

**AVOIMEN TERVEYS- JA VÄESTÖTIETOAINIESTON VISUALISOINTI
POWER BI -SOVELLUKSELLA**



Ammattikorkeakoulututkinnon opinnäytetyö

Hämeenlinna, Tietojenkäsittelyn koulutusohjelma

Syys, 2018

Nina Huikuri

Tietojenkäsittelyn koulutusohjelma
Hämeenlinna

Tekijä	Nina Huikuri	Vuosi 2018
Työn nimi	Avoimen terveys- ja väestötietoaineiston visualisointi Power BI -sovelluksella	
Työn ohjaaja	Lasse Seppänen	

TIIVISTELMÄ

Avoim data on kuluvan vuosikymmenen ilmiö. Käsite avoin data viittaa yleensä julkishallinnon avoimiin tietoaineistoihin. Myös tämä opinnäyte-työ tarkastelee julkishallinnon tietoaineistoja ja kartoittaa hyötyjä, joita tietoaineistojen avaamisen ennakoivaan tuovan niin kansantaloudellisesti kuin yhteiskunnallisestikin. Ilmiön uutuuden vuoksi kokonaisvaltaista tutkimustietoa avoimesta datasta on kuitenkin vielä varsin vähän saatavilla.

Työn teoriaosuudessa tarkastellaan myös tiedon visualisointia ja liiketoimintatiedon hallintaa (Business Intelligence, BI). Tavoitteena oli tutkia case-tutkimuksen avulla BI-välineen käyttöä avoimen datan visualisoinnissa. BI-välineistä esitellään lähemmin Microsoftin Power BI -työpöytäsovellus.

Opinnäytetyön käytännön osuudessa raportoidaan datan käsittelyä ja visualisointia Power BI -työpöytäsovelluksella. Aineiston teemana on terveyteen ja väestöön liittyvä avoimesti julkaistu tieto. Aineistojen latausta kehitettiin sekä tiedostomuotoisena että avoimen ohjelmointirajapinnan kautta.

Power BI -työpöytäsovellus soveltuu avoimen tiedon käsittelyyn ja visualisointiin. Sovelluksen parhaaksi anniksi osoittautuivat vuorovaikutteiset ja helppokäyttöiset visualisoinnit, joita voi jakaa pilvipalvelun kautta. Avointa dataa löytyy runsaasti ja Suomi onkin avoimen datan julkaisijoiden kärkimaita. Avoimen datan käytettävyyttä ja löydettävyyttä helpottaisi avoimen datan keskitetty jakaminen, aineiston tunnistetiedot ja jakelumaattien ajantasaisuus.

Avainsanat avoin tieto, tiedon visualisointi, liiketoimintatiedon hallinta, Power BI

Sivut 44 sivua, joista liitteitä 5 sivua

Degree Programme in Business Information Technology
Hämeenlinna

Author	Nina Huikuri	Year 2018
Subject	Visualization of the Open Health and Population Data with the Power BI Application	
Supervisor	Lasse Seppänen	

ABSTRACT

The purpose of the thesis was to focus on open data and discover what kind of researches have been done. The main focus is on the public organizations and what kind of benefits the open data can bring both economically and socially. The open data phenomenon is rather new and it was difficult to discover extensive researches about this subject.

The thesis has a short introduction about visualization of information and business intelligence. The purpose was to inspect how well business intelligence application fits for the visualization of the open data. The thesis includes an introduction about the Microsoft Power BI -application too.

The thesis also includes a report from the handling and the visualization of the open data with the Power BI -application. The theme of the open data is health and population register. The data was downloaded both via the API-interface and datafiles.

The tests showed that the Power BI -application is suitable for handling and visualization of the open data. You can create interactive and user-friendly graphics and share them via a cloud-service.

There are plenty of open data available and Finland is well positioned in international comparisons. It would help to use and exploit the data if it would be shared centrally, with identifier information and via up-to-date file versions.

Keywords Open data, visualization of information, business intelligence, Power BI

Pages 44 pages including appendices 5 pages

SISÄLLYS

1	JOHDANTO.....	1
2	TYÖN TAVOITTEET	2
3	AVOIN DATA	3
3.1	Avoimen datan määritelmä.....	3
3.2	Avoimien tietoaineistojen käyttöluvat ja julkaisu.....	4
3.3	Datan avaaminen julkisissa organisaatioissa	4
3.4	Avoimen datan jatkojalostuksen hyödyt	6
3.5	Avoimen datan yhteiskunnalliset hyödyt.....	6
3.6	Avoimien tietoaineistojen tilanne Suomessa.....	8
3.7	Avoimen datan keskitetty jakelualusta	8
4	AVOIN OHJELMOINTIRAJAPINTA.....	10
4.1	Ohjelmointirajapinnan määritelmä.....	10
4.2	Ohjelmointirajapinnan toiminta käytännössä	10
5	TIEDON VISUALISOINTI	12
5.1	Visualisoinnin toteutus käytännössä.....	12
5.2	Tiedon visualisointivälineet.....	12
6	LIIKETOIMINTATIEDON HALLINTA (BI).....	14
6.1	BI-raportit.....	14
6.2	BI-välineiden koontinäytöt eli käyttöliittymät tietoon	15
6.3	Power BI	16
6.3.1	Datan käsittelyominaisuuksia.....	17
6.3.2	DAX-kieli	17
7	PILVIPALVELUT TUOVAT JOUSTAVUUTTA RAPORTOINTIIN.....	18
8	KÄYTETYT MENETELMÄT	19
8.1	Opinnäytetyössä käytetyt aineistot	20
8.2	Aineiston lataus ja käsittely Power BI-sovelluksella	21
8.3	Tiedonkeruun riittävyys ja tuloksien tulkinta.....	22
9	DATAN LATAUS THL:N TIKU-JULKAISUJÄRJESTELMÄSTÄ	23
9.1	Tiedostomuotoisen datan lataus	23
9.2	Datan lataus rajapintakutsuna THL:n TIKU-julkaisujärjestelmästä.....	24
9.2.1	CSV-rajapintakutsu	25
9.2.2	JSON-rajapintakutsu	26
9.3	Parametrien asetus THL:n aineistoille.....	27
10	DATAN KÄSITTELY POWER BI -TYÖPÖYTÄSOVELLUKSESSA	29

10.1 Taulujen yhdistäminen ja liittäminen.....	30
10.1.1 Taulujen yhdistäminen "Merge"	30
10.1.2 Taulujen liittäminen "Append"	31
10.2 Tietojen suodatus	32
10.3 Mittarin muodostaminen DAX-funktiolla.....	33
11 DATAN VISUALISOIMINEN POWER BI -SOVELLUKSELLA	35
12 KOONTINÄYTÖT JA RAPORTIN JAKAMINEN PILVESSÄ	38
13 TULOKSET	39
13.1 Tiedostomuotoisen datan lataus	39
13.2 Rajapintakutsut datan latauksessa	40
13.3 Power BI -työkalu	40
13.4 Avoimet tietoaineistot	41
14 JOHTOPÄÄTÖKSET	44
LÄHTEET	45

Liitteet

Liite 1	KÄYTETYT THL:N AINEISTOJEN CSV-RAJAPINTAKUTSUT
Liite 2	VÄESTÖAINEISTOJEN LATAUSOSOITTEET
Liite 3	HAVAINTOLOMAKE

1 JOHDANTO

Valitsin avoimen datan opinnäytetyön aiheeksi, koska aihe on kuluvan vuosikymmenen ilmiö ja avoimen datan vaikutukset on arvioitu ennakkoon yhteiskunnallisesti merkittäviksi. Avoin data helpottaa kansalaisten arkielämää ja lisää organisaatioiden toiminnan läpinäkyvyyttä. Avoimen tiedon avulla yhteiskunta tulee entistä demokraattisemmaksi. (Valtiovarainministeriö 2015a, 1–2.)

Halusin saada ilmiöön ymmärryksen ja kartoittaa, mitä tutkimukseen perustuvaa tietoa avoimesta datasta on saatavilla ja mitä niissä on kyseisestä ilmiöstä todettu. Valitsin esimerkkiaineistoiksi avoimia aineistoja Terveystieteiden ja hyvinvoinnin laitokselta, Kuntaliitolta, pääkaupunkiseudun kuntien ja Tilastokeskuksen tuottamaa aineistoa Avoindata.fi -palvelusta. Aihe on myös omakohtainen siinä mielessä, että olen työskennellyt Terveystieteiden ja hyvinvoinnin laitoksessa (jatkossa THL) tietoaineistojen Hydra-tietomallin parissa. Hydra-tietomuoto on THL:ssa kehitetty avoimien tietoaineistojen tietomuoto.

Tietoa on nykyään paljon tarjolla, mutta miten tätä tietoa voisi tuoda esille ja visualisoida helposti ymmärrettävään ja helposti käytettävään muotoon? Edellä esitetty kysymys oli mielessäni alkaessani tekemään tätä opinnäytetyötä. Halusin myös kartoittaa, mitä tiedon visualisointi on ja miten ja millä periaatteilla sitä käytännössä tehdään. BI-välineillä on perinteisesti visualisoitu organisaation omaa, sisäistä tietoa, mutta miten BI-väline soveltuu avoimen datan käsittelyyn ja visualisointiin? Myös tätä halusin selvittää työssäni.

Toiminnallinen opinnäytetyöni jakautuu kahteen osaan: teorian ja käytännön osioon. Teoriaosassa selvitän avoimen datan teoriaa, tutkimustietoa avoimen datan hyödyistä, tiedon visualisointia ja liiketoimintatiedon hallinnan (BI) määritelmää ja käyttökohteita. BI-välineistä esittelen tarkemmin Microsoftin Power BI -työpöytäsovelluksen. Työn käytännön osuudessa raportoin kokeistani, joita suoritin Power BI -työpöytäsovelluksella avoimille tietoaineistoille.

Asetin opinnäytetyölleni alla olevat tutkimuskysymykset, joista ensimmäiseen vastataan työn teoriaosuudessa ja kahteen viimeiseen työn käytännön osuudessa:

- Mitä on avoin data ja mitä sen hyödyistä on todettu tutkimuksissa?
- Miten avointa dataa tarjotaan käytettäväksi?
- Miten avoin terveys- ja väestötietoaineisto soveltuu BI-välineellä käsiteltäväksi ja visualisoitavaksi?

2 TYÖN TAVOITTEET

Opinnäytetyö on tyypiltään toiminnallinen opinnäytetyö. Vilkka ja Airaksinen määrittelevät toiminnallisen opinnäytetyön tavoitteeksi mm. ammatillisen käytännön toiminnan ohjeistamisen ja opastamisen sekä toiminnan järjestämisen tai järjeistämisen. Toiminnallisessa opinnäytetyössä yhdistyvät käytännön toteutus ja raportointi tutkimusviestinnän keinoin. (Vilka & Airaksinen 2003, 9.)

Työn tarkoituksena oli kartoittaa avoimen datan saatavuutta, käytettävyyttä, käsittelyä ja BI-välineen tarjoamia tiedon visualisointi mahdollisuuksia. Edellä mainittuja asioita lähestyttiin kokeissa case-strategialla ja kokeiltiin käytännössä avoimen tiedon visualisointia Power BI-työpöytäsovelluksella. Tarkoitus oli tutkia, miten BI-väline soveltuu avoimen tiedon käsittelyyn ja tuottaa tästä käytäntöön perustuvaa tietoa sekä ohjeistusta ja mahdollisia suosituksia avoimen datan latauksesta ja käsittelystä.

Kuten kirjassa *Kehittämistyön menetelmät – uudenlaista osaamista liiketoimintaan* todetaan, case-tutkimusta luonnehtii pyrkimys tuottaa syvälistä ja yksityiskohtaista tietoa. Tutkimuksessa tarkastellaan, kuinka jokin on mahdollista tai kuinka jokin tapahtuu. Kyse ei ole siitä, kuinka yleistä jokin on. (Ojasalo, Moilanen & Ritalahti 2014, 37, 52–53.) Opinnäytetyön käytännön osuudessa on ollut tavoitteena tutkia, miten avoimen tietoa-ineiston lataus ja käsittely tapahtuu Power BI-sovelluksella sekä miten tämä työväline soveltuu saatavilla olevien avoimien terveys- ja väestötietoa-ineistojen visualisointiin.

3 AVOIN DATA

Yhdysvallat julkaisi avoimen datan periaatteet vuonna 2008, jonka jälleen julkishallinnon avoimet tietoaaineistot tulivat keskustelun aiheeksi myös Suomessa (Valtiovarainministeriö 2015b, 18). Termillä avoin data viitataankin yleensä julkishallinnon avoimiin tietovarantoihin (Salo 2014, 43). Julkisen organisaation keräämät, tuottamat ja hallinnoimat tietoaaineistot voidaan luokitella avoimiksi, mikäli tietoaaineistot ovat luettavissa koneellisesti, saatavilla maksuttomasti ja kenen tahansa vapaasti käytettävissä, muokattavissa sekä eteenpäin jaettavissa. Uudelleen käyttö voi olla sekä yksityistä että kaupallista. (Valtiovarainministeriö 2015a, 1.)

3.1 Avoimen datan määritelmä

Kirjassa *Issues in open research data* määritellään kolme seikkaa, joiden perusteella tiedon saatavuutta ja avoimuutta voidaan arvioida. Määriteltäessä tietoaaineiston avoimuutta voidaan dataa tarkastella kolmen seuraavan kysymyksen näkökulmasta: Ovatko tiedot luettavissa koneellisesti (tekstiä kuvan sijasta)? Ovatko tiedot saatavissa avoimessa tiedostoforfaatissa (CSV)? Ovatko tiedot avoimesti lisensioitu (Creative Commons lisenssi)? Tämän lisäksi kirjassa todetaan, että datan liitännäistiedot ovat tärkeitä datan jatkohyödyntämisen kannalta. (Moore 2014, 133, 136.)

Pelkkä datan avaaminen ei kuitenkaan riitä. Datan täytyy olla myös helposti saatavissa, ja tämän turvaamiseksi tarvitaan yhteisiä standardeja, portaaleja, rajapintoja ja käyttöehtoja. Datan ollessa hajallaan hyödyt jäävät vähäisiksi verrattuna hyötyihin, joita voidaan saavuttaa keskitetysti hallinnoitulla ja helposti saatavilla olevalla datalla. (Salo 2014, 43–44.)

Valtiovarainministeriön Avoimen tiedon ohjelman 2013–2015 loppuraportissa todetaan, että datan on oltava rakenteisessa muodossa, jotta sitä pystytään yksilöimään, tekemään siihen poimintoja sekä yhdistelemään koneellisesti (Valtiovarainministeriö 2015b, 7). Helposti saatavissa olevaa avointa dataa voidaan yhdistellä myös muiden organisaatioiden tietoihin ja näin datalle voi syntyä aivan uusia sovelluskohteita ja innovaatioita, joita ei välttämättä osattu edes etukäteen kuvitella (Salo 2014, 43–44). Tietojen jatkokäyttö ja hyödyntäminen edellyttävät yhteen toimivia tietorakenteita ja metatietoja (Valtiovarainministeriö 2015b, 21).

Tieto on organisaation pääomaa, jonka arvo kasvaa tietoa edelleen jalostettaessa ja jaettaessa eteenpäin. Julkishallinto voi kartuttaa tietopääomaa helpottamalla hallussaan olevan tiedon käyttöä. Valtiovarainministeriön Avoimen tiedon ohjelman 2013–2015 loppuraportin mukaan tietoa voidaan luokitella neljään kategoriaan, joita ovat data, informaatio ja tieto tai tietämys sekä ymmärrys ja viisaus. (Valtiovarainministeriö 2015b, 7, 18.)

3.2 Avoimien tietoaineistojen käyttöluvat ja julkaisu

Avoimien tietoaineistojen käyttöluvaksi on laadittu julkisen hallinnon suositus (JHS 189). Suosituksessa todetaan, että julkishallinnossa sovelletaan käyttöilupaa Creative Commons Nimeä 4.0. Käyttöluvan perusteella tietoa-aineiston käyttäjä voi vapaasti kopioida, muokata ja jakaa tietoa-aineistoa. Julkisen hallinnon neuvottelukunta JUHTA on hyväksynyt suosituksen otettavaksi käyttöön. (Valtiovarainministeriö 2015b, 15.)

Käyttö lupa antaa käyttäjälleen mahdollisuuden kopioida, levittää ja jatkojalostaa dataa. Datasta voi luoda uusia aineistoja ja hyödyntää myös kaupallisessa tarkoituksessa. Lisenssi edellyttää, että datan lähde mainitaan käyttöluvassa ilmoitetulla tavalla sekä ilmoitetaan, mikäli dataan on tehty muutoksia. Jatkokäyttäjä ei voi myöskään asettaa datalle ehtoja tai teknisiä rajoitteita, jotka estävät kolmatta osapuolta tekemästä mitään, mitä lisenssi oikeuttaa. (creative commons n.d.)

Toinen lisenssi, jota voidaan käyttää julkishallinnon avoimissa tietoa-aineistoissa, on CCO-käyttö lupa. Tätä voi käyttää silloin, kun datan yksilöiminen tai tuottajan tietojen ilmoittaminen ei ole tarpeellista. CCO-käyttö lupa merkitsee käytännössä, että tietojen julkaisija luopuu kaikista yksinoikeuksistaan dataan. (Valtioneuvoston kanslia 40/2017, 49.)

Avointa tietoa voidaan jakaa tiedostojen, katselu- ja latauspalvelun ja ohjelmointirajapinnan kautta (Valtioneuvoston kanslia 40/2017, 52). Avointa dataa voidaan julkaista verkkosivulla, jossa toimii katselu- ja latauspalvelu. Katselu- ja latauspalvelua käytetään yleensä käyttöliittymän tai rajapinnan kautta. Aineistoja voi ladata katselu- ja latauspalvelun kautta paikallisesti toiselle koneelle tai palvelimelle. (Valtioneuvoston kanslia 40/2017, 54-55.) Suuri osa avoimesta datasta julkaistaan ohjelmointirajapintojen kautta (Valtioneuvoston kanslia 40/2017, 41). Ohjelmointirajapinta on tehokas tiedonsiirto- ja jakamistapa erityisesti siinä tapauksessa, kun aineisto on suuri, reaaliaikaista tai sitä päivitetään usein (Valtioneuvoston kanslia 40/2017, 55).

