pystytään koostamaan koontinäyttöjä, paketoimaan raportteja sovelluksiksi ja myös tekemään pienimuotoista raporttikehitystä. PBI Report Server -verkkoportaalisissa pystyy tekemään suorituskykyilmaisimia (KPI, Key Performance Indicator), jotka näkyvät käyttäjille omassa kansiossaan.

3.2 Power BI toimintamallit

Power BI:n visualisointien helppokäyttöisyydestä johtuen Power BI mielletään yrityksissä monesti täysin itsepalveluraportointivälineeksi, jossa ei tarvita perinteisen tietohallinto-organisaation tukea lainkaan. Moniin tehtäviin tämä sopiikin mainiosti, varsinkin kun tietolähteenä ovat yksittäiset tiedostot, joiden tietoja pitää visualisoida esimerkiksi esitystä varten. Suuremmissa organisaatioissa halutaan kuitenkin monesti keskittää raporttikehitystä yhtenäisyyden ja ympäristön hallittavuuden takia.

Kummallakin tavalla on omat etunsa ja haittansa, joten yrityksissä tehdyt strategiapäätökset vaikuttavat osaltaan siihen, minkälaista toimintamallia yrityksissä käytetään. Yleensä myös yrityksen organisointirakenne vaikuttaa käytettyyn toimintamalliin, sillä keskitetyssä raportoinnissa tietohallinto on yleensä vahvassa roolissa, kun taas itsepalveluraportoinnissa raportoinnin kehitys ja toteutus on vahvasti liiketoimintavetoista. Yrityksissä myös itsepalveluraportointi (Self-Service reporting) voi tarkoittaa eri organisaatioissa eri asioita ja tämä on hyvä huomioida suunniteltaessa Power BI:n käyttöönottoa yrityksessä. (Enho 2018) Knight, Knight, Pearson ja Quintana (2018) jakavat kirjassaan Microsoft Power BI Complete Reference Power BI:n käytön kolmeen eri malliin: keskitettyyn, itsepalveluvisualisointiin ja itsepalveluraportointiin. Seuraavassa kuvataan tarkemmin Knightin ym. (2018) Power BI toimintamallijakoa.

Keskitetty BI -raportointi (Corporate BI)

Keskitetyssä BI -toimintamallissa tietohallinto omistaa sekä Power BI tietomallin että myös Power BI raportit ja koontinäytöt. Tällöin tietohallinto on vastuussa sekä tietomallin että raporttien ja koontinäyttöjen kehityksestä, julkaisusta ja ylläpidosta. Liiketoiminnan käyttäjät ovat raporttien kuluttajia, eli hyödyntävät olemassa olevia raportteja, koontinäyttöjä sekä muita Power BI sisältöjä. Tässä toimintamallissa tietohallinto pystyy toteuttamaan keskitettyjä, yhdenmukaisia ja suorituskykyisiä tietomalleja sekä raportteja, jolloin kaikille käyttäjille on sama sisältö tarjottuna yhtenäisellä julkaisutavalla. Liiketoiminta saa haluamansa tiedot helposti valmiiksi koottuna eri tietolähteistä yhdenmuotoisena. Tämä toimintatapa on liiketoiminnan loppukäyttäjien kannalta yksinkertaisin ja helpoin tapa kuluttaa raportteja, koska heidän ei tarvitse tehdä itse tietomalleja eikä raportteja. He voivat keskittyä valmiiksi muokatun tiedon hyödyntämiseen, tehdä siitä analyysejä sekä käyttää tietoa päätöksentekonsa tukena. Liiketoiminnan pääkäyttäjät eivät kuitenkaan pysty itse tekemään raportteja tai koontinäyttöjä, vaan siinä heidän on aina turvaututtava tietohallinnon osaamiseen. Tämä voidaan kokea kankeana toimintamallina liiketoiminnan käyttäjien näkökulmasta.

Keskitetty BI -toimintamallin ansiosta tietohallintoon keskittyy syvällistä osaamista Power BI -ratkaisujen toteutuksesta ja henkilöstö tietohallinnossa on muutenkin teknologisesti osaavampaa kuin keskimäärin muussa liiketoimintayksiköissä. Tällöin tietohallinnossa yleensä pystytään toteuttamaan vaativampia ja suorituskyvyltään tehokkaampia Power BI -ratkaisuja kuin liiketoimintayksiköissä loppukäyttäjien toimesta. Keskitetty BI -toimintamalli on toisaalta tietohallinnolle kuormittavaa, sillä sekä tietomalli- että raporttikehitys ovat molemmat työllistäviä tehtäviä ja liiketoiminnalta tulleita muutospyyntöjä ei välttämättä saada kovin nopealla tahdilla toteutettua. Toinen negatiivinen seikka keskitetyssä toimintamallissa on se, ettei tietohallinnossa ole välttämättä samaa liiketoimintaosaamista ja tietämystä kuin liiketoimintayksiköissä. Tällöin kehitystyössä joudutaan turvautumaan liiketoimintayksiköiden osaamiseen tietämyksen osalta.

Itsepalveluvisualisointi (Self-Service Visualization)

Keskitettyä BI -raportointia hieman joustavampi toimintamalli on itsepalveluvisualisointi. Siinä mallissa tietohallinto omistaa Power BI tietomallit ja vastaa nii-

den ylläpidosta, kehityksestä ja julkaisusta, mutta liiketoiminta omistaa tietomalleista tehdyt raportit ja koontinäytöt ja vastaa niiden kehityksestä ja ylläpidosta. Tämän toimintamallin etuna on liiketoiminnan pääkäyttäjille joustavampi tapa päästä itse toteuttamaan raportteja ja koontinäyttöjä, mutta kuitenkin yhtenäisestä ja tietohallinnon ylläpitämästä tietomallista. Tällöin liiketoiminnan pääkäyttäjillä on mahdollisuus tehdä itse erilaisia visualisointeja ja raportteja oman tarpeensa mukaan ilman tietohallinnon työpanosta.

Tietohallinnon näkökulmasta tietomalli pysyy yhtenäisenä ja keskitettynä, jolloin liiketoiminnassa käytetään samoja mittareita ja laskentakaavoja. Visualisoinnissa on kuitenkin käytettävissä vain tietomallissa tarjolla olevat tiedot, joten jos liiketoiminta haluaa tietomalliin uusia toiminnallisuuksia tai tietoja, niitä on pyydettävä tietohallinnolta.

Itsepalveluraportointi (Self-Service BI)

Kaikkein joustavin tapa Power BI toimintamalleista on itsepalveluraportointi, jolloin liiketoiminta omistaa sekä Power BI tietomallit että raportit ja koontinäytöt sekä vastaa niiden kehityksestä ja ylläpidosta. Tietohallinnon tehtävänä on vain hallinnoida välttämätöntä Power BI infranstruktuuria ja tarjota liiketoiminnalle tekninen alusta Power BI:n käyttöön. Itsepalveluraportointimallissa liiketoiminnalla on vapaus toteuttaa Power BI tietomallit haluamallaan tavalla tarjolla olevista tietolähteistä ja muokata niistä täysin vapaasti oman tarpeensa mukaan. Tietolähteet ovat tällöin tyypillisesti tiedostopohjaisia Excel-tiedostoja tai eri lähdejärjestelmistä csv-tiedostomuotoon tulostettuja tiedostoja. Tietomallien laatijoilla ei yleensä ole mahdollisuutta muokata lähdejärjestelmistä tulevaa dataa kuin muokkaamalla sitä PBI Desktopin kyselyissä. Itsepalveluraportoinnissa tietojen lataaminen tietomalliin on tyypillisesti manuaalista, eikä lähdejärjestelmistä tuotettuja tietolähteitä voida aina automatisoida.

Itsepalveluraportointimalli edellyttää liiketoiminnan käyttäjiltä monipuolista PBI Desktop -ohjelman ja DAX-kielen osaamista sekä vaatii käyttäjiltä teknistä tietämystä ja ymmärrystä esimerkiksi tietokantojen taulujen välisistä relaatioista. Teknisen osaamisen lisäksi käyttäjillä pitäisi olla kyvykkyys rakentaa tietomallit suorituskykyisiksi ja dimensionaalisen mallinnuksen hyvien käytäntöjen mukaisiksi.

Usein liiketoiminnassa toimivilla henkilöillä ei ole riittävää osaamista ja kyvykkyyttä tietomallien teknisessä rakentamisessa kuin yksittäisissä ja pienissä toteutuksissa. Itsepalveluraportointimallin vaarana on myös tiedon pirstaloituminen ja tiedon eri totuuksien leviäminen eri organisaatioiden kesken, kun jokainen toimintayksikkö tekee omanlaisensa tietomallin omilla muotoiluillaan ja laskentakaavoillaan.

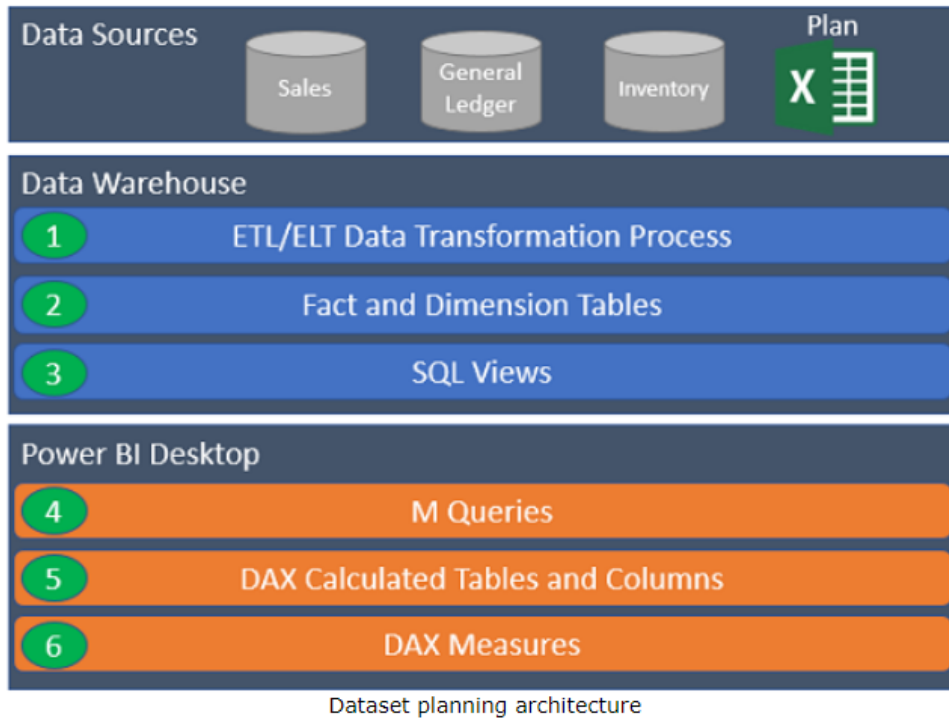
Kaikki yllä kuvatuista toimintamalleista ovat käyttökelpoisia perusteltuihin käyttötarkoituksiin eri organisaatioissa, eikä yksikään ole parempi kuin muut. Monesti yrityksessä on käytössä näiden yhdistelmiä riippuen käyttötarpeesta, joka on aivan perusteltua monissa organisaatioissa. Yrityksissä pitäisi kuitenkin käyttää aikaa ja miettiä käyttötarpeita ennen Power BI -ratkaisujen käyttöönottoa, ettei ratkaisusta tulisi hajanainen ja vaikeasti hallittava kokonaisuus. Vaikka Power BI mahdollistaa tietojen yhdistämistä, eri lähdejärjestelmistä tietojen hakemista ja tiedon muokkausta sekä tietomallien rakentamisen, se ei kuitenkaan korvaa yrityksissä tietovarastoja. (Knight ym. 2018.) Esimerkiksi tietojen historiatiedon tallentaminen kuuluu tietovarastoon, eikä Power BI -ratkaisun hoidettavaksi.

