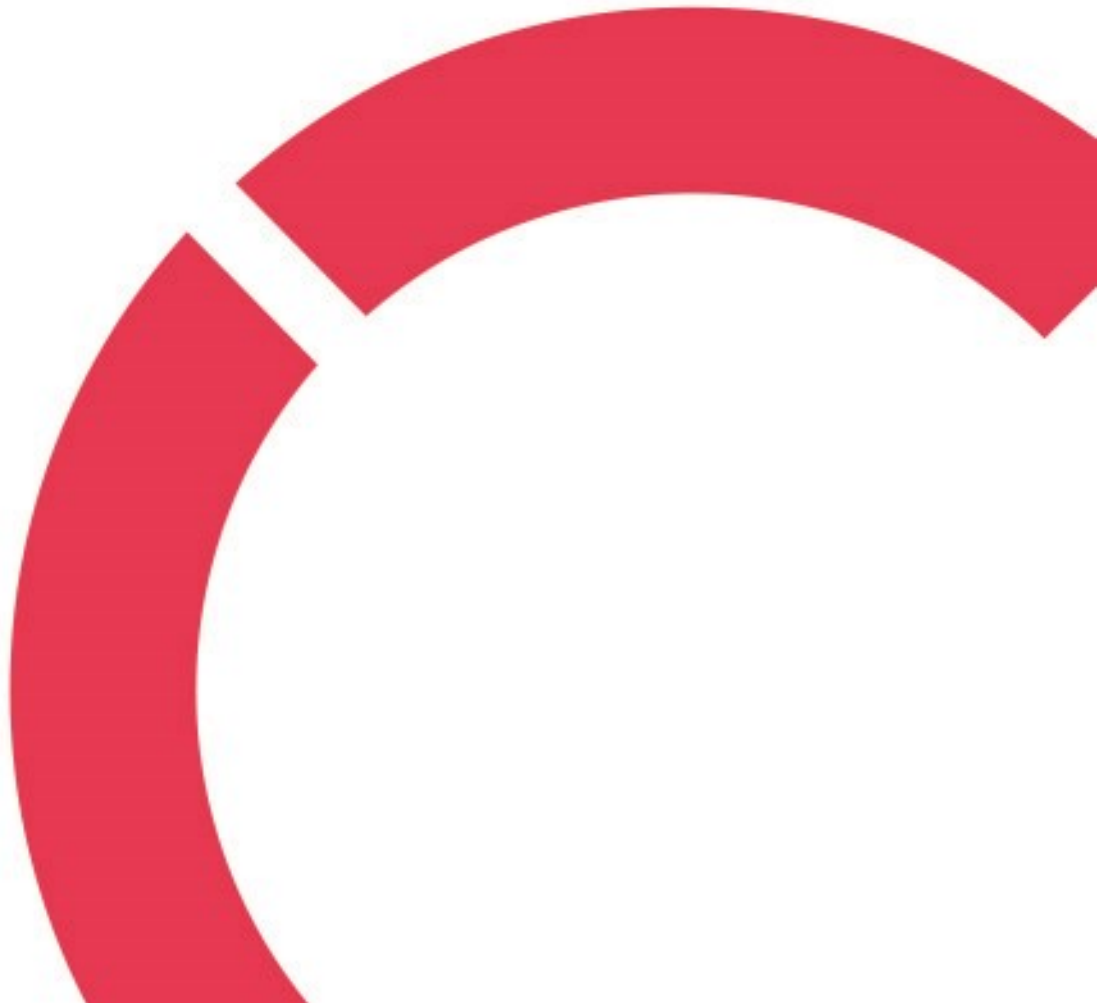


Vesa Sainio

**RAPORTTIEN TUOTTAMINEN MICROSOFT POWER BI:LLA SO-
SIAALI- JA TERVEYSPALVELUIDEN TIEDOLLA JOHTAMISEEN**

**Opinnäytetyö
CENTRIA-AMMATTIKORKEAKOULU
Tieto- ja viestintäteknikan koulutus
Maaliskuu 2022**



Centria-ammattikorkeakoulu	Aika maaliskuu 2022	Tekijä/tekijät Vesa Sainio
Koulutus Tieto- ja viestintätekniikka, insinööri AMK		<input checked="" type="checkbox"/> AMK <input type="checkbox"/> YAMK
Työn nimi RAPORTTIEN TUOTTAMINEN MICROSOFT POWER BI:LLA SOSIAALI- JA TERVEYSPALVELUIDEN TIEDOLLA JOHTAMISEEN		
Työn ohjaaja Jari Isohanni		Sivumäärä 50 + 1
Työelämäohjaaja Kalle Tornberg		
<p>Tässä opinnäytetyössä tutkittiin julkisen sote-palvelunjärjestäjän tiedolla johtamisen tarpeisiin tuotettavien raporttien tekemistä sellaisissa käyttötapauksissa, joissa raporttien tausta-aineistona hyödynnetään potilastietoja. Tavoitteena oli laatia tilaajan tietotarpeet täyttävä raportti, joka julkaistaan rajatulle käyttäjäjoukolle. Raportin tuotantoprosessin kaikissa vaiheissa on otettava huomioon henkilötietoja sisältävien tausta-aineistojen asianmukainen käsittely. Opinnäytetyössä käytettiin raporttien tuottamisen työvälineenä Microsoftin Power BI -ohjelmistoa.</p> <p>Opinnäytetyössä esitellään lainsäädännöllisiä perusteita ja rajoitteita, joiden puitteissa henkilötietoja voidaan käyttää sote-alan tietojohdamisen tausta-aineistona. Osa lainsäädännön vaateista henkilötietojen suojaamisessa voidaan toteuttaa tietovarastojen arkkitehtuurin suunnittelussa ja toteutuksessa. Muilta osin vastuu tietojen asianmukaisesta käsittelystä jää tiedolla johtamisen raportteja tuottavan asiantuntijan vastuulle. Opinnäytetyön pääpaino olikin sen kuvaamisessa, miten henkilötietoja sisältävää aineistoa tulisi eri työvaiheissa käsitellä, kun raportti tuotetaan ja julkaistaan Power BI:lla.</p> <p>Raportin tuottaminen alkaa aina tietotarpeiden tunnistamisesta. Seuraavia raportin tuotantoprosessin vaiheita ovat tietojen poiminta ja raportointiaineiston koostaminen, tietomallin rakentaminen ja optimointi sekä toiminnan sisältöön liittyvien laskentakaavojen määrittely. Varsinaisen muotonsa raportti saa, kun raporttisivujen ulkoasu suunnitellaan ja valitaan aineistoon sopivat visualisaatiot, jotka tuovat aineiston olennaisimmat tiedot helposti käyttäjän nähtäviksi. Ennen julkaisua raportin visualisaatiot tulee testata, jotta varmistutaan niiden tarkoituksenmukaisesta ja tietosuojan turvaavasta toiminnasta.</p> <p>Power BI -ohjelmiston tietojen analysoinnin helppoutta lisäävät ominaisuudet voivat helppouden sijasta olla haitta, kun raportointiaineistossa olevia tietoja ei tietosuojasyistä haluta kaikilta osin antaa käyttäjän tarkasteltaviksi. Olennainen osa raportin tuotantoprosessia onkin pyrkimys täyttää toimeksiantossa annettu tietotarve mahdollisimman kattavasti, mutta samalla estää yksittäisten henkilöiden suora tai epäsuora tunnistaminen raportista.</p>		

Asiasanat DAX, henkilötieto, Power BI, raportointi, SQL, tiedolla johtaminen, tietojohdaminen, sosiaali- ja terveydenhoitopalveluiden tuottaminen

ABSTRACT

Centria University of Applied Sciences	Date March 2022	Author Vesa Sainio
Degree programme Bachelor of Engineering, Information Technology		
Name of thesis PRODUCTION OF REPORTS FOR KNOWLEDGE-BASED MANAGEMENT ON SOCIAL WELFARE AND HEALTH CARE SECTOR WITH MICROSOFT POWER BI		
Centria supervisor Jari Isohanni	Pages 50 + 1	
Instructor representing commissioning institution or company Kalle Tornberg		
<p>The aim of this thesis was to study the production of business intelligence reports for knowledge-based management on public social welfare and health care sector in use cases where the report's underlying data contains sensitive personal data. The aim of the thesis was also to produce a report to meet the informational needs of selected administrators of the hospital and publish it to a limited group of administrators. Because the report's underlying data contains sensitive personal data the processing of data must be conducted using appropriate means during every step of the development process. In this thesis the software used to produce the report is Microsoft Power BI.</p> <p>This thesis shows the legal grounds and limitations based on the law that are applied when sensitive personal data is being used as underlying data when producing reports for knowledge-based management on social welfare and health care sector. Part of the legal requirements of protecting and processing sensitive data can be fulfilled by using special architecture in data warehouse that is being used for storing the data. The rest of responsibility of proper handling is left to the report developer. Therefore, the main emphasis of this thesis is to illustrate how to handle sensitive data in the different steps of the development process during production of Power BI report.</p> <p>The production of a report begins with a recognition of needs for knowledge. The next steps in report production are gathering the needed information, building and optimizing a data model and adding relevant business-related measures to the data model. The report starts to get its form when the layout for the report pages is made, and correct types of visualizations are chosen to show the relevant data of the report to the consumer. All visualizations must be tested for appropriate function from the viewpoint of data privacy and user-friendliness before publishing the report.</p> <p>Power BI has many features to help report users to analyze data. However, features that help in detailed data analyzation are not a good thing when report developer does not want to make the data visible in all details to the report user in order to preserve data privacy. Essential part of report production in this context is the goal to fulfill the users' need of knowledge as well as possible, but at the same time prevent the identity of individuals from being exposed directly or indirectly from the data viewed on the report.</p>		
Key words BI, business intelligence, DAX, health care services, knowledge-based management, personal data, Power BI		

KÄSITTEIDEN MÄÄRITTELY

BI (BUSINESS INTELLIGENCE)

Systemaattista tiedon hankintaa, tallennusta ja analysointia. Liiketoiminnassa syntyvää dataa muutetaan liiketoiminnan tarkoitusta palvelevaksi hyödylliseksi tiedoksi.

DATA

Tietoa, joka on koneluettavassa muodossa. Datasta saatavaa hyötyä voidaan parantaa rakenteistamalla dataa tai lisäämällä siihen metatietoja. Tiedon jalostusasteena alin luokka.

HILJAINEN TIETO

Henkilölle kokemuksen kautta kertynyt tieto, joka on laadultaan subjektiivista ja siksi usein vaikeaa muuntaa kattavasti eksplisiittiseksi tiedoksi

RAKENTEETON DATA

Koneluettavaa tietoa, jota ei ole valmiiksi jäsennelty erillisten metatietojen avulla

RAKENTEELLINEN DATA

Koneluettavaa tietoa, johon on tietoon liitettyjen metatietojen avulla luotu tiedon jäsentelyn mahdollistavia rakenteita. Potilas- ja asiakastietojen käsittelyssä tämä vaatii yhteisesti sovittuja periaatteita tiedon kirjaamiselle.

TIETO

Merkkijono, fakta, havainto tai tulkinta. Tietoa voidaan jalostaa, eli pyrkiä hyödyntämään tietoa entistä paremmin. Tietoa voidaan luokitella jalostusasteensa mukaan alimmalta tasolta ylimpään seuraavasti: data, informaatio, tietämys, ymmärrys, viisaus

RELAATIOTIETOKANTA

Tietokanta on tietokoneen tallentama kokoelma tietoa, joka koostuu tauluista ja sarakkeista. Relatio-tietokannoissa eri taulujen välille luodaan yhteyksiä taulujen tietoja yhdistävien avainsarakkeiden avulla.

**TIIVISTELMÄ
ABSTRACT
KÄSITTEIDEN MÄÄRITTELY
SISÄLLYS**

1 JOHDANTO	1
2 MITÄ ON TIETOJOHTAMINEN?	3
2.1 Tietojohtaminen käsitteenä	3
2.2 Tietojohtaminen soite-alalla	4
3 HENKILÖTIETOJEN HYÖDYNTÄMISEN LAINSÄÄDÄNNÖLLISET PERUSTEET SOTE-ALAN TIETOJOHTAMISESSA	6
3.1 Henkilötieto.....	6
3.2 Pseudonymisoitu henkilötieto	7
3.3 Anonymisoitu tieto ja tietojen aggregointi.....	7
3.4 Sote-alan tiedolla johtaminen ja henkilötietoja sisältävien asiakastietojen käsittelyä koskeva lainsäädäntö.....	9
3.4.1 Asiakastietojen toisiokäyttö.....	9
3.4.2 Käyttöoikeus asiakastietoon	10
4 TIEDOLLA JOHTAMISEN ARKKITEHTUURI SOTE-ALALLA	11
4.1 VIRTAA-arkkitehtuuri	11
4.2 Pseudonymisoidun tiedon käyttö eri raportointiaineistoissa	12
5 TIEDOLLA JOHTAMISEN RAPORTOINNIN KÄYTETTÄVIÄ TYÖKALUJA	14
5.1 Microsoft SQL Server Management Studio (SSMS)	14
5.2 Microsoft Power BI	15
5.2.1 Tietojen tuonti ja tallennus Power BI -tiedostoon	16
5.3 Tietojen yhdistäminen toisiinsa ja tietomallin rakenne.....	20
5.4 Power BI -apuohjelmat.....	20
5.4.1 DAX Studio	21
5.4.2 Tabular Editor.....	21
5.5 Power BI -raportin tuottamisessa käytettäviä SQL- ja DAX- funktioita	22
5.5.1 SQL-kyselyn rakenne.....	22
5.5.2 SQL aggregointifunktiot.....	23
5.5.3 HASHBYTES-funktio.....	24
5.5.4 DAX -lausekkeet.....	Virhe. Kirjanmerkkiä ei ole määritetty.
6 SOTE-ALAN TIEDOLLA JOHTAMISEN RAPORTIN TUOTTAMINEN MICROSOFT POWER BI:LLA	27
6.1 Tiedolla johtamisen raportoinnin tarpeet julkisen soite- palveluiden järjestämisen eri järjestelmätasolla	27
6.2 Tuotettavan raportin kuvaus ja tuotantoprosessin vaiheet	29
6.3 Tietotarpeen tunnistaminen tai toimeksiannon saanti	30
6.4 Tietojen poiminta tietokannasta ja aineiston koostaminen.....	31
6.4.1 Faktataulun poiminta aggregoimalla	31
6.4.2 Dimensiotaulun poiminta	32
6.4.3 Henkilötunnisteita sisältävän faktataulun poiminta aggregoimalla.....	32
6.4.4 Faktataulun poiminta ilman aggregointia	33

6.5 Tietomallin rakentaminen	34
6.5.1 Kalenteridimension lisääminen tietomalliin	35
6.6 Laskentakaavojen määrittäminen	36
6.7 Raportin ulkoasun suunnittelu ja toteuttaminen	37
6.8 Raportin testaaminen ja optimointi	43
6.9 Raportin julkaiseminen	44
7 JOHTOPÄÄTÖKSET	46
LÄHTEET	48
LIITTEET	
KUVAT	
KUVA 1. Yksinkertaistettu kuvaus tiedolla johtamisen VIRTIA-arkkitehtuurimallista	12
KUVA 2. Henkilötunnuksen hash -arvon käyttö faktatauluissa	13
KUVA 3. Tietojen suodattaminen osittajalla	16
KUVA 4. Tietojen noutamisen valintanauha ja tietolähteen valintaruutu	17
KUVA 5. Yksittäisen kyselyn käsittely Tabular -taulukkomallissa	17
KUVA 6. Valintaruutu, jolla tuodaan tietoja SQL Server -tietokannasta	18
KUVA 7. M-kielen käyttäminen tietojen tai sarakkeiden muokkaamiseen Power Query -editorissa....	19
KUVA 8. DAX-kielen käyttäminen uuden sarakkeen luomiseksi	19
KUVA 9. Esimerkki Power BI- tiedoston tietomallista	20
KUVA 10. Yksittäisen tietotaulun kokoa ja ominaisuuksia voi tarkastella DAX Studion Vertipaq Ana-lyzerin avulla	21
KUVA 11. Raportin tietomallin tarkastelua Best Practice Analyzerilla	22
KUVA 12. Koodin 3 tuottamia hash -arvoja henkilötunnisteelle	24
KUVA 13. Tietojohdamisen raportoinnin ominaisuuksia asetettuna Laihosen esittämiin hyvinvointisek- torin tiedonhallinnan järjestelmätasoihin	28
KUVA 14. Power BI-raportin tuottamisen vaiheet ja käytettävät ohjelmistot	29
KUVA 15. Kuvakaappaus raporttitiedoston tietomallin rakenteesta	34
KUVA 16. Terveyden- ja sairaanhoidon toimialueen raporttinäkymän vasen puoli	37
KUVA 17. Terveyden- ja sairaanhoidon toimialueen raporttinäkymän oikea puoli	38
KUVA 18. Kirjanmerkin asetusten asettaminen	39
KUVA 19. Terveyden- ja sairaanhoidon toimialueen valmis raporttinäkymä	41
KUVA 20. Kuva-elementtiin toiminnoksi määritetty kirjanmerkki ja sivupalkin toiminnot toteuttavia kirjanmerkkejä	42
KUVA 21. Taloustietokuvaajan datapisteestä nousevia lisätietoja työkaluvihjetoiminnon avulla	42
KUVA 22. Kuvakaappaus suorituskyvyn analysointityökalun näkymästä	43
KUVA 23. Suodatinpaneelin piilottaminen käyttäjältä ja yksittäisen suodattimen lukitseminen	45
KOODIT	
KOODI 1. FROM-lause LEFT OUTER JOIN -poiminnalla päätaulusta ja yhdestä lisätaulusta	23
KOODI 2. Yksittäisten potilaiden määrän laskeminen COUNT(DISTINCT()) -yhdistelmällä	24
KOODI 3. HASHBYTES-funktion käyttö	24
KOODI 4. CALCULATE-funktion käyttö	25
KOODI 5. Useiden suodatinparametrien antaminen CALCULATE-funktiolle	25
KOODI 6. DAX-kielen IF ja SWITCH-lauseita	26
KOODI 7. Esimerkki suodatinvalinnan tilaa tarkastelevien funktioiden käytöstä	26
KOODI 8. Faktataulun koodiavainten selitteiden poiminta SELECT DISTINCT -lauseella	32

KOODI 9. SQL-poiminta: potilasmäärä kuukausittain.....	33
KOODI 10. Taloustietojen poiminta ilman aggregointia.....	33
KOODI 11. Kalenteri-taulun luominen DAX-lauseella	35
KOODI 12. Käyntien yhteismäärän laskentakaava	36
KOODI 13. Perusterveydenhoidon käyntien laskentakaava.....	36
KOODI 14. Perusterveydenhoidon vuoden 2021 käyntien laskentakaava	36
KOODI 15. Mittareiden uusiokäyttöä uusien mittareiden luomisessa CALCULATE-funktiolla.....	36
KOODI 16. Kuvaajassa käytettävän mittarin määrittäminen SWITCH-lauseella.....	39
KOODI 17. Kuvaajassa käytettävän otsikon määrittäminen SWITCH-lauseella.....	40
KOODI 18. Kumulatiivinen kuukausittaisten kustannusten laskeminen.....	40
KOODI 19. Otsikon määrittäminen dynaamisesti DAX-lauseella	41

TAULUKOT

TAULUKKO 1. DAX-lauseella muodostetun kalenteritaulukon ensimmäiset rivit	35
---	----

1 JOHDANTO

Tässä opinnäytetyössä tutustutaan Microsoft Power BI -ohjelmistolla tuotettujen tiedolla johtamisen raporttien tuottamiseen sote-alalla sellaisissa käyttötapauksissa, joissa käytettävä tausta-aineisto sisältää henkilötietoja. Opinnäytteen tuloksena tuotetaan taloustietoja ja potilaiden käyntitietoja yhdistelevä raportti Keski-Pohjanmaan sosiaali- ja terveystieteiden tutkimuskeskuksen Soiten johdon käyttöön.

Opinnäytetyön tavoitteina on:

- Selvittää, millä lainsäädännöllisillä perusteilla ja minkälaisissa tapauksissa sote-palvelujen järjestäjän tietokannoissa olevaa henkilötietoja sisältävää dataa saa käyttää pohja-aineistona julkisen sosiaali- ja terveystieteiden tutkimuskeskuksen järjestävän kuntayhtymän tai hyvinvointialueen tietojohdantamisessa.
- Kartoittaa sellaisia teknisiä ja työnkulkuun liittyviä ratkaisuja, joita raportteja tekevä tai raportointiaineistoja valmisteleva asiantuntija voi käyttää työnkulun eri vaiheissa tietoa-aineistoja käsiteltäessä, jotta organisaation tietotarpeet saadaan täytettyä voimassa olevan lainsäädännön puitteissa. Tässä tarkastelun kohteena ovat Microsoftin Business Analytics -tuoteperheen tuotteet Microsoft SQL Server Management Studio, Microsoft Power BI sekä Microsoftin raporttien ja tietojoukkojen jakamiseen käytettävä pilvipalvelu osoitteessa app.powerbi.com.
- Tutkia, miten Microsoft Power BI -ohjelmistolla saadaan tuotettua suuresta tietomäärästä tehokkaasti toimivia raportteja optimoimalla tietomallin rakennetta sekä minimoimalla tietojoukon kokoa tilanteeseen sopivalla aggregoinnilla.
- Tuottaa yllä mainittuja lainsäädännön vaatimuksia vastaava tehokkaasti toimiva ja päivittyvä raportti, jolla voidaan dynaamisesti päivittyvien visualisointien kautta tarkastella palvelualueittain koostettuja tietoja Keski-Pohjanmaan sosiaali- ja terveystieteiden tutkimuskeskuksen Soiten yksiköiden käynneistä, hoitopäivistä ja kustannuksista kuukausi- ja vuositasolla. Raportti julkaistaan Soiten organisaation sisäisessä tiedolla johtamisen järjestelmässä.

Tässä opinnäytetyössä tiedolla johtamista tarkastellaan raportteja valmistelevan tai tuottavan analytiikan näkökulmasta. Tässä näkökulmassa painottuvat tiedolla johtamisen raportoinnissa käytettävän pohjadataan käsittelyyn ja raporttiaineiston koostamiseen sekä itse raporttiedoston muokkaamiseen liittyvät tekniset kysymykset sekä olennaisten tietojen esille tuontiin vaadittavat tietotekniset taidot.

Henkilötietojen, ja erityisesti terveyttä koskevien tietojen käsittely on tiukasti säänneltyä toimintaa. Terveydenhoitoalalla terveyteen liittyvien tietojen käyttäminen on luonnollisesti välttämätöntä hoidon toteuttamiseksi ja siksi erityissääntelyllä sallittua. Vuonna 2019 voimaan tulleen toisiolain nojalla sote-palveluiden tuottamisessa kerättyä tietoa on kuitenkin saanut käyttää myös muuhun tarkoitukseen, kuin mitä varten tiedot on alun perin kerätty. Terveystietojen käyttö tietojohdantamisessa on toisiolain pohjalta mahdollista, mutta tietojen käytölle on asetettu tiukkoja ehtoja.