Valtioneuvoston raportissa *Avoimen datan hyödyntäminen ja vaikuttavuus* todetaan avoimen datan jakelun olevan pirstaleista. Pirstaleisesti jaettu avoin tieto tekee tiedon löytämisestä ja hyödyntämisestä haasteellista. Raportissa todetaan, että tietoa-aineistojen hyödyntämistä voitaisiin lisätä lisäämällä avoimen datan hallinnoijien ja hyödyntäjien vuorovaikutusta. (Valtioneuvoston kanslia 40/2017, 2.)

3.3 Datan avaaminen julkisissa organisaatioissa

Julkinen sektori kokoaa ja tilastoi eri aloja koskevaa tietoa tietokantoihin ja rekistereihin. Julkishallinnon organisaatiot myös päivittävät, ylläpitävät ja tarjoavat tätä tietoa eteenpäin. (Valtioneuvosto 2011, 2.) Valtioneuvoston kanslian raportissa todetaan, että tietoa-aineistojen avaaminen on 2010-

luvun megatrendi (Valtioneuvoston kanslia 40/2017, 40). Julkisia tietovarastoja on alettu avaamaan viime vuosina niin Suomessa kuin muuallakin maailmassa sekä kansalaisten että yritysten kuin koko yhteiskunnan käyttöön. On arveltu, että tietojen avaamisen hyödyt ovat suuremmat kuin tiedon avaamisesta syntyvät kustannukset. Taustalla on myös julkisuusperiaate sekä ajatus, että pääsääntöisesti verovaroin kerätyt tietovarannot tulee olla maksutta kansalaisten käytettävissä. (Valtiovarainministeriö 2015b, 7.)

Valtioneuvoston periaatepäätös julkisen sektorin digitaalisten tietoaineistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä 3.3.2011 asettaa tavoitteeksi, että julkisen sektorin tietovarannot ovat laajasti koko yhteiskunnan käytössä. Tällä tavoitellaan innovaatio- ja tutkimustoiminnan lisääntymistä ja tuotteiden, palvelujen sekä markkinoiden digitalisaation kehittymistä. Periaatepäätöksen tavoitteena on myös hallinnon tehokkuuden, vaikuttavuuden ja läpinäkyvyyden lisääminen sekä kansalaisten osallistumisen mahdollistaminen päätöksentekoon. (Valtioneuvosto 2011, 1–2.)

Valtiovarainministeriö asetti vuonna 2013 kaksivuotisen ohjelman, jonka tavoitteena on avata julkiset tietovarannot mahdollisimman nopeasti ja laajasti. Päämääränä on, että vuoteen 2020 mennessä kaikki merkittävät julkishallinnon tietovarannot on avattu ja ne ovat käytettävissä koneluettavassa muodossa. Tämän lisäksi aineistot ovat saatavilla maksutta ja niiden uudelleen käyttö on merkitty selkein käyttöehdoin. Myös Euroopan komissio on julkaissut joulukuussa 2011 avoimen datan strategian. (Valtiovarainministeriö 2015a, 8.)

Avoimen tiedon ohjelman tavoitteena on, että tietovarantojen avaaminen vakiintuu osaksi hallinnon toimintaa. Päämääränä on, että tietovarantojen avoimuus otetaan huomioon jo tietojärjestelmien suunnittelu-, kehittämis- ja hankintavaiheessa. Jatkossa painotetaan tietovarantojen hyödyntämistä ja tieto-osaamisen kasvattamista koko yhteiskunnassa. (Valtiovarainministeriö 2015b, 10.)

Julkisen tiedon avoimuuden kansainvälisessä vertailussa Suomi on sijoittunut hyvin erilaisilla mittareilla. Mittareilla mitataan julkisen hallinnon tietoaineistojen avoimuutta kuin myös avoimen tiedon uudelleenkäyttöä. (Valtiovarainministeriö 2015b, 13.) Valtioneuvoston kanslian maaliskuussa 2017 julkaiseman raportin *Avoimen datan hyödyntäminen ja vaikuttavuus* mukaan aineistojen avaaminen on hyvin pitkällä organisaatioissa, jotka hallitsevat keskeisiä rekisteritietoja. Aineistoista, joiden avaamiselle ei ole lakisääteistä estettä, oli avointa dataa keskimäärin 69 prosenttia. (Valtioneuvoston kanslia 40/2017, 13.)

3.4 Avoimen datan jatkojalostuksen hyödyt

Suomessa on maailman parhaat terveydenhuollon rekisterit, toteaa Sitran yliasiamies Mikko Kosonen Helsingin Sanomien Vieraskynä-kirjoituksessa *Terveystiedoista on tullut Suomen kansallisaarre* (Kosonen 2017). Kosonen toteaa, että kohta arvokkain raaka-aineemme voi olla terveyteen ja hyvinvointiin liittyvä tieto, jota emme ole vielä tähän mennessä tunnistanee riittävän hyvin, saati jatkojalostaneet siitä uutta omaisuutta. Kosonen painottaa julkisen ja yksityisen sektorin yhteistyötä sekä datan jatkojalostamista ja datan tarjoamista Suomessa toimivien tutkijoiden ja yritysten jatkokäyttöön. (Kosonen 2017.)

Data sisältää informaatiota, jota jalostamalla voidaan saada tietoa ja ymmärrystä sekä menneisyyteen, nykyisyyteen että tulevaan. Tiedon avulla tehdään myös päätöksiä, ja datan hyödyntämisellä päästään ymmärtämättömyydestä ymmärrykseen ja arvailuista tarkkoihin ennusteisiin. Datasta hyödynnettävän informaation avulla siirrytään intuitiivisesta päätöksenteosta tiedolla johtamiseen. (Salo 2014, 31.)

Tiedolla johtaminen on kirjattu myös THL:n strategiaan, jossa todetaan seuraavaa: ”Tarjoamme tietopohjan ja tiedolla johtamisen välineitä”. Myös tiedon jalostusasteen, tietoaineistojen ja tuotteiden hyödynnettävyyden parantaminen on kirjattu THL:n strategiaan. (THL 2016.)

Pelkkä tiedon olemassaolo ja varastointi ei tuo arvoa, vaan lisäarvo syntyy jalostuksen kautta. Perinteisesti tieto on ollut harvojen etuoikeus ja vallan väline, mutta meneillään oleva digitaalinen muutos tuo tiedon kaikkien ulottuville helpottamaan arkea ja päätöksenteon tueksi. (Markkula & Syväniemi 2015, 20, 36.)

3.5 Avoimen datan yhteiskunnalliset hyödyt

Julkisten toimijoiden avoimien tietoaineistojen vaikutus voidaan jakaa taloudellisiin ja yhteiskunnallisiin vaikutuksiin. Avoimen datan odotetaan vaikuttavan uusien tuotteiden ja palveluiden kehittymiseen, organisaatioiden toiminnan tehostumiseen ja tuottavuuden kasvuun. Avoimien tietoaineistojen odotetaan myös lisäävän uutta yritystoimintaa. Julkisen sektorin odotetaan saavan kustannussäästöjä tietovarantojen avaamisesta. Muita yhteiskunnallisia vaikutuksia ovat hallinnon ja päätöksenteon avoimuuden lisääntyminen sekä toiminnan läpinäkyvyys. Myös kansalaisten vaikutusmahdollisuudet ja osallisuuden odotetaan lisääntyvän. (Valtiovarainministeriö 2015a, 2.)

Tietoaineistojen avaamisella tavoitellaan myös monipuolistettavan koulutuksen ja tutkimuksen käytettävissä olevia tietoaineistoja (Valtiovarainministeriö, 2015b, 9). Tietoaineistojen avaaminen ja saatavuus mahdollistaa erilaisten ilmiöiden tutkimisen entistä helpommin. Kerättyjen aineistojen

ja niiden pohjalta tehtyjen selvitysten ja tutkimuksen jakamisella voidaan karsia päällekkäistä työtä. (Valtiovarainministeriö 2015a, 20.)

Julkisten tietovarantojen avaamisen vaikutus on arvioitu mittavaksi. Avoimien tietoaineistojen yhteiskunnallisten ja taloudellisten vaikutuksien mittaaminen on kuitenkin vaikeaa, eikä vaikutuksia pystytä tällä hetkellä arvioimaan kattavasti saatavilla olevan tiedon pohjalta. Vaikutuksien arviointiin tarvittaisiin seurantamallia ja mittaristoa sekä systemaattista tiedonkeruuta. Tämän lisäksi tarvittaisiin tietoa avoimeen dataan liittyvistä parhaista käytännöistä ja keinoja, miten nämä parhaat käytännöt saadaan leviämään. Avoimen datan yhteiskunnallisista tai taloudellisista vaikutuksista ei ole vielä juurikaan tutkimustietoa eikä tämänkaltaista laajaa tutkimusta ole tiettävästi tehty vielä missään maassa. (Valtiovarainministeriö 2015a, 1–3, 8.)

Tietovarantojen avaamisen tarkastelun ohella tarvitaan myös tietoa tiedon hyödyntämisen ja vaikuttavuuden näkökulmasta. Lisäksi tarvitaan seurantatietoa ja ymmärrystä avoimen tiedon vaikuttavuudesta. Edellä mainitut edellyttävät laadullista ja määrällistä tutkimusta. Valtiovarainministeriön julkaisussa todetaan, että avoimen tiedon hyödyntämistä ja menetelmien parantamista tulisi viedä eteenpäin osana laajempaa tietojohdamista. (Valtiovarainministeriö 2015b, 20–21.)

Valtioneuvoston kanslian maaliskuussa 2017 julkaisemassa raportissa *Avoimen datan hyödyntäminen ja vaikuttavuus* todetaan, että avointa dataa innovaatiotoiminnossaan hyödyntävät yritykset tekevät tavara- ja/tai palveluinnovaatioita huomattavasti muita yrityksiä useammin. Edellä mainittua esiintyy etenkin niissä yrityksissä, joissa digitaalisilla tuotteilla on suuri merkitys yrityksen liiketoiminnassa. (Valtioneuvoston kanslia 40/2017, 24.)

On todettu, että avoimen tiedon hyödyntäminen innovaatioiden tuottamisessa on yhteydessä innovaatioyhteistyöhön julkisten toimijoiden kanssa. Tämän lisäksi avointa dataa hyödyntävillä yrityksillä on keskimäärin muita useampia yhteistyötahoja kuin avointa dataa käyttämättömillä. (Valtioneuvoston kanslia 40/2017, 23.) On kuitenkin huomattava, että tällä hetkellä olemassa olevat aineistot avoimen datan hyödyntämisestä ovat varsin suppeita ja on vaikea arvioida missä määrin avoin data vaikuttaa kriittisesti innovaation syntyyn. (Valtioneuvoston kanslia 40/2017, 19, 22).

3.6 Avoimien tietoaisteojen tilanne Suomessa

6Aika-strategia on Suomen kuuden suurimman kaupungin hanke, jonka tavoitteena on kehittää älykkäitä ja avoimia palveluita. Hankkeeseen kuuluvat Helsinki, Espoo, Tampere, Vantaa, Oulu ja Turku. Strategian tavoitteena on saada aikaan uutta osaamista, työpaikkoja ja liiketoimintaa. Strategialla edistetään avoimia toimintamalleja, joilla pyritään kannustamaan koko kaupunkiyhteisön osallistumista palveluiden kehittämistyössä. Toimiva kaupunkiyhteisöön kuuluu kaupunkilaisia, yrityksiä, viranomaisia sekä tutkimus- ja kehitystoimijoita. (6Aika n.d.)

6Aika-hankkeen fyysisillä ja virtuaalisilla avoimilla innovaatioalustoilla voi testata kehitettäviä palveluita ja tuotteita, ja avointa dataa ja rajapintoja tarjotaan uusien palveluiden kehittämiseen. 6Aika-kaupungeilla on ollut lisäksi monia pilottihankkeita sekä kolme yhteistä kärkihanketta. 6Aika-innovaatioiden kehittämisellä parannetaan koko Suomen kilpailukykyä tarjoamalla merkittävät markkinat myös kansainvälisesti. Osoitteessa Data-Business.fi on nähtävillä esimerkkejä avoimen datan hyödyntämisestä sovelluksissa, palveluissa ja muissa tarkoituksissa. Sivulta löytyy yhteensä yli 100 näytettä. (6Aika n.d.)

Suomi on perinteisesti pärjännyt hyvin avoimen datan kansainvälisissä vertailuissa (Valtiovarainministeriö 2015b, 12-13). Avoimen tiedon vaikuttavuus –esitutkimuksessa todetaan, että etenkin paikkatiedot, säätiedot, väestötiedot, yritystiedot, liikennetiedot ja lainsäädännölliset tiedot ovat hyödynnetyimpiä julkisen sektorin avaamista tietoaisteista (Valtiovarainministeriö 2015a, 9).

World Wide Web Foundationin yllä pitämässä Open Data Barometrissa Suomi sijoittui sijalle 20 vuonna 2016. Suomen sijoituksessa on tapahtunut pienoista laskua edelliseen tulokseen verrattuna. Tarkasteltaessa Suomen tietoja lähemmin, voidaan todeta, että Suomesta löytyy avointa dataa, mutta avoimena datana tarjottavissa aineistoissa on puutteita etenkin tietojen yksilöimisessä tarvittavien tunnisteiden suhteen. (World Wide Web Foundation n.d.)

3.7 Avoimen datan keskitetty jakelualusta

Suomessa keskitettynä avoimen datan jakelualustana toimii Avoindata.fi-portaali. Portaalissa julkaistaan julkishallinnon avointa dataa, mutta myös yritykset voivat julkaista avointa dataansa palvelussa. Palvelun tavoitteena on edistää hallinnon läpinäkyvyyttä ja avoimen tiedon löydettävyyttä sekä saatavuutta. Palvelu perustettiin osana Valtiovarainministeriön avoimen tiedon ohjelmaa vuonna 2014. Palvelu pyrkii tukemaan liiketoimintaa, koulutusta ja tutkimusta sekä sähköisten palveluiden kehittämistä ja tietojärjestelmien yhteen toimivuutta. Avoindata.fi-palveluun tallennettu tieto on vapaasti käytettävissä ja tiedon tallennuksen yhteydessä valitaan millä ehdoilla tietoa voi hyödyntää uudelleen. Avoimet tietoaisteistot siirtyvät

Avoindata.fi-palvelusta Euroopan unionin avoimen datan portaaliin. (Avoindata.fi 2018.)

Palvelun tavoitteena on, että keskeisimmät julkishallinnon tietoaaineistot ovat julkaistu avoimena ja ovat maksutta saatavilla yhteiskunnan käytössä. Julkisen hallinnon tieto on julkisuuslain mukaan avointa, mikäli sitä ei ole laissa erikseen määritelty salattavaksi. Avoin data ei saa sisältää henkilö-tietoja tai liikesalaisuuksia. Data ei myöskään saa sisältää kolmansien osapuolien tekijänoikeuksia tai muuta erikseen määriteltyä salassa pidettävää tietoa. Palvelu sisältää avoimien tietoaaineistojen lisäksi myös ohjeita järjestelmien yhteen toimivuuden edistämiseksi, tietoa kehittämistyöstä ja aiheeseen liittyvästä koulutuksesta. Palveluun tuodaan tietoa sekä manuaalisesti että automaattisesti muista avoimen datan portaaleista. Avoindata.fi-palvelua kehittää ja ylläpitää Väestörekisterikeskus. (Avoindata.fi 2018.)

Avoindata.fi-palvelussa oli vuoden 2017 alussa yli 1600 tietoaaineistoa. Alustalle tallennettua dataa voi hyödyntää uudelleen joko koneluettavan ohjelmointirajapinnan kautta tai lataamalla aineistoa suoraan verkkosivulta. Portaalin avoimet tietoaaineistot kootaan Euroopan avoimen datan portaaliin, joka sisältää yli 600 000 aineistoa. (Valtioneuvoston kanslia 40/2017, 53.)

4 AVOIN OHJELMOINTIRAJAPINTA

Ohjelmointirajapinnan (Application Programming Interface, API) avulla määritellään, mitä tietoja tai palveluita ohjelma tai järjestelmä tarjoaa sekä miten niitä tarjotaan toisille ohjelmointirajapintaa kutsuville ohjelmille tai järjestelmille. Avoin ohjelmointirajapinta (Open Application Programming Interface, Open API) on ohjelmointirajapinta, jonka kaikki ominaisuudet ovat julkisia ja rajapinnan tarjoamia toimintoja voi käyttää vapaasti ilman rajoittavia ehtoja. Tämä mahdollistaa sen, että kuka tahansa voi rakentaa sovelluksen, joka hyödyntää ohjelmointirajapinnan tarjoamia palveluita. Rajapinnan hyödyntämisestä ei tarvitse pyytää erillistä hyväksyntää tai maksaa lisenssimaksuja. (Avoin rajapinta 2017.)

4.1 Ohjelmointirajapinnan määritelmä

Avoimen ohjelmointirajapinnan hyödyntäminen edellyttää vapaasti verkossa saatavilla olevaa ohjelmointirajapinnan kuvausta ja dokumentaatiota. Rajapinnan kautta tarjottavien tietoa-aineistojen tiedot ja niiden tietorakenne on käytävä ilmi dokumentaatiosta. Näiden lisäksi ohjelmointirajapintaa on voitava käyttää vapaasti omien sovellusten tekemisessä. Vapaan ohjelmointirajapinnan tarjoamien tietojen tai palveluiden käyttö ei siis saa edellyttää rajapinnan tarjoajan toimia. Vapaan ohjelmointirajapinnan tulee olla myös testattavissa ja tätä varten tarjolla on oltava vähintään testiaineisto. (Avoin rajapinta 2017.)

Ohjelmointirajapinta määrittelee, miten tietoja tai palveluita tarjotaan toisille sovelluksille tai järjestelmille, jotka ovat kutsuneet rajapintaa. Rajapinta voi toimia kahdella eri tavalla. Rajapinta voi olla joko datarajapinta, jonka kautta pystyy lukemaan dataa tai toiminnallinen rajapinta, joka sisältää laskenta-algoritmeja tai mahdollistaa datan muokkaamisen rajapinnan kautta. (Avoin rajapinta 2017.)