3.3 Power BI tietolähteet ja tiedon muokkaus

Lukuun ottamatta pieniä ja yksittäisiä itsepalveluraportointi -mallilla toteutettuja Power BI -toteutuksia useimmiten tietolähteenä toimii joko yrityksen tietovarasto, monidimensionaalinen tai tabular -paikallistietovarasto tai vastaava tietolähde, mihin lähdejärjestelmistä on ladattu tieto ETL-prosessin läpi. Tämä toimintatapa mahdollistaa automaattiset tietolataukset, yhtenäisen tiedon ja hallitun rajapinnan liiketoimintatiedon hallinnan työkaluille, kuten Power BI:lle. Power BI:hin tehdyt suorat tietolähteet operatiivisten järjestelmien tietokantoihin eivät ole suositeltuja, koska ne häiritsevät operatiivisen järjestelmän suorituskykyä sekä aiheuttavat operatiivisten järjestelmien päivityksen yhteydessä muutoksia tietolähteiden rakenteisiin. (Knight ym. 2018)

Tietovarastointia koskevissa luvuissa käsiteltiin tiedon latauslogiikkaa operatiivisista lähdejärjestelmistä ETL-prosessin kautta tietovarastoon ja eri tietovarastointiarkkitehtuureja. Kuvassa 10 havainnollistetaan tiedon jalostumisen eri kerrokset

tiedon lähteiltä tietovaraston kautta Power BI tietolähteeksi. (Knight ym. 2018) Tiedon muokkausta voidaan tehdä kuudessa eri kerroksessa, sekä tietovarastossa että PBI Desktopissa.



KUVA 10. Tiedon muokkauksen kerrokset tietovarastossa ja PBI Desktopissa. (Knight ym. 2018)

PBI Desktopissa tehty tiedon muokkaus mahdollistaa monipuolisen tiedon hyödyntämisen joustavasti liiketoiminnan tarpeisiin, vaikka tietovarastossa tieto ei olisiakaan toivotussa muodossa. PBI tietomalliin voidaan rakentaa oma logiikkakerroksensa liiketoimintaraportoinnin tarpeisiin, jolloin liiketoiminta saa tarvitsemansa raportit tarpeidensa mukaan. Tässä ratkaisussa on kuitenkin vaarana tiedon jäljitettävyyden ja muokausprosessien ylläpidon hajautuminen tiedon jalostamisen eri kerroksiin. Tieto tulisi muokata ja jalostaa mahdollisimman lähellä sen lähdettä tai tietovarastointiprosessissa, jolloin tiedon muokkauksen ylläpito tehdään vain kerran, kun tietorakenteet muuttuvat. (Knight ym. 2018.) Knight ym. (2018) suosittelee, että tiedon muokkaus tehtäisiin mahdollisimman lähellä tiedon alkuperäistä lähdettä ja että Power BI:tä käytettäisiin vain tiedon analysointiin ja visualisointiin. He kuitenkin toteavat, että tämä toimintamalli harvoin toteutuu tässä muodossa johtuen tietohallintoressurssien rajallisuudesta ja tiukoista projekti aikatauluista. (Knight ym. 2018)

Tietolähteiden arkkitehtuuriin liittyy myös tietojen latauksen arkkitehtuurivalinta, miten tietoa ladataan tietomalleihin raportoitavaksi. Arkkitehtuurissa suunnitellaan tietojen lataustiheys, ladattavan datan määrä ja myös se, ladataanko tieto tietomalliin (Import) vai käytetäänkö lähes reaaliaikaista yhteyttä (DirectQuery) lähteenä oleviin tietokantoihin tai kuutioihin. Lähes reaaliaikaisen yhteyden rajoituksena on aiemmin ollut se, ettei tietoa ole voinut tuoda tietomalliin kuin yhdestä lähes reaaliaikaisen yhteyden tietolähteestä. Lokakuun 2018 PBI Desktop päivityksessä on julkaistu uusi tietomallityyppi, Yhdistelmämalli, joka mahdollistaa eri tietolähdetyyppien yhdistämisen yhteen tietomalliin. (Microsoft n.d.) Microsoft suosittelee tietojen lataamista Power BI:hin aina, kun se on mahdollista, jolloin hyödynnetään Power BI:n tehokasta kyselytoimintoa parhaiden käytäntöjen mukaisesti. Aina tämä ei kuitenkaan ole mahdollista ja toivottavaa esimerkiksi tietosuojasyistä, jolloin lähes reaaliaikaisen yhteyden käyttäminen tulee perustelluksi vaihtoehdoksi.

Tietojen tuomisen arkkitehtuuriratkaisu riippuu aina toteutettavasta ratkaisusta, yrityksen arkkitehtuuriratkaisusta ja monesta muusta käytännön toimintaa määrittävästä seikasta. Tässä opinnäytetyössä ei käydä asiaa tämän syvällisemmin läpi, koska asia ei ole olennainen opinnäytetyön aiheen kannalta. Tampereen kaupungilla teknologiavalintoihin tehtyjen päätösten kannalta asiaan liittyvät olennaiset ja merkitykselliset seikat käsitellään tässä opinnäytetyössä nykytilan kuvausta, johtopäätöksiä ja pohdintaa koskevissa luvuissa niiltä osin, kuin on tarpeen.

4 NYKYTILA TAMPEREEN KAUPUNGILLA

4.1 Yleiskuva tietovarasto- ja raportointiympäristöstä

Tampereen kaupungilla SAP BW -järjestelmästä raportoitavasta sosiaali- ja terveystietojen kokonaisuudesta käytetään termiä SAP BW Toimintaraportointi, joka kuvaa hyvin raportoitavan tiedon hyödyntämistarkoitusta. SAP BW:n kautta tuotetaan sote-tietoa etupäässä toiminnan tarpeisiin ja sen seurantaan. Raportointiratkaisu on suunniteltu alun perin operatiivisten järjestelmien yhteiseksi organisaatio-, asiakas-, palvelu- ja toimenpide -informaation tietovarastoksi. Ratkaisu koostuu kahdesta eri kerroksesta, EDM-välitietokannasta (EDM, Enterprise Data Management) ja SAP BW -tietovarastosta sekä näihin liittyvistä tiedon siirtoon, harmonisointiin ja muokkaukseen liittyvistä vaiheista, jotka on toteutettu eri tekniikoilla.

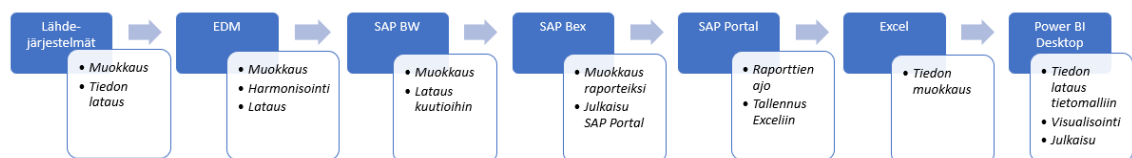
Tiedon siirtoon lähdejärjestelmistä käytetään SAP XI integrointityökalua, tiedon latauksessa ja harmonisoinnissa käytetään PL/SQL -proseduureja ja ajastettuja komentoriviskriptejä. Tiedon rajapintana SAP BW -latauksille toimivat pääsääntöisesti materialisoidut tietokantanäkymät. Löytyneen dokumentaation perusteella tulkitsin, että alkuperäisen suunnitelman mukaan EDM olisi toiminut tietovarastona ja SAP BW:tä oltaisiin hyödynnetty raportointiin, jolloin valmiiksi EDM:ssä harmonisoitu tieto oltaisiin voitu siirtää BW-kuutioihin yksinkertaisten liittymien avulla.

Tampereen kaupungin raportointi- ja tietovarastoympäristöä ollaan rakennettu ja kehitetty aluksi projekteissa palvelualuekokonaisuuksittain, jolloin työssä on ollut kiinteästi mukana sisäistä henkilöstöä niin palvelualueilta kuin tietohallinnosta ja ulkoista osaamista järjestelmätoimittajien puolelta. Raportointi- ja tietovarastoympäristön kehitys projekteissa auttoi alkuvaiheessa tuomaan palvelualueille kohtuullisessa ajassa tarjolle raportteja ja rakentamaan samaan aikaan perustaa tietovarastoratkaisulle. Kokonaisuus rakentui projekteittain eri palvelualueille ja projektien päätyttyä toimintaraportointi kattoi laajimmillaan terveydenhuollon puolelta mm. perusterveydenhuollon, erikoissairaanhoidon, osastohoidon, vanhus-

tenhoidon, suun terveydenhoidon, kotihoidon ja omaishoidon toimintaraportoinnin. Sosiaalipuolelta raportointi kattaa mm. yksilö- ja perhehuollon, toimeentulotuen, ostopalvelujen palveluseteliraportoinnin ja vammaispalveluiden toimintaraportoinnin. Kokonaisuus on vuosien saatossa elänyt sen mukaan, minkälaista toimintaa kaupungin järjestämispalveluihin kuuluu milloinkin. Esimerkiksi eräät toimintojen siirrot kaupungin tuottamasta palvelusta sairaanhoitopiiriin vastuulle ovat vaikuttaneet olennaisesti myös raportoinnin laajuuteen ja tarpeeseen.

Viime vuodet SAP BW toimintaraportointia ollaan pääsääntöisesti vain ylläpidetty, eikä isompia kehityshankkeita ole pystytty toteuttamaan resurssien niukkuudesta johtuen. Osaltaan myös organisointia ja toimintojen ulkoistuksia koskevat päätökset ovat vaikuttaneet mahdollisuuksiin ylläpitää ja kehittää täysipainoisesti raportoinnin ja tietovarastoinnin kokonaisuutta koko kaupunkia koskevana toimintana. Raportointi on keskittynyt pääsääntöisesti luvussa 2.1.3 Tietojohtaminen kuvatun tiedon eri tasojen alimmalle, operatiiviselle tasolle, jossa se palvelee etupäässä operatiivisen toiminnan tarpeita ja seuranta. Yksittäisiä ratkaisuja ollaan toteutettu tiedon tuottamiseksi taktisen tason tietotarpeita varten, mutta nämä eivät ole vielä täysin automatisoituja ja ne palvelevat vain yksittäisiä tiedon hyödyntäjiä. Power BI -raportointiratkaisun käyttöönottopäätöksen teon jälkeen SAP BW -toimintatietoa ollaan mahdollisesti hyödynnetty pistemäisesti yksittäisiin raportointiratkaisuihin eri puolilla organisaatiota.

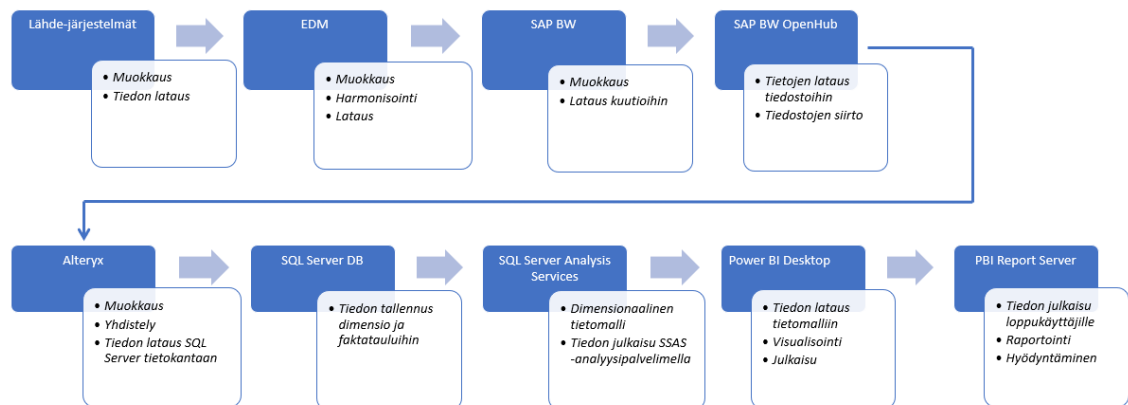
Nykytilanteessa Power BI -käyttäjät joutuvat ajamaan SAP BW toimintaraportteja SAP Portaalin kautta ja tallentamaan ne Excel-tiedostoiksi saadakseen sote-tietoa yhtenäisen tietovarastoratkaisun kautta Power BI tietomalliinsa. Raporttien ajaminen SAP Portaalista, tallentaminen Excel-muotoon ja tiedon muokkaus Power BI -tietomalliin soveltuvaksi ovat manuaalista työtä eikä työ ole tällä hetkellä automatisoitavissa. Kuvassa 11 kuvataan prosessi sote-tiedon kulusta lähdejärjestelmistä loppukäyttäjän omaan Power BI tietomalliin.