Opinnäytteen luvussa 2 esitellään tietojohdantamista käsitteenä sekä sitä, mitä erityispiirteitä tietojohdantamisella on julkisella sote-alalla. Sote-palveluiden asiakastietojen toisiokäytön lainsäädäntöä ja niihin liittyviä määritelmiä käsitellään luvussa 3. Luvussa 4 puolestaan tutustutaan suomalaisen sote-alan tietojohdantamiseen kehitettyyn viitearkkitehtuuriin. Luvussa 5 esitellään yleisesti raportin tausta-aineiston käsittelyyn käytettäviä työkaluja ja menetelmiä sekä erityisesti tietojen karkeistamiseen käytettäviä kommentoja SQL- ja DAX-kielillä. Luvussa 6 tuotetaan valmis raportti ja käydään läpi yksityiskohtaisesti sote-alan tiedolla johtamisen raportin tuottamisen työvaiheita. Lopuksi tutustumme siihen, kuinka raportointitarpeen käyttötapaus ja käytettävä aineisto huomioiden tehty raportti voidaan tietoturvallisesti jakaa oman organisaation sisällä oleville loppukäyttäjille huomioiden mahdollisesti toisistaan poikkeavat käyttöoikeudet eri tietoihin.

Opinnäytetyön pääpaino on hyvien teknisten ratkaisujen etsimisessä pohjatietojen käsittelyyn, raportin tuottamisen työnkulkuun, raportin optimointiin sekä jäännösriskin minimoimiseen raportointiaineistossa. Työn rajaamiseksi moniin tähän aihepiiriin kiinteästi liittyviin seikkoihin tutustutaan tässä työssä vain suppeasti tai ei lainkaan. Tällaisia seikkoja ovat esimerkiksi julkisen sote-palveluiden tietojohdantamisen järjestelmien arkkitehtuuri, tietojärjestelmiin tallennettavan tiedon kirjaamiseen liittyvät vaatimukset, datan laadun varmistaminen ja kehittäminen, digitaalisten palveluiden saavutettavuusvaatimukset sekä visuaalisesti ja tarinallisesti hyvän raportin rakentamisen periaatteet.

Henkilötietojen käsittely on hyvästä syystä tiukasti lailla säädeltyä toimintaa. Tästä huolimatta henkilötietoja sisältäviä aineistoja on mahdollista käyttää sosiaali- ja terveyspalveluiden tuottajan tietojohdantamisen raportoinnissa, kunhan tietoja käsitellään ja tuodaan nähtäville oikeilla tavoilla. Tätä kautta voidaan saada paljon arvokasta tietoa, joka hyödyttää sekä terveydenhoitojärjestelmää itseään että sen asiakkaita.

2 MITÄ ON TIETOJOHTAMINEN?

Sosiaali- ja terveystalveluiden johtamisessa jatkuvana suurena haasteena on lisääntyvän palvelutarpeen ja käytettävissä olevien resurssien tasapainottaminen. Tarvitaan siis tulevaisuuteen suuntaavia arvioita sekä palvelutarpeista että tarpeen täyttämiseen tarvittavista rahallisista panostuksista. Toisaalta johtamisessa halutaan analysoida myös aikaisemmin tehtyjen päätösten vaikutuksia ja yleistä kehityksen suuntaa. (Klemola, Uusi-Illikainen & Askola 2014, 11–13.) Ovatko kustannukset nousussa vai laskussa, mihin suuntaan palveluiden tarve on kehittymässä eri sektoreilla ja mihin suuntaan oman alueen väestön hyvinvointi on kehittymässä? Muun muassa näihin vaikeisiin kysymyksiin voidaan hakea vastauksia tietojohdamisen kautta.

2.1 Tietojohdaminen käsitteenä

Tietojohdaminen on johtamisen ajattelumalli, jota käytetään erityisesti tietointensiivisillä aloilla toimivissa organisaatioissa. Sen tavoitteena on kehittää organisaation suorituskykyä ja työn sujuvuutta. Tähän pyritään muodostamalla tietojohdamisen kautta organisaation käytettävissä olevasta tiedosta kokonaiskuva, jota käytetään liiketoiminnan parantamiseen. (Laihonen ym. 2013, 10–13.)

Tietojohdaminen pyrkii vastaamaan tietoyhteiskunnan johtamisen haasteisiin kolmesta näkökulmasta. Ensinnäkin sen avulla pyritään ymmärtämään, miten tiedosta luodaan arvoa erilaisissa liiketoimintaympäristöissä. Toiseksi saadun ymmärryksen perusteella voidaan tarkastella sitä, millaisia johtamisen käytäntöjen tulisi olla, jotta tärkeät tietoresurssit tunnistetaan ja niitä käytetään ja johdetaan oikealla tavalla. Kolmas näkökulma on käytännön johtamistyökalut, eli minkälaisilla tieto- ja viestintäteknologisilla työkaluilla aineettomat tietoprosessit ja niistä saatava tieto saadaan paremmin hallintaan, jotta sitä voidaan hyödyntää organisaation johtamisen tukena. (Laihonen ym. 2013, 7–9.)

Lisäksi tietojohdamisessa voidaan nähdä olevan kaksi pääsuuntausta: liiketoiminnan johdollinen suuntaus, joka näkee tiedon yrityksen menestystekijänä, ja tietotekninen suuntaus, jossa korostetaan tietojärjestelmien merkitystä tiedon hallinnassa ja käsittelyssä. (Laihonen ym. 2013, 7–9.) Tietojohdamista onkin sote-alalla jaettu näiden pääsuuntausten mukaan kahteen osaan: tiedolla johtamiseen ja tiedon johtamiseen. Tiedolla johtamisella tarkoitetaan tiedon hyödyntämistä, eli kysymyksiä siitä, miten tietoja esitetään ja miten niitä käytetään päätöksenteon tukena eri tasoilla. Tiedon johtamisella puolestaan

tarkoitetaan tiedon hankkimiseen, luokitteluun, varastointiin ja suojaamiseen liittyvien prosessien suunnittelua ja toteuttamista. Tiedon johtamisen alle kuuluu myös sopiminen kirjaamiskäytännöistä. (Leskelä ym. 2019, 14; Leskelä ym. 2018, 6–12; Laihonen ym. 2013, 32.)

Tietojohtamisessa on siis monia ulottuvuuksia ja se vaatii monenlaista osaamista. Liiketoiminnan ymmärryksen ja johtamistaitojen lisäksi on sisäistettävä myös tiedon merkitys organisaatiolle sekä tunnettava riittävässä määrin tieto- ja viestintäteknologiaa käytettävissä olevan tiedon hyödyntämiseksi. Tietojohtamisen syntyaikoina tietojohtamisen painotus oli tiedon hallinnassa, eli tiedon tuottamisessa, käsittelyssä ja jakamisessa. (Laihonen ym. 2013, 10–11.) Tärkeää oli se, että tieto oli olemassa ja sitä tarvitsevien saatavilla. Organisaatioiden käytettävissä olevan tiedon määrän kasvaessa tietojohtamisen merkitys on kuitenkin keikahtanut tiedon hyödyntämiseen (Laihonen ym. 2013, 10). Samalla tietojohtaminen on myös kasvattanut rooliaan, ja se on nykyisin ajattelumalli, joka läpäisee koko organisaation toiminnan ja toiminnan eri tasot operatiivisesta toiminnasta yrityspuolella konsernin globaalin mittakaavan tai julkisen sote-palvelutuotannon ollessa kyseessä valtakunnallisen tason tarkasteluun.

2.2 Tietojohtaminen sote-alalla

Tietojohtamista ja johtamista ylipäätään tehdään usealla eri tasolla. Sitran Sosiaali- ja terveystieteiden tietojohdamisen käsikirjassa (Klemola ym. 2014) tietojohtaminen on jaettu kolmeen tasoon. Operatiivisella tasolla johdetaan päivittäistä toimintaa, taktisella tasolla tarkastellaan keskipitkän aikavälin muutoksia ja strategisella tasolla johdetaan ja kehitetään organisaation toimintaa pitkällä aikavälillä. Sote-alalla tarkastelujakson katsotaan Sitran mukaan olevan 5–10 vuotta tulevaisuuteen. (Klemola ym. 2014, 11–13.)

Laihonen (2009) puolestaan on jakanut sote-alan tietojohtamisen viiteen järjestelmätasoon. Alimmalla järjestelmätasolla on operatiivinen toiminta ja ylimpänä valtakunnallinen informaatio-ohjaus. Näiden tasojen väliin jäävät organisaation johtamiseen, palveluiden tilaamiseen ja kunnallisen tai alueellisen päätöksenteon tiedonhallinnan järjestelmätasot. (Laihonen 2009, 442–445.) Eri järjestelmätasoilla tavoitteet ovat erilaisia ja näin myös tiedolla johtamiseen liittyvät tietotarpeet poikkeavat toisistaan selkeästi. Tarkastelen tiedolla johtamista eri järjestelmätasoilla raporttien tuottamisen näkökulmasta tarkemmin luvussa 6.1.

Sote-alan tietojohdamisen erityispiirteenä on, että liiketoiminnan analysointiin käytettävien raportointiaineistojen lähdeaineisto koostuu suurelta osin ihmisten henkilötiedoista. Kyse ei ole ainoastaan ihmisten nimistä, osoitteista ja henkilötunnuksista, vaan myös hyvin arkaluontoisista ja yksityiskohtaisista terveyteen liittyvistä tiedoista. Lainsäädäntö asettaakin tiukat rajoitukset sille, kuinka tietoja täytyy käsitellä, jotta niitä voi käyttää julkisten sosiaali- ja terveystalveluiden tuottamisen tietojohdamisen tarkoituksiin.

3 HENKILÖTIETOJEN HYÖDYNTÄMISEN LAINSÄÄDÄNNÖLLISET PERUSTEET SOTE-ALAN TIETOJOHTAMISESSA

Tässä luvussa esitellään, millä lainsäädännöllisillä perusteilla sosiaali- ja terveystalouden tuottamisessa kerättyjen henkilötietojen hyödyntäminen tietojohdamisessa on mahdollista. Ensin tutustutaan henkilötiedon määritelmään sekä siihen, mitä tarkoittavat henkilötietojen tietynlaisella käsittelyllä muodostettu pseudonymisoitu ja anonymisoitu tieto.

3.1 Henkilötieto

Tietosuojavaltuutetun mukaan henkilötiedoiksi luetaan kaikki sellaiset tiedot, jotka ”liittyvät tunnistettuun tai tunnistettavissa olevaan henkilöön”. Tällaisia tietoja voivat olla esimerkiksi henkilön nimi, osoite, henkilötunnus, paikannustiedot tai potilastiedot. (Tietosuojavaltuutettu.)

EU:n yleinen tietosuojajasetus määrittelee henkilötietojen käsittelyä koskevia yleisiä periaatteita, joita kutsutaan myös tietosuojaperiaatteiksi. Tietosuojaperiaatteita ovat esimerkiksi vaatimukset tietojen käsittelystä ”lainmukaisesti, asianmukaisesti ja rekisteröidyn kannalta läpinäkyvästi”. Tietoja saa kerätä ainoastaan tiettyä tarkoitusta varten ja lisäksi tietojen keräämisessä ja käytössä on pyrittävä käyttötarkoituksensa mukaan tietojen minimointiin sekä kaikissa tilanteissa käsiteltävä tietoja luottamuksellisesti ja turvallisesti. (EU:n yleinen tietosuojajasetus 2016/679, 5. artikla.)

Henkilötietoja on aina kerättävä ja käsiteltävä tietosuojaperiaatteiden mukaisesti. Tietosuojaperiaatteet puolestaan vaativat tietojen käsittelyltä luottamuksellisuutta ja tarkasti rajattua käyttötarkoitusta. Millä ehdoilla henkilötietoja sisältävää aineistoa sitten voitaisiin käyttää esimerkiksi tietojohdamisessa tai tieteellisiin tutkimuksiin, eli muihin tarkoituksiin, kuin mitä varten tiedot on kerätty? Miten henkilötietoja sisältävää aineistoa tulisi käsitellä, jotta tietoja voitaisiin käyttää hyödyksi myös tiukasti rajatun tietojen tietyn käyttötarkoituksen vuoksi käsittelevien luotettujen henkilöiden piirin ulkopuolella?

Näistä kysymyksistä ensimmäiseen etsitään vastauksia luvuissa 4 ja 6. Jälkimmäistä kysymystä käsitellään tämän luvun seuraavissa alaluvuissa.

3.2 Pseudonymisoitu henkilötieto

Aineistoja käsiteltäessä henkilötiedoista voidaan poistaa suoraan henkilön tunnistamiseen mahdollistavat tunnisteet, kuten nimi, osoite tai henkilötunnus. Näitä tietoja voidaan myös korvata jollain yksittäiseen henkilöön viittaavalla pseudotunnisteella, kuten esimerkiksi henkilötunnuksen suojatulla hash-arvolla. Tällaista aineiston käsittelyä kutsutaan pseudonymisoinniksi. (Sosiaali- ja terveysministeriö 2020, 1.) EU:n yleisen tietosuoja-asetuksen määritelmän mukaan pseudonymisoinnilla ”tarkoitetaan henkilötietojen käsittelemistä siten, että henkilötietoja ei voida enää yhdistää tiettyyn rekisteröityyn käyttämättä lisätietoja” (EU:n yleinen tietosuoja-asetus 2016, 4. artikla).

Aineiston pseudonymisointi siis poistaa mahdollisuuden henkilön suoraan tunnistamiseen, mutta lisätietojen lisääminen aineistoon saattaa mahdollistaa yksittäisen henkilön tunnistamisen. Pseudonymisointi ei tarkoita tiedon karkeistamista, eli tietojen tarkkuus ei heikkene: oikea tunniste ainoastaan vaihdetaan pseudotunnisteeseen. (Sosiaali- ja terveysministeriö 2020, 1.) Koska yksittäisen henkilön tunnistaminen aineistosta voisi olla lisätietoja lisäämällä vielä mahdollista, pseudonymisoitu tieto määritellään edelleen henkilötiedoksi (EU:n yleinen tietosuoja-asetus 2016, resitaali 26; Sosiaali- ja terveysministeriö 2020, 1). Sosiaali- ja terveysministeriön asettaman toissijaisen käytön tietosuojan asiantuntijaryhmän mukaan pseudonymisointi ei siis anna mahdollisuutta julkaista aineistoja, vaan ainoa hyväksyttävä julkaistavan aineiston suojauskeino on aineiston anonymisointi (Sosiaali- ja terveysministeriö 2020, 1.)

3.3 Anonymisoitu tieto ja tietojen aggregointi

Sosiaali- ja terveysministeriön asettaman toissijaisen käytön tietosuojan asiantuntijaryhmä on linjannut, että anonymisoituna voidaan pitää ainoastaan sellaisia aineistoja, joista ei voida tunnistaa yksittäistä henkilöä suoraan tai välillisesti, eikä myöskään tehdä päätelmiä yksittäisestä henkilöstä. Anonymisoitava aineisto on käsiteltävä niin, että siitä ei enää voida yhdistellä yksittäisen henkilön tietoja muihin aineistoihin. Lisäksi tulee huolehtia siitä, että aineiston palauttamisen takaisin sellaiseen muotoon, josta yksittäinen henkilö voisi olla mahdollista tunnistaa, tulisi olla mahdotonta tai se tulisi tehdä hyvin vaikeaksi toteuttaa. (Sosiaali- ja terveysministeriö 2020, 1–2.)

Anonymisointi voidaan toteuttaa karkeistamalla eli aggregoimalla aineistoa. Aggregoinnissa aineistosta muodostetaan tietoja yhdistämällä eri perusteilla summattua tietoa. Aggregointi ei kuitenkaan automaattisesti merkitse sitä, että aineistoa voitaisiin käsitellä anonymisoituna aineistona. Laajan tietoa-ineiston aggregointi liian tarkalla tasolla saattaa johtaa siihen, että yksittäinen henkilö voi olla siitä tunnistettavissa epäsuorasti jonkin tiedon, kuten harvinaisen sairauden perusteella. Tunnistamiseen saattaa riittää myös eri tietojen yhdistäminen, esimerkiksi yhdistämällä tieto henkilön asuinkunnasta ja ikäryhmästä. Aggregointi, eli tietojen karkeistaminen olisi tehtävä riittävän yleiselle tasolle, jotta yksittäisen henkilön tunnistaminen ei olisi enää mahdollista. (Kinanen 2021a.)

Käytännössä aggregointia toteutettaessa on pyrittävä käyttämään mahdollisimman yleisellä tasolla olevia tietojen luokituksia summaamisen perustana. Esimerkiksi henkilön ikätiedot voidaan aineistossa luokitella kolmeen luokkaan: 0–17-vuotiaat, 18–64-vuotiaat ja yli 65-vuotiaat. Näin karkealla luokittelulla yksittäisen henkilön paljastumisen riski ikänsä perusteella pienenee merkittävästi verrattuna yhden tai viiden ikävuoden porrastuksella tehtyyn luokitteluun.

Siltä varalta, että jokin taho tarkoituksellisesti yrittäisi tunnistaa henkilöitä anonymisoidun aineiston perusteella, tulee EU:n yleisen tietosuojasetuksen (2016, resitaali 26) mukaan anonymisoinnin riittävyyttä tarkasteltaessa arvioida tunnistamiseen todennäköisesti vaadittavia rahallisia kustannuksia ja aikaa. Tarkastelussa täytyy ottaa huomioon myös tietojen käsittelyyn käytettävissä olevan teknologian kehitys (EU:n yleinen tietosuojasetus 2016, resitaali 26). Käytännössä tämä tarkoittaa sitä, että anonymisoinnin riittävää tasoa tulee tarkastella aika ajoin uudelleen teknologian kehittyessä. Vaikka nykyteknologiaa käyttäen käsitellyn aineiston tiedot saattavat olla riittävästi anonymisoituja, tilanne voi muuttua tulevaisuudessa.

Tietosuojaperiaatetta ei sovelleta anonymisoituihin tietoihin (EU:n yleinen tietosuojasetus 2016, resitaali 26). Toisaalta kuitenkin anonymisoidun aineiston käyttö edellyttää sitä, että riittävän anonymisoinnin taso määritellään tapauskohtaisesti (Kinanen 2021a). Yksittäisen henkilön paljastumisriskiä ei juuri mitenkään voida laskea nolnaan ja samalla säilyttää luodun aineiston käyttökelpoisuus. Aineiston käsittelyn jälkeen jää siis aina jäljelle jonkinlainen paljastumisen jäännösriski, vaikka käsitellyn aineiston arvioitaisiinkin olevan riittävällä tasolla anonymisoitu (Kinanen 2021a).

Jäännösriskin minimointiin ja aggregoinnin toteuttamiseen Power BI:llä luotavissa raporteissa tutustutaan myöhemmin luvussa 6.4.

3.4 Sote-alan tiedolla johtaminen ja henkilötietoja sisältävien asiakastietojen käsittelyä koskeva lainsäädäntö

Sote-alan tiedolla johtamisen raporttien tuottamisen ja niiden käyttöoikeuksien jakamisen oikeille henkilöille täytyy tapahtua lainmukaisesti. Sote-palveluiden tuottamisessa kerätyn tiedon käyttö tiedolla johtamisen tarpeisiin perustuu lakiin sosiaali- ja terveystietojen toissijaisesta käytöstä (552/2019). Tästä laista käytetään myös nimitystä toisiolaki. Lisäksi tutustutaan lyhyesti myös potilastietojen ensisijaista käyttöä ohjaavaan lakiin sosiaali- ja terveydenhuollon asiakastietojen sähköisestä käsittelystä (784/2021). Tästä laista käytetään myös nimitystä asiakastietolaki.

3.4.1 Asiakastietojen toisiokäyttö

Asiakastietojen käyttö muuhun tarkoitukseen kuin siihen, mitä varten ne on alun perin kerätty, on mahdollista tietyin ehdoin. EU:n yleisen tietosuojasetuksen 6:n artiklan mukaan henkilötietojen käsittelyn lainmukaisuus voidaan perustella rekisterinpitäjän lakisääteisellä velvoitteella, yleistä etua koskevan tehtävän täyttämiseksi tai rekisterinpitäjälle kuuluvan julkisen vallan käyttämisellä (EU:n yleinen tietosuojasetus 2016, 6 artikla).

Kansallisessa lainsäädännössä sosiaali- ja terveystietojen tuottamisessa kerättyjen tietojen käyttäminen tietojohdallisessa perustuu lakiin sosiaali- ja terveystietojen toissijaisesta käytöstä (Toisiolaki 552/2019). Lain tavoitteina on toisaalta mahdollistaa sote-palvelutuotannossa tallennettujen tietojen tietoturvallinen ja tehokas käyttö ja toisaalta tietojen käsittelyyn liittyen turvata niiden henkilöiden oikeuksia, joiden tietoja on kerätty. (Toisiolaki, 1 §.)

Toisiolain 2 § nojalla tallennettuja sote-asiakastietoja voidaan käyttää seuraaviin käyttötarkoituksiin, vaikka niitä ei olisi alun perin tallennettu mainitussa tarkoituksessa:

- 1) tilastointi;
- 2) tieteellinen tutkimus;
- 3) kehittämis- ja innovaatiotoiminta;
- 4) opetus;
- 5) tietojohdaminen;
- 6) sosiaali- ja terveydenhuollon viranomaisohjaus ja -valvonta; sekä
- 7) viranomaisen suunnittelu- ja selvitystehtävä (Toisiolaki, 2 §.)

Jotta sosiaali- ja terveydenhuollon julkinen palvelunjärjestäjä voisi soveltaa toiminnassaan toisiolakia, sen on järjestettävä organisaatiossaan kuntoon toisiolain 10 §:ssa listatut palvelut. Näillä palveluilla, tai toiminnoilla, rakennetaan ympäristö toissijaisen käytön tietoturvaliselle käytölle ja mahdollistetaan

parempaa tietojen yhdistelyä ja esikäsittelyä. Lisäksi on huolehdittava pseudonymisoidujen henkilötunnusteiden hallinnasta siten, että pseudonymisoidut aineistot voidaan tarvittaessa palauttaa takaisin tunnusteelliseksi. (Toisiolaki, 10–17 §.) Pseudonymisoidujen aineistojen käytössä on myös huomioitava, että pseudonymisoidut tiedot tulee pseudonymisoida uudelleen eri tunnisteilla jokaista tietojen luovutusta kohden (Toisiolaki, 14 §, 4. mom.) Pseudotunnusteiden suojaamista käsitellään tarkemmin luvuissa 4.2 ja 5.5.3.