Avoimen rajapinnan määritelmä ei edellytä, että rajapinnan kautta saatavien tietojen täytyisi välttämättä olla avointa dataa tai saatavilla maksuttomasti. Rajapinta voi olla avoin, vaikka sen kautta tarjottaviin tietoihin olisi pääsy vain rajatulla joukolla. Esimerkiksi tiettyä henkilöä koskevat aineistot voidaan tarjota saatavaksi avoimen rajapinnan kautta, mutta tähän tarvitaan henkilön oma suostumus (my data). (Avoin rajapinta 2017.)

4.2 Ohjelmointirajapinnan toiminta käytännössä

Ohjelmointirajapinnan avulla rakennetaan yhteyksiä järjestelmien, laitteiden ja sovellusten välille. Rajapinta sisältää käskyjä ja toimintoja, joilla voi esimerkiksi hakea tietoa tai käyttää järjestelmää ulkopuolelta. API-rajapinnan dokumentaatio määrittelee, mitä toimintoja ohjelmointirajapinnassa on tarjolla ja miten niitä käytetään. (Kotilainen 2017.)

Verkon API-rajapinta toteutetaan pääasiassa kahdella tavalla. Yleisemmin käytetään REST-rajapintaa (representational state transfer), jossa käytetään yleensä JSON-tiedostomuotoa. REST-rajapintatoteutuksessa sovellus kutsuu rajapintaa. Nykyään käytössä on myös ns. reaaliaikainen rajapinta. Reaaliaikaisessa rajapinnassa API lähettää tietoa koko ajan ilman, että sovelluksen tarvitsee sitä pyytää. (Kotilainen 2017.)

Verkko (world wide web) perustuu kolmeen teknologiaan, URL-nimeämiskäytäntöön, HTTP-protokollaan ja HTML-dokumenttimuotoon (Richardson & Amundsen 2013, 1). URL identifioi jonkun resurssin (kotisivun, kuvatiedoston jne.) web-palvelimella. Sovellus lähettää web-palvelimelle HTTP-pyyntö (HTTP request) ja HTTP:n GET-metodi, pyytää palvelimelta URL-osoitteen identifioiman resurssin sisällön. Web-palvelin lähettää vastauksen (response) sovellukselle, joka on esimerkiksi jotain web-palvelimella hallussa olevaa dataa. (Richardson & Amundsen 2013, 2–3, 6.)

5 TIEDON VISUALISOINTI

Tiedon visualisoinnilla pyritään asian tiivistämiseen, ymmärryksen lisäämiseen, yksinkertaistamiseen ja nopeaan viestin perille menoon. Tiedon visualisointi ei ole siis vain kuvien lisäämistä, vaan nimenomaan ymmärryksen lisäämistä. Tiedon visualisoinnin avulla voidaan tuoda esille ja korostaa haluttuja asioita viestissä. Tieto hukkuu helposti tekstin ja numeroiden joukkoon, mutta visuaalisilla elementeillä voidaan esimerkiksi korostaa poikkeamia ja vertailla asioita. Tehokeinoina käytetään värejä, grafiikkaa ja numeroita. (Rintakoski 2017.)

Kirjassa *Tieto näkyväksi* tiedon visualisointi määritellään näin: Visualisointi on grafiikka, joka paljastaa uusia piirteitä aineistosta. Visualisointi ei painotu ainoastaan tiedonvälitykseen vaan myös uuden tiedon löytämiseen. Visualisoinnin avulla lukija voi itse löytää aineistosta kiinnostavia piirteitä. (Koponen & Hildén & Vapaasalo 2016, 21.)

5.1 Visualisoinnin toteutus käytännössä

Visualisoinnit tuotetaan nykyään lähes aina tietokoneella ja ne ovat useimmiten vuorovaikutteisia. Visualisointien pohjalla oleva aineisto on aina abstraktia numerotietoa kuten tilastollista tietoa tai muuta näköaistin tavoittamatonta tietoa. Sanaa visualisointi tai visualisoiminen käytetään kuvaamaan sekä visualisoinnin lopputulosta että itse prosessia, jossa data muutetaan visuaaliseen muotoon. Visualisoinnin suunnittelija tekee aineistolle esitysrakenteen, jota voidaan hyödyntää sellaisenaan tai vähän muokattuna muiden samanlaisten, rakeenteellisten tietoaineistojen visualisointiin. (Koponen ym. 2016, 21–23.)

Visualisointi voi kuvata monenlaisia relaatioita visualisoinnin pohjalta olevasta aineistosta. Relaatiot voidaan ryhmitellä seuraavaan viiteen pääryhmään: lukumäärä tai suuruus, järjestys, kategoria, aika ja sijainti. Visualisointia suunniteltaessa olennaisin kysymys on, mitä vertailuja visualisoinnilla halutaan mahdollistaa. Esitystapa on valittava niin, ettei se johda lukijaa harhaan tai vie huomiota kuvattavana olevan aineiston todellisilta piirteiltä. Visualisoinnin tulee kiinnittää lukijan huomio merkittäviin seikkoihin ja esittää tosia väitteitä todellisesta maailmasta selkeällä tavalla. (Koponen ym. 2016, 25–26, 32.)

5.2 Tiedon visualisointivälineet

Tiedon visualisointia käytetään esimerkiksi liiketoiminnallisen tiedon, tieteellisen ja tilastollisen tiedon graafisessa esittämisessä. Tiedon visualisointityövälineet sisältävät kaavioita ja muita visuaalisia elementtejä, joita käytetään esimerkiksi tuloskorttien ja koontinäyttöjen luomiseen. Koontinäytöissä tiedot ovat yhdistetty ja järjestetty yhdelle näytölle, josta niitä

voidaan tarkastella yhdellä silmäyksellä ja tarvittaessa porautua syvemmälle tietoon. (Sharda & Delen & Turban 2014, 129–130, 143.)

Yleisimpiä vertailuja, joita tyypillisesti tehdään visualisointiohjelmissa, ovat vertailut aikaisempiin arvoihin, ennustettuihin arvoihin, tavoiteltuihin arvoihin, vertailuarvoihin ja keskimääräisiin arvoihin. Tämän lisäksi on tärkeää ilmaista, onko luku hyvä vai huono, ja mihin suuntaan pitkäaikainen trendi on kehittymässä. Näitä ilmaistaan yleensä esimerkiksi liikennevalosymbolein tai muilla visuaalisilla elementeillä ja värikoodauksilla. (Sharda ym. 2014, 146.)

6 LIIKETOIMINTATIEDON HALLINTA (BI)

Kansainvälinen ICT-alan tutkimus- ja konsultointiyritys Gartner vakiinnutti käsitteen liiketoimintatiedon hallinta (Business Intelligence, BI) käyttöön 1990-luvun keskivaiheilla, vaikka käsite onkin todellisuudessa tätä vanhempi. BI-tuotteet alkoivat kuitenkin muotoutumaan 1990-luvun keskivaiheille mennessä lähelle nykyisiä moniulotteisia raportteja, jotka mahdollistavat porautumisen yksityiskohtiin ja sisältävät analyysityökaluja. 2005 vuoden jälkeen BI-välineet ovat alkaneet sisältää myös tekoälyominaisuuksia ja entistä tehokkaampia analytiikan ominaisuuksia. (Sharda ym. 2014, 33.)

Modernien BI-välineiden tavoite on tuoda oikea tieto oikeaan aikaan oikeaan paikkaan. BI käsittää tiedon, jota kootaan eri lähteistä ja muokataan ja varastoidaan tietovarastoissa, tiedon analysoinnin, jonka tuloksia hyödynnetään toiminnan johtamisessa ja seurannassa sekä käyttöliittymän tietoon, joka pitää sisällään koontinäyttöjä ja portaaleja. (Sharda ym. 2014, 33.)

BI-välineet auttavat muokkaamaan monimuotoisesta tiedosta uutta informaatiota päätöksenteon tueksi. Ne sisältävät monipuolisia raportointi- ja tiedon analysointiratkaisuja, mittaristoja, web-portaaleja ja mobiiliratkaisuja. Tavoitteena on kohdentaa oikea tieto oikeille ihmisille oikea-aikaisesti. (Hovi & Hervonen & Koistinen 2009, 73–74.)

6.1 BI-raportit

Raportit ovat olennainen osa päätöksentekoa ja organisaation tiedon hallintaa. Nykyään raportit käyttävät entistä enemmän hyväkseen visuaalisia elementtejä tiedon sisällön esille tuomiseksi. Koontinäytöissä voi olla esimerkiksi värikoodattu avainlukujen mittaristo, joka auttaa raportin tarkastelijaa kiinnittämään huomiota oikeisiin asioihin. (Sharda ym. 2014, 119, 125–127.)

Raporttien lähteinä käytetään sekä sisäisiä että ulkoisia tietolähteitä, joiden tiedot yhdistetään yhtenäiseksi raportiksi. Raportointiprosessi sisältää tiedon kyselyn eri tietolähteistä. Tietolähteet voivat olla tuotettuja eritietomallisina, ja raportointiprosessi pitää sisällään myös tiedon tuottamisen luettavaksi ja helposti ymmärrettäväksi raportiksi. Yhtenäinen, helposti luettavissa ja käytettävissä oleva raportti mahdollistaa vaihtoehtojen ja mahdollisuuksien kartoittamisen sekä tietoon pohjautuvan päätöksenteon. Kriittisintä tässä prosessissa on se, miten eri tietolähteistä saadaan koostettua tietoa, jota voidaan hyödyntää toiminnassa. (Sharda ym. 2014, 119, 124.)

BI-välineet vastaavat tiedon jatkojalostamisen, analysoinnin ja hyödyntämisen haasteisiin. Ne toimivat helpoimmillaan point-and-click -tapaisesti

ja tieto on esitetty visuaalisesti ja helpossa muodossa. Välineiden avulla voi nostaa esille järjestelmiin kerätyn informaation suuresta tietomassassa päätöksentekijöiden käyttöön. BI-ratkaisujen näkökulma on informaation analyttisessä hyödyntämisessä sekä informaation esittämisessä ja jake- lussa päätöksenteon tukea varten. BI-välineet tuottavat tietoa sekä val- miiksi mietityssä formaatissa että mahdollistavat ennalta määrittelemät- tömien tietojen haun. Ne sisältävät helppokäyttöisiä kysely- ja raportointi- ominaisuuksia. Käyttäjien mahdollisuus hakea itse tietoa ja jakaa sitä mui- den tietoa tarvitsevien kanssa on yksi keskeisimpiä tavoitteita BI-ratkaisuja luotaessa. (Hovi ym. 2009, 74, 80–82.) BI-ratkaisuja käytetään myös jul- kishallinnon organisaatioissa, joissa BI-välineillä hallitaan muuta kuin liike- toimintaan liittyvää tietoa (Hovi ym. 2009, 78).

6.2 BI-välineiden koontinäytöt eli käyttöliittymät tietoon

BI-järjestelmien koontinäytöille on yhteistä seuraavat piirteet: visuaaliset elementit, jotka korostavat dataa ja datassa olevia poikkeuksia, helppo- käyttöisyys, mahdollisuus datan kokoamiseen eri tietolähteistä, mahdolli- suus porautua syvemmälle datan lähteisiin, vaativat vain vähän, jos ollen- kaan, muokkausta tai koodausta toteutusta, ylläpitoa tai kehitystä varten. (Sharda ym. 2014, 148–149.)

Palvelun tai tuotteen käyttöliittymän käytettävyyttä voidaan mitata sillä, miten hyvin tuotteen toimintoja voidaan käyttää haluttuun lopputulok- seen. Tuote tai palvelu on käyttökelpoinen, jos sen toiminnot ovat hyödyllisiä ja toimintoja on sujuva käyttää. ISO 9241 -standardin osa 11 määritte- lee käytettävyyden kokonaisuudeksi, jossa tarkastellaan, miten hyvin tuo- tetta tai palvelua voidaan käyttää tiettyjen tehtävien suorittamiseen ja ta- voitteiden saavuttamiseen tietyssä ympäristössä. Käytettävyyttä voidaan tarkastella mm. seuraavien mittarien näkökulmasta: opittavuus, tehok- kuus, muistettavuus, virheettömyys, miellyttävyyys ja hyödyllisyys. (Virtu- aaliammattikorkeakoulu n.d.)

Opittavuudella tarkoitetaan sitä, kuinka helposti käyttäjä suoriutuu tuo- teen tai palvelun perustoiminnoista ensimmäisellä käyttökerralla. Tehok- kuus tarkoittaa puolestaan sitä, kuinka nopeasti jo opitut asiat voidaan suorittaa uudestaan. Muistettavuutta voidaan mitata sillä, miten helposti toiminnot muistetaan jälkeen päin. Virheettömyydellä tarkoitetaan sitä, kuinka paljon käyttäjät tekevät virheitä toiminnoissaan ja miten hyvin pal- velu niistä toipuu. Miellyttävyydellä mitataan käyttäjän käyttötyytyväi- syyttä ja hyödyllisyydellä sitä, kuinka hyvin palvelu tai tuote sopii työhön, johon se on tarkoitettu. Listaan voi myös lisätä käyttöliittymän ymmärret- tävyyden: käyttäjä voi päätellä helposti, mitä tuotteella voi tehdä ja miten haluttuun lopputulokseen päästään, sekä esteettömyyden: palvelu on kaikkien käyttäjien käytettävissä. (Virtuaaliammattikorkeakoulu n.d.)

Koontinäytössä esitetylle tiedolle on tärkeää asettaa seuraavia kysymyksiä: Onko data luotettavaa, ajankohtaista tai puuttuuko siitä jotain? On

myös tärkeää miettiä mittaavatko mittarit oikeasti sitä mitä niiden oletetaan mittavaan. Tiedon ymmärrettävyyden vuoksi myös tiedon metatiedot ovat tärkeässä asemassa (mistä tieto on peräisin, millaisia laatuongelmia datassa mahdollisesti on, onko data ajantasaista?). (Sharda ym. 2014, 148– 149.)

6.3 Power BI

Microsoftin Power BI ratkaisu sijoittui kansainvälisen ICT-alan tutkimus- ja konsultointiyritys Gartnerin helmikuussa 2017 tekemässä tutkimuksessa *Magic Quadrant for Business Intelligence and Analytics Platforms* nelikenttävertailun Johtavat-kategoriaan. Tutkimuksen mukaan työkalun etuja ovat edullisuus alansa markkinoilla, jakelu Office 365-tuotteessa, helppokäyttöisyys, moninaiset analyysimahdollisuudet, tuotteen jatkuva kehittäminen sekä vahva kumppanien, jälleenmyyjien ja yksittäisten käyttäjien yhteisö. Tuotteen heikkouksiksi tutkimuksessa puolestaan todettiin tuotteen keskeneräisyys käsiteltävien tietojen volyymin ja toiminnallisuuksien suhteen. Microsoft jatkoikin tuotteensa parantelua kuukausittain julkaisutavalla versiolla vuonna 2017. (Gartner 2017.)

Tutkimuksessa todetaan, että BI-tuotteiden ostajat haluavat laajentaa nykyisten BI-tuotteiden käyttöä ja mahdollistaa tuotteiden käytön organisaation kaikille käyttäjille sekä analysoida entistä monipuolisempia ja monimutkaisempia tietolähteitä. Ratkaisun joustavuus ja helppokäyttöisyys ovat kriittisiä ostospäätöstä tehtäessä. (Gartner 2017.)

Power BI on työväline, jolla voi koota, yhdistää, mallintaa ja rikastuttaa hajallaan olevista tietolähteistä dataa ja tehdä niistä raportteja, visualisointeja sekä analyysejä. Tietolähteinä voi olla niin internetissä olevat tietolähteet, paikallinen tietokanta tai vaikkapa paikalliset tiedostot. Työkalua voi käyttää hyväkseen tiedon analysoinnissa kuin myös apuna päätöksenteossa. (Microsoft 2018.)

Power BI -tuote sisältää työpöytäsovelluksen nimeltään Power BI Desktop, Saas-palveluna toteutettavan sovelluksen nimeltään Power BI service ja mobiiliversion nimeltään Power BI apps. Näiden kolmen roolit ovat yleensä seuraavat raportin luomisessa ja jakamisessa: Tietoaineisto ladataan Power BI Desktop-versioon, jossa datasta muodostetaan raportti. Sen jälkeen raportti julkaistaan Power BI service-sovelluksessa, jossa voidaan luoda visualisoinneista koontinäyttö (dashboard), joka voidaan jakaa Power BI Mobile -versiossa katseltavaksi puhelimesta tai tabletista. (Microsoft 2018.)

Raporteissa voidaan käyttää vaikka yksittäistä Excel-tiedostoa tai yhdistellä dataa monesta eri tietolähteestä esimerkiksi HTML-taulukosta, Excelin työkirjasta, reaaliaikaisesti SQL-tietokannasta, pilvipalvelusta jne. Tietolähteistä tehdyistä visualisoinneista kootaan koontinäyttö (dashboard), joka voidaan jakaa yksittäiselle käyttäjälle tai käyttäjäryhmille. (Microsoft 2018.)

6.3.1 Datan käsittelyominaisuuksia

Power BI -sovelluksen avulla datasta voi rakentaa omia tarpeita vastaavan tietomallin muuttujien nimeämis- ja tietotyyppikäytäntöineen. Sovelluksella voi suorittaa laskutoimituksia ja muodostaa datasta erilaisia tunnuslukuja. Tunnusluvut, mittarit ja visualisoinnit liitetään raporttiin, joka voi sisältää useamman sivun. Datan käsittelyssä ja tarkastelussa voi käyttää apuna datan suodatusominaisuuksia sekä raporttien tarkastelussa Excelistä tuttua osittajaa. (Aspin 2018, 2–3.)

Käytännön elämässä tarvitsee usein yhdistellä hajallaan olevia tietolähteitä ja Power BI -sovellus sisältää Merge Queries-operaation tätä varten. Mikäli tauluissa on yhdistävä kenttä tai kentät, taulut voidaan helposti yhdistää Power BI -sovelluksen kyselyeditorissa. Yhdistämiseen on kuusi erilaista liittämisehtoa. (Aspin 2018, 235, 240–241.) Eri tietolähteistä tuotuja tauluja voidaan myös liittää toisiinsa, mikäli tauluilla on sama määrä sarakkeita, sarakkeet ovat samassa järjestyksessä sekä niiden tietotyypit ja nimet ovat samoja. Liittäminen tapahtuu Power BI:n Append Queries-toiminnolla. Taulujen alkuperän ei kuitenkaan tarvitse olla samasta tietoformaatista. (Aspin 2018, 235, 246–247.)