KUVA 11. Sote-tiedon Power BI -latauksen prosessi loppukäyttäjät.

Power BI:tä ollaan kokeilumielessä testaamassa myös vastaanotto toiminnan palvelualueella ja kokeilussa hyödynnetään SAP BW:stä saatavaa avopalveluiden vastaanotto toiminnan soite-tietoa, joka on tuotettu SAP BW OpenHub siirto-ohjelman avulla tiedostoihin. Tässä kokeilupilotissa hyödynnetään taktisen tason analysointiin tuotettuja tiedostoja, jotka muokataan ja ladataan itsepalveluanalysointiin tarkoitetulla Alteryx -ohjelmistolla SQL Server tietokantaan. SQL Server -tietokannasta tiedot luetaan sisään SQL Server Analysis Services dimensionaaliseen tietomalliin, johon on liitetty myös muita tietolähteitä. Power BI Desktopilla on otettu tietolähdeyhteys SSAS tietomalliin, josta tiedot tuodaan raportoitavaksi ja visualisoitavaksi Power BI:hin. Raportit julkaistaan pilotissa PBI Report Serverille käyttäjien hyödynnettäväksi.

Johtuen nykytilanteen tietovarastoarkkitehtuurista ja teknologiasta, tässä pilottiratkaisussa on paljon manuaalista lataustyötä välissä, työ on yksittäisen henkilön varassa, eikä prosessi ole teknisesti automatisoitavissa. Kuvassa 12 on kuvattuna prosessi tiedon kulusta lähdejärjestelmistä PBI Report Serverillä oleviin valmiisiin visualisointeihin saakka.



KUVA 12. Sote-tiedon Power BI -latauksen prosessi vastaanotto toiminnan pilotissa.

Avopalveluiden vastaanotto palveluiden kokeilupilotissa käytetty arkkitehtuuri on suuntana tavoiteltava, vaikka teknisesti sen toteuttaminen on vielä puutteellista ja arkkitehtuurillisesti epävakaa pohjalla. Tekniset puutteet ja arkkitehtuurin epä johdonmukaisuus johtuvat nykyisen ympäristön kankeudesta ja soveltumattomuudesta soite-tiedon tietolähteeksi Power BI raportointiin. Osaltaan myös organisoinnista johtuvat resurssien puutteet eri organisaatioyksiköissä aiheuttavat

pullonkauloja sote-tiedon parempaan hyödynnettävyyteen. Palvelualueilla ei ole resursoitu riittävästi osaavaa henkilöstöä tiedon analysointiin ja raportointiin, jotka toimisivat myös pääkäyttäjinä raporttien kehityksessä. Palvelualueilta puuttuvat nimetyt tietovastaavat, jotka tuntevat syvällisesti substanssitetiedon ja toimisivat laadun varmistajina sekä raporttien sisällön hyväksyjinä.

4.2 Tietovarastointi

Tampereen kaupungin tietovarastointi- ja raportointiympäristöstä ei ole olemassa juurikaan dokumentaatiota tai se on vanhentunutta ja puutteellista. Tämä vaikeuttaa nykytilan selvittämisessä tehtyjen päätösten jäljitettävyyttä. Vuosien varrella on tehty muutamia auditointeja (vuosina 2010 ja 2016) koskien tietovarasto- ja raportointiympäristöä, joissa on kuvattu sen hetkistä nykytilaa ja tehty ehdotuksia ympäristön kehittämiseksi. Perustuen olemassa olevaan dokumentaatioon, omiin havaintoihin ja näiden auditointien tuottamiin kuvauksiin, Tampereen kaupungin tietovarastointi- ja raportointiympäristön nykytilan arkkitehtuuri poikkeaa siitä, miten alkuperäisessä suunnitelmassa arkkitehtuuri oli kuvattu käytettävän.

Alkuperäisestä suunnitelmasta poiketen tietovaraston painopiste on vaihtunut EDM:stä SAP BW:hen, jolloin EDM:n rooli on nykyisin välitietovarasto, jonka kautta pääosa tiedosta tuotetaan SAP BW:hen ladattavaksi. EDM:n nykyroolia voisi kuvailla myös tilannekantana. EDM:ssä tehdään paljon tiedon muokkausta ja yhtenäistämistä tiedon latauksen eri vaiheissa. EDM-kannan mallinnus on tähti- ja lumihiihtalemallin sekoitus, osittain dimensionaalisen mallinnuksen piirteitä käsittävä tietokanta.

Johtuen juuri tietokantamallinnuksen epä johdonmukaisuudesta, EDM -välitietovaraston fakta- ja dimensiotaulukojen välillä ei ole käytössä vierasavaimia. EDM -välitietovarasto sisältää piirteitä Inmonilaisesta keskitetystä tietovarastosta, mutta ei kuitenkaan puhtaasti ole sen kaltainen. EDM -välitietovarastosta puuttuu esimerkiksi perustietojen historiointi, kaikki tapahtumatiedot eivät sisällä historiatietoa, kaikki tapahtumat eivät ole aikaan sidottuja, eikä taulujen välillä ole vierasavaimia. Samankaltaisuuksia keskitettyyn tietovarastoon puolestaan löytyy siinä,

että EDM -välitietovarastoon on keskitetty ja koottu tietoa useammasta eri lähdejärjestelmästä ja että tietoa on siellä joiltain osin yhtenäistetty. Tyypillisiä EDM -välitietovarastossa tehtäviä rikastamisia ovat esimerkiksi asiakkaan iän ja ikäluokan laskenta, hoitopäivien ja keskeytyspäivien lukumäärän laskenta ja puuttuvien päivämäärien täydentäminen. EDM -välitietovarastossa tehdään myös eri lähdejärjestelmistä tulevien asiakastietojen harmonisointi yhden asiakasnumeron taakse.

Keskitetyn tietovaraston piirteitä sisältävästä, tilannekannan kaltaisesta EDM -välitietovarastosta tietoa ladataan pääsääntöisesti materialisoitujen näkymien kautta SAP BW -tietovarastoon. SAP BW -tietovarastoa taas voisi luonnehtia pikemminkin kokoelmaksi monidimensionaalisia paikallistietovarastoja, koska tiedot ladataan SAP BW:hen kuutioihin, jotka sisältävät yksittäisen aihealueen tietoa. Eri kuutioiden tietoja on haasteellista yhdistää yhtenäiseksi kokonaisuudeksi, joskin kuutioiden tietojen liittäminen raportointia varten yhteen tietokokonaisuuteen on mahdollista. Myös SAP BW:ssä voidaan tehdä ja tehdään tiedon muokkausta ja rikastusta, esimerkiksi kuutioihin tyypillisesti lisätään tietuelaskuri niihin faktankaltaisiin tietoihin, joissa ei luontaisesti ole mitattavaa arvoa käytössä. Samoin asiakkaan iän laskentaa tehdään myös SAP BW:ssä, jos sitä ei jo valmiiksi ole EDM -välitietovarastossa laskettu.

Tietovarastointiin kuuluva historiatiedon tallentaminen on Tampereen kaupungin ympäristössä toteutettu osittain EDM:ään ja osittain SAP BW:n puolelle. Historioinnissa ei olla käytetty standardoitua tekniikkaa, kuten esimerkiksi Kimballin ja Casertan (2004, 183-193.) Slowly Changing Dimension tyyppejä 1-3. Dimensioiden lataus tehdään pääsääntöisesti päälle päivittävänä latauksena (SCD1), mutta osaa dimensioita ladataan SCD2 -tyyppisen kaltaisella tekniikalla SAP BW:hen sisäänrakennetulla tekniikalla. EDM:ssä historiointiin ei käytetä Slowly Changing Dimension -tekniikoita, vaan siellä dimension kaltaisten tietojen lataukset ovat päälle kirjoittavia. EDM:ään on tallennettu osittain historiatapahtumia erillisiin historianäkymiin, jotka ovat tarvittaessa ladattavissa SAP BW:hen erikseen manuaalisesti toteutettavilla latauksilla. SAP BW:n puolella tapahtumatiетojen historiointia on toteutettu kopioimalla kuutioita historiakuutioiksi, jolloin vanhat tapahtumatiedot jäävät käytettäviin erillisiin historiakuutioihin. Tietojen historiointista ei ole ollut suunnitelmaa tietovarastoarkkitehtuuria laatiessa, eikä siitä ole

olemassa standardia, miten historiointi toteutetaan ja missä. Historioinnin toimintatapa on muotoutunut vuosien aikana ja toteutusta on tehty niihin kuutioihin, missä sitä on eniten tarvittu. Historiointi ei kata kaikkea sote-tietoa, vaan pelkästään osaa ladattavista kuutioista.

Tampereen kaupungilta on puuttunut tietovarastointi- ja raportointiympäristön kehityksessä selkeä tietovarastostrategia sekä arkkitehtuurin että ylläpito-organisaation osalta, joka on aiheuttanut ympäristön pirstaloitumista ja kehityksen lyhytjänteisyyttä. Koska tietojen latausprosessi on pitkä, kaikki vaiheet eivät ole näkyvissä Tampereen kaupungin henkilöstölle ja dokumentointi on puutteellista, raportointiin tulevat tietopyynnöt ja selvityspyynnöt ovat hitaita ja työläitä toteuttaa. Ympäristön monikerroksisuudesta ja pirstaleisuudesta johtuen monet pienetkin muutokset saattavat olla työläitä ja hankalia toteuttaa. Organisaation organisoinnista johtuneet resurssipuutteet ovat osaltaan aiheuttaneet pitkiä viiveitä työpyyntöjen toteutukseen.

Tietovaraston mallinnukseen ei olla käytetty mitään mallinnustyökaluja, vaan tau-luja ja näkymiä on lisätty käyttäjien toiveiden mukaan. Mallinnustyökalun puuttu-minen on osaltaan johtanut dokumentoinnin puutteeseen ja pistemäiseen kehittä-miseen, koska tietovarastosta puuttuu luonnollinen paikka dokumentoida ympäristöä ja latausprosesseja. Tietokantaohjelmistona EDM -välitietovarastossa käytetään Oraclen relaatiotietokantaa. EDM -välitietovaraston tiedon poiminnat lähdejärjestelmistä ja lataukset on toteutettu useassa eri kerroksessa PL/SQL -proseduureilla, joita on ajastettu komentoriviskripteillä tietojen automaattisten la-tausten toteuttamiseksi. Tampereen kaupungin ratkaisussa ei ole käytetty erillistä ETL-työkalua, jolla tietolatauksia oltaisiin toteutettu, vaan kaikki tiedon muokkaus tapahtuu joko PL/SQL -proseduureilla tai SAP BW:ssä sen omilla tekniikoilla. Tampereen kaupungin tietojohdamisen puutteellisesta toimimisesta on seurannut kriittisen osaamisen ulkoistamista EDM -ylläpidon osalta, koska ulkoistuspäätök-sessä ei otettu riittävästi huomioon sisäistä toimintatapaa, osaamista ja kyvyk-kyyksien tuntemista.