Toisiolaki vaatii, että tietoja on käsiteltävä tietoturvallisesti ja tietosuojamääräyksiä noudattaen. Tietojen käyttöympäristöjen tietoturvallisuutta tulee parantaa kartoittamalla riskejä ja niihin tulee toteuttaa asianmukainen pääsynhallinta ja valvonta. (Toisiolaki, 18 §.)

3.4.2 Käyttöoikeus asiakastietoon

Käyttöoikeudet sähköisesti tallennettujen asiakastietojen käsittelyn tulee lain mukaan rajoittaa ammatilliselle ”vain työtehtävissään tarvitsemiinsa välttämättömiin asiakastietoihin, joihin hänellä on tiedonsaantioikeus”. Käyttöoikeuksien myöntäminen asiakastietojen käyttöön ja käyttöoikeuksiin liittyvä kirjaaminen ja seuranta ovat palvelunantajan, eli sosiaali- tai terveyspalveluiden järjestäjän vastuulla. (Asiakastietolaki 784/2021, 15 §.) Palveluiden järjestäjällä tulee siis olla tarkat tiedot siitä, mihin asiakastietoihin kullakin työntekijällä on käyttöoikeuksia. Myönnettävät käyttöoikeudet perustuvat sosiaali- ja terveysministeriön antamaan asetukseen, jossa määritellään tarkemmin, mitä tietoja ammattihenkilöt ovat työtehtävänsä tai antamansa palvelun perusteella oikeutettuja käyttämään (Asiakastietolaki, 15 §, 2 mom.).

Asiakastietolaki velvoittaa myös palveluiden järjestäjän keräämään lokitietoja kaikesta asiakastietojen käytöstä. Tämän niin sanotun luovutuslokirekisterin pitäminen mahdollistaa sen, että asiakkaat voivat pyytää tietoja omien tietojensa käytöstä. Asiakas voi tietopyynnöllä voi selvittää, kuka tai ketkä ovat katsoneet hänen tietojensa tietojärjestelmästä ja mistä syystä. (Asiakastietolaki, 25–26 §.)

Edellä mainitussa asiakastietolaissa asiakastietoja tarkastellaan tietojen ensisijaisen käytön, eli palveluiden tuottamisen näkökulmasta. Tästä huolimatta lain vaatimuksia käyttöoikeuksien minimoimisesta ja kirjaamisesta voi mielestäni hyvin soveltaa myös tietojohdamisessa, vaikka se luokitellaankin tiedon toisiokäytöksi. Käyttöoikeuksien rekisteröinti ja raportin jakelun rajaaminen pienentävät vahinkoja sellaisessa valitettavassa tapauksessa, että jäännösriski olisi realisoitunut tai arvioitu väärin.

4 TIEDOLLA JOHTAMISEN ARKKITEHTUURI SOTE-ALALLA

Asiakastietojen käyttöön tiedolla johtamisessa liittyy riskejä. Näitä riskejä voidaan pienentää erilaisilla toimenpiteillä tietojen käsittelyn eri vaiheissa. Tässä luvussa esitetään tapoja, joilla riskejä voidaan pienentää raportointiaineiston varastointiin liittyvällä arkkitehtuurilla.

Julkisen sote-alan tiedolla johtamiseen on ryhdytty kehittämään kansallisia toimintamalleja Sosiaali- ja terveysministeriön Toivo-ohjelmassa (Sote-uudistus). Osa Toivo-ohjelmaa on Virta-hanke, jonka puitteissa on kehitetty kansallista arkkitehtuurimallia sote-asiakastietojärjestelmissä olevan tiedon käsittelyyn ja varastointiin (Sote-uudistus; Kinanen 2021a; Kinanen 2021b).

4.1 VIRT A-arkkitehtuuri

Virta-verkoston VIRT A-arkkitehtuuri on suunniteltu nimenomaisesti tietojohdamisen arkkitehtuurimalliksi ja sen suunnittelussa on haluttu alusta lähtien huomioida henkilötietojen asianmukainen käsittely (Kinanen, 2021b, 4). Kuvassa 1 on yksinkertaistettu kuvaus VIRT A-arkkitehtuurimallista. Mallissa tietoja siirretään automatisoidusti lähdejärjestelmistä tietovarastoon.

Aluksi tiedot siirretään tietoaltaaseen, johon kerätään tietoja eri lähdejärjestelmistä. Tietoaltaan ensimmäisen kerroksen jälkeen tiedot pseudonymisoidaan automatisoidusti ja näin käsitellyt tietotaulut tallennetaan tietoaltaan toiseen kerrokseen. Tämän jälkeen tiedot siirretään tietovarastoon, josta eri lähdejärjestelmistä koottuja tietoja voidaan ryhtyä yhdistämään toisiinsa. Lopulta tietovarastoa käytetään lähteenä pseudonymisoitujen tai anonymisoitujen tietojoukkojen luomiseen ja näitä tietojoukkoja puolestaan käytetään tuottamalla raportteja esimerkiksi dataa visualisoivilla ohjelmistoilla. (Kinanen 2021a; Kinanen 2021b, 5.)



KUVA 1. Yksinkertaistettu kuvaus tiedolla johtamisen VIRTIA-arkkitehtuurimallista (mukaiillen Kinanen 2021b, 5)

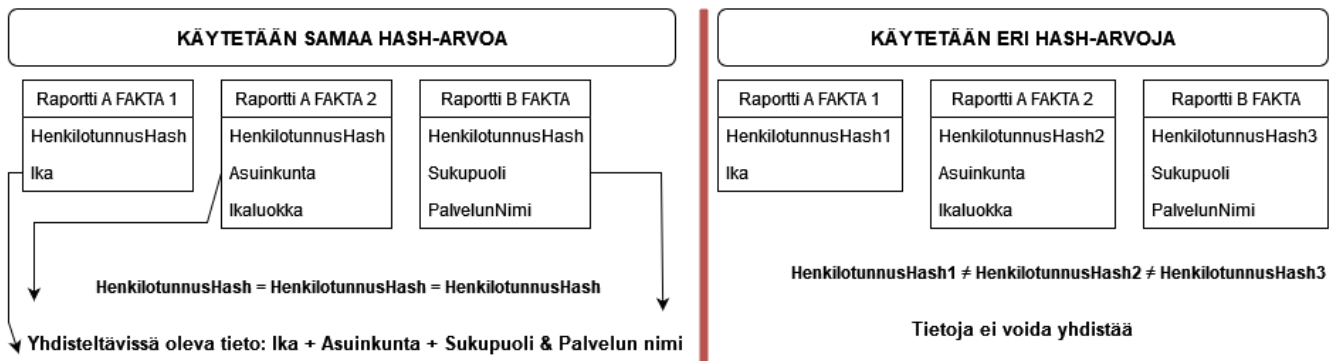
Tietovarasto toimii tietojen yhdistämisen keskuksena, johon kaikki raporteissa käytettävä tieto ensin ohjataan. Tietovarastoon voidaan kerätä tietoja myös muista lähteistä kuin palveluntuotannossa käytävistä asiakastiedon tallennusjärjestelmistä. Muita käyttökelpoisia lähteitä voisivat olla esimerkiksi taloushallinnon ja henkilöstöhallinnon järjestelmät tai vaikkapa luettelot toimipisteiden GPS-koordinaateista karttavisuaalisointeja varten.

Koska raportoinneissa käytettävä aineisto otetaan käyttöön tietovarastosta, käytetään raporttien taustaineistona valmiiksi pseudonymisoitua tietoa. Vahinko henkilötunnuksen lipsahtamisesta raportointiaineistoon voidaan estää siis jo sillä, että raportointiin käytettävissä tauluissa ei käytetä henkilötunnuksia.

4.2 Pseudonymisoidun tiedon käyttö eri raportointiaineistoissa

Luvussa 4.1 kuvatun VIRTIA-arkkitehtuurin mukaan raportointiin käytettävät tiedot otetaan käyttöön tietovarastosta, joka sisältää pseudonymisoiduilla yksilötunnisteilla varustettua tietoa. Tätä tietoa voidaan harkinnan mukaan käyttää raportoinnissa, mutta tällöin on pyrittävä varmistamaan, että pseudotunnisteita ei voida yhdistellä muiden raportissa käytettävien pseudotunnisteiden kanssa. (Kinanen 2021a; Kinanen 2021b, 25–27.)

Otetaan kuvitteellinen esimerkki (KUVA 2, vasen puoli): Jos tietokantaan tallennettua henkilötunnuk- sen hash-arvoa käytetään useassa raportin faktataulussa, faktataulujen tietoja voidaan yhdistää toisiinsa keskenään yhtenevän hash-arvon perusteella. Jos toisessa faktataulussa tunnisteeseen on liitetty ikä ja toisessa asuinkunta, saadaan tiedot yhdistämällä selville kyseiseen hash-arvoon liittyvä tieto ikä (72 vuotta) + asuinkunta (jokin pieni kunta). Jos myöhemmin tuotetaan raportti esimerkiksi mielenterveys- palveluiden käytöstä sukupuolen (mies) mukaan käyttäen samoja tietovaraston hash-tunnuksia, voi- daan aiemman raportin tietoa yhdistää uuden raportin tietoon. Jos esimerkkitapauksemme 72-vuotias mies asuu pienessä kunnassa, jossa hänen ikäluokkansa asukkaita on neljä ja heistä kolme on naisia, voidaan hänen henkilöllisyytensä selvittää epäsuorasti ja saada näin selville tietoja hänen terveydenti- lastaan.



KUVA 2. Henkilötunnuksen hash-arvon käyttö faktatauluissa

Jotta tietojen yhdistely eri faktataulujen välillä ei olisi mahdollista, tulisi eri tauluissa käyttää eri pseudotunnisteita (KUVA 2, oikea puoli). Käytännössä tämä voidaan tehdä ajamalla tietovarastosta poimittu hash-arvo uudelleen hash-funktion läpi. (Kinanen 2021a.) Hash-funktion käyttöä yksilötunnisteiden suojaamiseen tarkastellaan tarkemmin luvussa 5.5.3.

5 TIEDOLLA JOHTAMISEN RAPORTOINNISSA KÄYTETTÄVIÄ TYÖKALUJA

Tässä luvussa esitellään tiedolla johtamisen raporttien tuottamisessa käytettäviä Microsoftin ohjelmistoja. Näihin kuuluvat tietokantakyselyiden tekemiseen liittyvä SQL Server Management Studio, tietojoukkojen rakentamiseen ja tietojen visualisointiin käytettävä Power BI Desktop-ohjelma apuohjelmineen sekä raporttien julkaisualusta Power BI service. Lisäksi tutustutaan lyhyesti Power BI -raportin tuottamisessa käytettäviin kieliin SQL:ään ja DAX:iin. SQL-osiossa keskitytään erityisesti tietojen anonymisointiin käytettävien aggregointifunktioiden käyttöön sekä pseudotunnisteiden suojaamisen mahdollistavaan hash-funktioon.

5.1 Microsoft SQL Server Management Studio (SSMS)

Raportin tietolähde on hyvin usein relaatiotietokannassa. Relaatiotietokannoista tarvittavat tiedot kerätään lähettämällä tietokantapalvelimelle kysely. Palvelin vastaa palauttamalla kyselyssä listatut tiedot käyttäjän kyselyssä määrittelemässä muodossa.

SQL Server Management Studio on Microsoftin ohjelmisto, jolla voidaan suorittaa SQL-relaatiotietokantojen hallintaan liittyviä tehtäviä graafisen käyttöliittymän kautta. Ohjelmistolla voidaan tehdä monenlaisia tietokantapalvelimien ja tietokantojen hallintaan, valvontaan ja ylläpitoon liittyviä tehtäviä. (Microsoft 2020; Microsoft 2021a.)

Power BI -raportteja luotaessa tätä ohjelmistoa käytetään aineiston poiminnassa tarvittavien SQL-kyselyiden luomiseen, testaamiseen ja muokkaamiseen. SSMS-ohjelmassa käytetään Microsoftin omistamaa SQL-tietokantakielen laajennusta nimeltä Transact-SQL (T-SQL). T-SQL sisältää samat toiminnallisuudet kuin SQL, mutta lisäksi siinä on laajenuksena omia komentoja. Merkittävä ero on myös se, että toisin kuin SQL, joka käsittelee kyselyn rivi kerrallaan, T-SQL käsittelee koodin proseduraalisesti lohkoina. (Clark 2021.)

SQL Server Management Studion avulla SQL-kyselyiden toimivuutta on helppoa ja nopeaa testata, kunnes kysely tuottaa halutut tiedot halutussa muodossa. Ohjelman avulla on myös helppo nähdä näytteitä eri tauluissa olevista tiedosta. Tästä on paljon apua monimutkaisempia tietojoukkoja luotaessa,

kun etsitään eri tauluista toisiaan vastaavia avainsarakkeita ja niiden välityksellä yhdistetään eri tauluissa olevia tietoja toisiinsa.

5.2 Microsoft Power BI

Tietolähteestä noudettavia tietoja voidaan muokata helpommin ymmärrettävään muotoon muodostamalla tiedoista taulukoita, matriiseja, pylväs- tai ympyräkaavioita, karttoja tai muita graafeja. Tietojen esilletuontiin voidaan käyttää Microsoft Power BI -ohjelmistoperhettä, jonka avulla taulukkomuotoisista tiedoista voidaan muodostaa käyttäjän valintojen mukaan dynaamisesti päivittyviä raporttinäkymiä.

Microsoft Power BI koostuu valikoimasta ohjelmia ja verkkosovelluksia. Raportin valmistelutyö tehdään Power BI Desktop -ohjelmalla ja raportteja tai valmisteltuja tietojoukkoja voidaan tarkastella ja jakaa pilvipalvelu Power BI servicen kautta. Tuoteperheeseen kuuluu myös mobiililaitteille suunnattu sovellus Power BI Mobile. Power BI servicessä voidaan tarkastella raportteja yksittäin tai hyödyntää käytettävissä olevia raportteja yhdistelemällä niiden tietoja koontinäyttöihin (englanniksi dashboard). Koontinäyttöjä on myös mahdollista jakaa muiden käyttäjien kanssa. (Mikä Power BI on?) Power BI Desktop sisältää yli 30 erilaista visualisointia. Lisäksi visualisointeja voi ladata lisää AppSource-verkkopalvelusta (Tietojen mallinnus ja visualisointi).

Power BI:n käyttö perustuu käyttäjäkohtaisiin käyttöoikeuksiin, joita on kolmea tyyppiä. Ilmainen Power BI- käyttöoikeus mahdollistaa muiden tuottaman sisällön tarkastelun, mutta julkaiseminen on mahdollista ainoastaan upottamalla raportti julkiselle verkkosivulle. Maksulliset käyttöoikeustyyppit Power BI Pro ja Power BI Premium tarjoavat vastineeksi enemmän vaihtoehtoja raporttien julkaisemiseen sekä myös näille käyttöoikeustasoille varattuja keinoja raporttiaineiston sisältöjen suojaamiseen. (Microsoft 2021b; Microsoft 2021c.)

Tietojen tarkastelu Power BI:llä eri visualisaatioiden kautta perustuu tietojen suodattamiseen käyttäjän tekeminen valintojen perusteella. Käyttäjä voi klikata visualisoinnin osaa tai erillistä osittajatyökalua ja näin määrittellä millä, perusteella visualisoinnin tiedot suodatetaan. (Tietojen mallinnus ja visualisointi.)

Parillinen	Luku
<input type="checkbox"/> Ei	1
<input type="checkbox"/> Kyllä	2
<input type="checkbox"/> Ei	3
<input type="checkbox"/> Kyllä	4
<input type="checkbox"/> Ei	5
<input type="checkbox"/> Kyllä	6
<input type="checkbox"/> Ei	7
<input type="checkbox"/> Kyllä	8

Parillinen	Luku
<input checked="" type="checkbox"/> Ei	1
<input type="checkbox"/> Kyllä	3
<input type="checkbox"/> Ei	5
<input type="checkbox"/> Kyllä	7

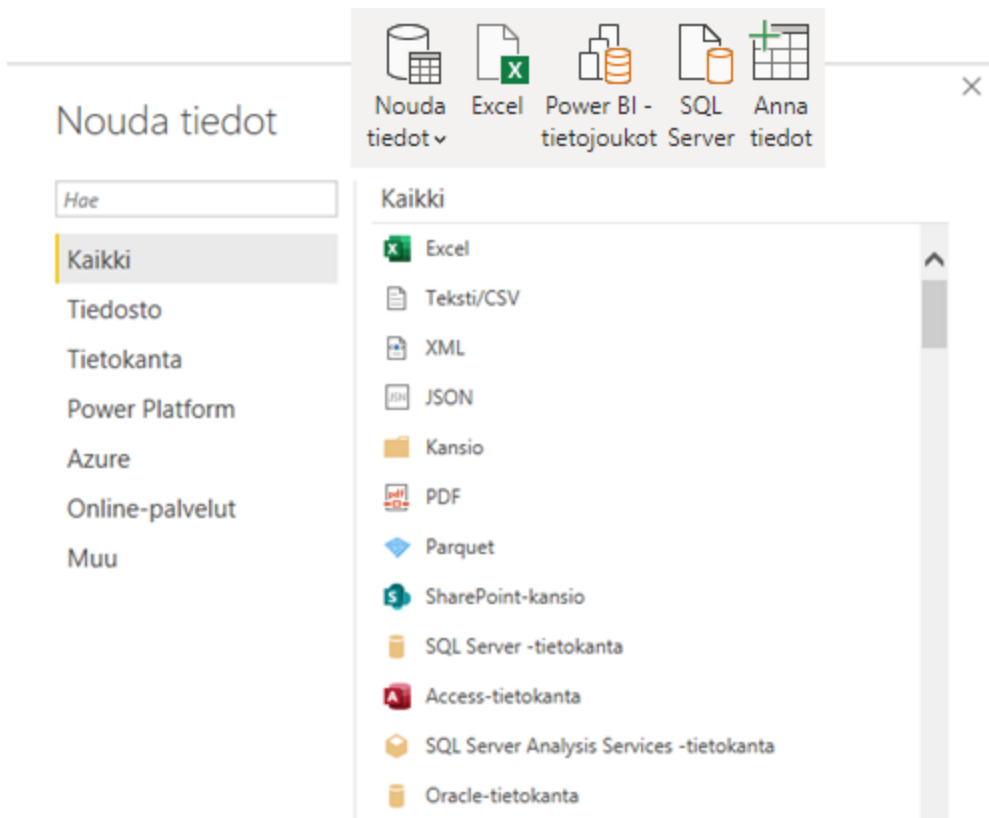
Parillinen	Luku
<input type="checkbox"/> Ei	2
<input checked="" type="checkbox"/> Kyllä	4
<input type="checkbox"/> Ei	6
<input type="checkbox"/> Kyllä	8

KUVA 3. Tietojen suodattaminen osittajalla

Raportin kehittäjän kannalta tärkeitä raportin elementtejä ovat kirjanmerkit ja painikkeet. Niiden avulla raporttisivun sisälle tai niiden välille voidaan luoda valmiita polkuja erilaisiin näkymiin. Käyttäjä pääsee siirtymään kehittäjän luomaan näkymään yksinkertaisesti painamalla raportilla painiketta, johon on kiinnitetty toiminnoksi siirtyminen kirjanmerkkiin tai toiselle raporttisivulle. Myös raporttiin lisättyihin kuviin voidaan kiinnittää toimintoja samalla tavalla kuin painikkeisiin. (Saxton 2021.)

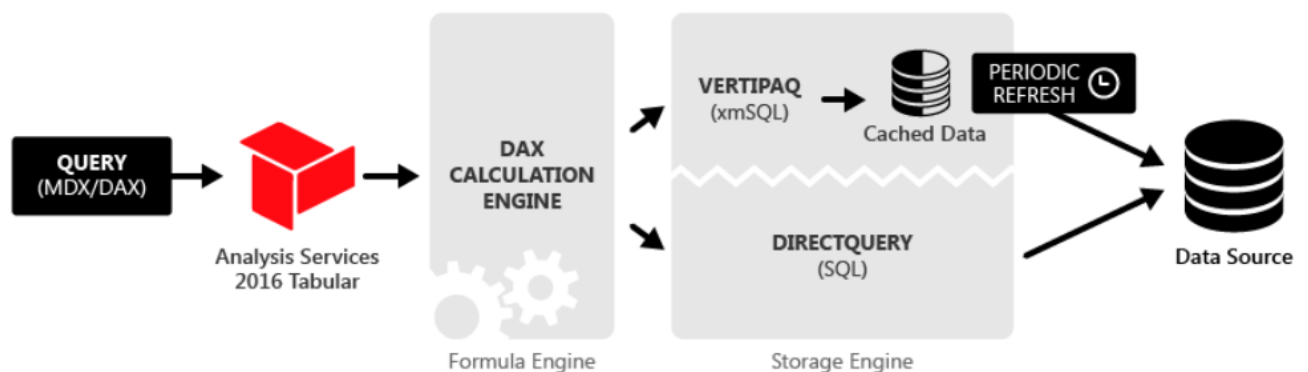
5.2.1 Tietojen tuonti ja tallennus Power BI -tiedostoon

Power BI:hin voidaan tuoda kattavasti eri muodoissa olevia tietolähteitä eri tiedostomuodoista erilaisiin tietokantamalleihin ja pilvipalveluihin (KUVA 4). Tietolähde voi siis olla joko raporttikehittäjän työasemalla, verkkokansiossa, tietokannassa tai pilvipalvelussa. Tietolähteenä voi käyttää myös aikaisemmin app.powerbi.com -pilvipalveluun ladattuja tietojoukkoja. ”Anna tiedot” -valinnalla käyttäjä voi syöttää pienehkön taulukon verran tietoja suoraan Power BI:hin. Kaikkiaan Power BI voidaan yhdistää kymmeniin erityyppisiin tietolähteisiin. Kuvassa 4 näkyy joitakin esimerkkejä käytettävissä olevista tietolähdetyypeistä.



KUVA 4. Tietojen noutamisen valintanauha ja tietolähteen valintaruutu

Russon ja Ferrarin (2020) mukaan Power BI tallentaa ja käsittelee raporttitiedostoon tuotavia tietoja taulukkomallissaan (englanniksi Tabular-model). Taulukkomallissa tietoja käsitellään kaavamoottorin (formula engine) ja varastointimoottorin (storage engine) kautta. Jokainen visualisointi muodostuu siihen liittyvän kyselyn perusteella. Kysely kulkee kaavamoottorin kautta varastointimoottorille, joka palauttaa vastauksen kyselyyn. (Russo & Ferrari 2020, 546.)