6.3.2 DAX-kieli

Power BI laskee automaattisesti lukujen aggregaatteja, mutta sovelluksessa on myös mahdollista tehdä omia monipuolisia laskutoimituksia DAX-kielen (Data Analysis Expressions) avulla. DAX-kaavat käyttävät samoja funktioita, operaatioita ja syntaksia kuin Excel-kaavat, mutta DAX-funktiot ovat suunniteltu relaatiotietomalliselle datalle. DAX-kieli sisältää 200 funktiota yksinkertaisista laskutoimituksista monimutkaisiin tilastollisiin funktioihin. (Microsoft 2017.)

DAX-kielen ansiosta raporteissa voi suorittaa monipuolisia laskutoimituksia ja tehdä näistä uudelleen käytettäviä mittareita. DAX-kieli muistuttaa voimakkaasti Excelin kaavoja, mutta DAX-funktiossa viittaus tehdään soluosoitteen sijaan sarakkeeseen. DAX-kieli sisältää loogisia funktioita (IF, AND, OR jne.) sekä informaatiofunktioita (ISBLANK, ISNUMBER jne.), yhteenvetofunktioita (SUM, AVERAGE, MIN, COUNTAX jne.), muotoiluun liittyviä toimintoja (esimerkiksi päivämäärät) sekä mahdollistaa taulujen sarakkeiden yhdistämisen. (Kilpeläinen 2017.)

7 PILVIPALVELUT TUOVAT JOUSTAVUUTTA RAPORTOINTIIN

Viime vuosien trendinä tietotekniikassa on ollut siirtyä tietotekniikan käyttöön palveluna. Yhä pienempi osa teknisestä kokonaisuudesta on omilla laitteilla ja palveluita kulutetaan internet-verkon yli. Kansainvälisen ICT-alan tutkimus- ja konsultointiyrityksen Gartnerin vuosittain listaamien kymmenen strategisen tekniikan joukossa pilvipalvelut ovat olleet vuodesta 2008. (Salo 2014, 92.)

Pilvipalveluissa esimerkiksi sovellukset tai palvelut tarjotaan käyttöön verkon kautta ja käyttäjän ei tarvitse huolehtia palveluiden toiminnasta, ylläpidosta tai tietää, missä palvelu fyysisesti sijaitsee (Salo, 2010, 16). Pilvipalveluita voi käyttää päätelaiteriippumattomasti ja niiden käyttö onnistuu niin työasemalla, kannettavalla kuin mobiililaitteellakin, kunhan vain laite on kytketty verkkoon. Palvelut mukautuvat käytettävään päätelaitteeseen. (Salo 2010, 17–18.)

Päätelaite- ja paikkariippumattomuus helpottavat ja yksinkertaistavat monia palveluiden käyttötapauksia. Samoihin palveluihin päästään käsiksi toimiston työasemalla, älypuhelimella ja kotitietokoneella, joten paikkariippumattomuus tarjoaa vapauden tehdä töitä missä tahansa. (Heino 2010, 46–47.)

Microsoftin Azure-pilvessä jaettu Power BI -raportteja voi tarkastella Power BI apps -sovelluksen kautta Windows-, iOS- tai Android-mobiililaitteella (Aspin 2018, 2). Power BI -sovellukseen voi myös ladata pilveen varastoitua dataa esimerkiksi erilaisista online-palveluista, pilvessä sijaitsevista SQL-tietokannoista ja tietovarastoista sekä big data -alustoista (Aspin 2018, 125).

8 KÄYTETYT MENETELMÄT

Opinnäytetyössä on käytetty case-lähestymistapaa. Avoimeen dataan on perehdytty ensin teorian tiedon ja aiheesta viime vuonna julkaistun kansallisen Valtioneuvoston kanslian toteuttaman tutkimuksen avulla. Aineistona on käytetty THL:n internet-sivuilta TIKU-julkaisujärjestelmästä löytyviä avoimia tietoaaineistoja, Kuntaliiton sekä Avoindata.fi-palvelusta löytyvää pääkaupunkiseudun kuntien ja Tilastokeskuksen tuottamaa avointa väestötietoaaineistoa. Kuten Kananen toteaa kirjassaan *Case-tutkimus opinnäytetyönä*, Case-tutkimuksessa tutkimusongelman vastaus kerätään eri lähteistä ilmiön kokonaiskuvan saamiseksi. Tiedonkeruun ja aineiston hajanaisuus tekee tutkimuksesta nimenomaan case-tutkimuksen. (Kananen 2013, 77–78.)

Case-lähestymistavassa on olennaista miettiä keneltä tieto ja aineisto kerätään tutkimusongelman ratkaisemiseksi (Kananen, 2013, 76). Opinnäytetyön yksi tärkeimmistä lähteistä oli Valtioneuvoston kanslian vuonna 2017 julkaisema raportti *Avoimen datan hyödyntäminen ja vaikuttavuus*. Kyseisessä tutkimuksessa kartoitetaan 12 keskeisen tietovarantojen julkaisijan avoimen datan tilannetta sekä yrityksille suunnattua innovaatiotutkimusta. THL on yksi näistä edellä mainitun tutkimuksen keskeisistä tietovarantojen julkaisijasta ja Avoindata.fi toimii Suomen keskitettynä avoimen datan jakelualustana.

Kokeessa on käytetty Microsoftin Power BI -työpöytäsovellusta Windows-ympäristössä. Ohjelma on valittu kokeisiin, koska Power BI on arvioitu kansainvälisen ICT-alan tutkimus- ja konsultointiyritys Gartnerin vuonna 2017 tehdyssä tutkimuksessa kolmen johtavan BI-välineen joukkoon. Microsoftin tuotteet ovat myös laajalti käytössä eri organisaatioissa ja sovellus löytyi myös Hämeen ammattikorkeakoulun Office 365 -tuotteesta.

Käytännön osuudessa menetelminä on käytetty omia havaintoja Power BI -työpöytäsovelluksella työskennellessä ja tietoa visualisoidessa. Kokeita varten on ensin laadittu lomake, joka sisälsi kymmenen kohtaa liittyen aineiston lataukseen Power BI-sovelluksella sekä aineiston parissa työskentelyyn. Kohtien tarkoitus oli kohdistaa havaintoja BI-välineen käytettävyyteen sekä aineiston lataamisen, parametrien asettamiseen aineistolle ja sen parissa työskentelyyn sekä valmiin tuotoksen jakamiseen pilvipalveluissa. Lopussa kokeen onnistumisia arvioidaan ja analysoidaan opinnäytetyön Tulokset-osuudessa.

Kananen kirjoittaa kirjassa *Case-tutkimus opinnäytetyönä*, että Case-tutkimuksessa ilmiötä havainnoidaan sen luonnollisessa ympäristössään ja tutkija on aina osallistuvan havainnoijan roolissa. Tällaisen havainnoinnin etuna on tilanteen aitous ja tiedonkeruumenetelmistä tämä onkin yksi tehokkaimmista menetelmistä. (Kananen 2013, 81, 89.) Ojasalo ym. puolestaan toteaa, että havainnoimalla selvitetään, mitä tutkimuskohde tekee,

miten sitä käytetään tai mitä tutkimuskohteessa tapahtuu. Havainnointi on aina järjestelmällistä ja se kohdistuu ennalta määrättyyn kohteeseen ja tulokset pyritään rekisteröimään välittömästi. (Ojasalo, Moilanen & Ritalahti 2014, 114–115.)

Kananen toteaa case-tutkimusta tarkasteltavan usein tutkimusstrategiana eikä niinkään erillisenä menetelmäkokonaisuutena. Menetelmällä ei ole omaa metodologiaansa vaan se hyödyntää laadullista ja määrällistä tutkimusta. Case-menetelmässä tutkimuksen kohteena on usein yksi, tämänhetkinen ilmiö, johon perehdytään syvällisesti omassa, luonnollisessa kontekstissaan ja monia tietolähteitä käyttäen. Tutkimuskohteesta pyritään saamaan syvälinen ymmärrys ja antamaan hyvä kuvaus, mutta tutkimustuloksia ei voida yleistää. Tulokset pätevät vain tutkitun tapauksen osalta. (Kananen 2013, 23, 28, 54, 77.)

Myös tämän opinnäytetyön käytännön tulokset pätevät kyseessä oleviin THL:n, Kuntaliiton ja Avoindata.fi-palvelusta ladattujen pääkaupunki- ja pääkaupunkiseudun väestödatojen osalta. Tuloksissa on kuitenkin tietoa, jota voidaan hyödyntää myös yleisellä tasolla, kuten tietoa Power BI-ohjelmasta ja erilaisten avoimien tietoaaineistojen dataformaattien latauksesta ja käsittelystä yleensäkin Power BI-työpöytäsovelluksella.

Työssä tarkastellaan myös avoimen datan rajapintaa sekä teorian että käytännön näkökulmasta. Kokeissa on hyödynnetty THL:n avointa rajapintaa aineistojen latauksessa sekä kokeiltu asettaa rajapintakutsuihin parametreja ladattavan aineiston rajaamiseksi eri ulottuvuuksien suhteen. Kokeissa käsiteltävien avoimien tietoaaineistojen teemaksi valittiin terveys ja väestö. Aineistoa ladattiin eniten THL:n TIKU-julkaisujärjestelmästä, mutta kokeissa on myös käytetty Kuntaliiton, pääkaupunkiseudun kuntien ja Tilastokeskuksen tuottamaa avointa väestöaineistoa. Muut julkishallinnon avoimet datat on jätetty käytännön osuuden ulkopuolelle, mutta niitä sivutaan työn teoriaosuudessa.

8.1 Opinnäytetyössä käytetyt aineistot

Kokeissa käytetyt aineistot ovat THL:n Erikoissairaanhoidon jonotilanne, Lastensuojelun määräaikojen seuranta, AvoHILMO: Käyntien odotusajat perusterveydenhuollossa (käynti toteutunut), Perusterveydenhuollon asiakkaat ja Rokotusrekisteri: Lasten rokotuskattavuus. Kuntaliiton aineistona on käytetty Kuntajaot ja asukasluvut 2000–2017 aineistoa ja Avoindata.fi-palvelusta Tilastokeskuksen tuottamaa Helsingin väestö viisivuotiskäryhmittäin, siviilisäädyn ja sukupuolen mukaan 2004 alkaen sekä Pääkaupunkiseudun väestöennuste 2012–2021 aineistoa, joka sisältää Helsingin, Espoon ja Vantaan laatimat ennusteet sekä Tilastokeskuksen ennusteen Kainaisen osalta.

8.3 Tiedonkeruun riittävyys ja tuloksien tulkinta

Kanasen mukaan tiedonkeruun riittävyys on sitä, että ilmiöstä on kerätty aineistoa sen verran, jotta siitä voidaan tehdä oikeita ja varmoja johtopäätöksiä. Kvalitatiivisessa tutkimuksessa tämä tarkoittaa saturaatiota. (Kananen, 2013, 133.) Kananen toteaa, että ensisijaisena tavoitteena on löytää aineiston avulla ratkaisu tutkimusongelmaan ja vastaukset ongelmasta johdettuihin tutkimuskysymyksiin. Aineistosta haetaan selitystä ilmiölle, tyypillistä kertomusta, toiminnan logiikkaa sekä samanlaisuutta tai erilaisuutta. (Kananen 2013, 107.)

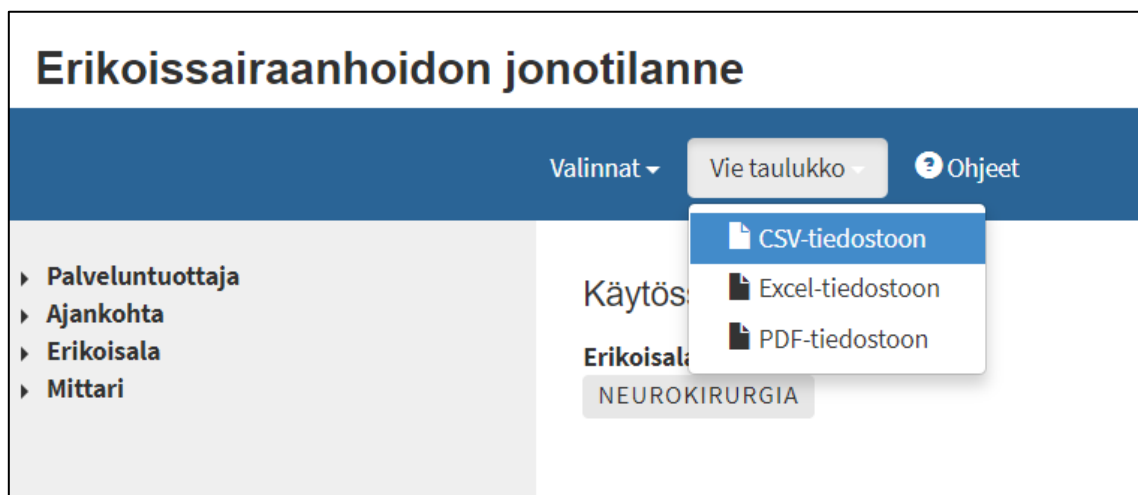
Hirsijärven ym. kirjassa *Tutki ja kirjoita* todetaan, että tulosten tulkinnassa tutkija pohtii analyysin tuloksia ja tekee niistä omia johtopäätöksiä. Se on myös aineiston analyysissä esiin nousevien merkitysten pohdintaa. (Hirsijärvi, Remes & Sajanvaara 2006, 213.)

Ojasalo ym. mukaan merkityksen luomisen taktiikoita tutkimustulosten tulkinnassa ovat mm. toistuvien rakenteiden tunnistaminen, uskottavan selityksen näkeminen, ryhmittely, laskeminen, kontrastien ja vertailujen tekeminen, yksityiskohtien sisällyttäminen yleiseen, muuttujien välisten linkkien tunnistaminen, loogisen tapahtumaketjun kehittäminen sekä käsitteellisen tai teoreettisen johdonmukaisuuden luominen. (Ojasalo, Moilanen & Ritalahti 2014, 143–144.)

Vaikka Kananen toteaa, että havainnointi on yksi tehokkaimmista tiedonkeruumenetelmistä, sen ongelmana voidaan pitää sitä, että tutkija ei aina välttämättä tiedä, mihin asioihin havainnoinnissa kiinnitetään huomiota. Lisäksi osallistuvassa havainnoinnissa tutkija on itse mukana ilmiössä ja voi näin vaikuttaa olemassaolollaan tutkittavan ilmiön toimintaan. (Kananen 2013, 81, 88.)

9 DATAN LATAUS THL:N TIKU-JULKAISUJÄRJESTELMÄSTÄ

THL:n avointa tietoaainestoa voi ladata joko tiedostomuodossa tai rajapinnan kautta organisaation Internet-sivuilta löytyvästä TIKU-julkaisujärjestelmästä. Tiedostomuotoisena ladattaessa valittavina ovat CSV-, Excel ja PDF-muoto, ja rajapinnan kautta ladattaessa, joko CSV tai JSON. Käytännön osuudessa kokeiltiin näitä kaikkia menetelmiä lukuun ottamatta PDF-tiedoston latausta.



Kuva 1. TIKU-julkaisujärjestelmästä löytyy Vie taulukko-toiminto, jonka kautta voi ladata dataa kolmessa eri tiedostomuodossa.

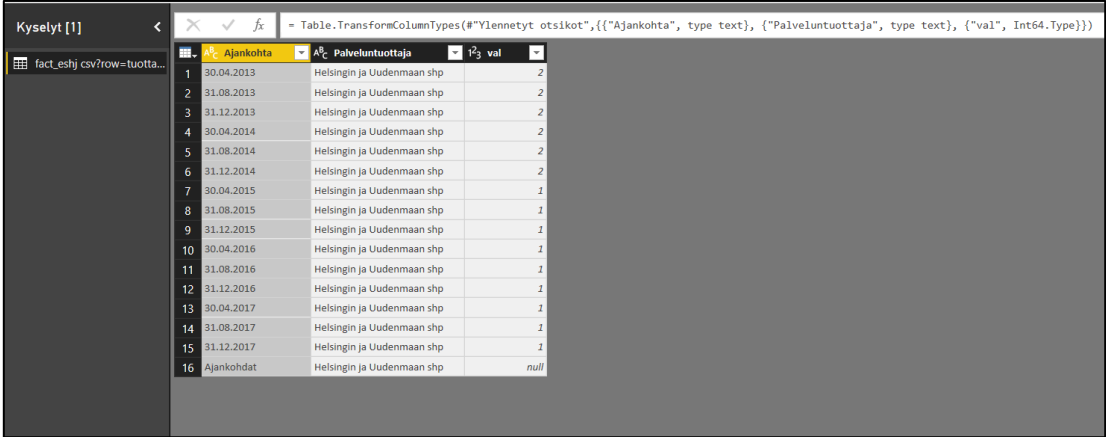
THL:n internet-sivuilla on ohjeistus avoimesta rajapinnasta. Ohjeistuksessa esitellään avoimen datan muoto, periaatteet ja rajapinnan kuvaus. Ohjeistuksessa on myös listaus datan rajapinnassa esiintyvistä tietokentistä ja niiden selitykset sekä esimerkit rajapintakutsuista. Rajapinnan kautta voi ladata koko aineiston tai vain osia siitä. Osia ladattaessa on kutsuun laitettava halutut parametrit esimerkiksi alueulottuuden suhteen. Aineiston kaikki ulottuvuudet voi ladata omalla kutsulla samoin kuin ulottuvuuden solmun lapsialkiot. Aineistot on ryhmitelty aihealueittain ja aihealueen kuvauksen saa esille omana rajapintakutsunaan. Avoimet tietoaainestot ovat Hydra-tietomuotoa ja yksi hydra sisältää yhden tai useamman faktatiedoston. Hydran sisältämät faktatiedostot saa selville rajapintakutsuna. Ohjeissa oli esimerkit JSON-tiedostojen lataukseen.

9.1 Tiedostomuotoisen datan lataus

Ensimmäisillä kerroilla esiintyi vaikeuksia ladata CSV-muotoinen tiedosto. CSV-tiedosto kyllä latautui, mutta se oli täysin tyhjä. Excel-tiedosto latautui puolestaan moitteettomasti ja jonkin ajan kuluttua alkoi CSV-muotoinen latauskin onnistumaan.