Tietovaraston suunnittelussa ja toteutuksessa on korostunut operatiivisen, eli alimman tason tietotarpeet toiminnan seurantaan palvelevaksi ratkaisuksi. Tieto-varaston kehittäminen on perustunut yksittäisten käyttäjien tietopyyntöihin, jolloin

tietovaraston kokonaisuus on pirstaloitunut ja keskittynyt palvelemaan entistä pienempiä käyttäjäryhmiä. Selkeää kokonaiskehittämistä taktisen tai operatiivisen tason tarpeisiin ei olla tehty, jolloin kokonaiskuvan ja summatason tiedon saaminen tietovarastosta on haastavaa.

4.3 Liiketoimintatiedon hallinta

SAP BW -tietovarastoa voidaan pitää niin tietovarastointiin kuin liiketoimintatiedon hallintaan kuuluvana osana, mutta johtuen toimintaympäristöstä, Tampereen kaupungilla sen katsotaan kuuluvan enemmän tietovarastointiin kuin raportointiin. Varsinaiseen liiketoimintatiedon hallintaan tässä opinnäytetyössä katsotaan kuuluvaksi Tampereen kaupungin ympäristössä raportoinnin osuus tiedon hyödyntämisestä. Liiketoimintatiedon hallintaan kuuluu vakioraportit SAP Portaalista ajettuina, pääkäyttäjäraportit, tiedon analysointi SAP Analyzer -ohjelmalla, SAP BW:stä tuotettavien tiedostojen tietojen hyödyntäminen eri työkaluilla, automaattisesti käyttäjille toimitettavat raportit sekä sote-tiedon hyödyntäminen Power BI -ohjelmalla.

Pääosa SAP BW:n kautta tarjolla olevasta sote-tiedosta on tehty SAP BEx Query Designer -raporttikehittimellä valmisraporteiksi ja julkaistu raportit SAP Portaaliin käyttäjien ajettavaksi. Loppukäyttäjäraportit ovat käyttäjien käynnistämiä parametroituja valmisraportteja, joihin on tehty valmiiksi aihekohtaisia tietojen suodatuksia ja muokkauksia. Näitä suodatuksia käyttäjät eivät pysty muuttamaan, mutta käyttäjät voivat tehdä ajoparametreihin haluamiaan valintoja sekä voivat raportilla suodattaa haluamiaan tietoja raportille valitusta tietojoukosta. Käyttäjillä on mahdollisuus hyödyntää SAP BW kuutioiden tietoja valitsemalla haluamiaan ominaisuuksia taulukon riveille tai sarakkeille raportilla tarjolla olevista ominaisuuksista. Tämä monidimensionaalisen kuution mahdollistama toiminnallisuus ollaan koettu erityisen hyödylliseksi sote-tietojen ad-hoc -raportointiin, sillä loppukäyttäjätkin pystyvät analysoimaan raportilla tarjolla olevaa tietojoukkoa eri näkökulmista oman tarpeensa mukaan.

SAP Portaalista käynnistettävistä toimintaraporteista suurin osa on perusraportteja, jotka eivät sisällä asiakkaan henkilötietoja, vaan pelkästään tilastollista kappalet- ja lukumäärätietoa. Nämä raportit ovat käytettävissä perusraportoijan oikeuksilla. SAP Portaalissa on käytettävissä myös joitain pääkäyttäjäraportteja, jotka sisältävät toimintatietoa asiakaskohtaisella tasolla. Nämä pääkäyttäjäraportit ovat suunnattu etupäässä yksiköiden esimiehille ja toiminnan kehittäjille. Pääkäyttäjäraportteihin on omat käyttöoikeusroolinsa, jotka pitää erikseen anoa esimiehen toimesta. Pääkäyttäjäraportteilla tieto on yksityiskohtaisempaa ja yksiköiden esimiehet voivat tehdä tarkempia asiakaskohtaisia listauksia tilastointitarkoituksiin. SAP Portaalista käynnistettävien pääkäyttäjäraporttien lisäksi järjestelmästä on ajettavissa raportteja, joita käyttävät pääsääntöisesti edistyneet, tiedon raportointiin keskittyneet käyttäjät, jotka tarvitsevat vielä monipuolisempia ominaisuuksia raporteille. Tällaisia raportointitarpeita ovat esimerkiksi vuosittaisten tilastojen laatiminen, Kuusikko-työryhmien tarpeisiin tehtävä vuosittainen raportointi sekä yksikköjen johtajien tarpeisiin tehdyt erilaiset koosteraportit.

Käyttäjien itsensä ajamien raporttien lisäksi lähdejärjestelmien tietokannoista ja EDM-välitietovarastosta ajetaan joitain automaattisia raportteja lähetettäväksi käyttäjien sähköpostiin. Nämä raportit on tehty PL/SQL -proseduureilla ja ajastetuilla skripteillä, jotka poimivat lähdejärjestelmistä ja EDM -välitietovarastosta sellaisia listauksia, joiden tietoa ei SAP BW:hen saakka ole tuotu. Näiden käyttäjille automaattisesti toimitettavien listausten lisäksi lähdejärjestelmistä ja EDM -välitietovarastosta ajetaan ajoittain manuaalipoiminnoilla listauksia, jotka toimitetaan salattuna sähköpostina niitä pyytäneille käyttäjille. Tyypillisiä muutaman kerran vuodessa ajettavia listauksia ovat esimerkiksi erikoissairaanhoidon hoitotakuulistaukset, eri ikäluokkien terveystarkastuksia koskevat poiminnot, varusmiespalvelukseen astuvien terveystarkastuslistaukset sekä erilaiset satunnaiset tietopyynnot. Tämä raportointiosio on ongelmallinen liiketoimintatiedon hallinnan kannalta, sillä kaikki näihin raportteihin tehtävät muutokset sekä manuaaliset ajot on tilattava ulkoiselta toimittajalta, koska Tampereen kaupunki on ulkoistanut osa-alueen ylläpidon ulkopuoliselle toimittajalle.

Liiketoimintatiedon hallintaan kuuluu myös taktisen tason tiedon hyödyntäminen. Tämän alueen liiketoimintatiedon hallinta on hajautunut eri palveluyksikköihin yk-

sittäisille tiedon hyödyntäjille. Strategisen tason tietojohdamista varten on perustettu Tietojohdamisen Konsernipalveluyksikkö, joka keskittyy johdon tietotarpeiden palvelemiseen, mutta selkeää liiketoimintatiedon hallinnan ohjausta yksikkö ei toteuta. Taktisen ja strategisen tason tietotarpeisiin kohdennettua liiketoimintatiedon hallintaa sote-tiedon osalta palvelevat sama tietovarasto ja raportointiratkaisut, kuin operatiivista raportointia. Liiketoimintatiedon hallinta on näiltä osin erkaantunut operatiivisen ja taktisen sekä strategisen tason silloihin, sillä strategisella tasolla liiketoimintatiedon hallintaa toteutetaan irrallaan muusta tietovarastointi ja raportointiympäristöstä. Taktisen tason liiketoimintatiedon hallinta on satunnaista, tälle tasolle liiketoimintatiedon hallinta on lähinnä yksittäisten tietopakettien toimittamista yksittäisille tiedon hyödyntäjille nykyisestä SAP BW:stä.

Yksi haasteellisimmista raportointiin kohdistuvista tarpeista on erilaisten satunnaisten ja kertaluonteisesti toteutettavien tietopyyntöjen toteuttaminen nyky-ympäristössä, joka on tarkoitettu tuottamaan jatkuvaa tietoa operatiivista toimintaraportointia varten. Yksittäisiä ad-hoc -tietopyyntöjä varten SAP BW:ssä joudutaan usein rakentamaan uusia tietorakenteita ja kuutioita sekä tekemään SAP BEx Query Designer -raporttikehittimellä uusia raportteja, jotka palvelevat vain satunnaista ja yksittäistä tarvetta varten. Kehitys kuitenkin on tehtävä aina koko kehityspotken kautta, jolloin uudet rakenteet tehdään ensin kehitykseen, viedään siitä testiin testattavaksi ja siitä eteenpäin tuotantoon.

Nämä satunnaista tarvetta varten tehdyt tietorakenteet ja raportit jäävät kehitys-, testi- ja tuotantoympäristöihin senkin jälkeen, kun tarvittava tieto on toimitettu pyytäjille. Vuosien myötä näitä satunnaisia tietotarpeita varten tehdyt tietorakenteet ovat lisääntyneet ja nyky-ympäristöstä on vaikea enää todentaa, mitkä tietorakenteet ovat välttämättömiä perusraportoinnin kannalta ja mitkä ovat jääneet roikkumaan satunnaisten tietopyyntöjen takia. Tuotantoympäristössä on myös lukuisia testiraportteja ja kokeiluversioita varsinaisista tuotantoraporteista, ja näiden testiraporttien lukumäärä kasvaa jokaisen yksittäisen tietopyynnön toteutuksen jälkeen.

Raportointiympäristö on monimutkaistunut vuosien varrella raportointialueiden lisääntyneenä, joidenkin raportointialueiden poistuttua käytöstä sekä yksittäisten,

satunnaisten tietopyyntöjen myötä, jolloin sen hallittu ylläpito ja kehitys ovat käyneet haastavaksi. Selkeästi mallinnetun ja toteutetun tietovaraston sekä ad-hoc kyselyihin suunnitellun työkalun puuttuessa kehitys on johtanut SAP BW -tietovaraston tietorakenteiden ja raportointikerroksen raporttien hallitsemattomaan laajenemiseen.

Sote-tietojen ad-hoc analysoinnissa Power BI -työkalulla on nyky-ympäristössä haasteellinen tilanne, sillä käytettävissä oleva tieto on erkaantunut niin EDM -välitietovarastoon kuin SAP BW -ympäristöön. EDM tarjoaa yksinkertaisemman rajapinnan PBI Desktop tietomallille, koska relaatiotietokantaan on helppo tehdä tietolähdeyhteys sekä käyttää PBI Desktopin tarjoamia tiedon muokkaustoimintoja. EDM tarjoaa osittain enemmän tietoa, kuin mitä SAP BW:hen on tuotu, mutta ei kuitenkaan tarjoa kaikista tiedoista yhtä kattavaa historiatietoa kuin taas SAP BW:hen on tallennettu. Tämä aiheuttaa tilanteen, jossa PBI Desktopia pystytään hyödyntämään ad-hoc raportointityökaluna helposti vain niiden sote-tietojen osalta, mistä EDM:stä löytyy koko historian kattava tieto, tietopyyntöön ei tarvita kovin montaa ulottuvuutta ja datamäärät ovat kohtuulliset. Toisen rajoittavan asian PBI Desktopin hyödyntämisessä sote-tiedon ad-hoc -raportointiin tekee käytettävän työaseman muisti- ja suorituskyky. Suoraan EDM:ää tietolähteenä käyttävään ad-hoc -raportointiin voidaan käyttää 64-bittistä versiota PBI Desktopista, jolloin sen suorituskyky isommilla tietomäärillä on parempi kuin 32-bittisen version.

PBI Desktopissa ad-hoc -raportoinnin tietolähteenä käytettävä SAP BW on haastavampi, koska se vaatii toimiakseen työasemaan asennettavan erillisen tietolähdeajurin, eikä PBI Desktopista voida hyödyntää kuin 32-bittistä versiota. SAP BW on kuitenkin Microsoftin tukema tietolähde, vaikka sen käyttöön on asetettu lukuisia rajoituksia muihin tietolähteisiin verrattuna. Myöskään monia SAP BW:n toimintoja ja ominaisuuksia ei tueta Power BI:ssä, josta johtuen kaikissa tapauksissa Power BI:ssä esitetyt tulokset eivät vastaa SAP-työkalussa nähtäviä tuloksia. (Microsoft n.d.) Näistä Microsoftin asettamista rajoitteista sekä SAP BW:n kuutioiden sisältämän suuren tietomäärän vuoksi Power BI:llä on äärettömän haastavaa toteuttaa ad-hoc -raportointia sote-tiedoista.