KUVA 5. Yksittäisen kyselyn käsittely Tabular-taulukkomallissa (Russo & Ferrari 2020, 546; Russo 2020)

Kyselyn vaatimat tiedot voidaan ottaa käyttöön kahdella eri tavalla. ”Tuontitilassa” (englanniksi Import) tiedoista luodaan kopio raporttiedostoon ja tietoja käytetään työaseman muistista. Tällöin tiedot tallennetaan käyttämällä Vertipaq -moottoria. Vertipaq -varastointimoottori pakkaa tietoja siten, että tarvittava muistin määrä pienenee, mutta tiedot ovat kuitenkin käsiteltävissä nopeasti. Vertipaq -moottori on monisäikeinen ja kykenee myös hyödyntämään työaseman resursseja tehokkaasti käyttämällä yhtäaikaaisesti useita prosessoriytimiä. (Russo & Ferrari 2020, 546–549.) Tietojen käsittely Vertipaq-moottorilla on siis useimmiten nopeaa suuriakin tietoaaineistoja käsiteltäessä. Tietojen uudelleenlataus suoritetaan tarpeen mukaan tai ne voidaan asettaa tapahtumaan ajastettuna Power BI-pilvipalvelussa (Microsoft 2021d).

Toinen vaihtoehto tietojen tuontiin on käyttää ”Direct Query” -tilaa. Tällöin Power BI hakee tiedot kyselyhetkellä suoraan tietolähteestä ennakkoon syötettyjen lähdekyselyiden ja käyttöoikeuksien perusteella. Tietoja ei tallenneta työasemalle, vaan tiedot ladataan palvelimelta uudelleen aina, kun käyttäjä visualisointia klikkaamalla luo uuden kyselyn. (Russo & Ferrari 2020, 546–549; Microsoft 2021d.) ”Direct Query” -tilassa tietojen hakemisen nopeus riippuu luonnollisesti työaseman ja tietokantapalvelimen välisestä tietoliikenneyhteydestä sekä palvelimen kuormituksesta.

SQL Server -tietokanta

Palvelin ⓘ

Tietokanta (valinnainen)

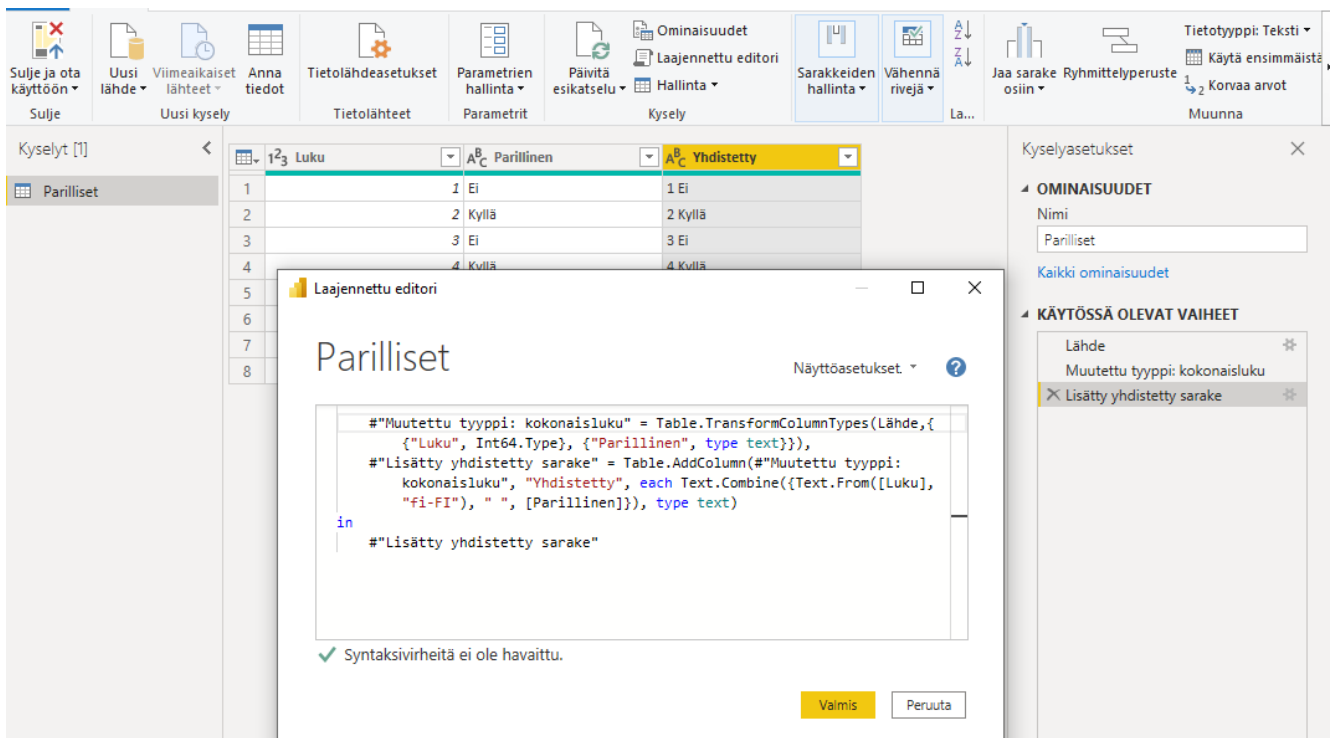
Tietoyhteystila ⓘ
 Tuo
 DirectQuery

⌵ Lisäasetukset
 Komennon aikakatkaisu minuutteina (valinnainen)

SQL-lauseke (valinnainen, edellyttää tietokantaa)

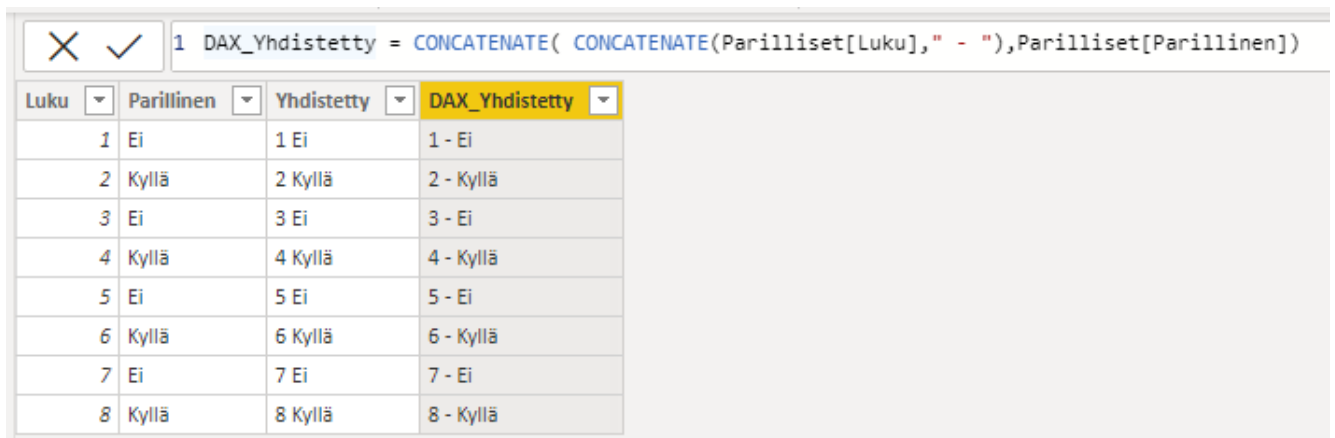
KUVA 6. Valintaruutu, jolla tuodaan tietoja SQL Server -tietokannasta

Tuotuja tietoja voi muokata Power BI:n Power Query -työkalulla. Tietojen asetuksia, kuten esimerkiksi sarakkeiden tietotyyppiä, voi muokata joko graafisen käyttöliittymän kautta klikkaamalla tai kirjoittamalla editoriin lausekkeita M-kielillä. Kuvassa 7 tuotuihin tietoihin on lisätty uusi sarake yhdistämällä kahden sarakkeen tiedot käyttämällä M-kieltä.



KUVA 7. M-kielen käyttäminen tietojen tai sarakkeiden muokkaamiseen Power Query -editorissa

Varastointimoottori vastaa lähdetietojen tuonnista ja varastoinnista. Tämän lisäksi visualisoinneissa tarvitaan tietoihin liittyviä laskentakaavoja eli mittareita. Ne lisätään tietomalliin käyttämällä DAX-kieltä. DAX, eli Data Analysis eXpressions on ohjelmointikieli, jota käytetään Microsoftin tuotteissa, kuten Power BI:ssä ja Power Pivot for Excelissä (Russo & Ferrari 2020, 1). DAX-kieltä voi käyttää luomaan mittareiden lisäksi myös uusia sarakkeita taulukkoihin. Kuvan 8 esimerkissä on yhdistetty sarakkeiden tiedot käyttämällä DAX-kieltä.

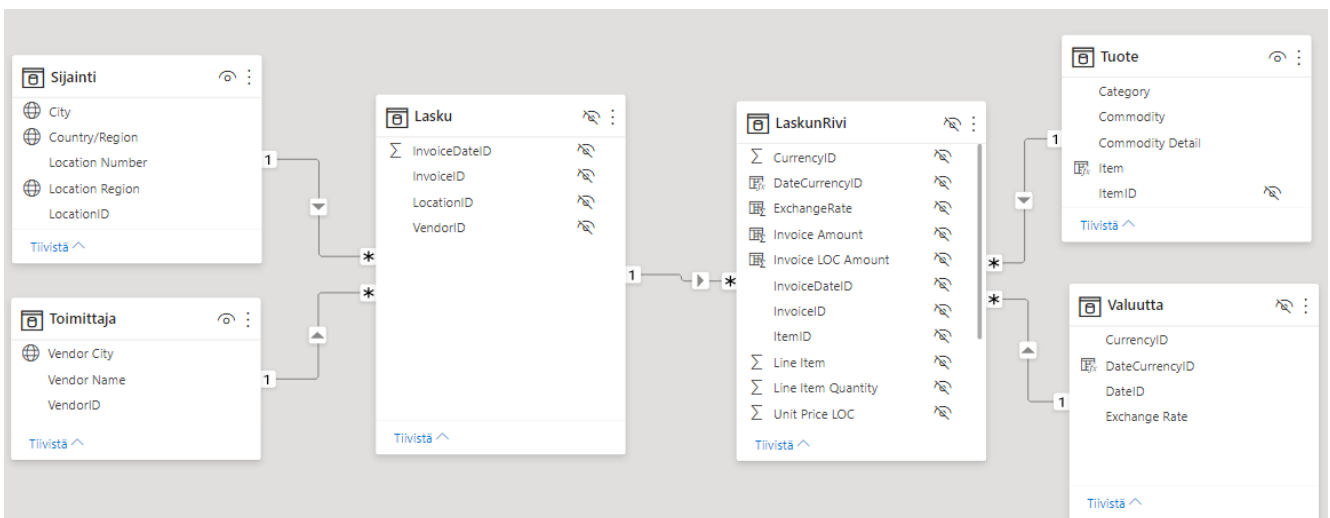


KUVA 8. DAX-kielen käyttäminen uuden sarakkeen luomiseksi

5.3 Tietojen yhdistäminen toisiinsa ja tietomallin rakenne

Tietomalliin voidaan tuoda tietoa tietokannoista tai muista lähteistä. Erilliset taulut kytketään toisiinsa avainsarakkeiden avulla. Avainsarakkeet muodostavat useimmiten yksi-moneen -suhdetta, eli dimensiotaulun ainoastaan yhdellä rivillä olevan avaimen arvo löytyy faktataulusta usealta eri riviltä. Power BI:n tietojen suodatus perustuu avainten kytkemisen lisäksi myös kytkemisten suuntaan. Yksinkertaisin ja tehokkain tietomallin rakenne on tähtimalli, jossa kaikki dimensiotaulut liitetään yhteen faktatauluun. (Ferrari & Russo 2017, 217–224.)

Kuvassa 9 on esimerkki raportin tietomallin rakenteesta. Kuvassa olevat nuolten suunnat osoittavat, kumpaan suuntaan Power BI suodattaa tietoja. Dimensiotaulut Sijainti ja Toimittaja suodattavat faktataulua Lasku yhden (1) suhde moneen (*) -kytkennällä. Dimensiotaulut Tuote ja Valuutta suodattavat faktataulua LaskunRivi. Tähtimallia noudattavat yhdistelmät Lasku–Sijainti–Toimittaja sekä LaskunRivi–Tuote–Valuutta on kytketty toisiinsa yhteisellä avaimella InvoiceID. Yhdessä koko malli muodostaa lumihjutaleeksi kutsutun mallin, jossa tähtimalliin liittyy lisää dimensioita, eivätkä taulut enää ole suorassa kytköksessä toisiinsa (Ferrari & Russo 2017, 222–223).



KUVA 9. Esimerkki Power BI- tiedoston tietomallista

5.4 Power BI -apuohjelmat

Power BI Desktopiin on saatavilla ohjelman käyttöliittymään integroitavia kolmannen osapuolen apuohjelmia, joiden avulla voidaan tehdä tehtäviä joihin Power BI ei itsessään taivu. Nämä apuohjelmat toimivat yhdistämällä itsensä avoimena olevan Power BI-tiedoston tietomalliin.

5.4.1 DAX Studio

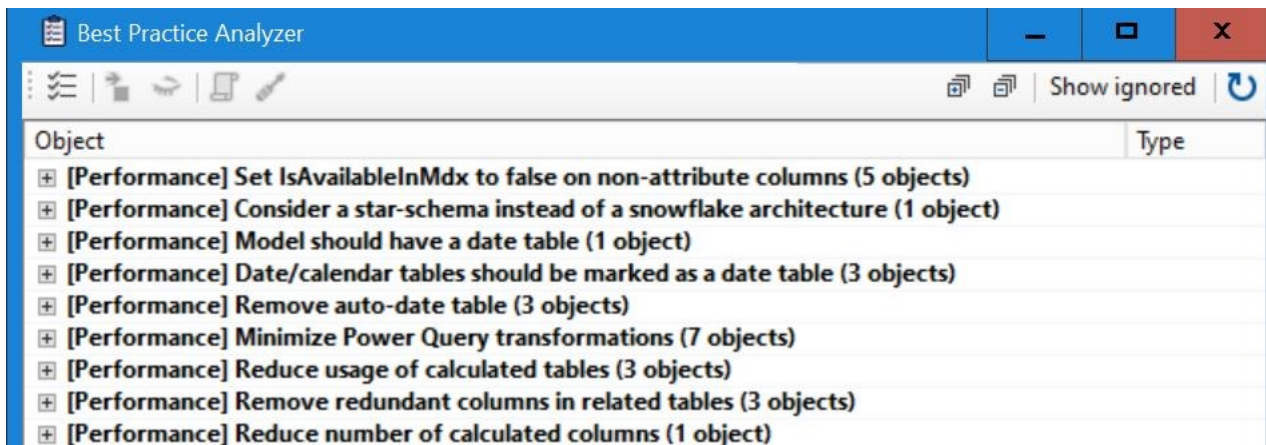
DAX Studio on apuohjelma DAX- lausekkeiden rakentamiseen ja testaamiseen. DAX Studio helpottaa pitkien DAX-lausekkeiden kirjoittamista ja mahdollistaa yksittäisten lausekkeiden ajamiseen vaadittavan ajan yksityiskohtaisen mittaamisen. (DAX Studio 2021.) Lisäksi DAX Studio sisältää Vertipaq Analyzer -työkalun, jolla voidaan tutkia yksityiskohtaisesti Power BI tietomallin tallentamiseen liittyviä tietoja. Työkalun avulla voidaan esimerkiksi selvittää, kuinka paljon resursseja eri tietomallin taulut ja niissä olevat sarakkeet käyttävät. Tämä tieto on tärkeää, kun tietomallin toimintaa halutaan optimoida toimimaan nopeammin. Ohjelman avulla voidaan paikallistaa, mitkä tietomallin osiot vievät eniten resursseja ja näiden tietojen perusteella voidaan lähteä poistamaan eniten resursseja vieviä tietoja, ellei niitä aivan ehdottomasti tarvita. (Russo 2021. "Data Model Size with Vertipaq Analyzer".)

Vertipaq Analyzer Metrics									
Tables Columns Relationships Partitions Summary									
Name	Cardinality	Table Size	Col Size	Data	Dictionary	Hier Size	Encoding	Data Type	RI Violations
▶ FAKTATAULU	50 472	4 288 002	4 206 682	749 136	2 997 970	459 576 Many	-	-	1
▲ Dimensiotaulu1	3 908	262 230	254 406	15 896	190 886	47 624 Many	-	-	1
Numerosarake	3 908	262 230	205 856	6 256	168 288	31 312	HASH	Double	-
AVAIN	3 908	262 230	23 600	7 824	120	15 656	VALUE	Double	-
Tekstisarake	22	262 230	18 754	360	18 170	224	HASH	String	-
kuukausi	12	262 230	2 616	1 088	1 384	144	HASH	Int64	-
Avain2	22	262 230	2 040	360	1 456	224	HASH	Double	-
vuosi	3	262 230	1 420	8	1 348	64	HASH	Int64	-

KUVA 10. Yksittäisen tietotaulun kokoa ja ominaisuuksia voi tarkastella DAX Studion Vertipaq Analyzerin avulla

5.4.2 Tabular Editor

Tabular Editor on apuohjelma Power BI:n tietomallin muokkaamiseen. Ohjelman avulla voi helposti ja nopeasti muokata taulukoiden tai sarakkeiden ominaisuuksia ja muodostaa DAXilla kirjoitetuista mittareista mittariryhmiä. Ohjelma sisältää myös työkalun Best Practice Analyzer. Sen avulla voidaan tarkastella tietomallin tilaa tietomallin rakentamisen hyvien käytäntöjen näkökulmasta ja muokata tehokkaasti tietomallin asetuksia skriptaamalla. (Otykier 2021. "Tabular Editor"; Best Practice Analyzer)



Object	Type
+ [Performance] Set IsAvailableInMdx to false on non-attribute columns (5 objects)	
+ [Performance] Consider a star-schema instead of a snowflake architecture (1 object)	
+ [Performance] Model should have a date table (1 object)	
+ [Performance] Date/calendar tables should be marked as a date table (3 objects)	
+ [Performance] Remove auto-date table (3 objects)	
+ [Performance] Minimize Power Query transformations (7 objects)	
+ [Performance] Reduce usage of calculated tables (3 objects)	
+ [Performance] Remove redundant columns in related tables (3 objects)	
+ [Performance] Reduce number of calculated columns (1 object)	

KUVA 11. Raportin tietomallin tarkastelua Best Practice Analyzerilla

5.5 Power BI -raportin tuottamisessa käytettäviä SQL- ja DAX- funktioita

Tässä luvussa esitellään lyhyesti useimmin raportin tuottamisessa käytettyjen SQL- ja DAX-funktioiden toimintaa. SQL-kieltä käytetään tietojen poimintavaiheessa määrittelemään mitä tietoja Power BI -tietomalliin sisällytetään. Lisää esimerkkejä SQL-lauseista esitellään luvussa 6.4.

DAX-kieltä käytetään Power BI -ohjelman sisällä luomaan liiketoimintalogiikkaan liittyviä laskenta-kaavoja tai muodostamaan tauluihin DAX-kaavalla laskettuja sarakkeita tai kokonaisia taulukoita. Lisää DAX-esimerkkejä esitellään luvussa 6.2.

5.5.1 SQL-kyselyn rakenne

Tietojen poiminnan suorittava kysely aloitetaan lähes aina SELECT-kutsulla, jolla määritellään kyselyyn sisällytettävät sarakkeet (Beaulieu 2009, 43). Sarakkeiden nimet luetellaan pilkulla erotettuna SELECT-kutsun alle. Sarakelista päättyy sanaan FROM, jonka jälkeen määritellään mihin kohteeseen tai kohteisiin kysely suuntautuu. FROM sanaa seuraa kyselyn päätaulu. Päätauluun voidaan liittää tietoja muista tauluista erilaisilla JOIN-komennoilla. (Beaulieu 2009, 81–86; 183–190.)

Tavallisimmin poiminnassa halutaan liittää muista tauluista tietoja päätaulusta poimittavien rivien jatkoksi. Tulosjoukkoon ei kuitenkaan haluta muista tauluista sellaisia tietoja, joita ei rivillä olevien

avainsarakkeiden perusteella voida liittää kuuluvaksi päätaulun riviin. Tällöin käytetään JOIN-komentoa LEFT OUTER JOIN. JOIN-komennolle määritellään yhdistettävän taulun sijainti sekä kyselyssä käytettävä nimitys yhdistettävälle taululle. ON-sanan jälkeen määritellään, mitä sarakkeita käytetään taulujen tietoja yhdistävänä avaimena. (Beaulieu 2009, 183–190.) Yhden taulun tietoja päätauluun lisäävä komentosarja rakentuisi siis alla olevan esimerkkikoodin mukaan:

```
FROM tietokanta.kayntitiedot_taulu kaynnit
LEFT OUTER JOIN tietokanta.yksikko_taulu yksikko
ON kaynnit.yksikkoKoodi=yksikko.Avain
```

KOODI 1. FROM-lause LEFT OUTER JOIN- poiminnalla päätaulusta ja yhdestä lisätaulusta

FROM-lauseen jälkeen kyselylle voidaan asettaa rajaavia ehtoja määrittelemällä rajoitukset WHERE-sanan jälkeen. Jos ehtoja on useita, käytetään niiden välissä operaattoreita AND tai OR. Jos samaan sarakkeeseen sisältyy monta ehtoa, kuten lista kelvollisista arvoista, voidaan ehdot asettaa sulkeisiin IN-operaattorin taakse. (Beaulieu 2009, 11; 71–72.)

5.5.2 SQL aggregointifunktiot

Tietojoukkojen luomisessa halutaan yleensä pienentää tulosjoukon kokoa mahdollisimman pieneksi, jotta se toimisi mahdollisimman tehokkaasti. Tähän voidaan käyttää SQL-kielen aggregointifunktioita. Aggregointi auttaa pitämään raporttitiedoston koon pienempänä, mutta erityisen tärkeää aggregointi on sote-alan tietojohdamisen raportoinnissa tietojen anonymisoinnin kannalta. Aggregoidessa osa tietokantaan tallennetuista tiedoista jää siirtymättä kyselyn tulosjoukkoon. Power BI-raporttien tuottamisessa tärkeimpiä aggregointifunktioita ovat COUNT ja SUM. COUNT laskee tulosjoukon arvojen lukumäärän ja SUM laskee arvojen summan.

Otetaan esimerkiksi taulukko, jossa on sarakkeessa T arvot 1, 2, 3, 4, 4, 6, 7, 12, 12, 29.

SUM-funktion käyttö sarakkeeseen T, eli SUM(T) tuottaisi tulosjoukkoon lukujen summan 80.