TIKU-julkaisujärjestelmästä ladattujen CSV- ja Excel-tiedostojen ero on siinä, että data on CSV-muodossa relaatiotietokannan taulunmukaisesti riveillä, kun taas Excel-tiedostossa data on sarakkeissa. Power BI -välineeseen ladattaessa ei ilmennyt suuria eroja näitä kahta eri tiedostomuotoa ladattaessa. Excel-formaatti vaatii hieman enemmän muokkausta, mutta tämä onnistui näppärästi sekä Power BI -sovelluksen kyselyeditorissa että myös Excelin Power Pivot -apuohjelmalla. Excelin Power Pivot -apuohjelmalla sai poistettua esimerkiksi sarakkeiden pivotoinnin helposti ennen Power BI -työpöytäsovellukseen latausta. Sama toiminto löytyy myös Power BI:n kyselyeditorista, joten on lähinnä makuasia, mitä ohjelmaa haluaa käyttää Excel-tiedostojen muokkaukseen.

Tiedoston latauksen yhteydessä voi valita Muokkaa-toiminnon, jolloin Power BI -ohjelma avaa Kyselyeditori-ikkunan. Ikkunan kautta voi poistaa mm. turhia rivejä ja sarakkeita, muuttaa muuttujien tietotyyppejä, määrittellä otsikkorivin, korvata arvoja ja tehdä muita datan muokkauksia.



	Ajankohta	Palveluntuottaja	val
1	30.04.2013	Helsingin ja Uudenmaan shp	2
2	31.08.2013	Helsingin ja Uudenmaan shp	2
3	31.12.2013	Helsingin ja Uudenmaan shp	2
4	30.04.2014	Helsingin ja Uudenmaan shp	2
5	31.08.2014	Helsingin ja Uudenmaan shp	2
6	31.12.2014	Helsingin ja Uudenmaan shp	2
7	30.04.2015	Helsingin ja Uudenmaan shp	1
8	31.08.2015	Helsingin ja Uudenmaan shp	1
9	31.12.2015	Helsingin ja Uudenmaan shp	1
10	30.04.2016	Helsingin ja Uudenmaan shp	1
11	31.08.2016	Helsingin ja Uudenmaan shp	1
12	31.12.2016	Helsingin ja Uudenmaan shp	1
13	30.04.2017	Helsingin ja Uudenmaan shp	1
14	31.08.2017	Helsingin ja Uudenmaan shp	1
15	31.12.2017	Helsingin ja Uudenmaan shp	1
16	Ajankohdat	Helsingin ja Uudenmaan shp	null

Kuva 2. Kyselyeditorinäköymä Power BI -sovelluksessa.

9.2 Datan lataus rajapintakutsuna THL:n TIKU-julkaisujärjestelmästä

THL:n avoimet tietoaineistot voi ladata myös REST-tyyppisen rajapinnan kautta sekä CSV- että JSON-muodossa. Rajapinta kutsuun voi asettaa parametrejä esimerkiksi aikaulottuvuuden suhteen. Kokeissa käytetyt rajapintakutsut aineistojen lataukseen on koottu liitteeseen 1 ja taulukko 1:een on koottu ohjeet rajapintakutsujen muodostukseen.



Kuva 3. Aineiston lataus rajapinnan kautta.

Power BI -sovelluksen Nouda tiedot-komennon lähteeksi valittiin Verkko ja kirjoitettiin rajapintakutsu URL-osoitteen riville kuten kuvassa 3. Aineiston lataus tapahtui rajapintakutsuna nopeasti ja rajapinnan kautta tapahtuvan latauksen etuna on se, että tiedostoja ei tarvitse tallentaa paikallisesti. Myös rajapintakutsuna ladattuja aineistoja muokattiin ennen visualisointien tekoja, kuten aikaisemmin tiedostomuotoisen datan parissa työskennellessä.

9.2.1 CSV-rajapintakutsu

Kokeissa kokeiltiin CSV-muotoisen datan latausta rajapintakutsuna. Rajapintakutsujen muodostamisessa käytettiin apuna THL:n internetsivuilta löytyvää ohjeistusta. Rajapintakutsun parametrit sai helpoiten selville TIKU-julkaisujärjestelmän osoiteriviltä. Ensin haettiin TIKU-käyttöliittymään näkyville haluttu aineisto, jonka jälkeen muokattiin vain hieman osoiterivillä näkyvää osoitetta rajapintakutsun muodostamista varten.

Kokeissa vertailtiin THL:n CSV- ja JSON-muotoisen aineiston latausta Power BI -ohjelmistoon. Aineistona käytettiin lasten rotavirusrokotteen rokote-kattavuus aineistoa. Rajapintakutsuun parametriksi asetettiin alueulottuvuudeksi Espoo ja aikaulottuvuudeksi syntymävuodet 2009-2011, näin ol-
len CSV-rajapintakutsu oli alla olevan mukainen.

https://sampo.thl.fi/pivot/prod/fi/vaccreg/coverage/fact_coverage_infant.csv?row=area-13044&column=vaccine-172606.&column=time-133277.133264.133270

Power BI:n Tiedot näkymässä kyseessä oleva aineisto näytti kuvan 4 mukaiselta.



Rokote	Aika	Alue	val
Rotavirusrokote	SyntymÄävuosi 2009	Espoo	72.7
Rotavirusrokote	SyntymÄävuosi 2010	Espoo	96.0
Rotavirusrokote	SyntymÄävuosi 2011	Espoo	94.6

Kuva 4. Lasten rotavirusrokotekattavuus -aineisto. Aineisto rajattu Espooseen ja syntymÄävuosille 2009-2011.

Kyseessä olevien rajausten jälkeen aineistosta latautui vain kolme tietuetta ja tiedot olivat relaatiotietokannan taulun mukaisessa järjestyksessä riveittäin. Latauksen jälkeen CSV-muotoisen datan kanssa työskentely oli helppoa ja vaati vain vähän muokkausta raporttia ja visualisointia tehdessä.

9.2.2 JSON-rajapintakutsu

Seuraavaksi kokeiltiin saman aineiston latausta JSON-rajapintakutsuna. THL:n ohjeistukset olivat laadittuja JSON-rajapintakutsujen tekoon. CSV- ja JSON-rajapintakutsut ovat muuten samanlaisia, mutta haluttu dataformaatti ilmaistaan joko päätteellä .csv tai .json. Yllä olevan aineiston lataus suoritettiin JSON-formaatissa seuraavalla rajapintakutsulla.

https://sampo.thl.fi/pivot/prod/fi/vaccreg/coverage/fact_coverage_infant.json?row=area-13044&column=vaccine-172606.&column=time-133277.133264.133270

Power BI -ohjelmiston kyselyeditorissa aineisto näytti kuvan 5 kaltaiselta eli tietueita oli JSON-muodossa kolminkertainen määrä. JSON-muotoisen aineiston käsittelyä ei jatkettu tämän enempää, vaan todettiin tässä kohta, että sarakemuotoinen aineisto sopii parhaiten Power BI -ohjelmistoon ja sitä voi suositella THL:n avoimien tietoaaineistojen käsittelyyn Power BI -ohjelmistolla.

	Value.dimension.time.category.index.133270	Value.dimension.time.category.label.133277	Value.dimension.time.category.label.133264
1	2	Syntymävuosi 2009	Syntymävuosi 2010
2	2	Syntymävuosi 2009	Syntymävuosi 2010
3	2	Syntymävuosi 2009	Syntymävuosi 2010
4	2	Syntymävuosi 2009	Syntymävuosi 2010
5	2	Syntymävuosi 2009	Syntymävuosi 2010
6	2	Syntymävuosi 2009	Syntymävuosi 2010
7	2	Syntymävuosi 2009	Syntymävuosi 2010
8	2	Syntymävuosi 2009	Syntymävuosi 2010
9	2	Syntymävuosi 2009	Syntymävuosi 2010

Kuva 5. JSON-formaatti ladattuna Power BI -ohjelmaan. Näkymä kysely-editorista.

9.3 Parametrien asetus THL:n aineistoille

Kaikkien rajapintakutsujen alkuosa on muotoa <https://sampo.thl.fi/pivot>. Rajapintakutsun loppuosaan voi myös lisätä haluttuja parametrejä kuten esimerkissä palveluntuottajan koodi 255447 (HUS) ja erikoisalan koodi 255520 (lastentaudit) sekä mittariulottuvuuden koodi 18646 (Mediaani, vrk, lähete käsitelty). Parhaiten nämä parametrit saa selville kuutiokäyttöliittymästä selaimen osoiterivin avulla hakemalla tarkasteluun halutun aineiston, mutta aineiston sisältämät ulottuvuudet voi myös ladata omalla rajapintakutsullaan. Rajapintakutsut muodostuvat alla olevan taulukon mukaisesti.

Taulukko 1. THL:n aineistojen rajapintakutsun muoto

Termi	Vastaavuus rajapintakutsussa ja sen tarkoitus
ympäristö	prod. Tarkoittaa tuotantoympäristöä erotuksena kehitys- ja testi-ympäristöstä. Julkisesti avoinna olevat aineistot ovat kaikki tuotantoympäristössä olevia aineistoja.
kieli	fi. Tarkoittaa suomenkielistä aineistoa. Ruotsinkielisten lyhenne on sv.
aihealue	Esimerkiksi vaccreg, rokotusrekisteri. Kaikki aineistot ovat ryhmitelty omiin aihealueisiinsa.
hydra	Esimerkiksi coverage. Tarkoittaa tässä ko. rekisterin kattavuusaineistoja. Yksi hydra sisältää joko yhden tai useamman fakta-aulun, joka pitää sisällään yhden loogisen tietokokonaisuuden tiedot.
faktataulun nimi	fact. Yksittäinen fakta-taulu jossain tietyyssä hydrassa, esimerkiksi rokotusrekisteri aihe-alueen kattavuusaineisto hydran fakta-taulu, jossa on tietoa lasten rokotuskattavuudesta

Yhdessä kokeessa käytetty erikoissairaanhoidon jonotilanne aineisto näyttää TIKU-julkaisujärjestelmään haettuna kuvan 6 kaltaiselta ja osoiteriviltä

näkee tarvittavat parametrit, jotka voi asettaa rajapintakutsuun taulukon 1 mukaisesti, jolloin rajapintakutsuksi muodostuu

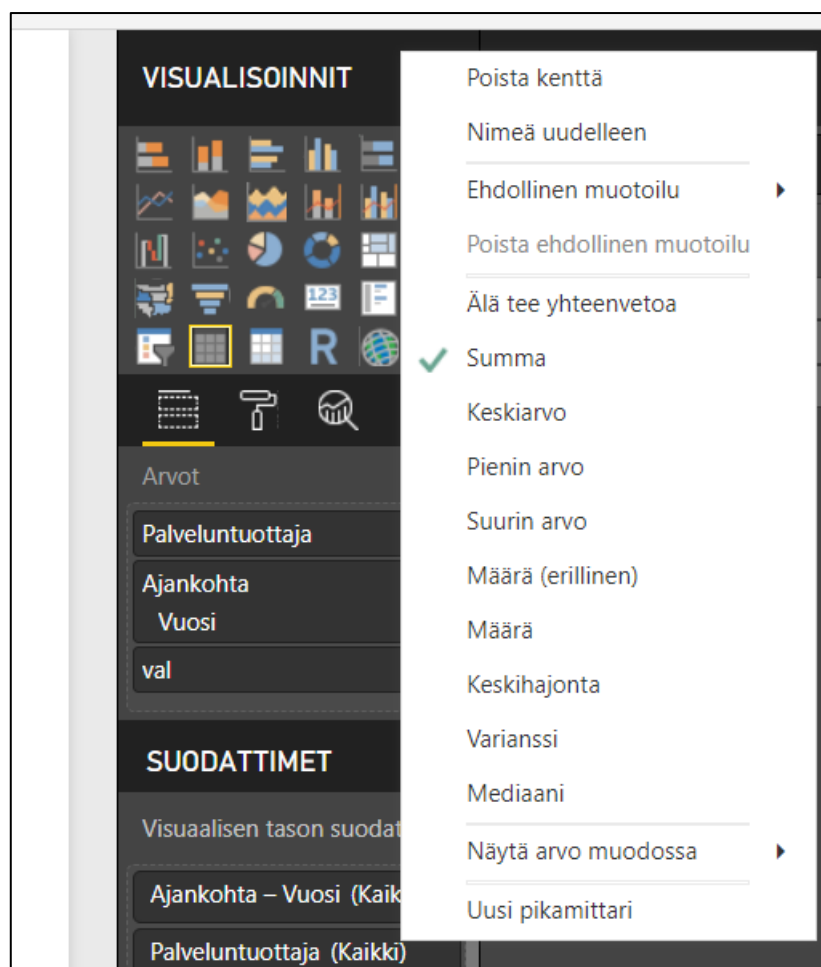
https://sampo.thl.fi/pivot/prod/fi/eshjono/eshj/fact_eshj.csv?row=tuottaja-255447.&column=time-430&filter=eala-255520&filter=measure-18646#

Mediaani, vrk, lähete käsitelty	30.04.2013	31.08.2013	31.12.2013	30.04.2014	31.08.2014	31.12.2014	30.04.2015	31.08.2015
Helsingin ja Uudenmaan shp	2	2	2	2	2	2	1	1

Kuva 6. Edellä mainittu rajapintakutsu lataa kuvanmukaisen erikoissairaanhoidon jonotilanne aineiston. Parametreiksi on asetettu palveluntuottaja HUS, erikoisalakoodiksi Lastentaudit ja mittari-olottuvuudeksi läheteiden käsittelyn mediaani vrk.

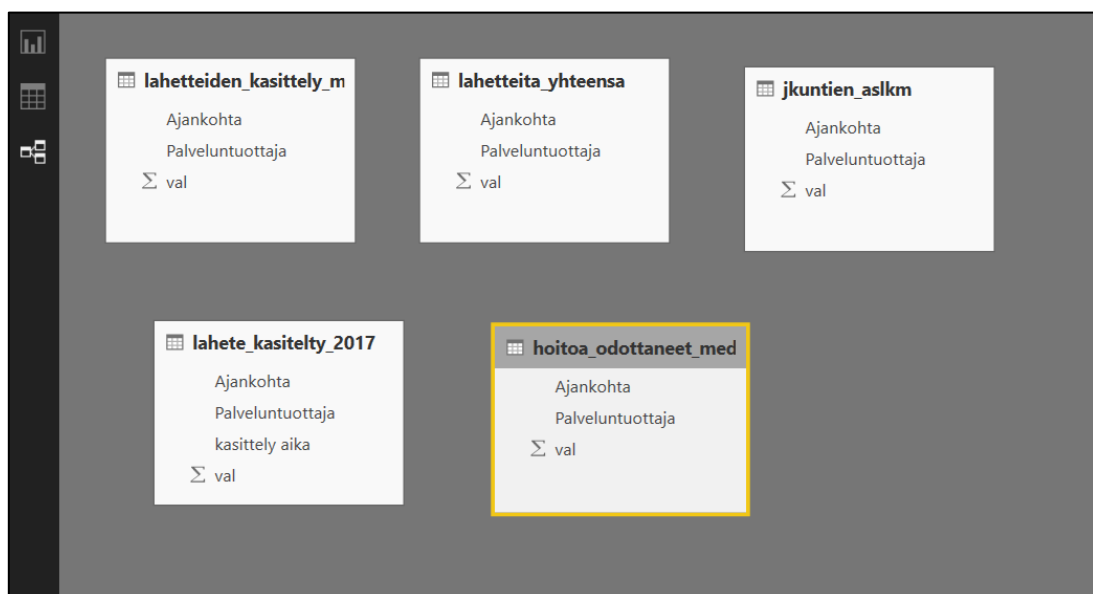
10 DATAN KÄSITTELY POWER BI -TYÖPÖYTÄSOVELLUKSESSA

Kokeissa käsitelty data vaati jonkin verran muokkausta ennen visualisointien laadintaa, ja nämä muokkaukset suoritettiin Power BI:n kyselyeditorissa. Tarvittaessa turhia rivejä poistettiin, muutettiin tietotyyppejä ja määriteltiin datamatriisin otsikkorivi. Aineistojen kanssa työskennellessä havaittiin Power BI:n päivämäärä-tietotyyppi erityisen käteväksi, sillä ohjelma pystyy laskemaan yhdellä klikkauksella tämän tyyppisestä tiedosta esimerkiksi summia ja mediaani-arvoja sekä muodostamaan automaattisesti hierarkiaa porautumista varten. Edellä mainitut pystyi tekemään nopeasti käyttöliittymästä käsin ilman kaavojen määrittelyä tai uusien sarakkeiden luomista.



Kuva 7. Power BI:n käyttöliittymästä pystyy valitsemaan yhdellä klikkauksella esimerkiksi lukujen yhteenlaskun.

Power BI-ohjelmaan pystyy lataamaan samaan raporttiin eri muotoista dataa useammasta eri lähteestä. Jokaisesta tietolähteestä muodostuu oma taulu ja taulujen väliset suhteet eli relaatiot voi määrittellä Suhteet-näkymässä.



Kuva 8. Power BI:n suhteet-näkymä.

Ladatut avoimet tietoaaineistot eivät sisältäneet pääavain- ja viiteavainarvoja. Tarvittaessa taulujen väliset relaatiot voi itse määritellä aineistoon.

10.1 Taulujen yhdistäminen ja liittäminen

Power BI -sovellus sisältää SQL-kyselykielestä tuttuja taulujen yhdistämis-toimenpiteitä (join). Yhdistä kyselyt -toimintoa kokeiltiin THL:n lastentautien erikoissairaanhoidon jonotilanne datan aineiston tauluille sekä Avohilmo: Perusterveydenhuollon asiakkaat aineistosta ladatuille tauluille.

Tauluja voi myös liittää toisiinsa Liitä kyselyt loppuun -toiminnolla, jolloin taulujen tietueet tulevat peräkkäin. Tämä edellyttää luonnollisesti, että taulujen tietorakenne on samanlainen. Tätä toimintoa kokeiltiin Lastensuojelun määräaikojen seuranta aineistoa käsiteltäessä.