Teoriassa SAP BW tietovarastoa voitaisiin pitää Power BI:n tietolähteenä sote-tiedon analysoinnissa, koska se on Microsoftin tukema tietolähde Power BI:lle. Käytännössä asia ei kuitenkaan ole näin yksinkertainen, sillä SAP BW:n tietovarasto sisältää kymmeniä eri sisältöisiä kuutioita, joiden tietosisältöä ei ole kuvattu missään metatietokerroksessa. Käyttäjien olisi tiedettävä, mistä kuutioista tietoa lataavat ja mitä tietoa ne sisältävät. Nykytilanteessa käyttäjät käyttävät valmiiksi toteutettuja standardiraportteja, joissa tieto esitetään ymmärrettävissä kokonaisuuksissa ja selkokielisillä kenttänimillä. SAP:in lisenssointipolitiikka asettaa myös rajoitteita SAP BW:n tietojen käytölle käytettäessä SAP BW:tä Power BI:n tietolähteenä.

Toinen haaste on teknologinen, SAP BW vaatii tietolähteenä toimiakseen 32-bittisen version PBI Desktopista, kun normaalikäytössä Microsoft suosittelee 64-bittisen version käyttämistä suorituskykyisempänä ohjelmana. 32-Bittinen PBI Desktop ei ole niin suorituskykyinen kuin 64-bittinen versio, eikä käyttäjien henkilökohtaisilla tietokoneilla ole välttämättä riittävästi suorituskykyä eikä muistia suurten sote-tietomallien analysointiin. Kolmannen haasteen tuo kaupungin it-ympäristö, jolloin käyttäjä ei pääse itse asentamaan tarvittavia ohjelmistoja ja ajureita tietokoneelleen, vaan käyttäjän on osattava pyytää oikeita ohjelmistoja ja ajureita tukipalvelun asennettavaksi.

4.4 Resurssien organisointi

Tietovarastointi- ja raportointiympäristön kehityksen alkuvaiheessa Tampereen kaupungilla oli omassa organisaatiossa henkilöt, jotka suunnittelivat, toteuttivat ja ylläpitivät koko tietovarastointi- ja raportointiympäristöä. Ulkoiset toimittajat toimivat projekteissa teknisessä roolissa arkkitehtuurin suunnittelijoina ja osittain myös toteuttivat ratkaisuja. Vuosien myötä organisaatio ja henkilöt ovat vaihtuneet ja etenkin EDM:ään liittyvät kriittiset tekijät on ulkoistettu Tampereen kaupungin ulkopuolelle, josta EDM ylläpito ja pienkehitys sittemmin on hankittu. Tämän työn ulkoistus kaupungin ulkopuolelle ja myöhemmin kilpailutus ja ylläpidon siirto toisen palvelutuottajan hoidettavaksi ovat aiheuttaneet osaamisen katoamista kaupungin sisältä ja tuonut kriittisen riippuvuuden ulkopuolisilta toimittajilta ostettavasta työpanoksesta liittyen EDM:n ylläpitoon ja pienkehitykseen. EDM:n

toteutus ja kehitysmalli on perustunut yksittäisten henkilöiden syvälliseen osaamiseen ympäristöstä, niin substanssista kuin teknologiasta, eikä dokumentointiin olla panostettu juurikaan. Ulkoistusten myötä osaamista on kadonnut niin ympäristön tuntemisen kuin substanssin osalta, jolloin nykyisin ulkoa joudutaan ostamaan paljon puhtaasti selvitystyötä, miten ympäristö toimii. Tampereen kaupungilla ei ole pääsyä eikä osaamista EDM -kyselyihin tai poimintaohjelmiin, joten tiedon jäljitettävyyteen joudutaan ostamaan kaupungin ulkopuolista työtä.

Tampereen kaupungilla tekninen tietovarastointi- ja raportointiosaaminen on nykyisin keskitetty tietohallinnon yhteen tiimiin niin ylläpidon kuin pienkehityksenkin osalta. Palvelualueilla on raporttien peruskäyttäjiä ja muutamassa yksikössä on yksittäisiä pääkäyttäjätyyppejä henkilöitä, jotka pystyvät hieman laajemmin tekemään pääkäyttäjäraportointia, mutta eivät kuitenkaan raporttikehitystä. Yhdellä palvelualueella on taktisen tason tiedon hyödyntämiseen erikoistunut henkilö, jolla on käytössään tiedon analysointiin tarkoitettu työkalu, mutta sote-tiedon osalta tieto toimitetaan tiedostoina SAP BW:stä. Strategiayksikössä, johon tietojohdamisen tiimi kuuluu, on myös sote-tiedon hyödyntämiseen erikoistunut henkilö, joka pystyy toteuttamaan myös yksittäisiä sote-tietoihin kohdistuvia tietopyyntöjä strategisen tason tietotarpeisiin. Strategiayksikössä on käytettävissä samat työkalut ja ympäristöt kuin muuallakin, eli pyynnöt toteutetaan SAP BW:n mahdollistamilla ratkaisuilla.

Substanssiosaaminen on keskittynyt palvelualueiden yksiköihin, mutta näissä yksiköissä ei kuitenkaan ole organisoitu tietovastaavia, joiden vastuulla olisi liiketoimintatiedon laadun ja oikeellisuuden hallinta. Palvelualueilla ei myöskään ole juuri teknistä osaamista tai kyvykkyyttä toteuttaa itse raportteja tai toimia raporttien teknisenä määrittelijänä nyky-ympäristössä. Siiloutuneesta ja eri toimintoihin erikoistuneista organisaatioista johtuen raportointitarpeiden määrittelyyn tarvitaan monesti henkilöitä useasta eri organisaatioyksiköstä sekä monesti myös ulkopuolisilta toimittajilta. Isommissa kehityshankkeissa, mutta väillä myös yksittäisten pyyntöjen osalta, toteutustyötä joudutaan ostamaan toimittajilta, koska kaupungilla ei ole riittävästi resursseja toteutustyöhön.

Nyky-organisaatiossa ei ole selkeästi määriteltyä vastuutahoa, jonka vastuulla olisi koko kaupungin tietovarastointistrategian ja arkkitehtuurin suunnittelu ja toteuttaminen niin strategisen, taktisen kuin operatiivisen tason tarpeita vastaavaksi. Kaupungilla tehdään kokonaisarkkitehtuurityötä, jonka yhteyteen olisi luonnollista yhdistää myös tietovarastointi- ja raportointiympäristön kuvauksen sekä arkkitehtuurin kehittämisen. Nykyinen organisaatio on jakautunut tiedon hyödyntämisessä strategisen tason tiedon hyödyntämiseen ja operatiiviseen raportointiin, mutta arkkitehtuuria ollaan toteutettu operatiivisen raportoinnin tarpeisiin.

5 TUTKIMUSTYÖN TULOKSET

5.1 Tietovarastointi ja raportointiympäristö

Kirjallisuuskatsauksessa tunnistettiin tietovarastoinnista kolme erityyppistä arkkitehtuuria, joilla jokaisella on erityispiirteensä:

- Yritystasoinen keskitetty tietovarasto
- Yhdenmukaistetut paikallistietovarastot
- Itsenäiset paikallistietovarastot

Näiden kolmen alaluvussa 2.2.2 Tietovarastojen arkkitehtuurit tarkemmin esitellyjen eri arkkitehtuurin lisäksi on yleistä hybridiarkkitehtuuri, jossa yrityksen tietovarasto on toteutettu soveltamalla usean eri arkkitehtuurin piirteitä. Yrityksessä toteutettavan tietovaraston arkkitehtuuriin vaikuttaa suunnitteluvaiheessa käyttäjien tarpeet, tietovaraston laajuus ja saatavissa olevat tiedot. Tutkimuksessa ilmeni, että kaikki tunnistetut tietovarastoarkkitehtuurit ovat käyttökelpoisia, mutta yrityksen pitää valita arkkitehtuuri omaan toimintaympäristöönsä, organisaatiokulttuuriin ja tietojohdamisen tarpeisiinsa sopivaksi. Kaikki toimialat ja yritykset ovat erilaisia ja niissä on erilainen toimintamalli sekä tietojohdamisen kyvykkyys, joten tietovarastoarkkitehtuurin valinta riippuu yrityksen omista lähtökohdista ja tarpeista.

Kirjallisuuskatsauksessa ei tunnistettu SAP BW:tä tietovarastoarkkitehtuurina eikä siitä löytynyt juuri mainintoja tietovarastoinnin yhteydessä. Tietovarastoinnin osalta ainoa aihetta käsittelevä materiaali löytyi Kimballin ja Casertan (2004, 103-105) kirjasta, jossa käsiteltiin SAP BW:n soveltuvuutta yritystasoisesta keskitetyn tietovaraston ja yhdenmukaistettujen paikallistietovarastojen arkkitehtuuriin. Molemmilla tapauksella Kimball ja Caserta tulivat tulokseen, että eivät suosittelut SAP BW:tä käytettävän kumpaankaan arkkitehtuuriin. (Kimball & Caserta 2004, 104)

Omat havainnot tukevat tätä Kimballin ja Casertan (Kimball & Caserta 2004, 104) suositusta, sillä tietojen yhdistely SAP BW:n omilla työkaluilla sekä SAP BW:n

sisällä olevien tietojen välillä, että varsinkin ulkopuolisten tietojen välillä on työstä ja kankeaa. Joustavampi ympäristö satunnaisille ad-hoc -tietopyynnöille olisi relaatiopohjainen tietovarasto, josta pystyisi esimerkiksi SQL-kielellä yhdistelemään tarvittavat tiedot. Toinen vaihtoehto SQL -kielelle tietojen poimintaan olisi esimerkiksi joku liiketoimintatiedon hallintaan erikoistunut kyselytyökalu.

Tampereen kaupungin tietovarastoarkkitehtuuri poikkeaa kirjallisuuskatsauksessa tunnistetuista arkkitehtuureista niin arkkitehtuurin kuin teknologiain osalta. Tietovarastoarkkitehtuurin epäjohdonmukaisuus ja hajautuminen kahteen eri teknologiaan: SAP BW ja EDM, ovat johtaneet haasteelliseen tilanteeseen liiketoimintatiedon hallinnan sekä tietojohdamisen osalta. Samoin tietovarastoarkkitehtuurin painottuminen operatiivisen tason tietotarpeita palvelevaksi on aiheuttanut taktisen ja strategisen tason tietotarpeiden toteuttamisen haastavaksi.

Tampereen kaupungilla on toteutettu tietovaraston ja raportoinnin auditointeja ulkopuolisten konsulttien toteuttamana. Auditoinneissa tuotiin esiin samoja ongelmia ja samansuuntaisia ratkaisuehdotuksia. Auditointien välillä oli kuitenkin kulunut useita vuosia aikaa, mutta arkkitehtuurin kehitystä ei juuri ollut tapahtunut, eikä niitä toimenpiteitä, mitä auditoinneissa ehdotettiin, oltu saatu vietyä eteenpäin. Tilanne on pysynyt kehityksen suhteen samankaltaisena viime auditoinnin jälkeen tähän päivään saakka.

5.2 Tietojohdaminen

Kirjallisuuskatsauksessa tunnistettiin, että tietojohdamiseen sisältyy tiedon teknisen tallentamisen näkökulman lisäksi ilmiön ymmärtämisen ja johtamisen käytäntöjen näkökulmat. Tietojohdaminen Tampereen kaupungilla on kehittynyt vuosien aikana ja tiedon arvon on ymmärretty olevan olennainen osa tulevaisuuden toimintatapojen suunnittelussa. Tietojohdaminen on kehittynyt tiedon seurannasta tiedon hyödyntämiseen johtamisen käytäntöjen näkökulmasta strategisella tasolla. Samaan aikaan kuitenkin tiedon saatavuuden arkkitehtuuri on jatkanut toimintaansa operatiiviseen raportointiin suuntautuneena palveluna.