COUNT(T) tuottaisi tuloksen 10, joka on sarakkeessa olevien arvojen lukumäärä. Pelkälle COUNT-funktiolla on harvemmin käyttöä, mutta COUNTin yhdistäminen yksittäisiä tuloksia merkitsevään DISTINCT-funktioon mahdollistaa koodin 2 esimerkin mukaisesti vaikkapa yksittäisten potilaiden (eli erilaisten henkilötunnisteiden) laskemisen sarakkeesta. (Beaulieu 2009, 143–148.)

```
COUNT(DISTINCT(henkilotunniste)) as henkiloLKM
```

KOODI 2. Yksittäisten potilaiden määrän laskeminen COUNT(DISTINCT()) – yhdistelmällä

Tällä tavoin saadaan siis laskettua yksittäisten potilaiden määrä, riippumatta siitä kuinka monta kertaa sama henkilötunniste esiintyisi sarakkeessa. Aggregointifunktioiden käyttö vaatii usein tiedon ryhmittelyä GROUP BY -lauseessa. Ryhmittelyllä määritellään summaamisen perusteet. (Beaulieu 2009, 143–152.) Esimerkiksi luvun 6.4.3 esimerkkikoodilla 9 lasketaan yksittäisten henkilötunnisteiden määrä jokaiselle vuosi-kuukausi-yhdistelmälle ja tulosjoukkoon palautetaan jokaiselle vuosi-kuukausiriville sitä vastaava tunnisteiden määrä.

5.5.3 HASHBYTES-funktio

T-SQL-ympäristössä yksilöllisiä henkilötunnuksia voidaan suojata ja eriyttää muissa raporttiaineistoissa käytettävistä tunnuksista käyttämällä HASHBYTES-funktiota. Funktiolle määritellään käytettävä hash-algoritmi sekä annetaan suojattavan sarakkeen nimi sekä suojattavaan merkkijonoon liitettävä merkkijono eli ”suola”. (Microsoft 2022a; Kelley 2014.)

Kun eri raportointiaineistoissa käytetään toisistaan poikkeavaa suolan arvoa, saadaan eri aineistoihin päätyvistä hash-arvoista erilaisia, vaikka taustalla oleva tunnus onkin sama. Alla olevalla koodilla 3 henkilötunnus 010101-XXXX ajetaan hashbytes-funktion läpi erilaisilla suolan arvoilla. Hash-funktion tulos esitetään kuvassa 12.

```
SELECT
  '010101-XXXX' as Henkilotunniste
  ,HASHBYTES('SHA2_256', ('010101-XXXX')) as EiSuolaa
  ,HASHBYTES('SHA2_256', ('010101-XXXX'+ 'suolaa')) as OnSuolaa
  ,HASHBYTES('SHA2_256', ('010101-XXXX'+ 'EnemmanSuolaa')) as EnemmanSuolaa
```

KOODI 3. HASHBYTES-funktion käyttö

	Henkilötunniste	EiSuolaa	OnSuolaa	EnemmanSuolaa
1	010101-XXXX	0x717AB44C2B53A1...	0x9977ECC2D61A0B...	0x22F860E67C01162C076A058F...

KUVA 12. Koodin 3 tuottamia hash-arvoja henkilötunnisteelle

5.5.4 DAX-lausekkeet

Power BI -raporttia tehtäessä tulee hyvin usein eteen tarve laskea jonkin sarakkeen arvoja suodattamalla tulosjoukkoa jokin toisen sarakkeen ominaisuudella. Tähän tarpeeseen vastaa hyvin yleisesti käytetty funktio CALCULATE. Funktio ottaa ensimmäiseksi parametrikseen laskettavan sarakkeen tai mittarin nimen ja toisena parametrinä annetaan suodatuksen määrittelevä sääntö. (Russo & Ferrari 2020, 123.)

Yhdistämällä CALCULATE-funktioon muita DAX-funktioita saadaan sen avulla tehtyä melko monimutkaisiakin sääntöjä suodatukselle. Koodiesimerkissä (KOODI 4) ensin muodostetaan mittari, joka laskee käyntien lukumääräsarakeesta yhteen kaikkien sellaisten rivien arvot, joilla Ika-sarakkeen arvo on 10. Seuraava mittari käyttää ensimmäisenä parametrinään ensimmäistä mittaria ja lisää suodatinparametriksi sukupuolen.

CALCULATE korvaa suodatinparametrin arvolla samaan sarakkeeseen kohdistuvat muista lähteistä johtuvat suodatukset. Jos halutaan säilyttää muualta tulevat suodatukset ja ainoastaan lisätä suodatussääntöjä CALCULATE-lauseella, voidaan suodatinparametrin osuus kääriä KEEPFILTERS-funktion sisään. (Russo & Ferrari 2020, 137.) Tätä on sovellettu koodiesimerkin 4 viimeisessä lausekkeessa:

```
10_VuotiaidenKaynnit = CALCULATE(
    SUM('Kaynnit'[lukumaara]), 'Kaynnit'[Ika] = 10)
10_VuotiaidenTyttöjenKaynnit = CALCULATE(
    10_VuotiaidenKaynnit, 'Kaynnit'[Sukupuoli]= "N")
10_VuotiaidenTyttöjenKaynnit = CALCULATE(
    10_VuotiaidenKaynnit, KEEPFILTERS('Kaynnit'[Sukupuoli]= "N"))
```

KOODI 4. CALCULATE-funktion käyttö

Samaan taulukkoon liittyviä suodatussääntöjä voidaan ketjuttaa suodatinparametriin käyttämällä ja-operaattoria (&&), tai-operaattoria (||) tai IN-operaattoria ja aaltosulkumerkkejä. (Russo & Ferrari 2020, 137–139.)

```
KaynnitPTH= CALCULATE([KaynnitYhteensa],
    KEEPFILTERS('Kaynnit'[kustannuspaikka] IN {1,2,3,4,5} &&
    'Kaynnit'[Kayntilaji] IN {"A","B"}))
```

KOODI 5. Useiden suodatinparametrien antaminen CALCULATE-funktiolle

Muista ohjelmointikielistä tutut ehtolauseet IF ja SWITCH ovat käyttökelpoisia myös DAX-kielessä. IF-lauseen parametrit ovat järjestyksessä ”totuustesti”, ”jos tosi” ja ”jos epätosi”. SWITCH-lauseen parametrit ovat lauseke tai mittari, arvo ja arvosta johtuva tulos. (Russo & Ferrari 2020, 36–37; 47–48.) Viimeisenä on parametrina ”muuten”, joka toteutuu, jos mikään lauseen arvoista ei toteudu.

Alla esimerkit IF- ja SWITCH-lauseista:

```
= IF([lampotila]>= 0, "Plussakeli", "Pakkassää")
= SWITCH([ViikonPvNro], 1, "Ma", 2, "Ti", 3, "Ke", 4, "To"
, 5, "Pe", 6, "La", 7, "Su"
, "Tuntematon päivännumero")
```

KOODI 6. DAX-kielen IF ja SWITCH-lauseita

DAX-kielessä on myös useita funktioita, joilla voi tarkastella suodatusvalinnan tilaa. Näistä esimerkiksi HASONEFILTER tarkistaa onko sarakkeella täsmälleen yksi valittu suodatin (DAX Guide 2022a). SELECTEDVALUE puolestaan palauttaa sarakkeesta arvon, jos sarakkeella on suodatuksen jälkeen vain yksi arvo (DAX Guide 2022b).

Kun alla esitetty mittarikoodi liitetään kuvaajan otsikkokenttään, voidaan mittarilla saada kuvaajan otsikko päivittymään dynaamisesti valitun kuluryhmän mukaan. Jos valittuja kuluryhmiä on useita, ehtolause saa arvon epätosi.

```
OtsikkoKuluryhma= IF(HASONEFILTER('Taloustiedot'[Kuluryhma]),
SELECTEDVALUE('Taloustiedot'[Kuluryhma]),"Useita kuluryhmiä va-
littu")
```

KOODI 7. Esimerkki suodatinvalinnan tilaa tarkastelevien funktioiden käytöstä

6 SOTE-ALAN TIEDOLLA JOHTAMISEN RAPORTIN TUOTTAMINEN MICROSOFT POWER BI:LLA

Tässä luvussa tutkitaan sote-alan tiedolla johtamiseen tehtävän Power BI -raportin tuottamisprosessia. Raportti tuotetaan Keski-Pohjanmaan sosiaali- ja terveystalvokuntayhtymän Soiten toimialueista vastaavien johtajien käyttöön. Raportin sisältö ja lähtökohdat kuvaillaan pääpiirteiltään luvussa 6.2 ja raportin tuottamisen vaiheita kuvaillaan seikkaperäisemmin luvuissa 6.3–6.9. Raportille tuotetaan neljä raporttinäkymää: yksi kullekin toimialueelle. Teknisesti raporttinäkymien tuottaminen eri toimialueille on keskenään niin samankaltaista, että tässä työssä raportin tuottamista tarkastellaan ainoastaan terveyden- ja sairaanhoidon toimialueen raportointinäkökulman osalta.

Ennen sitä tutustutaan kuitenkin vielä hieman siihen, mitä tiedolla johtamisen raportoinnissa tulee huomioida riippuen siitä, mille palveluiden järjestämisen järjestelmätasolle raportointia tehdään.

6.1 Tiedolla johtamisen raportoinnin tarpeet julkisen sote- palveluiden järjestämisen eri järjestelmätasoilla

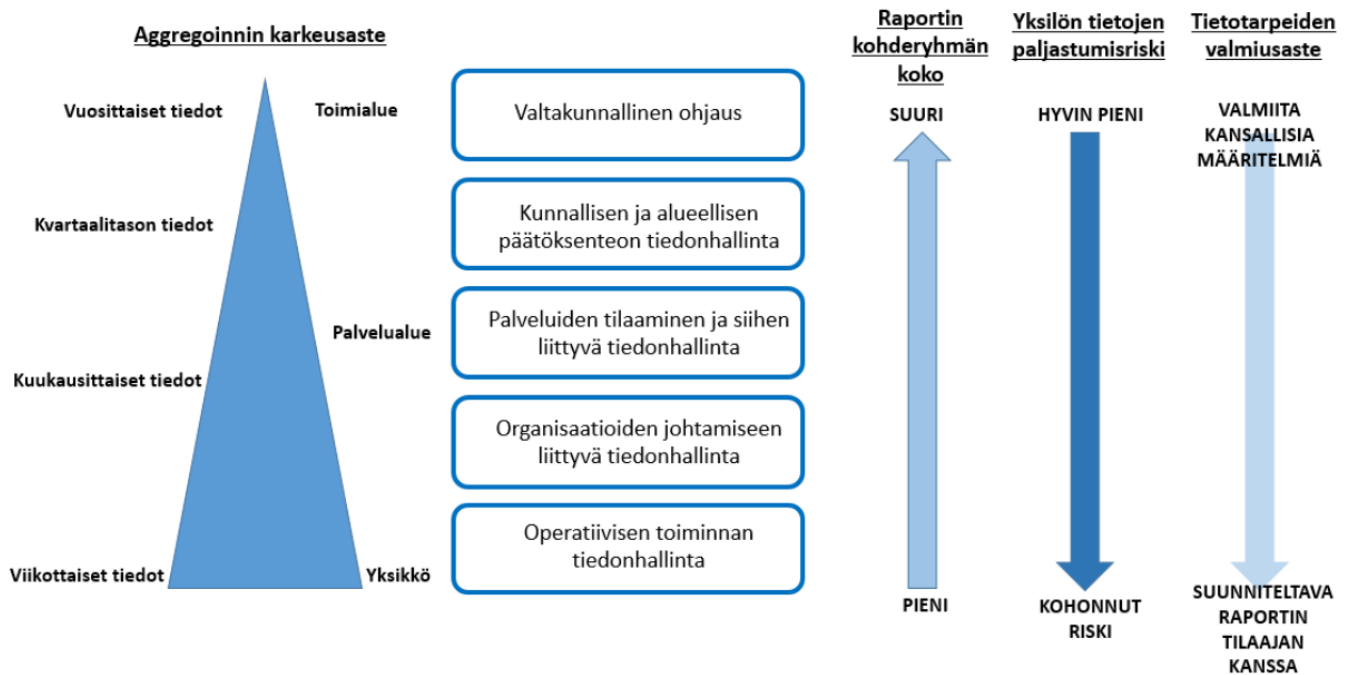
Julkisten sosiaali- ja terveystalvokuntayhtymien palvelutuotannossaan keräämiä tietoja voidaan käyttää hyödyksi sote- järjestelmän eri tasoilla. Tiedon hyödyntämisen tarpeet ja tavoitteet ovat erilaisia järjestelmätasosta riippuen.

Laihonen (2012, 6; 2009, 444–445) on jakanut sote-kentän tiedonhallinnan viiteen järjestelmätasoon. Kuvassa 13 esitetään Laihosen tiedonhallinnan järjestelmätasot, joiden yhteyteen olen lisännyt tiedolla johtamisen raportin tuottamiseen liittyviä tekijöitä. Laihosen mallissa ylimpänä on valtakunnallinen taso, eli Terveyden ja hyvinvoinninlaitoksen tai ministeriöiden antama ohjaus, jota tehdään perustuen vuosittain sote-palveluiden järjestäjiltä kerättyyn tietoon. Valtakunnan tasolla päätöksenteon tavoitteena on kansallisen hyvinvoinnin edistäminen. Tällä tasolla tiedot voidaan aggregoida hyvin karkealla tasolla, esimerkiksi vuosittain ja toimialueittain. Tästä syystä tietojen anonymisointi on verrattain vaikeaa toteuttaa. (KUVA 13.)

Seuraavana järjestämistasona on kunnallinen tai alueellinen taso, joka tulee vuoden 2023 alussa tarkoittamaan toimintansa aloittavia hyvinvointialueita ja niiden johtamista. Valtakunnallisen tason tie-

toja voidaan hyödyntää alueilla, kun halutaan verrata omaa aluetta muihin. Alueiden johtamiseen tarvitaan kuitenkin myös valtakunnan tasoa hieman yksityiskohtaisempia oman alueen vuosi- ja kvartaalitason tietoja. Kunta- ja aluetason päätöksenteon tietotarpeena on palvelutuotannon kustannustehokkuuden ja vaikuttavuuden mittaaminen.

TIETOJOHTAMISEN JÄRJESTELMÄTASOT JA RAPORTOINNIN OMINAISUUKSIA SOTE-ALALLA



KUVA 13. Tietojohtamisen raportoinnin ominaisuuksia asetettuna Laihosen (2012, 6; 2009, 444–445) esittämiin hyvinvointisektorin tiedonhallinnan järjestelmätasoihin.

Palveluiden tilaamisen ja hankinnan järjestelmätasolla tarvitaan vuosi- ja kvartaalitaso tietoja organisaation toiminnasta ja tarvittavien palveluiden tarpeesta, jotta tarvittavat tilaukset ja hankinnat saadaan toteutettua kustannustehokkaasti ja ajoissa. Palveluiden tilaamisen ja hankinnan taso on korkein taso, jolla tarvittavien tietojen kohderyhmä koostuu vielä kokonaisuudessaan sote-palveluja järjestävän organisaation työntekijöistä. Tämän tason ja alueellisen tason välissä kulkee siis raja, jonka jälkeen tietoja ryhdytään toimittamaan oman organisaation ulkopuolelle. Tämä raja täytyy huomioida, kun raportille valitaan sopivaa julkaisukanavaa.

Organisaation johtamisen tasolla tietotarpeet liittyvät tiedon hyödyntämiseen käytännön johtamistyössä. Tällä tasolla tarve on sekä seurata esimerkiksi palvelutarpeen kehittymistä ja resurssien riittävyyttä kuukausitasolla, että myös ennakoida tulevaa käytössä olevien tietojen pohjalta.

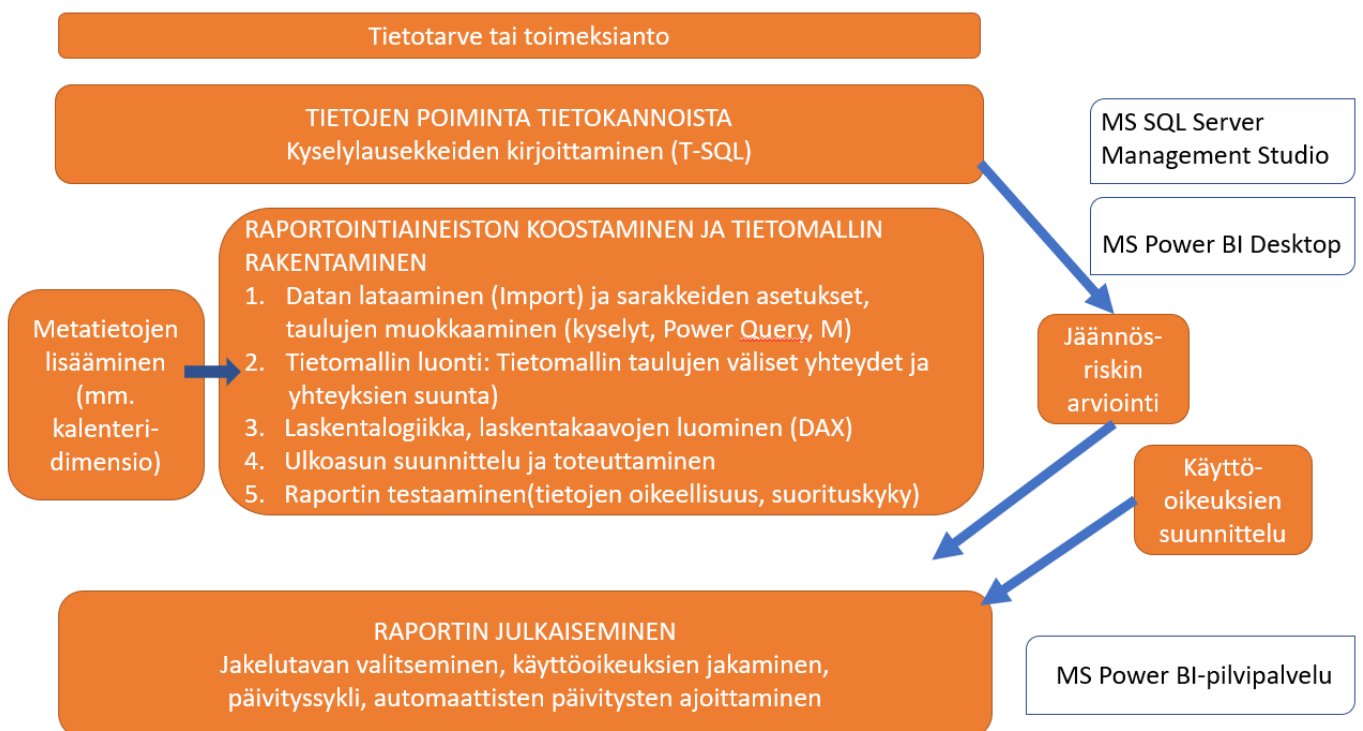
Alimpana järjestelmätasona Laihosen malli esittää operatiivisen toiminnan johtamisen (Laihonen 2012; 2009), eli yksittäisen toimintayksikön tai palveluiden järjestämisen vastuualueen johtamisessa tarvittavan tiedon. Tällä tasolla tiedolla johtamisen raportoinnin kohderyhmä voi olla hyvin pieni, mutta toisaalta tietotarpeet voivat olla hyvinkin laajoja ja yksityiskohtaisia. Operatiivisen johdon käyttöön tehtävissä raporteissa saatetaankin törmätä siihen, että toivotut tietotarpeet saattavat olla yksityiskohtaisempia kuin mitä raportille on mahdollista toteuttaa. Henkilötietojen anonymisointi johtaa aina tietojen osittaiseen häviämiseen ja tästä syystä kaikkia toiveita ei voida täyttää.

Järjestelmätasolta toiselle siirryttäessä tietotarpeet ja raportissa käytettävän tiedon aggregoinnin karkeus poikkeavat toisistaan suuresti. Mitä korkeammalle järjestelmätasolle tietoja tai tilastoja tuotetaan, sitä suurempi on tietojen vastaanottajien joukko. Korkeimmille järjestelmätasolle tuotetut raportit ovat helpompia anonymisoida, kuin alempien tasojen raportit. Operatiivisella tasolla tietotarpeena saattaa olla esimerkiksi tietojen erittely ammattiryhmittäin tai diagnoosien lukumäärien laskeminen. Näissä tapauksissa joudutaan tosissaan miettimään, miten tietopoiminta voidaan toteuttaa saavuttaen samalla riittävä anonymisoinnin taso. Toisaalta, jos jäännösriski toteutuu ja anonymiteetti onnistuttaisiin osittain purkamaan julkaistusta aineistosta, on vahinko sitä suurempi, mitä suuremmalle yleisölle raportti oli julkaistu. Raportin tuottamisen näkökulmasta on hyvä huomata myös se, että valtakunnan tason ja tulevaisuudessa pitkälti myös alueellisen tason tietotarpeet ovat valtakunnallisesti valmiiksi määritellyjä. Näihin tarpeisiin on siis saatavilla valmiita määritelmiä ja mittareita sekä yhteisiä ohjeistuksia tietojen koostamiseen. Raportin tuottaminen operatiivisen toiminnan tarpeisiin puolestaan vaatii jatkuvaa dialogia raportin toteuttajien ja käyttäjien välillä, sillä valmiita reseptejä räätälöintiä vaativiin raporteihin ei ole ennen kuin ne yhdessä luodaan.

6.2 Tuotettavan raportin kuvaus ja tuotantoprosessin vaiheet

Opinnäytetyön seuraavissa luvuissa tarkastellaan yksittäisen Power BI -raportin tuottamista sote-alan tiedolla johtamiseen. Tässä käyttötapauksessa rakennetaan raportti Keski-Pohjanmaan sosiaali- ja terveyspalvelukuntayhtymän Soiten toimialueista vastaavien johtajien käyttöön. Raportille yhdistetään tietoja toiminnasta, taloudesta ja organisaatorakenteesta.

Jokaiselle toimialueelle tehdään oma raporttinäköymä, jolla voi seurata kuukausitasolla tiettyjä toimialueen suoritteita (kuten käyntien tai leikkausten lukumääriä) sekä toiminnasta seuranneita kustannuksia kustannustyypeittäin (kuten henkilöstöstä tai ostetuista palveluista aiheutuvia kustannuksia). Käyttäjän on mahdollista tutkia näitä tunnuslukuja tarkemmin laskeutumalla organisaatiokaaviossa toimialueen alle kuuluvien vastuualueiden, palvelualueiden tai toimintayksiköiden (eli kustannuspaikkojen) tasolle. Lisäksi raporttinäköymään luodaan linkki erilliseen tuloslaskelma- näköymään. Kunkin toimialueen raporttinäköymästä on myös mahdollista siirtyä muiden toimialueiden näköymään. Kuvassa 14 esitetään yleisellä tasolla raportin tuottamisen vaiheet. Eri vaiheita kuvaillaan tarkemmin seuraavissa luvuissa.