10.1.1 Taulujen yhdistäminen "Merge"

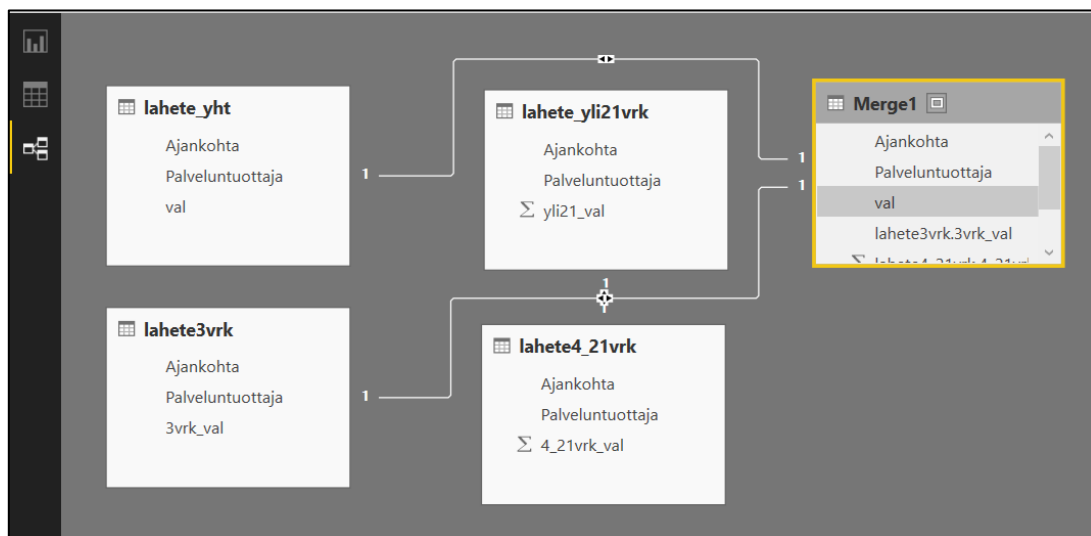
Ennen yhdistämistä aineistolle tehtiin tarvittavat tietotyyppien muokkaukset ja nimettiin taulujen arvo- eli "val"-muuttujat yksilöllisesti. Yhdistämisen jälkeen pystyttiin käyttämään yhdistettyjen taulujen tietoja samassa visualisoinnissa.

Yhdistä kyselyt -toiminnossa on määriteltävä liittäminen laji, ja liittäminen lajiksi valittiin vasen ulompi (kaikki ensimmäisestä ja vastaavat toisesta). Tässä kohtaa oli myös määriteltävä, mitkä sarakkeet tauluissa ovat vastaavia sarakkeita eli mitkä sarakkeet yhdistävät taulut toisiinsa. Toiminnossa taulut voi yhdistää joko suoraan toisiinsa tai yhdistää kyselyt uudeksi tauluksi, jonka nimeksi tulee automaattisesti Merge1.

	Ajankohta	Palveluntuottaja	val	lahete3vrk.3vrk_val	lahete4_21vrk.4_21vrk_val	lahete_yli21vrk.yli21_val
1	30.4.2013	Helsingin ja Uudenmaan shp	3495	1965	1338	52
2	31.8.2013	Helsingin ja Uudenmaan shp	6410	3730	2436	110
3	31.12.2013	Helsingin ja Uudenmaan shp	9835	5876	3697	147
4	30.4.2014	Helsingin ja Uudenmaan shp	3826	2328	1316	46
5	31.8.2014	Helsingin ja Uudenmaan shp	6764	4153	2361	107
6	31.12.2014	Helsingin ja Uudenmaan shp	10296	6410	3654	140
7	30.4.2015	Helsingin ja Uudenmaan shp	3660	2502	1083	14
8	31.8.2015	Helsingin ja Uudenmaan shp	6632	4665	1836	42
9	31.12.2015	Helsingin ja Uudenmaan shp	10061	7275	2656	52
10	30.4.2016	Helsingin ja Uudenmaan shp	3593	2623	875	14
11	31.8.2016	Helsingin ja Uudenmaan shp	6578	4881	1592	44
12	31.12.2016	Helsingin ja Uudenmaan shp	9970	7475	2386	52
13	30.4.2017	Helsingin ja Uudenmaan shp	3489	2721	718	9
14	31.8.2017	Helsingin ja Uudenmaan shp	6648	5223	1369	19
15	31.12.2017	Helsingin ja Uudenmaan shp	10105	7941	2085	30

Kuva 9. Taulujen yhdistäminen. Merge1 -taulussa näkyy kaikki mittari-sarakkeet jokaisesta taulusta.

Yhdistämisen jälkeen Power BI osaa luoda taulujen väliset suhteet automaattisesti muuttujan tietotyyppin perusteella. Ladatussa datassa taulujen väliset suhteet olivat yhden suhteen yksi, ja Power BI -ohjelman automaattisesti määrittelemiä taulujen suhteita pääsi tarkastelemaan Suhteet-näkymässä.



Kuva 10. Ladattujen taulujen ja Merge1-taulun suhteet Suhteet-näkymässä.

10.1.2 Taulujen liittäminen "Append"

Taulujen liittämis-toimintoa (Append) kokeiltiin käsiteltäessä THL:n Las-tensuojelun määräaikojen seuranta aineistoa. Eri vuosien aineistoa haluttiin liittää toisiinsa. Eri vuosien aineistoilla oli muuten sama tietorakenne, mutta aineiston latauksessa kävi ilmi, että vain viimeisin aineisto sisältää Ajankohta-muuttujan, mutta aikaisemmat aineistot eivät. Näin ollen aikaisemmille aineistoille jouduttiin luomaan Ajankohta-muuttuja Power BI:n

ehdollinen sarake -toiminnalla. Ehdollinen sarake -toiminnalla pystyy rakentamaan monipuolisia ehtoja, joiden avulla voi luoda uusia muuttujia aineiston tauluihin.

Lisää ehdollinen sarake

Lisää ehdollinen sarake, joka on laskettu muista sarakkeista tai arvoista.

Uuden sarakkeen nimi
Ajankohta

Sarakkeen nimi	Operaattori	Arvo	Tuloste
Jos	Mittari	ei ole yhtä suuri k...	Sitten
		ABC 123	ABC 123 1.10.2014-31.3.2015

Lisää sääntö

Muuten
ABC 123 null

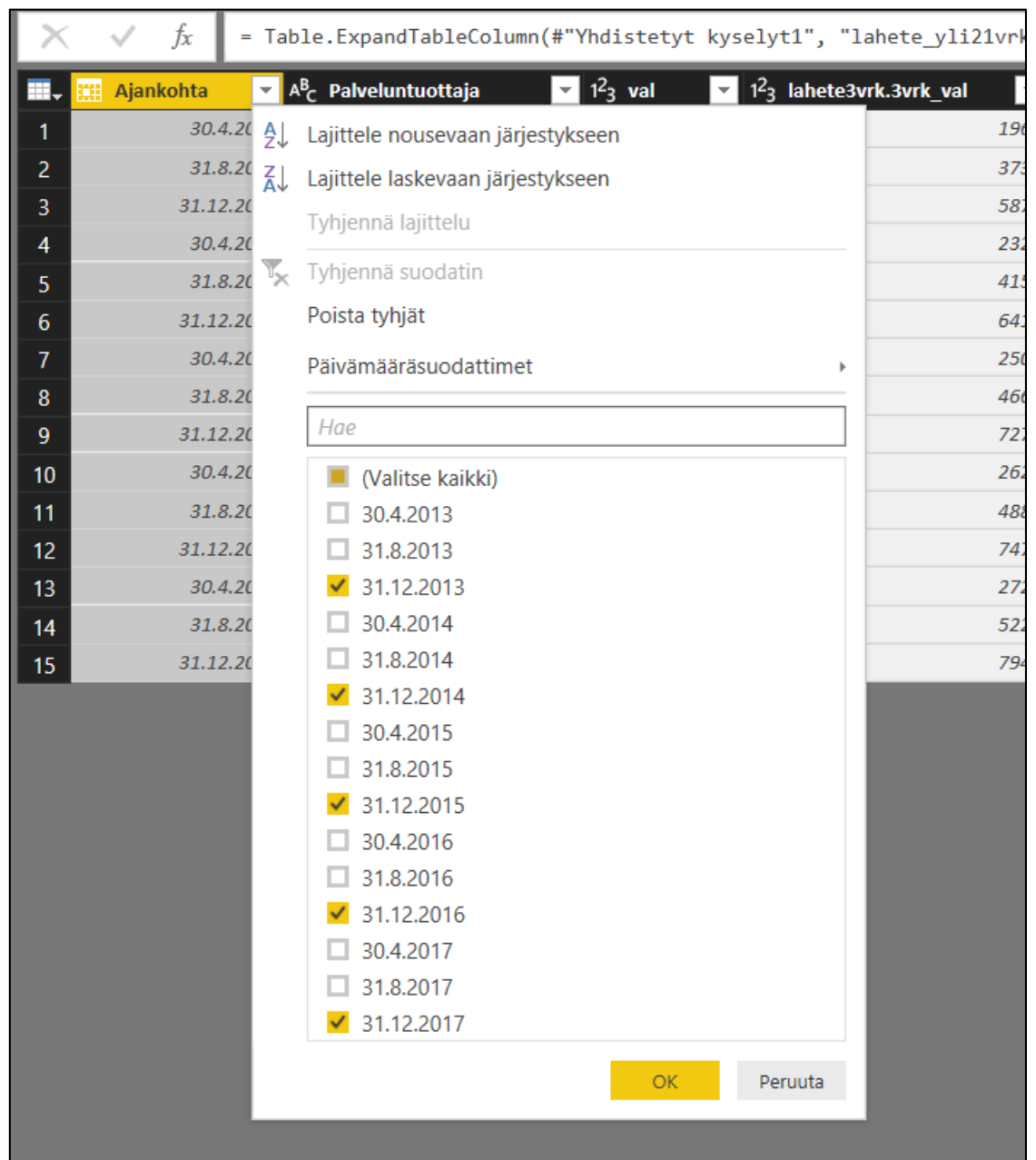
Kuva 11. Ajankohta-muuttujan luomiseen käytettiin Power BI:n ehdollinen sarake -toimintoa.

Liitä kyselyt -toiminnon suorittamisen jälkeen Power BI:ssa oli yksi taulu, joka sisälsi kaikki TIKU-julkaisujärjestelmästä ladatut aineistot ja Power BI oli nimennyt taulun automaattisesti Append1-tauluksi. Tämän jälkeen pystyttiin käyttämään aineiston tietoja samassa visualisoinnissa.

10.2 Tietojen suodatus

THL:n Lastentautien erikoissairaanhoidon jonotilanne datasta haluttiin tehdä taulukkovisualisointi, josta näkee läheteiden käsittelyajat vuosittain. Data sisälsi mittarin arvot kullekin vuodelle kumulatiivisesti lasketuna ja visualisointiin haluttiin ottaa mukaan kustakin vuodesta vain yksi viimeisin eli joulukuun arvo.

Dataa pystyy suodattamaan Power BI:n kyselyeditorissa graafisen työkalun avulla vastaavanlaisesti kuin Excel-ohjelman sarakkeita. Edellä mainittu poiminta pystyttiin tekemään helposti Suodata-toiminnan avulla ja käyttämään tämän jälkeen vain kunkin vuoden viimeisiä eli joulukuun arvoja taulukkovisualisointia tehdessä.



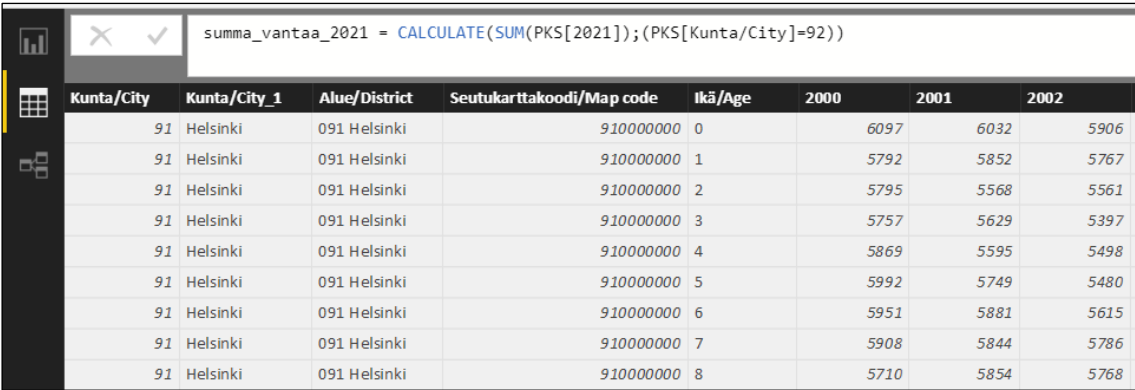
Kuva 12. Power BI -ohjelman kyselyeditori sisältää Suodatin-toiminnan, jossa pystyy poimimaan datasta halutut tietueet.

10.3 Mittarin muodostaminen DAX-funktiolla

Avoindata.fi-palvelusta ladattu CSV-tiedostopohjainen aineisto Pääkaupunkiseudun väestöennuste 2012-2021 sisältää pääkaupunkiseudun väkiluvun alueittain 2000-2011 ja väestöennusteen vuoteen 2021 asti. Aineisto sisältää Helsingin, Espoon ja Vantaan vuonna 2011 laatimat ennusteet sekä Tilastokeskuksen ennusteen Kauniaisen osalta.

Kyselyeditorissa tapahtuneen datan pienimuotoisen muokkauksen (tietotyyppien muuttaminen ja tietueiden suodattaminen) jälkeen kokeiltiin DAX-funktion käyttöä ja uuden mittarin luomista. Haluttiin laskea kunnittain asuvien 0-14-vuotiaiden lapsien määrä ja väestöennuste sekä käyttää tätä arvoa Power BI:n visualisoinnissa. Power BI laskee automaattisesti lukujen aggregaatteja, mutta visualisointiin haluttiin saada vain tämä yksittäinen summa-arvo, joten arvon laskemiseen käytettiin DAX-funktiota. DAX-funktion laskutoimituksissa käytettävät muuttujat ovat oltava tietotyybiltään luonnollisesti numeerisia.

Mittarin voi nimetä haluamakseen esimerkiksi summa_vantaa_2021. DAX-funktion syntaksi on muotoa FUNKTION_NIMI (ARGUMENTTI1;ARGUMENTTI2....). Power BI osaa ehdottaa automaattisesti sarakkeiden nimiä funktiota kirjoittaessa ja avautuvasta ikkunasta valittiin PKS-tilin 2021-vuoden sarake ja rajattiin laskutoimituksen koskemaan kutakin kuntaa. Tämän jälkeen mittari ilmestyi käyttöliittymässä Kentät-otsikon alle ja sitä pystyi käyttämään visualisoinneissa kuten muitakin muuttujia.

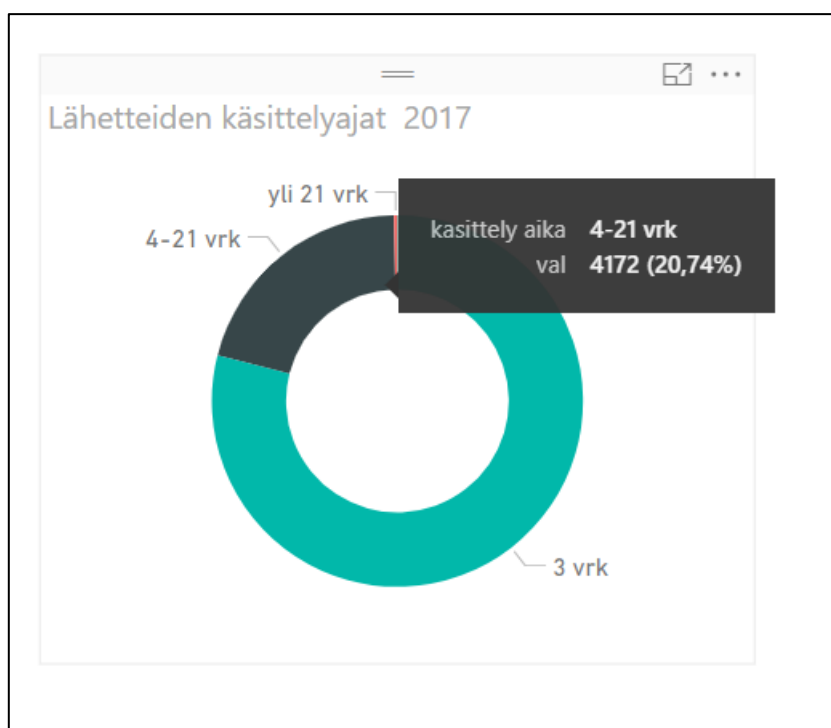


Kunta/City	Kunta/City_1	Alue/District	Seutukarttakoodi/Map code	Ikä/Age	2000	2001	2002
91	Helsinki	091 Helsinki	910000000	0	6097	6032	5906
91	Helsinki	091 Helsinki	910000000	1	5792	5852	5767
91	Helsinki	091 Helsinki	910000000	2	5795	5568	5561
91	Helsinki	091 Helsinki	910000000	3	5757	5629	5397
91	Helsinki	091 Helsinki	910000000	4	5869	5595	5498
91	Helsinki	091 Helsinki	910000000	5	5992	5749	5480
91	Helsinki	091 Helsinki	910000000	6	5951	5881	5615
91	Helsinki	091 Helsinki	910000000	7	5908	5844	5786
91	Helsinki	091 Helsinki	910000000	8	5710	5854	5768

Kuva 13. DAX-funktio laskee vantaan 0-14 vuotiaiden väestölisäyksen vuodelle 2021. Aineistoon suodatettu ensin 0-14 vuotiaiden lisäyksen lukumäärä.