Tampereen kaupungilla on tehty tiedolla johtamisen kehittämisselvitys 10/2012 - 01/2013 ulkopuolisen konsultin toimesta. Selvityksen tavoitteena oli mahdollistaa tiedolla johtamisen tehokas tukeminen ja kehittäminen oikeilla toimenpiteillä. Selvityksen teon jälkeen tietojohdantien Tampereen kaupungilla on kehittynyt, esimerkiksi sitä varten on perustettu oma tiiminsä Strategiayksikköön. Tämä kehityssuunta on ollut hyvä, vaikka tekemistä vielä riittää tietojohdantien jalkauttamisen suhteen. Tietojohdantien kehittyminen on näkynyt esimerkiksi siinä, että kaupunki on ottanut käyttöönsä kaikkia projekteja koskevan projektimallin, projektisalkkumallin ja ohjelmajohtamisen. Projektisalkkumallilla voidaan paremmin hallinnoida organisaatiossa meneillään olevia projekteja ja hankkeita sekä ylläpitää niiden välistä kommunikaatiota. Tietojohdantien kokonaisarkkitehtuuriin liittyviä hankkeita ja projekteja hallinnoidaan projektien ja projektisalkkujen avulla, mutta selkeät roadmapit ja tulevaisuuden tavoitetilat kaipaisivat vielä kovasti työpanosta.

Organisatorisesti tietojohdantia toteutetaan keskitetysti Konserniyksiköissä niin Strategia- ja kehittämissyksikössä kuin Tietohallinnossa sekä hajautettuna palvelualueilla. Tampereella ei olla määritelty selkeää tietojohdantien visiota eikä suunniteltu keinoja vision toteuttamiseen. Vision puute konkretisoituu monilla tavoilla, mm. päällekkäisenä ja rinnakkaisena työkalupakkina ja kyseenalaisina arkkitehtuuriratkaisuuina. Tietojohdantien on nykyisin keskittynyt tuottamaan taakse päin katsovaa tietoa toiminnan seurantaan, kun suuntana pitäisi olla eteenpäin ennakoivan tiedon tuottaminen ja hyödyntäminen.

Yhtenä syynä tietojohdantien puutteelliseen toimintaan näen hallinnollisen organisaatiomallin, joka on johtanut toimimaan siiloutuneesti ja edistämään oman organisaation sisäistä tavoitetta. Hallinto-organisaatiot ovat muuttuneet vuosien myötä useaan otteeseen, mutta vasta viimeisimmässä organisaatiomuutoksessa on edellytykset kokonaisvaltaisen arkkitehtuurin edistämiseksi. Onnistunut tietojohdantien toteuttaminen vaatii tiivistä keskustelua palvelualueiden ylimmän johdon sekä keskijohdon kanssa, jotta palvelualueiden tietotarpeet saataisiin näkyviin. Sama keskustelu pitäisi pysyä tiiviinä myös operatiivisten tietotarpeiden ylläpitämiseksi ja toteuttamiseksi. Näiden kaikkien kolmen tason tietotarpeiden perusteella tietojohdantien työkaluja ja arkkitehtuuria voidaan alkaa kehittää ja

luoda tarkoituksenmukainen arkkitehtuuri kaikkien tasojen tietotarpeiden tyydyttämiseksi.

5.3 Microsoft Power BI:n hyödyntäminen tiedon analysoinnissa

Kirjallisuuskatsauksessa tunnistettiin Power BI arkkitehtuurin kolme eri toimintamallia:

- Keskitetty BI -raportointi
- Itsepalveluvisualisointi
- Itsepalveluraportointi

Yrityksessä tehdyt strategiapäätökset, organisaatiomalli sekä liiketoimintatiedon hallinnan hankkeiden rahoitusmalli vaikuttavat valittavaan Power BI toimintamalliin. Samoilla taustatekijöillä on myös vaikutus tunnistettuihin Power BI käytön rooleihin:

- Suunnittelija
- Kehittäjä
- Kuluttaja eli loppukäyttäjä
- Järjestelmänvalvoja

Kirjallisuuskatsauksessa ilmeni, että Power BI:n käyttöönotto vaatii organisaatiolta pohdintaa ja perusteltuja päätöksiä toimintamallin valinnassa sekä kyvykkyyttä hoitaa Power BI käytön eri roolien vastuita. Tämä tuli esiin sekä tietovarastointia että Power BI:tä koskevan aineiston yhteydessä ja molemmissa painotettiin, että organisaatiossa pitää olla kyvykkyyttä eri roolien onnistuneeseen toimintaan.

Tehdyn sähköisen Power BI -kyselyn perusteella ei voi vetää johtopäätöksiä yleisimmistä käyttöympäristöistä, sillä kyselyyn tuli vastausaikana vain yksi vastaus. Tämä ei edusta kuin yhden käyttöympäristön tilaa, jolloin yleistyksiä ei tästä voida tehdä. Kyselyyn tulleiden vastausten vähäisestä lukumäärästä voinee vain vetää

löyhän päätelmän, että yrityksissä olevat Power BI -kehittäjät ovat niin työllistettyjä omissa töissään, ettei heillä riitä aikaa vastata ylimääräisiin kyselyihin.

Tampereen kaupungilla Microsoft Power BI:n käyttöönottopäätöksen jälkeen raportointiväline on otettu käyttöön erilaisiin käyttötarpeisiin eri yksiköissä ja käyttäjille on tarjottu koulutusta välineen käyttöönotossa. Nykyinen malli perustuu täysin käyttäjien itsepalveluraportointiin, vaikka nykyarkkitehtuuri ei tue sote-tietojen helppoa saatavuutta. Power BI:n käyttöönotossa ilman selkeää strategiaan perustuvaa käyttöönottosuunnitelmaa on kuitenkin vaarana työkalun jääminen yksittäiseksi irralliseksi ohjelmistoksi vailla organisaation tukea sen käyttöön. Power BI:n käyttöönotto vaatii niin organisaatiolta osaamista ja kypsyyttä hyödyntää välinettä kuin teknologiselta alustalta valmiutta tuottaa tietoa sellaisessa muodossa, missä sitä pystytään helposti hyödyntämään.

Täysin itsepalveluraportointiin siirtymisessä organisaation pitää olla riittävän kyvykäs ottamaan vastuulleen myös teknistä raporttikehitystä palvelualueille, mutta alustan ja arkkitehtuurin suunnittelu, toteutus ja kehitys on oltava keskitettynä jossain organisaation yksikössä, yleensä tietohallinnossa. Onnistuneen tietovarastointi- ja raportointiympäristön toteuttaminen ja eritoten sen ylläpito vaatii organisaatiossa usean eri tyyppisen osaamisen ja roolin sitoutunutta ja jatkuvaa työtä projektien jälkeenkin. Ylläpitoa ja kehitystä ei voida jättää vain IT-organisaation huoleksi, vaan kehityksen on oltava pitkäjänteistä ja johdonmukaista myös palvelualueiden puolelta.

Tampereen kaupungilla ei ole tehty Power BI:n käyttöönottopäätöksen jälkeen linjauksia, minkälainen toimintamalli otetaan käyttöön tai miten eri roolien organisointi nykyisessä organisaatiossa jalkautetaan. Nykyarkkitehtuurin puutteellisuuden vuoksi Power BI:n hyödyntäminen sote-tiedon analysoinnissa on satunnaista, eikä tarjoa varteenotettavaa työkalua yksittäisten ad-hoc -tietotarpeiden toteuttamiseen. Power BI:tä ei voida myöskään ottaa hallitusti täysimittaisesti käyttöön laajemmin sote-tiedon hyödyntämisen millään tasolla, koska nykyarkkitehtuuri ja -teknologia eivät tue tiedon hyödyntämistä Power BI -työkalulla.

Power BI:llä olisi kuitenkin potentiaalinen mahdollisuus toimia modernina raportointi- ja visualisointityökaluna sote-tietojen raportoinnissa, kunhan alusta-arkkitehtuuri suunniteltaisiin ja toteutettaisiin mahdollistamaan datan hyödyntäminen tällä välineellä. Kaupungilla pitäisi kehittää sekä arkkitehtuurin, teknologian että organisaation kyvykkyyteen liittyviä toimintoja, jotta tähän tavoitteeseen päästäisiin. Ensimmäinen työvaihe voisi olla tietovarastostrategian laatiminen ja Power BI käytötapauksen tunnistaminen tiedon hyödyntämisen eri tasoille. Power BI:n toimintamallit ja arkkitehtuurit pitäisi olla valittu tietojen käyttötarkoituksen ja -tarpeen mukaan, jolloin eri tiedon hyödyntämisen tasoille voisi olla määritelty erilaiset toimintamallit ja arkkitehtuurit. Esimerkiksi strategisen tason tiedon hyödyntämisessä voitaisiin arkkitehtuurillisesti siirtyä pilviteknologiaan ainakin julkaisun osalta, sillä tämän tason tieto ei enää sisällä yksilöiviä asiakastietoja, vaan on karkeammalla tasolla ja summatumpaa.

Yksi tapa edistää Power BI:n työkalun hallittua käyttöönottoa on kehittää Tampereen kaupungin kokonaisarkkitehtuuria kolmesta eri näkökulmasta: toiminta-, tieto- ja tietojärjestelmäarkkitehtuurit. Tällöin huomioitaisiin myös muut tietotarpeet kuin sote-tieto ja kehitettäisiin toiminta- ja tietojärjestelmäarkkitehtuurista kaikkia palvelualueita palveleva yhtenäinen arkkitehtuuri, jonka avulla pystyittäisiin vastaamaan myös tulevaisuuden haasteisiin tietojen analysoinnin saralla.

5.4 Kehitysehdotukset

Taulukkoon 4 on koottu kehityskohteita ja ehdotuksia Tampereen kaupungin tietovarastoinnin ja raportoinnin nykytilan kehittämiseen ja parantamiseen sekä Power BI -työkalun käytön täysimittaiseen hyödyntämiseen. Ehdotetut kehityskohdet eivät ole toteutus- tai tärkeysjärjestyksessä, vaan jokainen ehdotus vaatii vielä syvällisempää jatkojalostamista, määrittelyä ja priorisointia ennen niiden toteuttamista. Ehdotukset kehittämiskohteista on valittu tehtyyn taustatutkimukseen, nykytilan analysoinnissa esiin nousseisiin havaintoihin ja tutkijan omakohtaiseen kokemukseen perustuen.

Kehitysehdotusten jatkokehittäminen vaatii resursseja useasta kaupungin yksiköstä ja usean eri tahon yhteistyötä. Kehitysehdotusten toteuttaminen joidenkin

kohteiden osalta vaatii myös merkittävää rahallista panosta, jolloin niiden toteutus on arvioitava niitä käsittelevissä ohjausryhmissä ja projektisalkkuelimissä. Kehitysehdotuksia voidaan käyttää myös pohjana tietovarastoinnin ja raportoinnin tavoitetilan suunnitteluun ja tiekartan suunnittelun pohjana.

TAULUKKO 4. Kehitysehdotuksia tietovarastoinnin ja raportoinnin nykytilan kehittämiseen.