KUVA 14. Power BI-raportin tuottamisen vaiheet ja käytettävät ohjelmistot

6.3 Tietotarpeen tunnistaminen tai toimeksiannon saanti

Tiedolla johtamisen raporttia lähdetään työstämään perustuen organisaatiosta lähtevään tietotarpeeseen. Raportin tavoitteena on vastata tietotarpeeseen mahdollisimman hyvin. Tavallisesti tarve uuteen raporttiin ilmenee toimeksiantona organisaation sisältä. Uusia tietotarpeita voi tulla kuitenkin myös organisaation ulkopuolelta, esimerkiksi tarpeina etsiä vastauksia uusiin valtakunnallisiin toiminnan mittareihin.

Raportin laatijan on selvitettävä aluksi mahdollisimman tarkkaan mitä tietoja halutaan tarkastella, mitkä näistä tiedoista ovat käyttäjälle olennaisimpia tietoja, miltä aikajaksolta tietoja tarkastellaan ja minkälaiselle kohderyhmälle raportti julkaistaan. Raportin tilaajan kanssa on hyvä käydä keskustelua siitä, mitä tietoja tietokannoista on ylipäätään mahdollista ottaa käyttöön.

6.4 Tietojen poiminta tietokannasta ja aineiston koostaminen

Kun tietotarve ja raportille vaadittavat tiedot on selvitetty, seuraavana askeleena etsitään tarvittavat tiedot tietovarastosta. Tietojen poiminnan suorittavan SQL-kyselyn kirjoittamiseen tarvitaan ensinnäkin tiedot käyttöön otettavien tietojen taulujen ja sarakkeiden nimistä. Lisäksi saatetaan tarvita tiedot niiden avainsarakkeiden nimistä, joiden perusteella eri taulujen tietoja voidaan yhdistää toisiinsa.

SQL-kyselyn kirjoittaminen on raportin koostamisen tärkeimpiä vaiheita. Tavoitteena on poimia raportille tarvittavat tiedot oikealla tavalla laskettuna mahdollisimman tehokkaasti ottamalla tietokannasta käyttöön vain tietotarpeen tyydyttämiseen tarvittava minimimäärä tietoja. Pyrkimys minimaaliseen tietomäärään johtaa samalla mahdollisimman karkealla tasolla tehtävään aggregointiin, joka puolestaan tarkoittaa useimmiten parempaa anonymisoinnin tasoa käyttöön otettaville tiedoille.

6.4.1 Faktataulun poiminta aggregoimalla

Liitteessä 1 poimitaan tiedot sairaanhoitopäivien määristä vuodesta 2017 lähtien. Tämä taulu toimii yhtenä raportin pääasiallisista faktatauluista. Kyselyssä (LIITE 1) tauluun kaynnit yhdistetään tietoja neljästä muusta taulusta LEFT OUTER JOIN -lauseiden avulla.

Jotta aggregointi saataisiin pysymään mahdollisimman karkealla tasolla, otetaan poimintaan mukaan ainoastaan vuosi ja kuukausi sekä tietojen yhdistämiseen välttämättä tarvittavat kolme avainsaraketta: kustannuspaikkanumero, käyntityyppiavain sekä toimintotyyppiavain.

Näiden lisäksi CASE WHEN -lauseella luodaan erillinen muuttujasarake, jonka avulla voidaan määrittää, onko käynti liittynyt koronavirukseen.

Kyselyn kirjoittamisessa täytyy huomioida myös tietomalliin tuotavien muiden kyselyiden rakenne ja tarvittaessa lisätä kyselyihin sellaisia avainsarakkeita, joiden avulla eri kyselyt voidaan tietomallissa

kytkeä toisiinsa. Liitteessä 1 on saraketietoja yhdistävää CONCAT-funktiota käyttäen luotu uusi sarake VuosiKK, jota käytetään myöhemmin yhdistävänä avaimena kalenteridimensio-tauluun. Koodiesimerkin lopussa olevaan ryhmittelyyn tulee GROUP BY lauseessa sisällyttää kaikki SELECT-lauseeseen kuuluvat sarakkeet, lukuun ottamatta aggregointifunktion sisältävää SUM(lukumaara) -riviä. (LIITE 1.)

6.4.2 Dimensiotaulun poiminta

Faktataulun kumppaniksi tarvitaan dimensiotaulu, josta faktataulun avainten selitteet haetaan. Taulusta saadaan yksittäiset avain- selite- parit SELECT DISTINCT -komennolla. Seuraava koodiesimerkki (KOODI 8) antaa tulosjoukkona yhden rivin jokaista avain-selite-paria kohden. Dimensiotaulujen luomisessa tulee käyttää juuri SELECT DISTINCT -komentoa, sillä näin voidaan tietomallissa muodostaa yhden suhde moneen- kytkentöjä.

```
-- Poimitaan lista yksittäisistä Toimintotyypeistä
SELECT DISTINCT
    Avain as ToimintoAvain
    ,ToimintoTyyppi
FROM tietokanta.toimintoTyyppi_taulu
```

KOODI 8. Faktataulun koodiavainten selitteiden poiminta SELECT DISTINCT -lauseella

6.4.3 Henkilötunnisteita sisältävän faktataulun poiminta aggregoimalla

Seuraavassa koodiesimerkissä 9 poimitaan tiedot perustason terveydenhoidon käynneillä olleiden henkilöiden määrästä kuukausittain. Poiminta kohdistuu samaan kayntitiedot-tauluun kuin aikaisemmin. Nyt kuitenkin lasketaan ainoastaan yksittäisten henkilötunnisteiden määriä. Jotta henkilötunnisteiden summat saadaan rivitasolla mahdollisimman suuriksi, käytetään poiminnassa pienintä mahdollista määrää muita sarakkeita. Koska tämän taulun tietoa ei ole tarvetta suodattaa raportilla esimerkiksi kustannuspaikan tai käynnin tyyppin perusteella, ei näitä tietoja tarvita myöskään poiminnassa. Karkein mahdollinen aggregointi voidaan saavuttaa käyttämällä tarkimpana tietona pienintä suodatuksessa käytettävää aikayksikköä, joka on tässä tapauksessa kuukausi. Tietojen rajaamiseen tarvittavat ehdot on lueteltu WHERE-lauseessa. Koodiesimerkin tuottama tulosrivi on muotoa vuosi, kuukausi, VuosiKK, henkiloLKM.

```
--Poimitaan PTH-käynnillä olleiden henkilöiden määrät kuukausittain
SELECT DISTINCT
    vuosi, kuukausi,
    ,CONCAT(vuosi,'-',REPLACE(STR(kuukausi, 2), ' ', '0') ) as VuosiKK
    ,COUNT(DISTINCT(henkilotunniste)) as henkiloLKM
FROM tietokanta.kayntitiedot_taulu kaynnit
    LEFT OUTER JOIN tietokanta.yksikko_taulu yksikko
        on kaynnit.yksikkoKoodi=yksikko.Avain
    LEFT OUTER JOIN tietokanta.potilaskontakti_taulu kontakti
        on kaynnit.potilaskontaktiAvain = kontakti.Avain
WHERE yksikko.Lisatieto like 'Vastaanotto'
AND yksikko.Avain not IN (lista poisluettavista yksikköavaimista)
AND kontakti.Kontaktityyppi not like 'Peruuttamaton poisjäänti'
AND vuosi >= 2019
GROUP BY vuosi,kuukausi
ORDER BY vuosi,kuukausi
```

KOODI 9. SQL-poiminta: potilasmäärä kuukausittain

Yllä olevassa koodiesimerkissä yksittäisiä henkilötunnisteita ei sisällytetä kyselyn tulositykköön, vaan niiden lukumäärä ainoastaan lasketaan kuukausittain. JOIN-lauseissa mainittujen taulujen tietoja ei myöskään tuoda tulositykköön, mutta taulujen tietoja tarvitaan kyselyn suodatuksessa WHERE-lauseessa. (KOODI 9.)

6.4.4 Faktataulun poiminta ilman aggregointia

Tietoja voidaan poimia raporttiaineistoon luonnollisesti myös ilman aggregointeja.

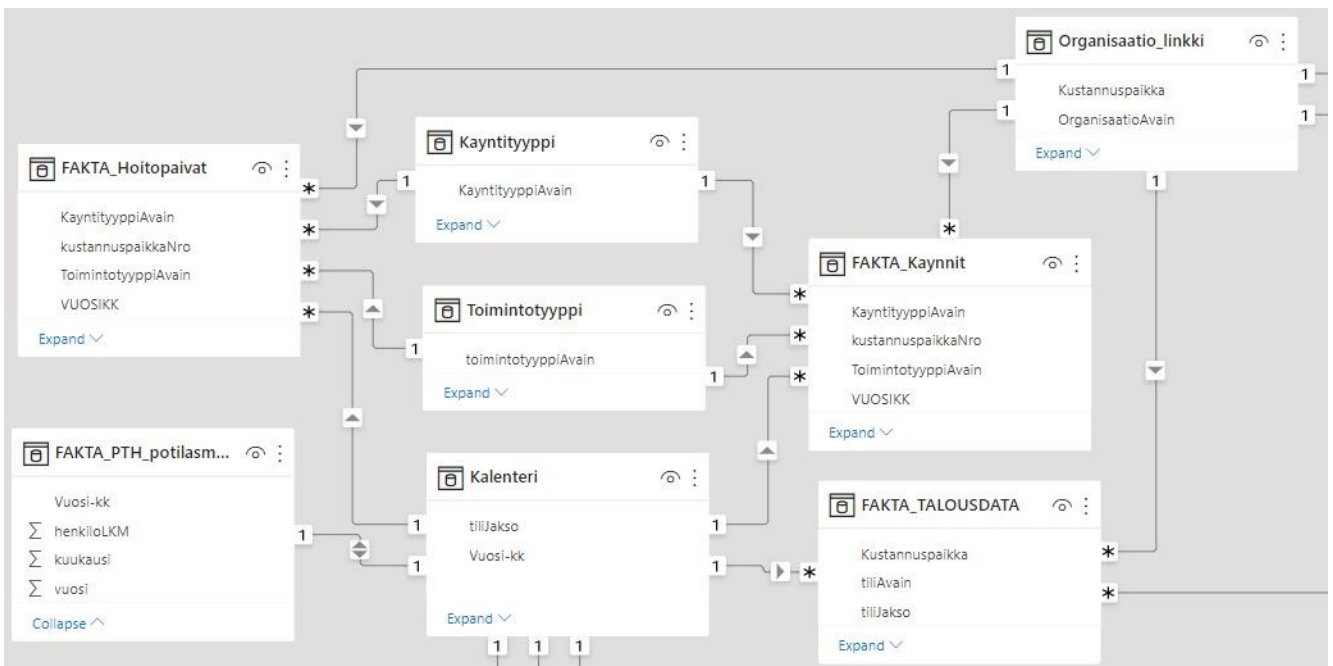
Koodiesimerkissä 10 taloustiedot poimitaan tietokannasta sellaisenaan, ilman aggregointia. Myös tässä kyselyssä luodaan tietojen yhdistämiseen käytettävä uusi sarake CONCAT-funktion avulla.

```
--Taloustiedot poimitaan rivitasolla, ei aggregoiden
SELECT
    Kustannuspaikka
    ,tiliJakso ,tiliNumero ,RiviSumma
    ,CONCAT(tiliNumero,'-',left((tiliJakso),4)) as NroJakso
FROM tietokanta.taloustiedot
WHERE tiliNumero >= 1111 AND tiliNumero < 9999
ORDER BY tiliJakso, Kustannuspaikka, NroJakso
```

KOODI 10. Taloustietojen poiminta ilman aggregointia

6.5 Tietomallin rakentaminen

Kuvassa 15 esitetään osa tuotettavan raportin tietomallin rakenteesta. Tietomallissa on kolme pääasiallista faktataulua, joihin on tallennettu tiedot käynneistä, hoitopäivistä ja taloustiedoista. Lisäksi taulussa FAKTA_PTH_potilasmaarat on tiedot yhdessä kuvaajassa käytettävistä perusterveydenhoidon potilaiden lukumääristä. Nämä tiedot poimittiin aiemmin koodilla 9. Luvussa 6.4.2 poimittu toimintotyyppi ja vastaavalla tavalla poimittu taulu käyntityyppi ovat dimensiotauluja. Kuvan ulkopuolelle jäävä talousdataan johtava kytkentä johtaa tilirakenteet sisältävään dimensiotauluun.



KUVA 15. Kuvakaappaus raporttitiedoston tietomallin rakenteesta. Kuvassa näkyvät ainoastaan avaimina toimivien sarakkeiden nimet.

Tietokannoista tuotavien fakta- ja dimensiotaulujen lisäksi hyvin usein on tarpeen luoda erillinen taulu kalenterille, jonka kautta reititetään tietojen ajallinen suodatus. Kalenteri-taulu kiinnittyy yhden suhdemoneen- tyyppisillä kytköksillä eri faktatauluihin ja mahdollistaa näin eri faktataulujen suodatuksen. Suodattavina avainkenttinä käytetään poiminnassa CONCAT-funktiolla luotuja Vuosi-kk -sarakkeita sekä talousdatassa olevaa tiliJakso-saraketta. (KUVA 15.) Kalenteritaulun muodostamista käsitellään tarkemmin seuraavassa luvussa.

Kuvan 15 ylänurkan taulu Organisaatio_linkki toimii aputaaluna yhdistämään tietoja kuvassa olevien taulujen ja kuvan ulkopuolelle jäävien organisaatorakenteeseen liittyvien taulujen välillä. Organisaatorakenteen tuontia tietomalliin ei käsitellä tässä työssä sen tarkemmin, mutta taulun tarkoituksena on yhdistää faktatauluihin tiedot organisaation hierarkiasta kustannuspaikka-avaimen kautta.

Nuolten suunnat kuvassa 15 osoittavat, kumpaan suuntaan Power BI suodattaa tietoja. Dimensiotaulut Kayntityyppi ja Toimintotyyppi suodattavat faktatauluja Hoitopaivat ja Kaynnit yhden (1) suhde mo-
neen (*) -kytkennällä. PHT_potilasmaara- taulun ja Kalenteri-aulun välillä kytkentä on tietopoimin-
nan toteutuksesta johtuen tyyppiä yhden suhde yhteen. Tällöin suodatus on automaattisesti kaksisuun-
taista. Kalenteritauluun on yhdistetty muutamia faktatauluja, jotka jäävät kuvan ulkopuolelle. Niiden
rakenne, kytkentä ja käyttötarkoitus ovat samanlaisia, kuin PTH_potilasmäärä- taululla.

6.5.1 Kalenteridimension lisääminen tietomalliin

Uusi taulukko voidaan luoda valitsemalla valintanauhasta Mallinnus ja Uusi taulukko. Avautuvaan nä-
kymään voidaan antaa DAX-koodi kalenteritaulukon luomiseksi seuraavasti:

```
Kalenteri = var Kuukausikalenteri =
ADDCOLUMNS (CALENDAR ("2017/1/1", TODAY ()), "KK-nro", MONTH ([Date]),
"KK-nimi", FORMAT ([Date], "MMMM"),
"KK-nimi-lyh", FORMAT ([Date], "MMM"),
"Vuosi", YEAR ([Date]),
"Vuosi-kk", CONCATENATE (YEAR ([Date]), CONCATENATE ("-", FORMAT ([Date], "MM"))),
"tiliJakso", CONCATENATE (YEAR ([Date]), FORMAT ([Date], "MM")))
return
SUMMARIZE (Kuukausikalenteri, [KK-nro], [KK-nimi], [KK-nimi-lyh], [Vuosi], [tili-
Jakso], [Vuosi-kk])
```

KOODI 11. Kalenteri-aulun luominen DAX-lauseella

Tuloksena saadaan kuukausitason kalenteri (TAULUKKO 1), aikavälille tammikuusta 2017 eteenpäin.
Vuosi-kk -saraketta käytetään potilaskäyntien tietojen yhdistämiseen ja tiliJakso -saraketta taloustieto-
jen yhdistämiseen.

TAULUKKO 1. DAX-lauseella muodostetun kalenteritaulukon ensimmäiset rivit

KK-nro	Vuosi	Vuosi-kk	KK-nimi	KK-nimi-lyh	tiliJakso
1	2017	2017-01	tammikuu	tammi	201701
2	2017	2017-02	helmikuu	helmi	201702
3	2017	2017-03	maaliskuu	maalis	201703
4	2017	2017-04	huhtikuu	huhti	201704
5	2017	2017-05	toukokuu	touko	201705
6	2017	2017-06	kesäkuu	kesä	201706
7	2017	2017-07	heinäkuu	heinä	201707

6.6 Laskentakaavojen määrittäminen

Tietomalliin poimittujen tietojen tuominen raportille käyttäjälle ymmärrettävässä ja organisaatiolle sopivassa muodossa vaatii laskentakaavojen määrittämistä. Useimmiten laskentakaavoja käytetään luomalla mittareita varsinaisten liiketoimintatietojen laskemiseen. DAX-kaavoja voi kuitenkin käyttää myös käytettävyyttä parantavien ominaisuuksien, kuten dynaamisesti päivittyvien kuvaajien otsikoiden ohjelmoimiseen. Aloitetaan laskemalla kaikkien potilaskäyntien määrä DAX-lauseella:

```
Kaynnit_total = SUM(FAKTA_Kaynnit[lkm])
```

KOODI 12. Käyntien yhteismäärän laskentakaava

Vastaavasti SUM -funktiolla summaamalla voidaan laskea myös esimerkiksi hoitopäiviä, PTH-potilas-määriä tai leikkausten tai synnytysten määriä omista tauluistaan. Luotuja mittareita voidaan antaa parametrinä toisille mittareille. Alla koodiesimerkissä 13 käytetään edellä luotua mittaria, mutta sille annetaan lisää suodattimia. PTH -käynteihin lasketaan mukaan vain tietyille kustannuspaikoille ja toimintotyypeille suodattuvat käynnit. Lauseessa käytetään myös KEEPFILTERS-funktiota, sillä käyttäjän antamat muut suodatusvalinnat halutaan säilyttää. Rakenteeltaan vastaavalla koodilla (KOODI 13) voidaan kustannuspaikkoja tai muita suodattimia vaihtamalla laskea myös erikoissairaanhoidon ja eri toimialueisiin kuuluvien erilaisten toimintojen käyntien yhteismääriä.

```
Kaynnit_PTH = CALCULATE([Kaynnit_total],
KEEPFILTERS( FAKTA_Kaynnit[kustannuspaikkaNro] IN { <luettelo kustannuspaikoista> } && FAKTA_Kaynnit[Toimintotyyppi] IN {<luettelo toimintotyypeistä> }
))
```

KOODI 13. Perusterveydenhoidon käyntien laskentakaava

Ainoastaan tietyn vuoden käyntien laskeminen onnistuu lisäämällä koodin 13 mittariin vuosisuodatin:

```
Kaynnit_PTH_2021 = CALCULATE([Kaynnit_PTH], FAKTA_Kaynnit[vuosi]=2021)
```

KOODI 14. Perusterveydenhoidon vuoden 2021 käyntien laskentakaava

Raporttiin halutaan yhdeksi elementiksi näkymä, josta voi seurata koronapandemiaan liittyviä sairauksien tai rokotukseen liittyviä potilaskontakteja. Tähän voidaan käyttää koodin 13 mittaria ja yksinkertaisesti antaa sille vain uusi suodatin seuraavasti:

```
PTH_Koronakontakti_Kaynnit = CALCULATE( [Kaynnit_PTH], FAKTA_Kaynnit[OnkoKoronakontaktiTaiRokote]=1)
```

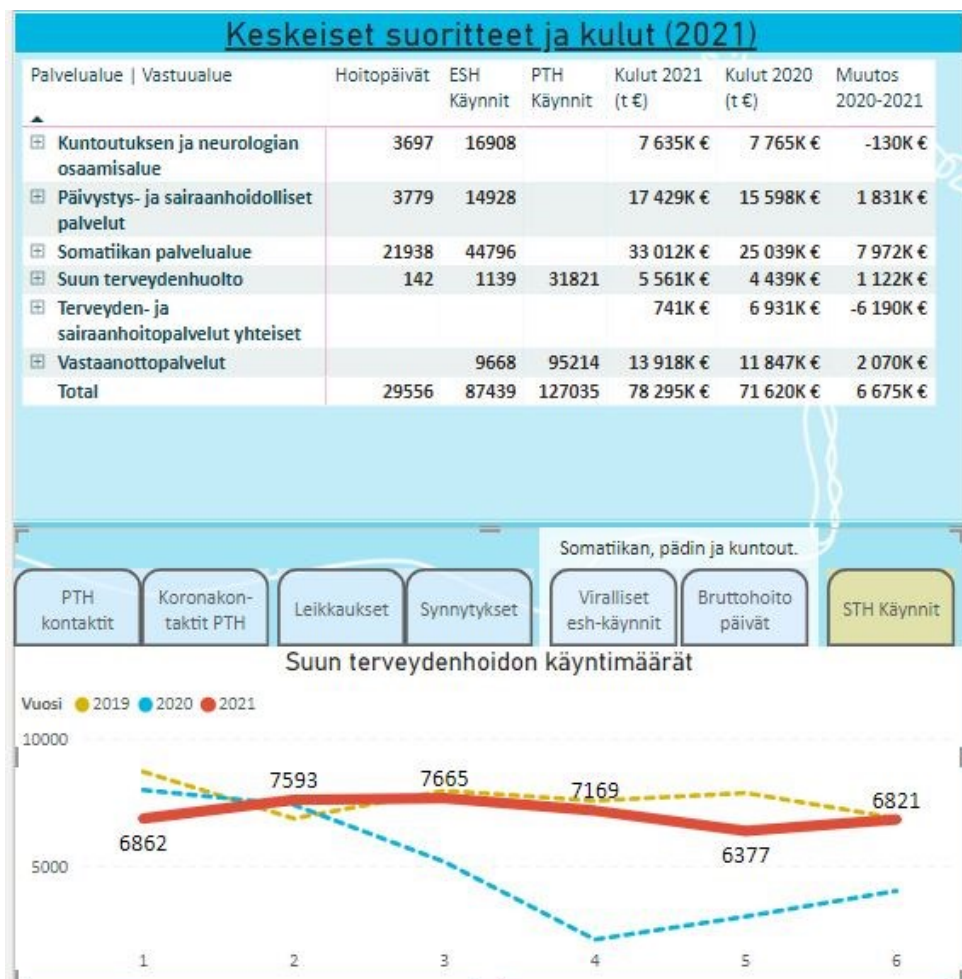
KOODI 15. Mittareiden uusiokäyttöä uusien mittareiden luomisessa CALCULATE-funktiolla

6.7 Raportin ulkoasun suunnittelu ja toteuttaminen

Raportilla esitetään tietoja käynneistä ja taloustiedoista. Tästä syystä raporttisivu jaetaan loogisesti kahteen osaan. Samaa kahtiajakoa voidaan käyttää kaikkien eri toimialueiden raporttinäkymillä, kunhan jokaiselle raporttisivulle tehdään oikeilla suodatinasetuksilla käyntitietoja suodattavat mittarit.