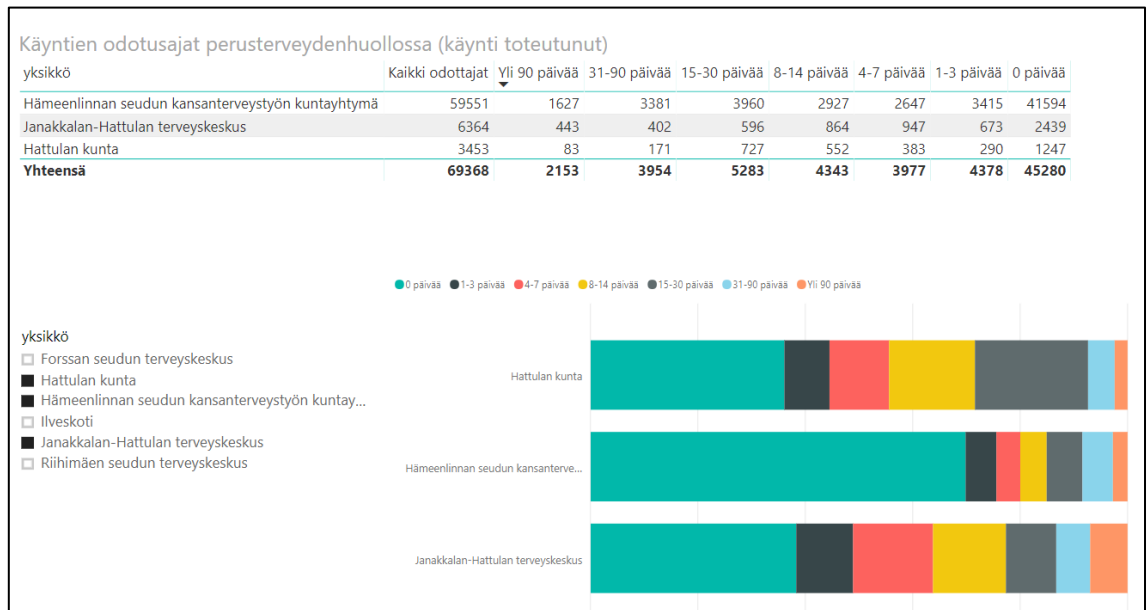
11 DATAN VISUALISOIMINEN POWER BI -SOVELLUKSELLA

Power BI -sovellus sisältää monipuolisia vuorovaikuttavia datan visualisointimahdollisuuksia. Visualisointimahdollisuuksista kokeiltiin kokeiden aikana taulukkoa, osittajaa, suppiloa, rengaskaaviota, pinottua palkkikaaviota, 100%:n pinottua palkkikaaviota, viivakaaviota, kortti- ja kartta-visualisointia. Visualisoinnit valittiin lähinnä datan luonteeseen sopivuuden ja esitystavan selkeyden mukaan. Visualisointien perustaitoihin pääsi melko nopeasti käsiksi ja visualisointien parhaimmaksi ominaisuudeksi osoittautui niiden vuorovaikuttavuus. Dataa pystyy lajittelemaan, korostamaan ja porautumaan syvemmälle pelkällä klikkauksella. Kaikki nämä ominaisuudet tulevat automaattisesti sovelluksesta ja perusvisualisointien teko on nopeaa ja vaivatonta.



Kuva 14. Rengaskaavio HUS Lastentautien lähetteen käsittelyajoista vuonna 2017. Lähde: Terveystieteiden ja hyvinvoinnin laitos 2018, CC BY 4.0

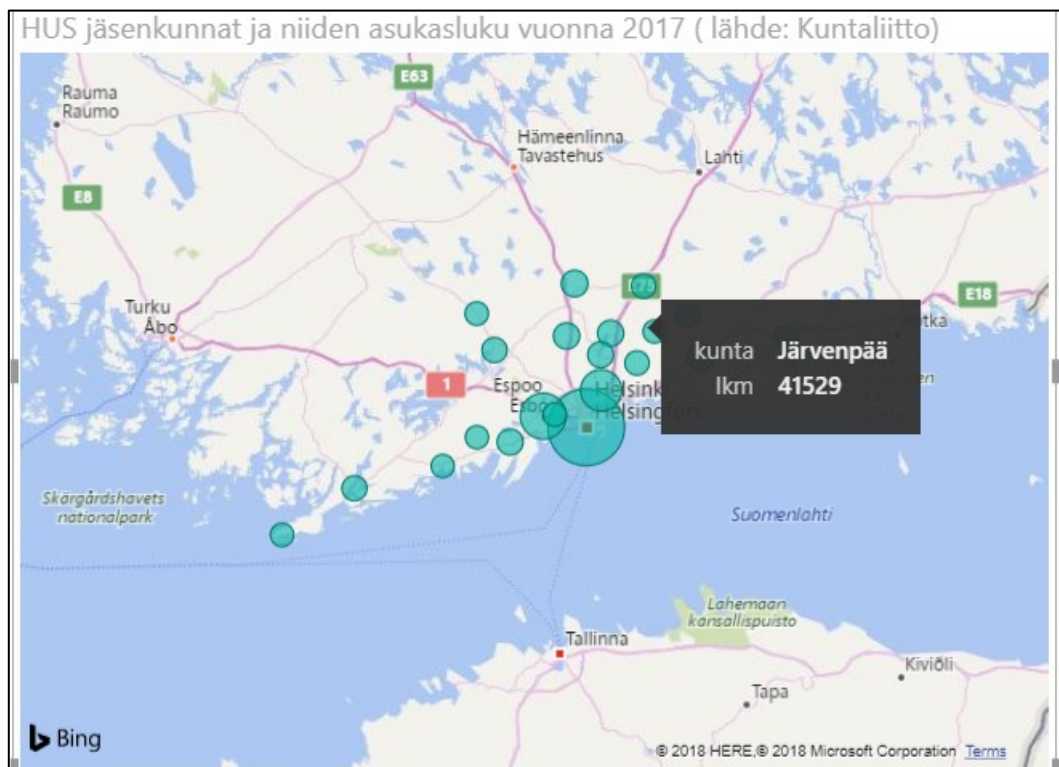
Raportissa tarkasteltavaa dataa pystyy myös rajaamaan helposti Excelin Pivot-taulukosta tutun osittajan avulla. Tällöin käyttäjä voi vain rastittaa raportissa olevan osittajan ruudusta esimerkiksi minkä palveluntuottajan tietoja halutaan ottaa tarkastelun alle. Osittajan ruutuja voi rastittaa yhden tai useamman. Osittajan valinnat voi nollata pyyhkekumi-kuvakkeesta. Raporttien vuorovaikuttavuus, intuitiivisuus ja helppokäyttöisyys osoittautuivat kokeissa Power BI -sovelluksen parhaiksi ominaisuuksiksi.



Kuva 15. Osittajan toiminta Power BI:n koantinäytöllä. Tässä valittu kolme palveluntuottajaa tarkastelun alle. Lähde: Terveiden ja hyvinvoinnin laitos 2018, CC BY 4.0.

Raporttiin pystyi helposti lisäämään kuvan tai organisaation logon. Visualisoinnin värejä pystyi myös muokkaamaan haluamakseen sekä värikartasta valitsemalla että määrittelemällä värin tarkan koodin. Myös fontin koko ja tyyli pystyi muuttamaan haluamakseen kuten myös muokkaamaan taustaa.

Karttavisualisoinnin teossa käytettiin lähteenä Kuntaliiton Internet-sivuilta löytyvää Kuntajaot ja asukasluvut 2000-2017_5 Excel-tiedostoa, joka ladattiin Power BI -työpöytäsovellukseen. Excel-tiedosto oli 97-2003 laskentataulukko ja tämä oli liian vanha tiedostomuoto Power BI -sovellukseen, joten tiedostomuoto oli ensin muunnettava Excelin 2010-versioon. Visualisointiin poimittiin mukaan HUS-jäsenkunnat Power BI -sovelluksen perusuodatusominaisuudella. Myös karttavisualisoinnissa pääsee tarkastelemaan yksittäisen muuttujan tietoa viemällä hiiren osoittimen arvopisteen päälle.



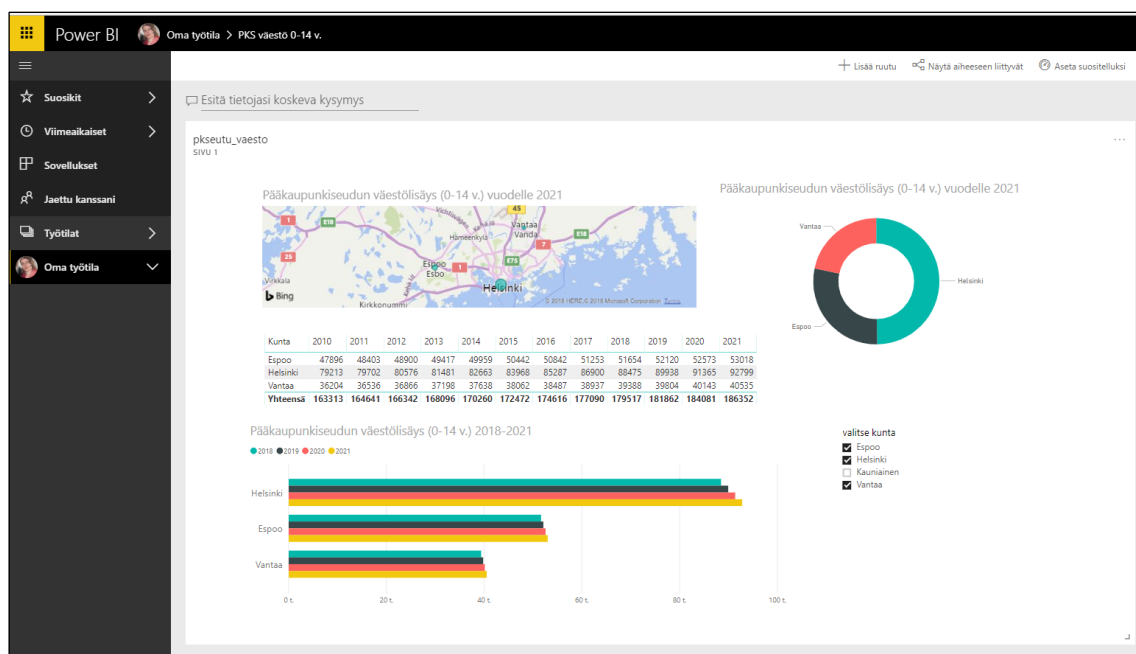
Kuva 16. Karttavisualisointi HUS:n jäsenkunnista. Lähde: Kuntaliitto.

12 KOONTINÄYTÖT JA RAPORTIN JAKAMINEN PILVESSÄ

Power BI raporteista voi koota koontinäyttöjä ja näitä voi jakaa Office 365 -tuotteissa muille kiinnostuneille. Lopuksi kokeiltiin aikaisemmin tehtyä Pääkaupunkiseudun väestölisäys 0-14 v. raportin koontinäytön jakamista. Koontinäyttöön voi liittää useampia raportteja ja esimerkiksi edellä mainitun Kuntaliiton karttavisualisoinnin sai liitettyä HUS Lastentaudit koontinäyttöön mukaan. Karttavisualisointi avautui sen linkkiä klikattaessa.

Koontinäyttöjen tekeminen ja jakaminen pilvessä tapahtui nopeasti. Raportin julkaisemisen etuja pilvessä ovat jo teoriaosuudessa todettu pääte- ja paikkariippumattomuus. Koontinäyttö myös skaalautui testattaessa tarkasteltavan näyttökoon mukaiseksi, ja käyttöliittymän vuorovaikutteisuus toimi kaikissa näyttökoissa. Raportin julkaisijan ei siis tarvitse huolehtia erikseen, millä päätelaitteella raportteja tarkastellaan, vaan raportti skaalautuu automaattisesti päätelaitteelle. Koontinäyttöä pääsee myös tarkastelemaan ennen julkaisua verkko- ja puhelinnäkymässä.

Raportin voi myös julkaista Internetiin muodostamalla raportille linkin tai upotuskoodin Office 365 -tuotteen avulla. Tämän jälkeen kuka tahansa internetin käyttäjä voi tarkastella raporttia julkiselta internet-sivulta. Raportista julkaistaan tällöin reaaliaikainen versio, joka pysyy synkronoituna Power BI-version kanssa eli raporttiin tehdyt muutokset Power BI:ssa päivittyvät automaattisesti julkaistuun julkiseen versioon.



Kuva 17. Koontinäyttö Office 365 -palvelussa. Pääkaupunkiseudun väestölisäys 0-14 vuotiaat. Lähde: Avoindata.fi -portaali (Espoo, Vantaa, Helsinki ja Tilastokeskus).

13 TULOKSET

Case-tutkimuksella ei ole omaa metodologiaansa eikä omia tiedonkeruuta tai analyysimenetelmiä. Sama koskee myös luotettavuustarkastelua, johon kuuluu reliabiliteetti ja validiteetti. (Kananen 2013, 114.) Case-tutkimuksen luotettavuuden kannalta on tärkeää, että esitetty pystytään perustelemaan ja näyttämään toteen (Kananen 2013, 122). Tutkimuksen validius merkitsee kuvauksen ja siihen liitettyjen selitysten ja tulkintojen yhteensovittuvuutta. Tutkimuksen luotettavuutta kohentaa tutkijan tarkka selostus tutkimuksen toteuttamisesta. (Hirsijärvi ym. 2006, 217.)

Tässä opinnäytetyössä tulosten toteen näyttämässä on käytetty järjestelmällistä dokumentointia havaintolomakkeen (liite 3) avulla. Kokeiden tulokset on koottu yhteen ja niistä on pyritty löytämään yhtenäiseviä seikkoja kokeiden ja aineistojen variaatiosta huolimatta sekä tekemään vertailuja eri aineistojen sekä tiedostomuotojen kesken. Havaintolomakkeesta on koottu analyysivaiheessa kirjallinen tiivistelmä. Ojasalon ym. mukaan analyysin tarkoituksena on informaatioarvon lisääminen, ja sanallisen ja selkeän kuvauksen luominen tutkittavasta asiasta (Ojasalo ym. 2014, 136).

Suoritettut kokeet ja visualisoinnit on perusteltu tietoaineistojen luonteesta tulleista seikoista. Kokeita on suoritettu yhteensä seitsemäntoista kappaletta. Kokeita on pyritty toistamaan niin, että niistä ei noussut enää merkittävää uutta tietoa eli käyttämään jo aiemmin mainittua saturatiota.

Case-tutkimuksessa tarkastellaan ilmiötä mahdollisimman monesta näkökulmasta ja pyritään saamaan ilmiöstä kokonaisvaltainen kuva (Kananen 2013, 34, 57). Tässä opinnäytetyössä on pyritty keräämään käsittelyyn useampia eri aineistoja eri julkaisijoilta ja eri muodoissaan sekä soveltamaan niiden käsittelyyn monipuolisesti Power BI:n toimintoja kulloisenkin aineiston luonteeseen sopivalla tavalla. Teoriatietoon on pyritty myös tutustumaan laajasti ja kattavasti sekä peilaamaan sitä käytännön kokeisiin.

13.1 Tiedostomuotoisen datan lataus

THL:n TIKU-järjestelmästä tietoa ladattiin sekä CSV- että Excel-tiedostomuodossa. Avoindata.fi-portaalista ja Kuntaliiton sivuilta ladattiin myös dataa CSV ja Excel-formaatissa. Lataus onnistui hyvin edellä mainituissa tiedostomuodoissa ja data oli myös käsiteltävissä Power BI -sovelluksella hyvin näissä kahdessa tiedostomuodossa.

THL:n TIKU-käyttöliittymän kautta pystyy tekemään dataan ensin haluamansa rajaukset ja sen jälkeen datan saa ladattu rajauksineen paikallisesti. Ladattuihin tietoaineistoon täytyi tehdä jonkin verran muokkauksia ennen raporttien ja visualisointien tekoa esimerkiksi rivien ja sarakkeiden pois-

toja, tietotyypin vaihtamista jne. Nämä onnistuivat sekä Excelin Power Pivot -apuohjelmalla että Power BI:n kyselyeditorissa hyvin samantyyppisesti. Kuntaliiton ja Helsingin väestö viisivuotisikäryhmittäin dataa lataessa havaittiin, että Excel 97-2003 version laskentataulukoiden lataus ei enää onnistu, vaan tiedosto on muutettava uudempaan versioon.

13.2 Rajapintakutsut datan latauksessa

Kokeiltiin myös datan latausta rajapintakutsuna THL:n TIKU-julkaisujärjestelmästä. Data latautui ongelmitta CSV- ja JSON-rajapintakutsuna. Tämän menetelmän etuina ovat datan nopea lataus, kun aineistoa ei tarvitse tallentaa ensin paikallisesti. Aineiston parametrit voi määrittellä suoraan rajapintakutsuun, jolloin data latautuu halutuilla rajauksilla. Parametrien määrittelyyn apuna käytettiin TIKU-käyttöliittymää, jonka osoiteriviltä saa suoraan haluamansa parametrit. Osoitteeseen lisättiin vain määrittely datan formaatista eli joko .csv tai .json.

Datan lataus onnistui molemmissa tiedostoformaateissa. Kokeissa todettiin THL:n aineistojen osalta, että CSV- tiedostomuodon käsittely on Power BI-sovelluksessa JSON-tiedostomuotoa helpompaa johtuen CSV-tiedostomuodon sarakemaisesta tietorakenteesta. Edellä olevien rajapintakutsujen tulosta vertaillaessa voi todeta, että JSON-muotoisen datan kanssa sai tehdä enemmän töitä esimerkiksi viemään datan tietueet taulukkoon ja datan tietomallia tarkasteltaessa JSON-muotoinen data sisältää paljon enemmän tietueita. JSON-tiedosto olisi siis edellyttänyt datan muokkausta ja siivoamista kyselyeditorissa ennen visualisointia, joten tässä tapauksessa voi suositella CSV-tietomuodon käyttöä, koska se on latauksen jälkeen ”valmiimmassa” muodossa eli ”sarakemuodossa”. Power BI -työkalulla on siis helpompaa työskennellä sarakemallisen datan kanssa.

THL:n TIKU-julkaisujärjestelmän rajapinta oli testien aikana vapaasti käytettävissä, eikä aineiston lataus edellyttänyt kirjautumista tai lupaa. Sivuilta löytyi myös kuvaus tietorakenteesta sekä esimerkkiaineisto. Näin ollen rajapinta täyttää sille määritellyt vaatimukset.

13.3 Power BI -työkalu

Kokeissa Power BI -työkalun eduiksi nousivat käyttöliittymän vuorovaikuteisuus ja helppous laskea esimerkiksi summia ja mediaaniarvoja ilman kaavojen määrittelyä. Päivämäärien tietotyyppiä tuli automaattisesti datan latauksen yhteydessä teksti, mutta tämä oli helppo vaihtaa päivämäärä-tietotyyppiä, jolloin pystyi hyödyntämään päivämäärä hierarkian etuja.