Kehityskohde	Alue	Miksi, mitä ja miten?	Huomioitavaa
Nykytilan tietovarastoarkkitehtuurin ja -teknologian selkeyttäminen	Tietovarasto	Nykytilan selkiyttämällä ja prosessien kuvaamisella jäädytetään nykytila ennalleen ja estetään sen hajaantuminen ennistä epäselvemmäksi. Tietovarastoarkkitehtuuri kuvataan ja eri toimintoille standardoidaan toimintaperiaatteet ja roolit, mitä asioita tehdään missäkin toiminnossa.	KA: Voidaan käyttää kuvaamisen apuna JHS 179 liitettä 3: Arkkitehtuurin nykytilan ja tavoitetilan kuvaaminen
Tavoitetilan arkkitehtuurin kuvaaminen kaikille tiedon tasoille	Tietojohdaminen	Tavoitetilan arkkitehtuuri kuvataan jokaiselle tiedon hyödyntämisen tasoille erikseen, jotta kaikkiin tietotarpeisiin on olemassa arkkitehtuuri, miten tietoa tuotetaan ja hyödynnetään. Kartoitetaan ensin tietotarpeet jokaiselle tasolle erikseen, kartoitetaan käytettävissä olevat tietolähteet, kartoitetaan käytettävissä olevat kyvykkyydet ja rakennetaan arkkitehtuuri näiden pohjalta.	KA: Voidaan käyttää kuvaamisen apuna JHS 179 liitettä 3: Arkkitehtuurin nykytilan ja tavoitetilan kuvaaminen
Tietovarastostrategian tekeminen	Tietojohdaminen	Tietovarastostrategian tulisi olla yhteydessä tietohallintostrategiaan sekä liiketoimintastrategiaan ja kuvata, mitä tietovarastoinnilla halutaan saavuttaa ja mitä tarpeita sillä toivotaan täytettävän.	KA: Voidaan käyttää kuvaamisen apuna soveltuvin osin JHS 179 liitettä 1: Strategian kuvaaminen strategiakartan avulla
Tavoitetilan resurssien organisoinnin suunnittelu	Tietojohdaminen	Tietovaraston ja raportoinnin ylläpito vaatii toimiakseen riittävästi jatkuvaa, asiantuntevaa resurssointia. Pelkkä ylläpito ei kuitenkaan riitä, vaan on oltava kyvykäs, osaava resurssi myös tietovaraston kehitykseen ja arkkitehtuurin jatkuvaan kehittämiseen tietotarpeita vastaavaksi. Kehitys pohjautuu tietovarastostrategiaan, joten sen yhteydessä on myös tehtävä suunnitelmat työn organisoimiseksi niin ylläpidon, kehityksen kuin arkkitehtuurin ja teknologian toteuttamiseksi.	KA: Voidaan käyttää kuvaamisen apuna soveltuvin osin JHS 179 liitettä 2: Liiketoimintamallit ja kyvykkyydet KA-suunnittelussa

Tiedon hyödyntäjien roolien selkeyttäminen eri tasoilla	Tietojohtaminen	Tiedon hyödyntäjien roolien selkeyttäminen jäsentää tiedon hyödyntämisen eri tasoja. Tällöin pystytään paremmin suunnittelemaan arkkitehtuuria ja teknologiaa vastaamaan paremmin eri roolien tarpeisiin hyödyntää tietoa.	KA: Voidaan käyttää kuvaamisen apuna soveltuvin osin JHS 179 liitettä 2: Liiketoimintamallit ja kyvykkydet KA-suunnittelussa
Tietovaraston kehityksen standardien kuvaaminen	Tietovarasto	Tietovaraston kehityksen standardointi edistää tietovarasto-arkkitehtuurin ylläpidettävyyttä ja hallittua kehitystä. Luomalla standardit, joiden mukaan kehitystä tehdään, selkeyttää ja jäsentää tietovarastostrategian toteutusta.	Voidaan hyödyntää esimerkiksi Kimballin (2008) hallintamallia kirjassa: The data warehouse lifecycle toolkit (kappaleet: Chapter 13: Deploying and Supporting the DW/BI System ja Chapter 14: Expanding the DW/BI System)
Palvelualueiden tietovastaavien perustaminen	Liiketoimintatiedon hallinta	Tiedon laatua ja johdonmukaista hyödyntämistä varten palvelualueilla tulisi nimetä tätä tehtävää hoitavat henkilöt. Palvelualueille tulisi vastuuttaa tiedon hyödyntämiseen ja kehittämiseen liittyviä toimia.	KA: Voidaan käyttää kuvaamisen apuna soveltuvin osin JHS 179 liitettä 2: Liiketoimintamallit ja kyvykkydet KA-suunnittelussa
Power BI toimintamalli	Power BI	Määrittelemällä käyttöönotettavan Power BI toimintamallin pystytään hallitusti ylläpitämään ja kehittämään Power BI -raportointia tiedon hyödyntämisen eri tasoilla.	Määrittelyssä voidaan hyödyntää esimerkiksi Knight ym. (2018) oppeja kirjassa: Microsoft Power BI Complete Reference (kappale: Planning Power BI Projects)
Power BI käytön roolit	Power BI	Määrittelemällä käyttöönotettavan Power BI käytön roolit pystytään hallitusti ylläpitämään ja kehittämään Power BI -raportointia tiedon hyödyntämisen eri tasoilla.	Määrittelyssä voidaan hyödyntää esimerkiksi Knight ym. (2018) oppeja kirjassa: Microsoft Power BI Complete Reference (kappale: Planning Power BI Projects)
Raportointivälineiden valinta tiedon hyödyntämisen eri tasoille sekä eri rooleihin	Liiketoimintatiedon hallinta	Yhdenmukaiset ja määritellyt raportointivälineet eri roolien tarpeiden mukaan selkeyttävät toimintaa. Määrittämällä järjestelmät ja työkalut toimenkuvien ja roolien perusteella välitetään pistemäisten ja tukemattomien ratkaisujen käyttöön otto yksiköissä.	KA: Voidaan käyttää kuvaamiseen apuna soveltuvin osin JHS 179 liitettä 5: KA-taulukot ja liitettä 6: KA-kuvausten visualisointi. Suunnittelussa voidaan hyödyntää esimerkiksi Kimballin (2008) oppeja kirjassa: The data warehouse lifecycle toolkit (kappale: Chapter 11: Introducing Business Intelligence Applications)

Tietojohtamisen vision määrittely	Tietojohtaminen	Selkeän vision määrittely antaa suunnan, minkälaista toimintamallia kohden tulisi edetä. Vision tehtävänä on varmistaa, että suunnittelun kohde ja suunta ovat kaiken aikaa selvät.	KA: Voidaan käyttää kuvaamisen apuna soveltuvin osin JHS 179 liitettä 1: Strategian kuvaaminen strategiakartan avulla
Tietovaraston jatkuvan kehityksen hallintamallin luonti	Liiketoimintatiedon hallinta	Tietovarastoa pitää kehittää jatkuvasti niin tietosisällöllisesti kuin arkkitehtuurillisesti. Hallitun ylläpitomallin määrittely ja johdonmukainen noudattaminen edesauttavat tietovaraston ylläpidettävyyttä, kehitystä ja tietosisällön laatua.	Voidaan hyödyntää esimerkiksi Kimballin (2008) hallintamallia kirjassa: The data warehouse lifecycle toolkit (kappaleet: Chapter 13: Deploying and Supporting the DW/BI System ja Chapter 14: Expanding the DW/BI System)
Tiedon lataamisen polun lyhentäminen ja muokkauksen tasojen standardointi	Tietovarasto	Nykyinen tiedon polku on liian pitkä ja mutkikas lähdejärjestelmistä loppuraporteille asti. Määrittelemällä arkkitehtuuri tavoitteenmukaiseksi, lyhennetään tiedon latauksessa tehtävien vaiheiden määrää ja kuvataan niiden eri käyttötarkoitukset.	Voidaan hyödyntää esimerkiksi Kimballin (2008) oppeja kirjassa: The data warehouse lifecycle toolkit (kappaleet: Chapter 4: Introducing the Technical Architecture ja Chapter 5: Creating the Architecture Plan and Selecting Products)
Roolijaon selkeyttäminen sisäiseen työhön ja toimittajilta ostettavaan työhön	Tietovarasto, Liiketoimintatiedon hallinta	Nykyinen toimintaympäristö on epäselvä, mitä toteutetaan itse ja mitä ostetaan ulkoa. Selkiyttämällä roolijakoa, pystytään keskittymään oman osaamisen kasvattamiseen valituissa toiminnoissa.	KA: Voidaan käyttää kuvaamisen apuna soveltuvin osin JHS 179 liitettä 2: Liiketoimintamallit ja kyvykkydet KA-suunnittelussa
Masterdatan määrittely keskeisten toimintojen osalta	Liiketoimintatiedon hallinta	Masterdatan määrittely edesauttaa jäsentämään eri tietolähteiden välisiä yhteyksiä ja määrittämään perustiedot keskeisistä toiminnoista. Määritetään keskeisten toimintojen osalta pää tietolähteet ja attribuutit, mitä tietoa tietovarastoon kerätään.	KA: Voidaan käyttää kuvaamiseen apuna soveltuvin osin (tietoarkkitehtuuri) JHS 179 liitettä 5: KA- taulukot ja liitettä 6: KA-kuvausten visualisointi.

6 JOHTOPÄÄTÖKSET JA POHDINTA

Teoreettisella ja yleisellä tasolla tietovarastointi ja liiketoimintatiedon hallinta on hyvin suoraviivainen prosessi, lähdejärjestelmistä poimitaan tietoa, sitä muokataan ja yhtenäistetään ja ladataan tietovarastoon ja lopuksi käyttäjät hyödyntävät tietoa BI-välineiden avulla. Toimivan ratkaisun toteuttamiseen ei kuitenkaan ole oikotietä tai hopealuotia, jonka avulla voitaisiin ratkaista kaikki vaiheet ilman systemaattista ja raakaa työtä. Jokainen vaihe vaatii omanlaista osaamista sekä sitoutuneen organisaation sen toteuttamiseksi, muuten ajaututaan tekemään pistemäisiä ratkaisuja, jotka eivät välttämättä toimi kokonaisuuden kannalta parhaalla mahdollisella tavalla. Organisaatioiden siirtyessä tapahtuneiden asioiden seurannasta ja takapeiliin katsomisesta entistä enemmän katsomaan tulevaisuuteen ja ennakoimaan asioita, tarvitaan organisaatioihin uudenlaista kyvykkyyttä hallitsemaan tiedon jalostuksen prosessia ja ymmärrystä eri valintojen merkityksen kokonaisuuden kannalta.

Varsinkin kuntaorganisaatioissa tietojohtaminen ja liiketoimintatiedon hallinta ovat suhteellisen nuoria oppeja, eikä julkisorganisaatioissa ole aiemmin ollut markkinatalouden kilpailun tuottamaa painetta seurata ja kehittää toimintoja kustannustehokkaammaksi. Suurten megatrendien aiheuttamat globaalit muutokset on yksi isossa mittakaavassa Suomeakin ja kuntia koskettava kehityssuunta, jonka vaikutukset on huomioitava ennakolta kuntien päätöksenteossa ja suunnittelussa. Toinen nyky-yhteiskunnan toimintaa nopeasti muuttava ilmiö on digitalisaatio, johon kuntien on vastattava yhä kiihtyvällä tahdilla. Kolmas kaikkia Suomen kuntia koskettava rakenteellinen muutos on viime vuosina ollut hallituksen reformi sosiaali- ja terveydenhuollon uudistamiseksi sekä maakuntauudistus.

Nämä kaikki luovat kunnille sellaisia vaateita, joihin ei perinteisillä menetelmillä pystytä vastaamaan. Kilpailutuksen ja valinnanvapauden aiheuttama tarve tehostaa toimintaa tuo enenevässä määrin tarpeen pystyä reaaliajassa seuraamaan organisaation toimintaa niin suorite- kuin talousmittareilla. Ilman jalostunutta liiketoimintatiedon hallintamenetelmää tähän ei monestikaan ole mahdollisuuksia varsinkaan suurissa organisaatioissa.

Tutkimuksessa tuli ilmi, että Tampereen kaupungin tietovarastointi- ja raportointiympäristö koostuu useasta eri osasta ja on hajaantunut useaan eri kerrokseen. Teknologialtaan ja arkkitehtuuriltaan ympäristössä on yhdenmukaisuuksia yleisesti tunnistettuihin tietovarastoarkkitehtuureihin, mutta myös suuria eroja löytyy. Poikkeuksellisimpana piirteenä kaupungin nykytilassa arkkitehtuurillisessa mielessä on sen epäjohdonmukaisuus ja toimintojen hajanaisuus eri kerroksien välillä. Selkeää tietovarastostrategiaa tai suunnitelmaa ei ole kuvattu, eikä siten myös olla noudatettu.