Muiden toimialueiden raporttisivun pohjaksi voi siis monistaa uudelle näkymäsivulle ensimmäisenä tehdyn näkymän ja tehdä siihen toimialuekohtaiset muutokset ja mittarit.

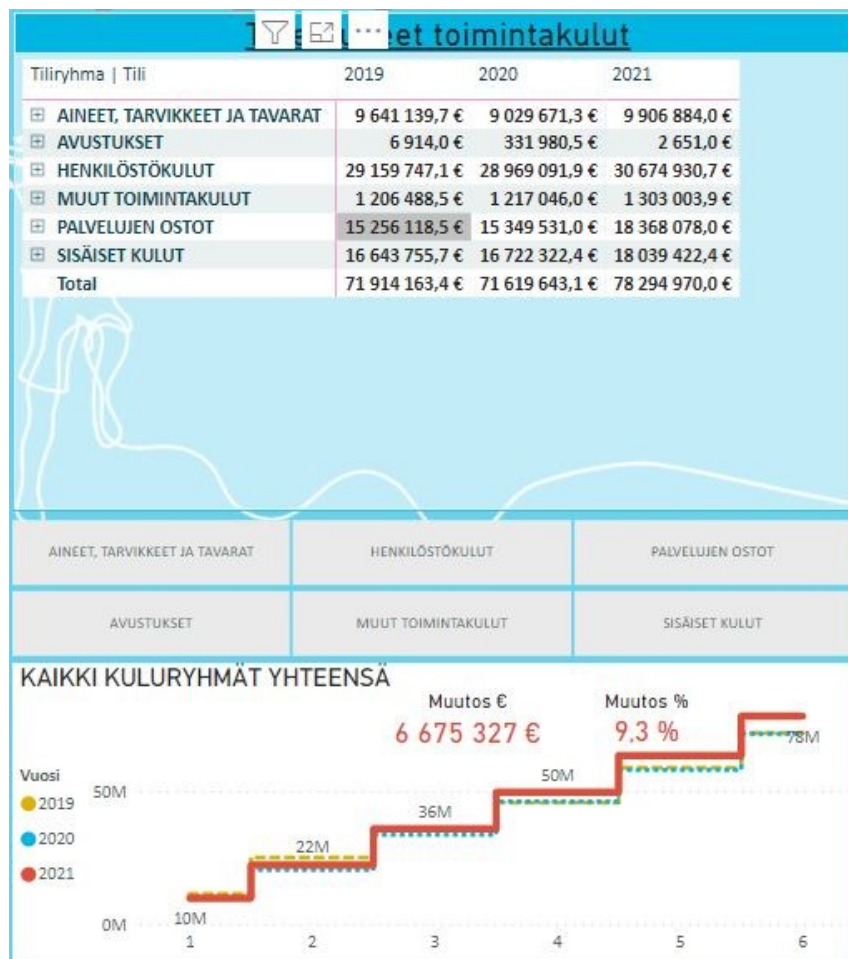
Raporttinäkymän vasemmalle puolelle kerätään ylös käynteihin liittyvät tiedot sekä muutama avainluku taloustiedoista taulukkomuodossa. Jotta taulukossa voisi siirtyä hierarkiassa alaspäin palvelualueelta vastuualueelle, valitaan visualisoinnin tyyppiä matriisi. Matriisin kiinnitetään tarvittavat käyntejä ja vuosittaisia kustannuksia laskevat mittarit, joiden luominen esiteltiin luvussa 6.6. Vasemmalle alas koostetaan kokoelma viivakuvaajia erilaisista käyntitiedoista. (KUVA 16.)



KUVA 16. Terveysten- ja sairaanhoidon toimialueen raporttinäkymän vasen puoli

Raportin oikealle puolelle koostetaan matriisiin näkymä taloustiedoista ja sen alapuolelle viivakuvaaja kulujen kuukausittaisesta kasvusta. Kuvaajan päälle asetetaan kaksi kortti- tyyppistä visualisointia, joi-

hin lasketaan kulujen muutos viime vuodesta kuluvaan vuoteen euroina ja prosentteina. Kuvaajan yläpuolelle lisätään vielä osittaja, jolla voi suodattaa tiliryhmiä. Oikealla puolella olevat taloustietojen visualisoinnit suodattavat kaikki toisiaan, mutta ne eivät suodata vasemmalla puolella olevia käynteihin liittyviä visualisointeja. Oletusarvoisesti kaikki visualisoinnit ovat vuorovaikutuksessa keskenään, jos niillä on jokin yhteinen tekijä. Visualisointien vuorovaikutusta toisiinsa voi kuitenkin muokata valitsemalla ensin muokattavan visualisoinnin ja sen jälkeen valintanauhasta Muotoile sekä Muokkaa vuorovaikutuksia. Kielto-merkki -symbolin klikkaaminen toisen visualisoinnin kohdalla kytkee vuorovaikutuksen muokattavaan visualisointiin pois päältä. (KUVA 17.)



KUVA 17. Terveys- ja sairaanhoidon toimialueen raporttinäkömön oikea puoli

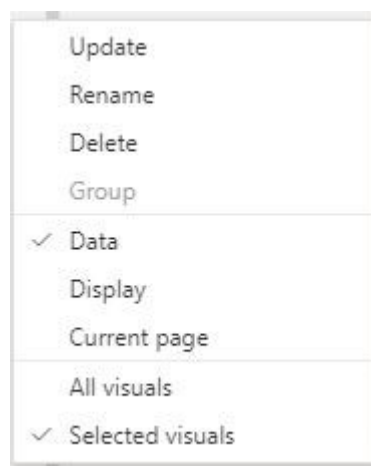
Terveys- ja sairaanhoidon näkömön halutaan useita kuvaajia eri käyntimääriä laskevista mittareista. Kaikkia haluttuja kuvaajia ei ole järkevää suloa sivulle vierekkäin, sillä niistä kaikista tulisi hyvin pienikokoisia ja niitä olisi vaikea lukea. Toisaalta jokaisen toimialueen näkömön laajuus halutaan pitää yhdessä sivussa. Tilaa voidaan säästää käyttämällä samaa visualisoinnin paikkaa usealle eri mittarille kirjanmerkkien, osittajan ja parin uuden mittarin avulla.

Luodaan ensin ”Anna tiedot” -toiminnolla uusi taulukko, jossa on arvot 1–7 sekä osittaja, joka viittaa tähän taulukkoon. Seuraavaksi luodaan rivi painikkeita ja sama määrä uusia kirjanmerkkejä. Lisäksi luodaan uusi viivakuvaaja-visualisointi. Viivakuvaajaan kiinnitetään mittariksi ehtolause (KOODI 16), joka valitsee kuvaajassa käytettävän mittarin osittajaan valitun arvon perusteella.

```
NaytaValittuKuvaaja =
SWITCH (
    SELECTEDVALUE (ValittuKuvaajaTesaan[SlicerID] )
    ,1,[PTH_potilaat]
    ,2,[Kaynnit_STH]
    ,3,[Kaynnit_ESH]
    ,4,CALCULATE ([Hoitopaivat_total],Organisaatio_Palvelualue[Nro] IN
        {<luettelo palvelualueiden numeroista})
    ,5,[PTH_Koronakontakti_Kaynnit]
    ,6,TESA_Leikkaukset [Leikkausten_LKM]
    ,7,TESA_Synnytykset [Synnytysten_LKM]
    , [PTH_potilaat]
)
```

KOODI 16. Kuvaajassa käytettävän mittarin määrittäminen SWITCH-lauseella

Luodut kirjanmerkit nimetään ja päivitetään ”Update” -toiminnolla valitsemalla kirjanmerkille asetukset ”Tiedot” ja ”Valitut visualisoinnit” sekä valitaan valituiksi visualisoinneiksi uusi viivakaavio sekä uusi osittaja. Jokaiseen kirjanmerkkiin tallennetaan osittajan valinnaksi eri arvo. Kirjanmerkit kiinnitetään toiminnoiksi oikeisiin painikkeisiin ja lopuksi vielä kirjanmerkkien ohjelmointiin käytettävä osittaja piilotetaan.



KUVA 18. Kirjanmerkin asetusten asettaminen

Jotta kuvaajassa näkyvän tiedon otsikko saadaan vaihtumaan oikein suhteessa klikattuun painikkeeseen, luodaan vielä toinen ehtolause, jolla määritellään näytettävä otsikko riippuen osittajassa olevasta arvosta. Tämä ehtolause (Koodi 17) kiinnitetään kuvaajan otsikoksi.

```
ValittuKuvaajaOtsikko_TESA =
SWITCH(
    SELECTEDVALUE(ValittuKuvaajaTesaan[TESA_SlicerID])
    ,1,"PTH kontaktissa olleiden henkilöiden lukumäärä"
    ,2,"Suun terveydenhoidon käyntimäärät"
    ,3,"Viralliset ESH-käynnit"
    ,4,"Bruttohoitopäivät"
    ,5,"PTH koronakontaktit"
    ,6,"Leikkausten lukumäärä"
    ,7,"Synnytysten lukumäärä"
    ,"PTH kontaktissa olleiden henkilöiden lukumäärä"
)
```

KOODI 17. Kuvaajassa käytettävän otsikon määrittäminen SWITCH-lauseella

Näillä toimenpiteillä saadaan raporttinäköymän vasen puoli toimimaan halutulla tavalla. Raportin oikealla puolella on kuitenkin vielä tekemistä. Taloustietojen viivakaavio saadaan näyttämään kumulatiivisesti kuukausittaisia summia vain valitulta aikaväliltä, kun siihen kiinnitetään koodin 18 mukainen mittari:

```
Kustannukset juokseva kk-vertailu =
CALCULATE(
    SUM(FAKTA_TALOUSDATA[Summa]),
    FILTER(
        ALLSELECTED(Kalenteri[KK-nro]),
        ISONORAFTER(Kalenteri[KK-nro], MAX(Kalenteri[KK-nro]), DESC)
    )
)
```

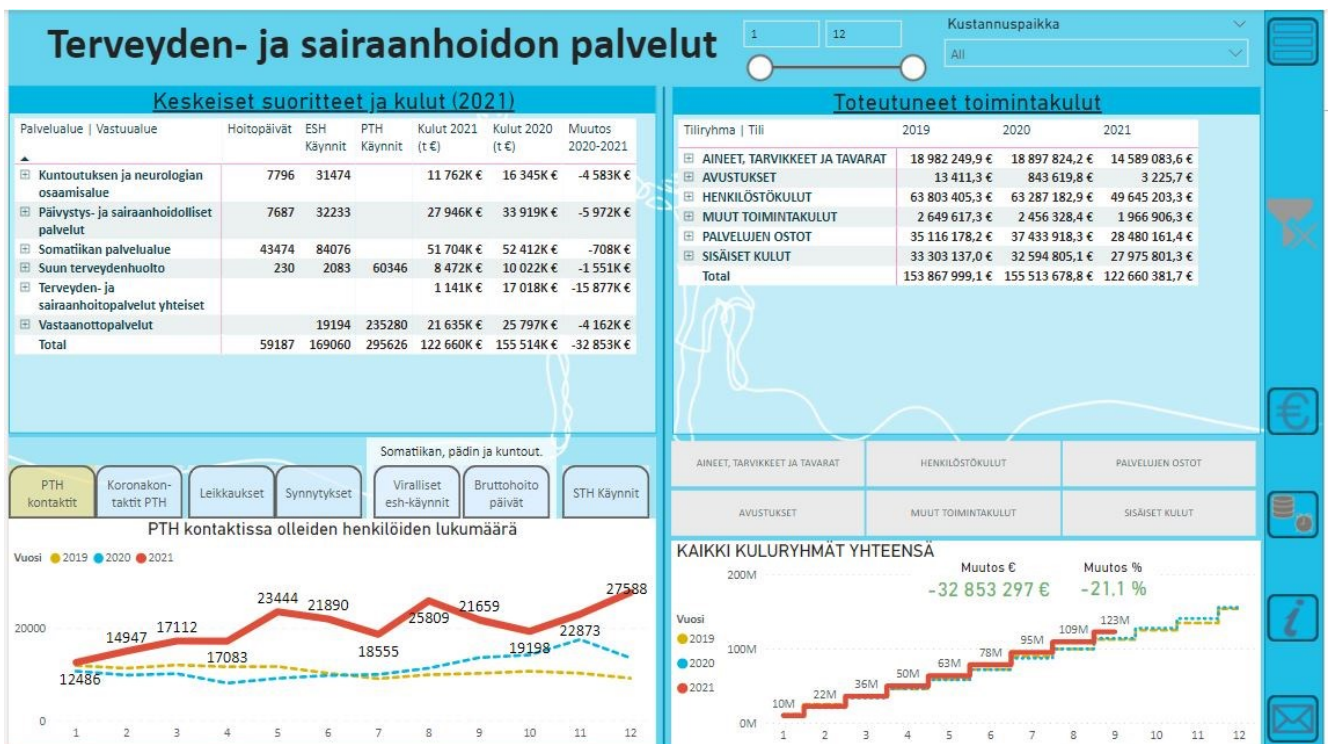
KOODI 18. Kumulatiivinen kuukausittaisten kustannusten laskeminen

Taloustieto-kuvaajaan voidaan asettaa dynaamisesti päivittyvä otsikko DAX-lauseella. Koodiesimerkissä 19, jos tiliryhmiä on valittuna vain yksi, näytetään tiliryhmän nimi. Jos kaikki kuusi tiliryhmää on valittu, näytetään ”kaikki yhteensä”- vaihtoehto. Muussa tapauksessa näytetään ”useita valittu”-vaihtoehto.

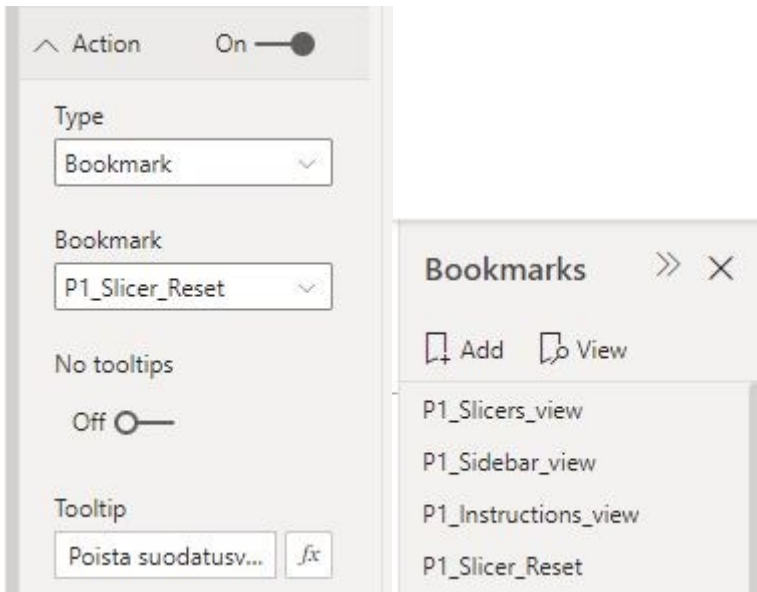
```
Tiliryhma_otsikkoon = IF(HASONEFILTER(Tilit[TiliryhmaNimi]), SELECTEDVALUE(Tilit[TiliryhmaNimi]), IF(DISTINCTCOUNTNOBLANK(Tilit[TiliryhmaNimi]) =6, "KAIKKI KULURYHMÄT YHTEENSÄ", "USEITA KULURYHMIÄ VALITTU"))
```

KOODI 19. Otsikon määrittäminen dynaamisesti DAX-lauseella

Kuvassa 19 esitetään valmis raporttinäkymän sivu kokonaisuudessaan. Raporttinäkymään lisätään yläreunaan osittajat kuukaudelle ja kustannuspaikalle. Nämä osittajat suodattavat kaikkia raportin taulukoita ja kuvaajia. Raporttinäkymän oikeaan reunaan lisätään vielä muotoelementeistä ja kuvista koostettu sivupalkki. Sivupalkin hampurilaisvalikosta aukeaa valintanäkymä muiden toimialueiden raporttinäkymiin siirtymiseen. Muihin kuviin kiinnitetään toiminnoiksi suodatinvalintojen palauttaminen, siirtyminen erilliseen tuloslaskelma- näkymään, ohjesivun avaaminen sekä palautesähköpostin lähettäminen sähköpostiohjelmalla valmiilla otsikolla varustettuna. Jokaiseen kuvaan lisätään myös tekstimuotoinen työkaluvihje, joka tulee esiin, kun käyttäjä siirtää hiiren osoittimen kuvan päälle. (KUVA 19.) Jokaista toimialueen raporttinäkymää kohti tarvitaan neljä kirjanmerkkiä, joiden kautta sivupalkin kuvakkeiden toiminnot toteutetaan (KUVA 20).

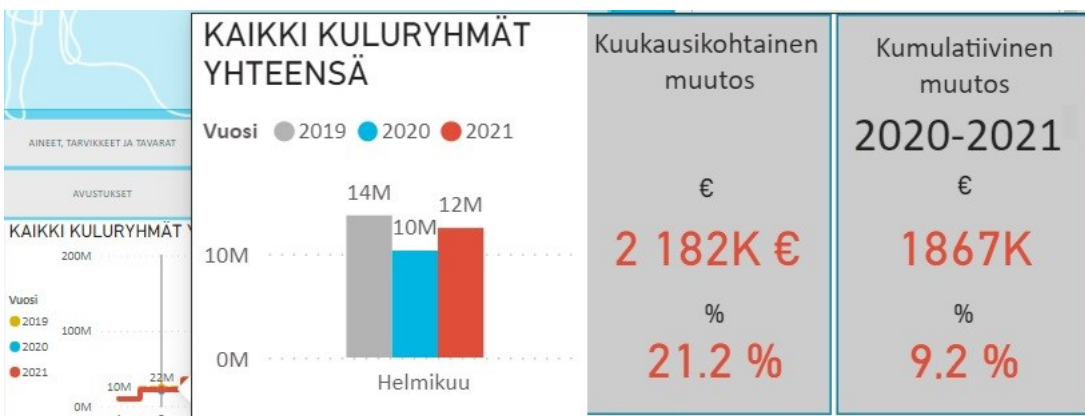


KUVA 19. Terveyden- ja sairaanhoidon toimialueen valmis raporttinäkymä



KUVA 20. Kuva-elementtiin toiminnoksi määritetty kirjanmerkki ja sivupalkin toiminnot toteuttavia kirjanmerkkejä

Raportin elementteihin voidaan lisätä myös erillisiä välilehtiä, jotka tulevat esiin työkaluvihjetoiminnon kautta. Visualisoinnille voidaan määrittää työkaluvihjeenä näytettäväksi erillinen välilehti. Työkaluvihje aktivoituu, kun käyttäjä siirtää hiiren osoittimen visualisoinnin datapisteen päälle. (KUVA 21.)

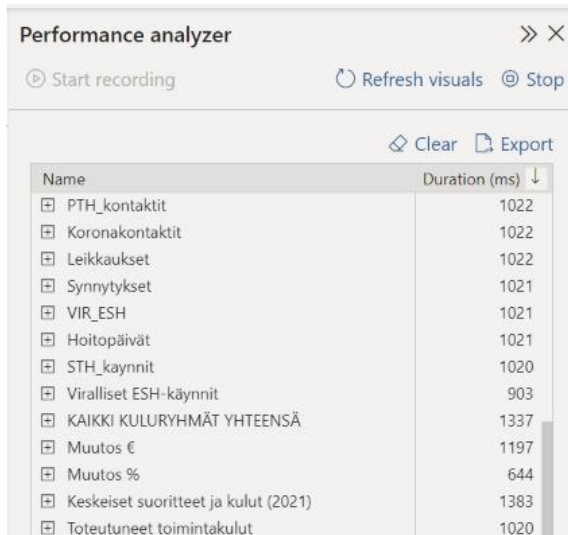


KUVA 21. Taloustietokuvaajan datapisteestä nousevia lisätietoja työkaluvihjetoiminnon avulla

6.8 Raportin testaaminen ja optimointi

Ennen julkaisua raportilla olevien tietojen oikeellisuus täytyy varmistaa jollakin tavalla. Jos organisaatiossa on raportoitu tai tilastoitu samankaltaisia tietoja ennenkin, voi tarkistamiseen käyttää aiemmin koottuja tietoja soveltuvin osin. Tämän lisäksi kehittäjän on testattava, että raportin visualisoinnit toimivat odotetulla tavalla sekä tarkastettava niiden keskinäiset vuorovaikutussuhteet. Erityisen tärkeää on lisäksi tarkastaa jokaisen visualisoinnin kohdalta, mitä tietoja ne antavat loppukäyttäjän käyttöön visualisoinnin ”Analysoi” -toiminnon kautta. Jos analysointitoiminto paljastaa liian yksityiskohtaisia tietoja, täytyy sarakkeita piilottamalla poistaa tietoja loppukäyttäjän ja analysointitoiminnon käytettävistä. Kun tietopöiminnan oikeellisuus ja visualisointien haluttu toiminnallisuus on varmistettu, voidaan ryhtyä tutkimaan mahdollisuuksia parantaa raporttitiedoston suorituskykyä.

Suorituskyvyn mittaamiseen voi käyttää Power BI:n sisäistä suorituskyvyn analyysityökalua, jonka saa esiin valintanauhan ”Näytä”-välilehdeltä. Sen avulla voi tutkia jokaisen raporttisivun elementin lataamiseen kuluva aikaa. (KUVA 22.) Jos jokin visualisointi latautuu hitaasti, voi syytä lähteä etsimään DAX-lausekkeesta tai tietomallin rakenteesta.



Name	Duration (ms)
PTH_kontaktit	1022
Koronakontaktit	1022
Leikkaukset	1022
Synnytykset	1021
VIR_ESH	1021
Hoitopäivät	1021
STH_kaynnit	1020
Viralliset ESH-kaynnit	903
KAIKKI KULURYHMÄT YHTEENSÄ	1337
Muutos €	1197
Muutos %	644
Keskeiset suoritteet ja kulut (2021)	1383
Toteutuneet toimintakulut	1020

KUVA 22. Kuvakaappaus suorituskyvyn analysointityökalun näkymästä taloustietojen faktataulun aggregoinnin jälkeen

DAX Studio -apuohjelmalla voi puolestaan tutkia tietomallin taulujen käyttämiä resursseja. DAX Studioissa tietomallin analysointiin pääsee ”Advanced”-välilehden kautta painikkeella ”View Metrics”. Tässä tapauksessa tietotaulujen resurssianalyysi paljasti taloustietojen vievän erityisen paljon resursseja. Tämä johtui siitä, että summasarakkeessa oli erittäin suuri määrä erilaisia arvoja. Taloustietojen

aggregointi kuukausitasolle pienensi merkittävästi taulun kokoa ja nopeutti huomattavasti taloustietoja käyttävien visualisointien lataamista.