Pitkä historia sote-tietojen operatiivisesta hyödyntämisestä luo hyvän pohjan kehittämään tietovarastointi- ja raportointiympäristöä tukemaan Power BI -työkalua. Valmiudet Power BI:n käyttöön on olemassa, kunhan alusta-arkkitehtuuria kehitetään johdonmukaisesti mahdollistamaan datan hyödyntäminen tällä välineellä. Haasteen alusta-arkkitehtuurin kehittämiseksi tuo kaupungin kehittämisen toimintamalli, joka on pitkälti palvelualue- ja projektivetoinen. Tämä aiheuttaa haasteita niin rahoituksen kuin yhtenäisen kokonaisarkkitehtuurin luonnin suhteen. Tähän haasteeseen voidaan vastata esimerkiksi tietojohdannon vision määrittelyllä, jossa voitaisiin luoda puitteet laajempaan tietovarastostrategian toteuttamiseen.

Opinnäytetyössä perehdyttiin laajasti aiheeseen liittyviin käsitteisiin ja teorioihin ja peilattiin niiden pohjalta Tampereen kaupungilla vallitsevaa nykytilaa. Tämän tuloksena opinnäytetyössä tunnistettiin 15 kehityskohdetta, joita jatkojalostamalla ja -kehittämällä kaupungin tietovarastointi- ja raportointiympäristöä saataisiin kehitettyä modernimpaan ja hallitumpaan suuntaan. Tietovarastointi- ja raportointiympäristön kehittäminen on jatkuva prosessi ja sen eteen pitää tehdä johdonmukaista ja jatkuvaa työtä useassa organisaation eri toiminnossa. Toivon tämän opinnäytetyön toimivan alkusysäyksenä laajamittaisempaan ja johdonmukaisempaan kehitystyöhön Tampereen kaupungin tietovarastointi- ja raportointiympäristössä.

Opinnäytetyössä tehdyn kyselytutkimuksen vähäisestä vastausmäärästä johtuen yritysten Power BI:n käytöstä ei voida tehdä yleistyksiä. Aihetta olisi hyvä tutkia laajemmin empiirisessä jatkotutkimuksessa.

LÄHTEET

Aho, M. 2011. Konstruktiio suorituskyvyn johtamisen kypsyyden arviointiin. Julkaisu 1000. Tampere: Tampereen teknillinen yliopisto. Väitöskirja.

Devlin, B. 2018. Thirty Years of Data Warehousing. Business Intelligence Journal. 23 (1), 12-24. Seattle: Data Warehousing Institute

Dixon, J. 2010. James Dixon's Blog. [blogi] Luettu 20.3.2019. Saatavissa: <https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/>

Enho, H. 2018. Power BI Hexcelligent. [blogi] Luettu 16.2.2019. Saatavissa: <https://hexcelligent.fi/power-bi/>

Hovi, A., Hervonen, H. & Koistinen, H. 2009. Tietovarastot ja business intelligence. Jyväskylä: Docendo.

Hovi, A., Koistinen, H. & Ylinen, J. 2001. Tietovarastot liiketoiminnan tukena. Helsinki: Satku.

Inmon, W. 2005. Building the data warehouse. Neljäs painos. Indianapolis: Wiley Publishing

Kianto, A. N.d. Tietojohtaminen – mitä, miksi ja miten? Tieto ja osaaminen kilpailuetuna. Bonnier Pro. Tulostettu 20.1.2019. Vaatii käyttöoikeuden. <http://www.bonnierpro.fi.elib.tamk.fi/fi/app/tieto-ja-osaaminen-kilpailuetuna/tietojohtaminen-mita-miksi-ja-miten>.

Kimball, R. 2008. The data warehouse lifecycle toolkit. Toinen painos. Indianapolis: Wiley Pub.

Kimball, R. & Caserta, J. 2004. The data warehouse ETL toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data. Indianapolis: Wiley.

Kimball, R. & Ross, M. 2013. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Kolmas painos. Somerset: John Wiley & Sons, Incorporated.

Klemola, K., Uusi-Ilkainen, J. & Askola, T. 2014. Sosiaali- ja terveystietojärjestelmien tietojohdamisen käsikirja. Helsinki: Sitra.

Knight, D., Knight, B., Pearson, M. & Quintana, M. 2018. Microsoft Power BI Complete Reference. Packt Publishing.

Kuusisto-Niemi, S., Ryhänen, M. & Hyppönen, H. 2018. Tieto- ja viestintäteknologian käyttö sosiaalihuollossa vuonna 2017. 2018. Helsinki: THL.

Laihonen, H. 2009. Terveystietojärjestelmän johtamisen tietovirrat. Tampere: Tampereen teknillinen yliopisto. Väitöskirja

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. Tietojohtaminen. Tampere: Tampereen teknillinen yliopisto - Tiedonhallinnan ja logistiikan laitos.

Maheshwari, A. 2015. Business intelligence and data mining. First edition. New York: Business Expert Press.

Microsoft. N.d. Power BI dokumentaatio. Ladattu 13.2.2019. <https://docs.microsoft.com/fi-fi/power-bi/>

Rangarajan, S. 2016. Data Warehouse Design – Inmon versus Kimball. The Data Administration Newsletter. Ladattu 31.1.2019. Vaatii käyttöoikeuden. <http://tdan.com/data-warehouse-design-inmon-versus-kimball/20300>

Reponen, J., Kangas, M., Hämäläinen, P., Keränen, N. & Haverinen, J. Tieto- ja viestintäteknologian käyttö terveydenhuollossa vuonna 2017: Tilanne ja kehityksen suunta. 2018. Helsinki: THL.

Sitra. 2018. Sote-tietopakettien käsikirja 2.1. Helsinki: Sitra

Sund, R., Nylander, O. & Palonen, T. 2004. Raa'asta rekisteriaineistosta terveyspoliittisesti relevanttiin informaatioon. Helsinki: Stakes. Ladattu 27.1.2019.

Surma, J., Górnjakowska, M. & Gee, P. 2011. Business intelligence: Making decisions through data analytics. New York, N.Y.: Business Expert Press.

LIITTEET

Liite 1. Power BI kyselylomakkeen kysymykset

1 (4)

Kyselyllä kartoitetaan taustatietoa yrityksissä käytettävästä Power BI -ympäristöstä, käyttökohteista ja tietolähdeteknologioista opinnäytetyötä varten. Kyselyn aineistoa käytetään kvantitatiivisen tutkimuksen pohjana tukemaan opinnäytetyössä esitettyjä teorioita tietovarastoinnin ja liiketoimintatiedon hallinnan alueilla.

Taustatietoa yrityksestä

1. Yrityksen toimiala:

- 1 Ammatillinen, tieteellinen ja tekninen toiminta
- 2 Teollisuus, mikä?
- 3 Rakentaminen, Kiinteistöalan toiminta
- 4 Tukku- ja vähittäiskauppa
- 5 Kuljetus ja varastointi
- 6 Majoitus- ja ravitsemistoiminta
- 7 Informaatio ja viestintä
- 8 Rahoitus- ja vakuutustoiminta
- 9 Hallinto- ja tukipalvelutoiminta
- 10 Julkinen hallinto
- 11 Terveys- ja sosiaalipalvelut, julkinen
- 12 Terveys- ja sosiaalipalvelut, yksityinen
- 13 Muu palvelutoiminta
- 14 Muu toimiala, mikä?

2. Yrityksen kokoluokka:

- 1 Alle 100 työntekijää
- 2 100 - 500 työntekijää
- 3 yli 500 työntekijää
- 4 Kokoluokka ei tiedossa

jatkuu

Power BI

2 (4)

3. Yrityksessä Power BI - käyttäjiä:

- 1 Alle 10 käyttäjää
- 2 10 - 100 käyttäjää
- 3 100 - 500 käyttäjää
- 4 yli 500 käyttäjää
- 5 Käyttäjämäärä ei tiedossa

4. Mikä on Power BI:n rooli yrityksen raportoinnissa:

- 1 Raportointi tehdään yksinomaan Power BI:llä
- 2 Käytössä muu raportointiympäristö ja Power BI toimii muun raportointijärjestelmän tukena
- 3 Power BI on otettu käyttöön yksittäisissä organisaatioissa
- 4 Power BI:tä käytetään satunnaisissa raportointitarpeissa
- 5 Power BI:n käyttöä pilotoidaan
- 6 Joku muu, mikä?

5. Minkälaista tietoa Power BI:llä yrityksessä raportoidaan?

6. Power BI – raporttien ja visualisointien julkaisukanava:

- 1 Power BI Service -pilvipalvelu
- 2 Power BI Report Server -raporttipalvelin
- 3 Power BI Embedded -palvelu Azuressa tai muu sisällön upotus
- 4 Power BI Mobile -mobiilisovellus
- 5 Joku muu, mikä?

7. Power BI – toimintamalli (Power BI deployment mode):

- 1 Keskitetty malli (Corporate BI), jossa sekä tietomallit että raportit ja visualisoinnit toteutetaan keskitetysti IT:n toimesta
- 2 Käyttäjien itsepalveluvisualisointimalli (Self-Service Visualization), jossa tietomallit toteutetaan IT:n toimesta, mutta käyttäjät saavat toteuttaa raportit ja visualisoinnit itse keskitetysti julkaistun tietomallin päälle.
- 3 Käyttäjien itsepalveluraportointi (Self-Service BI), jossa käyttäjät itse toteuttavat sekä tietomallit että raportit ja visualisoinnit.
- 4 Joku muu, mikä?

jatkuu

8. Minkälaisia tietolähteitä Power BI -tietomallissa käytetään?

- 1 Tiedostot (teksti/csv, Excel, XML ym.)
- 2 Operatiivisten järjestelmien relaatiotietokanta (SQL Server, Oracle, IBM Db2, ym.)
- 3 Tietovarasto, relaatiotietokanta (SQL Server, Oracle, IBM Db2, ym.)
- 4 Tietovarasto, pilvitietokanta (Azure SQL Server, Azure SQL Data Warehouse, Amazon Redshift, Snowflake, Azure Data Lake Storage, ym.)
- 5 SAP Business Warehouse tai SAP HANA (tai muu SAP -tietolähde)
- 6 Internet (Avoin data tai muut julkiset sivustot)
- 7 Pilvipalvelut (Facebook, Google Analytics, Salesforce, ym.)
- 8 Analysis Services Dimensionaalinen tietomalli (joko SQL Server Analysis Services tai Azure Analysis Services)
- 9 Analysis Services Tabular tietomalli (joko SQL Server Analysis Services tai Azure Analysis Services)
- 10 IoT tietolähde
- 11 Joku muu, mikä?

Tietovarasto

9. Jos tietolähteenä käytetään tietovarastoa, niin minkälainen arkkitehtuuri tietovarastossa on kyseessä?

- 1 Yritystasoinen keskitetty tietovarasto (EDW, Enterprise Data Warehouse 'Inmon')
- 2 Yhdenmukaistetut paikallistietovarastot (Data Warehouse Bus Architecture 'Kimball')
- 3 Itsenäiset paikallistietovarastot (Data Marts)
- 4 Hybridi, yhdistelmä yllä olevia
- 5 Joku muu, mikä?

10. Jos yrityksessä on käytössä tietovarasto, mitä teknologiaa tietovarastossa on käytetty?

1 Paikallinen relaatiotietokanta (SQL Server, Oracle, IBM Db2, ym.)

2 Pilvipalveluna toimiva tietokanta (Azure SQL Server, Amazon RDS: Oracle, MySQL SQL Server tai PostgreSQL, ym.)

3 Pilvipalveluna toimiva tietovarastokanta (Azure SQL Data Warehouse, Amazon Redshift, Snowflake, Azure Data Lake Storage, ym.)

4 Joku muu, mikä?