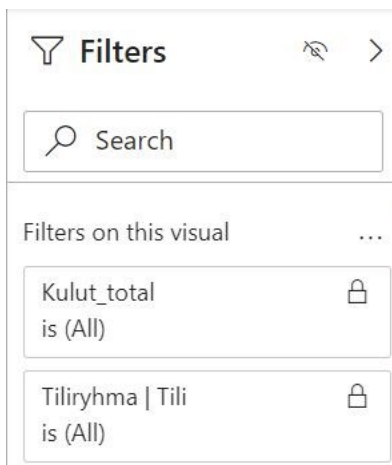
Ennen kuin raportti voidaan julkaista, täytyy vielä tehdä lopullinen arvio raporttiaineiston jäännösriskistä. Raportista ei tulisi saada esiin liian yksityiskohtaisia tietoja tai liian pieniä ja tarkkoja tulosjoukkoja terveystiedoista. Raporttia tulee testata siis myös tutkimalla erilaisten suodatusten kautta, kuinka pieniä tulosjoukkoja on mahdollista saada esiin. Jos jäännösriski katsotaan liian suureksi, on raporttia tai poimintaa muokattava, kunnes hyväksyttävä jäännösriskin taso saavutetaan. Tämä saattaa tarkoittaa myös raportille toivottujen ominaisuuksien karsimista tai käytettävyyden muokkaamista vähemmän käyttäjäystävälliseksi. Usein liian tarkkoja tietoja voidaan piilottaa käyttämällä visualisoinnissa TOP N-suodatusta, jonka avulla esiin tuodaan esimerkiksi vain viiden yleisimmän diagnoosin tiedot. Monimutkaisemmissa tapauksissa visualisoinneille voi olla tarvetta luoda räätälöityjä suodatuksia DAX-lausekkeilla, jotta liian pieniä tulosjoukkoja tai yksityiskohtaisia tietoja ei näytetä.

Raportilta loppukäyttäjälle paljastuvien tietojen jäännösriskiä voi pienentää myös piilottamalla taulujen sarakkeita, jolloin estetään loppukäyttäjän pääsy piilotettujen sarakkeiden yksityiskohtaiseen tarkasteluun. Raporttitiedoston asetuksista voidaan määrittää, saako loppukäyttäjä ladata itselleen raportin tietoja tai yhteenvetoja. Asetuksia pääsee määrittelemään valikkopolulla ”Tiedosto”, ”Asetukset ja vaihtoehdot”, ”Valinnat”, ”Raporttiasetukset” ja ”Vie tiedot”. Tässä valikossa turvallisin ratkaisu on valita vaihtoehto ”Älä salli käyttäjien viedä mitään tietoja palvelusta tai raporttipalvelimesta”.

6.9 Raportin julkaiseminen

Tämän opinnäytteen käyttötapauksessa tuottamassamme raportissa halutaan mahdollistaa kaikkien toimialueiden johtajien mahdollisuus tarkastella kaikkien toimialueiden tietoja samasta raporttitiedostosta. Tästä syystä samaan tiedostoon tehtiin raporttinäkymät eri toimialueille. Jos raportti haluttaisiinkin jakaa siten, että toimialueen johtajat näkisivät ainoastaan oman toimialueensa tiedot, olisi raportin jakamiseen kolme vaihtoehtoa. Ensimmäinen ja työläin vaihtoehto olisi tehdä toimialueille erilliset tietojoukot ja raporttitiedostot omine välilehtineen. Tässä tapauksessa usean tietojoukon muodostaminen aiheuttaisi väistämättä päällekkäisyyttä eri raporttien aineistoissa ja aiheuttaisi näin tarpeetonta tarvetta eri tietojoukkojen päivittämiselle julkaisun jälkeen. Tämä lisäisi haasteita myös raporttitiedostojen versionhallinnassa.

Toinen vaihtoehto olisi ensin tallentaa raporttitiedoston tietojoukko pilvipalveluun ilman visualisointeja ja tämän jälkeen liittää julkaistu tietojoukko tietolähteeksi erillisiin tiedostoihin, joihin luotaisiin toimialuekohtaiset raporttinäkymät. Muiden toimialueiden toimintaan liittyvät tiedot rajattaisiin raporttilta pois suodattamalla tietoja tiedostokohtaisilla suodattimilla. Tässä tapauksessa tehdyt tiedostokohdaiset suodatusvalinnat tulisi piilottaa loppukäyttäjältä poistamalla Suodattimet-paneeli näkyvistä klikkaamalla silmäkuvaketta tai lukitsemalla suodattimia klikkaamalla avoimen lukon kuvaa yksittäisessä suodattimessa (KUVA 23). Tätä vaihtoehtoa käytettäessä julkaistaisiin siis useita raporttitiedostoja, joilla olisi eri suodattimet ja visualisoinnit, mutta yhteinen tietojoukko. Pohjana käytettävää ilman visualisointeja pilvipalveluun ladattavaa tiedostoa ei julkaistaisi, vaan se toimisi ainoastaan tietojoukon varastona. Kaikkien raporttitiedostojen tietojen päivittäminen onnistuisi siis yhtä tietojoukkoa päivittämällä.



KUVA 23. Suodatinpaneelin piilottaminen käyttäjältä ja yksittäisen suodattimen lukitseminen

Kolmas mahdollinen toteutustapa erillisille toimialuekohtaisille raporteille olisi käyttää Power BI:n rivitason suojausta (englanniksi row level security, RLS). Tässä vaihtoehdossa eri käyttäjille määritellään erilaisia käyttöoikeusrooleja, joiden perusteella heillä on toisistaan poikkeavia oikeuksia nähdä raportin tietoja. Tässä tapauksessa visualisoinnit suodatetaan RLS-suodattimen kautta, joten näkymät voivat olla eri käyttäjille erilaisia. (Microsoft 2022b.)

Raportti voidaan julkaista organisaation sisällä käyttäjien saataville pilvipalvelun kautta. Jos raportin kohderyhmällä ei ole tunnuksia Power BI -pilvipalveluun, voidaan julkaisuun käyttää upotustoiminnon kautta myös jotain valmista organisaation sisäistä kanavaa, kuten Microsoftin Sharepoint -työtilaa tai Teams -kanavaa.

7 JOHTOPÄÄTÖKSET

Opinnäytetyön lopputuloksena tuotettiin julkaisuvalmis raportti neljän toimialueen potilaslukumääristä ja taloustiedoista. Raporttisivulla käytettiin kirjanmerkkien ja painikkeiden avulla samaa raporttisivun tilaa usean eri visualisoinnin esillepanoon, joten samalla jäi säästöön tilaa tulevia ajankohtaisia tietotarpeita varten.

Tietojoukon tiedot karkeistettiin tässä käyttötapauksessa anonymisoinnin toteuttamiseksi kuukausitasolle ja tiedot oli siksi verrattain helppo aggregoida tietolähteestä riittävän yleisellä tasolla hävittämättä raporttinäkymille olennaisia tietoja. Potilasmäärien laskennassa tietojen karkeutta pystyttiin lisäämään jättämällä yksittäiset potilastunnisteet pois tietojoukosta ja suorittamalla kuukausikohtainen laskenta jo poimintavaiheessa. Nämä toimenpiteet kuitenkin hieman rajoittivat raportin käytettävyyttä, kun tietoja ei voida enää tarkastella esimerkiksi viikkotasolla.

Raportin tietomallin tehokkuutta pyrittiin maksimoimaan tietomallin rakenteella luomalla tähtimallin luonteisia joukkoja toisiinsa kytkettävistä tiedoista. Erillisiä tähtimallisia joukkoja liitettiin toisiinsa sopivilla avainsarakkeilla. Sellaisille tiedoille, joille ei valmiiksi ollut toisiaan vastaavia avainsarakkeita, luotiin yhdistävä tekijä ajallista suodatusta varten itse luodulla avainsarakkeella sekä erillisellä kalenteritaululla. Tietomallin tietomäärän optimoinnissa tuli selkeästi ilmi, että suurin tekijä sarakkeen sisältämien tietojen kokoon on sarakkeessa olevien erilaisten arvojen lukumäärä, eli kardinaliteetti. Tässä tapauksessa taloustietojen kardinaliteettia ja samalla rivimäärää saatiin pienennettyä aggregoimalla tietoja ajallisesti.

Sosiaali- ja terveystietojen tuottamisessa syntyviä potilastietoja voidaan käyttää julkisten sote-palvelujärjestäjien tietojohtamisessa toisilain nojalla. EU:n tietosuojasetuksen ja toisilain vaatimusaineiston riittävästä anonymisoinnin tasosta ei kuitenkaan ole käytännön tasolla aina suinkaan helppo tehtävä, kun samalla halutaan säilyttää raportille poimittujen tietojen käyttökelpoisuus. Tietojen aggregointiin liittyvien kompromissien tekeminen onkin olennainen osa sote-alan tiedolla johtamisen raporttien tuottamista. Riittävästä anonymisoinnista ei voida tehdä kompromisseja, joten on tingittävä raportin käytettävyydestä ja tarvittaessa myös visualisointeihin nousevien tietojen tarkkuudesta.

Raportilta nähtävien tietojen analysointia voitiin rajoittaa piilottamalla taulukoita tai niiden sarakkeita loppukäyttäjältä. Näin kannattaa toimia varsinkin taulukkojen avainsarakkeiden tai pseudotunnisteiden

kohdalla. Lisäksi suodatinpaneeli voitiin poistaa loppukäyttäjän käytöstä ja tarjota raporttinäkymässä käytettäviksi vain muutamia suodattimia. Dynaamisesti päivittyvää raporttinäkymää on helpompaa hallita ja testata liian tarkkojen tietojen varalta, jos loppukäyttäjälle annetaan vain vähän vaihtoehtoja tietojen suodattamiseen. Erityistä tarkkuutta tietojen käsittelyssä ja visualisointien suodatusten suunnittelussa vaaditaan silloin, kun tietomallissa käytetään pseudotunnisteista dataa.

Raporttien käsittelyssä on myös huomattava, että arvio yksilön henkilöllisyyden paljastumisen jäännösriskistä on aina sidoksissa siihen, keiden käyttöön ja millä tavalla raportti julkaistaan. Tästä syystä raportit tulisi julkaisuhetkellä kiinnittää selkeisiin käyttöoikeuslistoihin ja käyttöoikeuksien laajentaminen vaatisi aina uutta arviota jäännösriskistä.

Lainsäädännön vaatimus riittävän anonymisoinnin tason arvioinnin tarkastamisesta teknologian kehityksessä vaatii paljon tietämystä ja vaikuttaa mielestäni resurssimielessä liian haastavalta tehtävältä pienelle hyvinvointialueelle. Tämän asian osalta vastuuta teknologian kehittymisen seurannasta tulisi osittain siirtää leveämmille hartioille, kuten sotealan tietojohdamisen kehittämisestä vastaaville valtakunnallisille verkostoille. Verkostot voisivat ylläpitää parasta ja ajankohtaisinta tietoa teknologisista muutoksista ja antaa niiden pohjalta hyvinvointialueille ohjeita tietojen käsittelyn toimintatapojen muuttamiseksi.

Power BI on erinomainen työkalu erilaisten aineistojen visualisointiin. Ohjelmiston kaikkien ominaisuuksien hyödyntäminen ei kuitenkaan ole sotealalla mahdollista, sillä se toisi käyttäjän lainsäädännön näkökulmasta liian tarkkojen tietojen äärelle. Power BI -raportoinnilla ei myöskään voida korvata sotealan yksilötason potilastietoja käsitteleviä raportointijärjestelmiä, jotka keräävät tietojen käytöstä lokitietoja.

Riittävästi karkeistettuun dataan perustuvilla Power BI -raportoinneilla tulevat hyvinvointialueet voivat kuitenkin valjastaa käyttöönsä valtavia määriä omassa hallussaan olevaa tietoa oman alueensa tilanteen seuraamiseksi ja tulevaisuuden päätöksenteon tueksi. Eri tietolähteiden tietoja yhdistelevien raporttien tarjoamat uudenlaiset näkymät ja tilannekuvat yhdistettynä sotealan ammattilaisten kokemukseen ja hiljaiseen tietoon antavat alalle hyvät mahdollisuudet tulevaisuuden haasteiden selättämiseen.

LÄHTEET

- Asiakastietolaki 784/2021. Saatavissa: <https://www.finlex.fi/fi/laki/alkup/2021/20210784>. Viitattu 16.1.2022.
- Beaulieu, A. 2009. Learning SQL. 2. painos. Sebastopol, CA: O'Reilly.
- Best Practice Analyzer. TE2. Saatavissa: <https://docs.tabulareditor.com/te2/Best-Practice-Analyzer.html>. Viitattu 11.2.2022.
- Clark, D. 2021. SQL vs. T_SQL: Understanding the Differences. Saatavissa: <https://www.dataquest.io/blog/sql-vs-t-sql/>. Viitattu 18.1.2022.
- DAX Studio. 2021. The ultimate tool for working with DAX queries. Saatavissa: <https://daxstudio.org/>. Viitattu 6.2.2022.
- DAX Guide. 2022a. HASONEFILTER DAX Function. Saatavissa: <https://dax.guide/hasonefilter/>. Viitattu 11.2.2022.
- DAX Guide. 2022b. SELECTEDVALUE DAX Function. Saatavissa: <https://dax.guide/selectedvalue/>. Viitattu 11.2.2022.
- EU:n yleinen tietosuoja-asetus 2016/679. Saatavissa: <https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=celex:32016R0679>. Viitattu 11.2.2022.
- Ferrari, A. & Russo, M. 2017. Analyzing Data with Microsoft Power BI and Power Pivot for Excel. 1.painos. Redmond, WA: Microsoft Press.
- Kelley, B. 2014. Add a Salt with the SQL Server HASHBYTES Function. Saatavissa: <https://www.mssqltips.com/sqlservertip/3293/add-a-salt-with-the-sql-server-hashbytes-function/>. Viitattu 11.2.2022.
- Kinanen, T-M. 2021a. VIRTÄ-arkkitehtuuri. Esitys tietojohdamisen verkoston kokouksessa 11.3.2021. Videotalenne. Saatavissa: <https://dreambroker.com/channel/63pxjpre/uit9psg0>. Viitattu 15.1.2022.
- Kinanen, T-M. 2021b. VIRTÄ-arkkitehtuuri. Esitys tietojohdamisen verkoston kokouksessa 11.3.2021. Power Point- esitys. Saatavissa: https://digifinland.fi/wp-content/uploads/2021/04/Virta-arkkitehtuuri_11.3.2021.pdf. Viitattu 15.1.2022.
- Klemola, K., Uusi-Illikainen, J., Askola, T. 2014. Sosiaali- ja terveystietojohdamisen käsikirja. Helsinki: Sitra. Saatavissa: https://media.sitra.fi/2017/02/27174607/Sosiaali_ja_terveyspalveluiden_tietojohdamisen_kasikirja-2.pdf. Viitattu 11.2.2022.
- Laihonen, H. 2009. Terveystietojohdamisen johtamisen tietovirrat. Kunnallistieteellinen aikakauskirja 37. Helsinki: Kunnallistieteiden yhdistys. Saatavissa: <https://urn.fi/URN:NBN:fi:ELE-1524467>. Viitattu 11.2.2022.

Laihonen, H. 2012. Tiedonhallinnasta tietojohdamiseen sote-sektorilla. Power Point-esitys. Saatavissa: <https://pikassos.fi/aineistot/send/40-pirkanmaan-sosiaalijohdon-tyoekokous-2-11-2012/127-laihonen-pirkanmaa-sote-johtajat-02112012>. Viitattu 11.2.2022.

Laihonen, H., Hannula, M., Helander, N., Ilvonen, I., Jussila, J., Kukko, M., Kärkkäinen, H., Lönnqvist, A., Myllärniemi, J., Pekkola, S., Virtanen, P., Vuori, V., Yliniemi, T. 2013. Tietojohdaminen. Tampereen teknillinen yliopisto, Tietojohdamisen tutkimuskeskus Novi. Saatavissa: <https://urn.fi/URN:ISBN:978-952-15-3058-6>. Viitattu 11.2.2022.

Leskelä, R-L., Haavisto, I., Jääskeläinen, A., Sillanpää, V., Helander, N., Laasonen, V., Ranta, T., Torkki, P. 2019. Tietojohdaminen ja sen kehittäminen: tietojohdamisen arviointimalli ja suosituksia maakuntavalmistelun pohjalta. Helsinki: Valtioneuvoston kanslia. Saatavissa <http://urn.fi/URN:ISBN:978-952-287-754-3>. Viitattu 11.2.2022.

Leskelä, R-L., Haavisto, I., Rissanen A., Lindh, M., Laasonen V., Manu S., Ranta, T., Helander, N., Jääskeläinen, A., Sillanpää, V., Torkki, P. 2018. Tulevaisuuden tietojohdaminen maakunnissa. Väliraportti 26.10.2018. Saatavissa: https://soteuudistus.fi/documents/1477425/3223876/tietojohdaminen_väliraportti_2018.pdf/. Viitattu 11.2.2022.

Microsoft. 2020. SQL tools overview. Saatavissa: <https://docs.microsoft.com/en-us/sql/tools/overview-sql-tools?view=sql-server-ver15>. Viitattu 18.1.2022.

Microsoft. 2021a. Download SQL Server Management Studio (SSMS). Saatavissa: <https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-ver15>. Viitattu 18.1.2022.

Microsoft. 2021b. Power BI -palvelun ominaisuudet käyttöoikeustyyppin mukaan. Saatavissa: <https://docs.microsoft.com/fi-fi/power-bi/fundamentals/service-features-license-type>. Viitattu 18.1.2022.

Microsoft. 2021c. Publish to web from Power BI. Saatavissa: <https://docs.microsoft.com/en-us/power-bi/collaborate-share/service-publish-to-web>. Viitattu 18.1.2022.

Microsoft. 2021d. Tietojen päivittäminen Power BI:ssä. Saatavissa: <https://docs.microsoft.com/fi-fi/power-bi/connect-data/refresh-data>. Viitattu 18.1.2022.

Microsoft. 2022a. HASHBYTES (Transact-SQL). Saatavissa: <https://docs.microsoft.com/en-us/sql/t-sql/functions/hashbytes-transact-sql>. Viitattu 11.2.2022

Microsoft. 2022b. Rivitason suojaus (RLS) Power BI:ssä Saatavissa: <https://docs.microsoft.com/fi-fi/power-bi/admin/service-admin-rls>. Viitattu 17.2.2022.

Mikä Power BI on? Saatavissa: <https://docs.microsoft.com/fi-fi/learn/modules/introduction-power-bi/2-what-power-bi>. Viitattu 18.1.2022.

Otykier, D. 2021. Tabular Editor. Saatavissa: <https://docs.tabulareditor.com/>. Viitattu 11.2.2022.

Russo, M. & Ferrari, A. 2020. The definitive guide to DAX: Business intelligence with Microsoft Power BI, SQL Server analysis services, and Excel. Second edition. Pearson Education, Inc.

Russo, M. 2020. Tuning query limits for Direct Query. Saatavissa: <https://www.sqlbi.com/articles/tuning-query-limits-for-directquery/>. Viitattu 11.2.2022.

Russo, M. 2021. Data Model Size with Vertipaq Analyzer. Saatavissa: <https://www.sqlbi.com/articles/data-model-size-with-vertipaq-analyzer/>. Viitattu 15.1.2022.

Saxton, A. 2021. Mastering Power BI bookmarks. Youtube-video. Saatavissa: <https://www.youtube.com/watch?v=xCMqWEvSkAs>. Viitattu 6.2.2022.

Sosiaali- ja terveysministeriö. 2020. Sosiaali- ja terveystietojen toissijaisen käytön tietosuojan asiantuntijaryhmä. Saatavissa: https://stm.fi/documents/1271139/2013549/Pseydonymisointi_anoymisointi_linjaus.pdf/. Viitattu 15.1.2022.

Sote-uudistus. Toivo-ohjelmassa kehitetään hyvinvointialueiden ja kansallisten viranomaisten tietojohdantamista. Sosiaali- ja terveysministeriö. Saatavissa: <https://soteuudistus.fi/toivo-ohjelma>. Viitattu 16.1.2022.

Tietojen mallinnus ja visualisointi. Saatavissa: <https://docs.microsoft.com/fi-fi/learn/modules/introduction-power-bi/2a-data-modeling-visualizations>. Viitattu 18.1.2022.

Tietosuojavaltuutettu. Mikä on henkilötieto? Verkkosivu. Saatavissa: <https://tietosuoja.fi/mika-on-henkilotieto>. Viitattu 11.2.2022.

Toisiolaki 552/2019. Saatavissa: <https://www.finlex.fi/fi/laki/alkup/2019/20190552>. Viitattu 16.1.2022.

```

-- Poimitaan tiedot hoitopäivistä vuodesta 2017 eteenpäin
SELECT
  vuosi
  ,kuukausi
  ,CONCAT(vuosi,'-',REPLACE(STR(kuukausi, 2), ' ', '0') )
    as VUOSIKK --Luodaan aika-avaintunniste tyyppiä VVVVKK
  ,kustannuspaikkaNro
  ,kontakti.KayntityyppiAvain
  ,kontakti.ToimintotyyppiAvain
  ,SUM(lukumaara) as lukumaara
  ,SELECT CASE WHEN lisatieto.lisatieto_1 like 'Koronakontakti'
    OR lisatieto.lisatieto_1 like 'Koronarokote' THEN 1 ELSE 0 END)
    as OnkoKoronakontaktiTaiRokote

FROM tietokanta.kayntitiedot_taulu kaynnit

LEFT OUTER JOIN tietokanta.potilaskontakti_taulu kontakti
  on kaynnit.potilaskontaktiAvain = kontakti.Avain

LEFT OUTER JOIN tietokanta.kayntityyppi_taulu kayntityyppi
  on kontakti.KayntityyppiAvain = kayntityyppi.Avain

LEFT OUTER JOIN tietokanta.lisatieto_taulu lisatieto
  on kontakti.LisatietoAvain=lisatieto.Avain

WHERE vuosi >= 2017 AND kaynnit.kayntilaji like 'Hoitopäivä'

GROUP BY
  vuosi, kuukausi
  ,CONCAT(vuosi,'-',REPLACE(STR(kuukausi, 2), ' ', '0') ) as VUOSIKK
  ,kustannuspaikkaNro
  ,kontakti.KayntityyppiAvain
  ,kontakti.ToimintotyyppiAvain
  ,SELECT CASE WHEN lisatieto.lisatieto_1 like 'Koronakontakti'
    OR lisatieto.lisatieto_1 like 'Koronarokote' THEN 1 ELSE 0 END)

```