Vuorovaikutteisuuden eduiksi todettiin muun muassa se, että käyttöliittymän objekteja on helppo siirtää raahaamalla näytöllä. Objekteja on vain kli-

kattava, jolloin sen paikkaa voi siirtää haluamaansa kohtaan. Tietoja voi lajitella joko nousevasti tai laskevasti klikkauksella. Klikkaamalla voi valita tarkasteluun yksittäisen kohteen esimerkiksi palkkikaaviosta palkin. Yksittäisen elementin tarkemmat tiedot saa esille viettäessä hiiren osoittimen sen päälle. Hierarkiaan pääsee tarkastelemaan syvemmälle omasta nuolipainikkeesta käsin. Suoritetuissa esimerkeissä tämä ei ollut ihan mielekäästä datan luoteen vuoksi, mutta Lähetteet yhteensä -visualisoinnissa tätä ominaisuutta pystyi demonstroimaan käytännössä. Osittajasta voi poistaa valinnat pyyhkekumin kuvakkeella, joka on käyttäjälle opittavaa ja ymmärrettävää.

Power BI sisältää myös datan käsittelyyn monipuolisia toimintoja esimerkiksi taulujen yhdistämiseen tarvittavat Yhdistä- ja Liitä kysely -toiminnot sekä datan suodatus-, etsi- ja korvaa sekä ehdollinen sarake -toiminnot. Sovellus sisältää myös DAX-kielen, jolla voi tehdä monipuolisia laskutoimintuksia.

Työkalu tarjoaa myös useita visualisointeja käytettäväksi ja perustoimintalogiikan oppii itsenäisesti melko nopeasti. Visualisoinnille voi myös määrittellä ehtoja graafisen käyttöliittymän avulla sekä tarjota käyttäjälle osittajan, jonka avulla voi tehdä rajauksia näytettäviin tietoihin pelkällä klikkauksella. Visualisointien värejä ja fontin kokoa pystyy muuttamaan haluamakseen, mikä on tärkeää esteettömyys-näkökulmasta tarkasteltuna.

Koontinäyttöön voi koota visualisointeja useammasta raportista ja jakaa näytön pilven kautta. Koontinäytöt skaalautuvat eri päätelaitteille ja niiden vuorovaikutteisuus säilyy jakamisen jälkeen myös pienemmillä näytöillä tarkasteltaessa.

Miinuspuoliksi kokeissa koettiin visualisointien käytännön toteutuksessa se, että visualisointi- ja kentät-osio tuntui toisinaan vähän kömpelöltä. Kenttä ei aina raahautunut oikealle kohdalleen. Visualisointien tekstit tulevat myös automaattisesti sen verran pienellä, että niitä on aina suurennettava. Myös erillistä SQL-editoria kaivattiin välillä, jolla voisi suorittaa aineistolle SQL-kyselykielisiä lauseita.

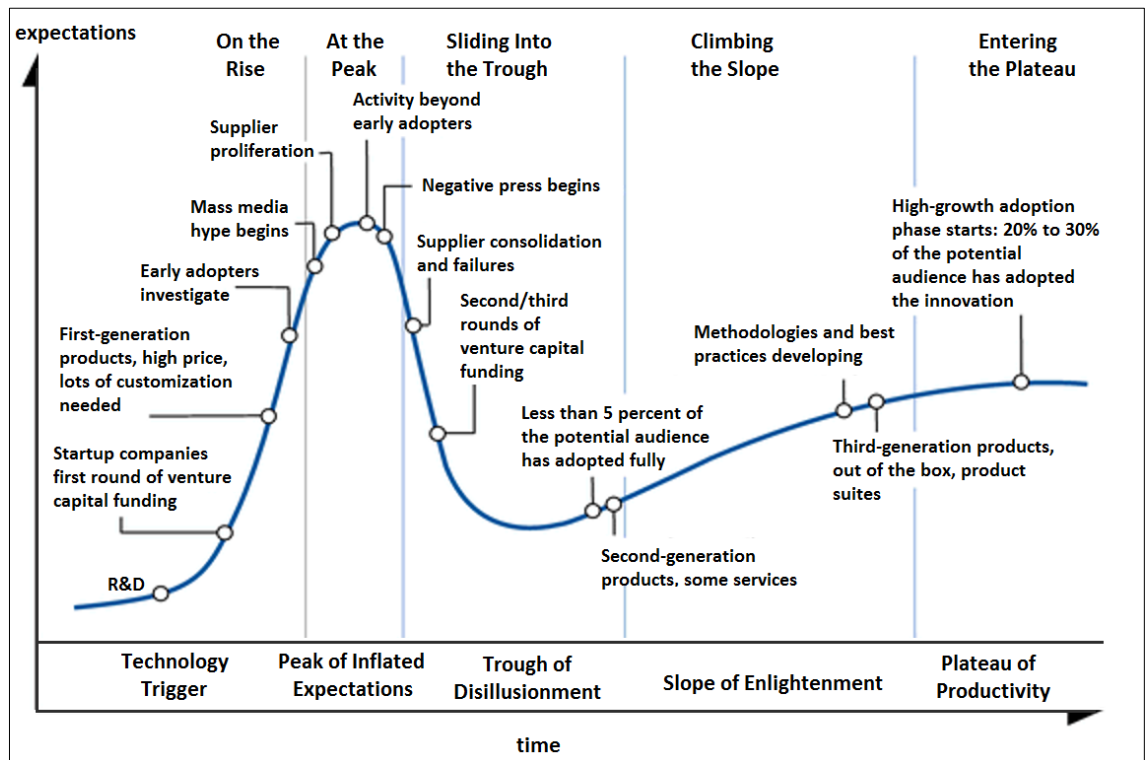
13.4 Avoimet tietoaaineistot

Avoimia tietoaaineistoja löytyi melko paljon sekä THL:n että Avoindata.fi-portaalista. Kaikki käsitellyt tietoaaineistot olivat lisensoitu CC BY 4.0 -lisenssillä. Tietoaaineistoja oli tarjolla useassa eri tiedostomuodossa sekä ladatavissa myös rajapinnan kautta. Aineistojen tietorakenteissa samoin kuin tiedostoformaateissa oli paljon vaihtelevuutta eri julkaisijoiden kesken. Myös rajapintaan tutustuminen vaatii perehtyneisyyttä ja paneutumista. THL:n rajapinnan osalta oli varmasti apua työkokemuksesta Hydratietomallin parissa. THL:n rajapintaan tutustuminen olisi ollut luultavasti haasteellisempaa asiasta ennalta täysin tietämättömälle.

Käsitelty data ei ollut aivan tasalaatuista eri vuosiin vertailtaessa myöskään saman julkaisijan osalta. Tämä on toki ymmärrettävää, kun ottaa huomioon, että avointa dataa on julkaistu vasta viime vuosien ajan ja toki aineistoille asetetut tarpeet muuttuvat luonnollisestikin ajan myötä. Kokeissa havaittiin, että mikäli data sisältää selkeästi erillään olevat raportointiulottuvuudet eikä raportointiulottuvuuksia ole yhdistetty samaan tietomatriisiin kenttään, niin data on helpompaa käsitellä.

Pari ladattua datatiedostoa olivat Excelin 97-2003 versiota ja ne osoittautuivat liian vanhoiksi Power BI -sovelluksen käsiteltäväksi. Excel 97-2003 tiedostomuoto oli muunnettava ennen käsittelyä 2010-versioon. Kaikki käsitellyt aineistot eivät myöskään sisältäneet tietueen yksilöimiseen tarvittavia tunnistetietoja ja joistakin aineistoista puuttui myös selkeä merkintä, milloin dataa on päivitetty.

Teoriatiedon perusteella avoimen datan hyödynnettävyys on vasta alkutaipaleella. Mikäli Suomen avoimen datan tilannetta tarkastellaan ”Innovaatiokäyrän” (kuva 18) näkökulmasta niin mielestäni avoin data sijoittuu innovaatiokäyrän aallonharjan kohtaan ”Early adopters investigate”. Tutkimusta ja kehitystä avoimen datan osalta on jo tehty viime vuosina ja dataa on myös alettu hyödyntämään sekä sen ympärille on käynnistetty erilaisia hankkeita kuten aikaisemmin mainittu 6Aika ja hackathon-tapahtumia. Silti avoimen datan hyödyntämisessä olisi vielä tehtävää, ja kuten Valtioneuvoston raportissa *Avoimen datan hyödyntäminen ja vaikuttavuus* todetaan, datan löydettävyyttä ja hyödynnettävyyttä helpottaisi datan entistä keskitetympi jakaminen. Myös datan avaintietojen ja yhteisien luokituksien lisääminen helpottaisi datan käsittelyä. World Wide Web Foundationin ylläpitämässä Open Data Barometrissa oli mainittu tunnistetietojen saatavuus kehityskohteeksi Suomen avoimissa tietoaaineistoissa. Yhteiset, vakiintuneet tunnistetiedot helpottaisivat datan yhdisteltävyyttä.



Kuva 18. Innovaatiokäyrän "aallonharja". (Brodnick 2016.)

14 JOHTOPÄÄTÖKSET

Avoim data on julkisesti saatavilla olevaa, koneluettavassa muodossa olevaa ja avoimeksi lisensoitua tietoa (Valtiovarainministeriö 2015a, 1). Avoin data ei saa sisältää luottamuksellista tietoa, henkilötietoja tai liikesalaisuuksia (Valtioneuvoston kanslia 40/2017, 59). Kaikki kokeissa käsittelemäni aineistot oli koneluettavassa muodossa ladattavissa Power BI -sovellukseen joko erillisen julkaisujärjestelmän välityksellä tai tiedostomuotoisena Internet-sivulla. Kaikki aineistot olivat myös lisensoitu selkeästi avoimiksi tietoaaineistoiksi Creative Commons –lisenssillä, ja ne sisälsivät vain julkiseksi tarkoitettua tietoa. Avoin data on ilmiönä uusi, ja sen vuoksi aiheesta ei ole vielä saatavilla kattavaa tutkimustietoa. Avoin datan hyödyt on kuitenkin ennakoitu mittaviksi niin yhteiskunnallisesti kuin taloudellisesti. (Valtiovarainministeriö 2015a, 1-2.)

Avointa tietoa tarjotaan ladattavaksi sekä ohjelmointirajapinnan välityksellä että tiedostomuotoisena (Valtioneuvoston kanslia 40/2017, 52). Kokeilin opinnäytetyöni käytännön osiossa avoimen datan latausta sekä tiedostomuotoisesti että rajapinnan välityksellä. THL:n TIKU-julkaisujärjestelmän rajapintakutsu oli nopea ja helppo keino saada tietoa latautumaan ja rajapintakutsuun oli myös helppo määritellä parametreja. Rajapinta oli myös vapaasti käytettävissä kokeiden aikana ja sisälsi tarvittavan dokumentoinnin ja kuvauksen, joten rajapinta voidaan todeta Avoinmen rajapinnan määrittelyn mukaiseksi. Myös tiedostomuotoinen lataus sekä THL:n TIKU-julkaisujärjestelmästä että Avoindata.fi-portaalista ja Kuntaliiton sivuilta onnistuivat. Datan latauksessa sarakemallinen data on helpoiten käsiteltävissä Power BI -sovelluksessa ja tiedostomuotoisissa versioissa uusimmat (2010) versiot. Datan käsittelyä ja hyödyntämistä helpottaa, mikäli datan liitännäistiedot esimerkiksi päivytyspäivämäärä ovat kunnossa ja data sisältää tietueiden yksilöimiseen vaadittavat tunnistetiedot sekä selkeästi erillään olevat raportointitulottuvuudet.

Power BI -työkalu soveltui tässä opinnäytetyössä käytettyjen avoimien terveys- ja väestötietoaaineistojen visualisointiin ja käsittelyyn. Työkalulla pystyi korostamaan sekä kiinnittämään huomiota datassa oleviin seikkoihin. Power BI -työkalun käyttöliittymällä pystyi tarkastelemaan visualisointeja helppokäyttöisesti ja lajittelemaan sekä tekemään vertailuja. Hiirellä klikkaamalla pystyi korostamaan yksittäistä tietoelementtiä ja saamaan näkyville enemmän tietoa hiiren osoittimella avattavan pikkuikkunan tarkentavan tekstin avulla. Koontinäyttöjä voi myös jakaa Microsoftin Azure-pilvipalvelun kautta tarkasteltavaksi eri päätelaitteille.

LÄHTEET

- 6Aika. Uutta kaupunkia yhdessä. Haettu 9.4.2018 osoitteesta <https://6aika.fi/6aika-avoimia-ja-alykkaita-palveluja/>
- Aspin, A. (2018). *Pro Power BI Desktop – Interactive data analysis and visualization for the desktop*. Staffordshire: Apress.
- Avoin rajapinta. Versio 1.1. (2017). Haettu 24.1.2018 osoitteesta <http://avoinrajapinta.fi/about/>
- Avoindata.fi (2018). Tietoa verkkopalvelusta. Haettu 10.4.2018 osoitteesta <https://www.avoindata.fi/fi/about>
- Brodnick, R. (2016). Innovations in strategy crafting. Blogijulkaisu 12.5.2016. Haettu 18.4.2018 osoitteesta <http://strategycrafting.blogspot.fi/2016/05/curves-that-matter-s-curve.html>
- Creative Commons (n.d). Nimeä 4.0 Kansainvälinen (CC BY 4.0). Haettu 14.2.2018 osoitteesta <https://creativecommons.org/licenses/by/4.0/deed.fi>
- Gartner (2017). Magic Quadrant for Business Intelligence and Analytics Platforms. Haettu 23.2.2018 osoitteesta <https://www.gartner.com/doc/reprints?id=1-3TXXSLV&ct=170221&st=sb>
- Heino, P. (2010). *Pilvipalvelut*. Hämeenlinna: Kariston Kirjapaino Oy.
- Hirsijärvi, S., Remes, P. & Sajavaara, P. (2006). *Tutki ja kirjoita*. Helsinki: Kustannusosakeyhtiö Tammi
- Hovi, A., Hervonen, H. & Koistinen, H. (2009). *Tietovarastot ja business intelligence*. Jyväskylä: Docendo
- Kananen (2013). *Case-tutkimus opinnäytetyönä*. Jyväskylä: Jyväskylän ammattikorkeakoulu.
- Kilpeläinen, T. (2017). Liiketoiminta, suunnittelu ja testaus moduulin verkkoinfo, Moodle. Hämeen ammattikorkeakoulu. Haettu 13.4.2018 osoitteesta <https://moodle.hamk.fi>
- Koponen, J., Hildén, J. & Vapaasalo, T. (2016) *Tieto näkyväksi – Informaatiomuotoilun perusteet*. Helsinki: Aalto-yliopiston julkaisusarja.
- Kosonen, M. (2017). Terveystiedoista on tullut Suomen kansallisaarre. *Helsingin Sanomat* 10.11.2017, A5.

Kotilainen, S. (2017). API tuo rahaa. *Tivi* elokuva, 18–25.

Markkula, T. & Syväniemi, A. (2015). *Analytiikkamatka datasta tietoon ja tiedolla johtamiseen*. Suomen Liikekirjat.

Microsoft (n.d.). Introducing Power BI. Haettu 8.3.2018 osoitteesta <https://docs.microsoft.com/en-us/power-bi/guided-learning/getting-started#step-1>

Microsoft (2017). Tutorial: Create your own measures in Power BI Desktop. Haettu 13.4.2018 osoitteesta <https://docs.microsoft.com/en-us/power-bi/desktop-tutorial-create-measures>

Moore, S. (2014). *Issues in open research data*. London: Ubiquity Press Ltd. Haettu 10.11.2017 osoitteesta <https://www.ubiquitypress.com/site/books/10.5334/ban/>

Ojasalo, K., Moilanen, T. & Ritalahti, J. (2014). *Kehittämistyön menetelmät – Uudenaista osaamista liiketoimintaan*. Helsinki: Sanoma Pro Oy.

Richardson, L. & Amundsen, M. (2013). *RESTful Web APIs*. Sebastopol: O'REILLY.

Rintakoski, S. (2017). Erotu joukosta tiedon visualisoinnin avulla. Avoin luento 22.11.2017, Skype.

Salo, I (2010). *Cloud computing – Palvelut verkossa*. Porvoo: Bookwell Oy.

Salo, I. (2014). *Big data & pilvipalvelut*. Jyväskylä: Docendo.

Sharda, R., Delen, D. & Turban, E. (2014). *Business Intelligence – A Managerial Perspective on Analytics*. Harlow: Pearson.

Terveyden ja hyvinvoinnin laitos (2016). Strategia. Haettu 9.1.2018 osoitteesta <https://www.thl.fi/en/thl/strategia>

Terveyden ja hyvinvoinnin laitos (2017). THL:n avoimen datan rajapinnan kuvaus. Haettu 30.1.2018 osoitteesta <https://yhteistyotilat.fi/wiki08/dislay/THLKA/THL%3An+avoimen+datan+rajapinnan+kuvaus>

Valtioneuvosto (2011). Valtioneuvoston periaatepäätös julkisen sektorin digitaalisten tietoaisteiden saatavuuden parantamisesta ja uudelleenkäytön edistämisestä 3.3.2011. Haettu 14.1.2018 osoitteesta [https://www.lvm.fi/documents/20181/782861/Ehdotus+valtioneuvoston+periaatepaatokseksi+-+Julkinen+tietoaiste+\(3.3.2011\).pdf/50cd10fc-824b-407c-aabe-2f56e333e108?version=1.0](https://www.lvm.fi/documents/20181/782861/Ehdotus+valtioneuvoston+periaatepaatokseksi+-+Julkinen+tietoaiste+(3.3.2011).pdf/50cd10fc-824b-407c-aabe-2f56e333e108?version=1.0)

Valtioneuvoston kanslia (2017). Avoimen datan hyödyntäminen ja vaikuttavuus, Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 40/2017. Haettu 11.1.2018 osoitteesta http://tietokayttoon.fi/documents/10616/3866814/40_avoimen+datan+16032017.pdf/0444467d-5400-4f0c-8728-2447cef039ad?version=1.0

Valtiovarainministeriö (2015a). Avoimen tiedon vaikuttavuus – esitutkimus, Valtiovarainministeriön julkaisuja – 15a/2015. Haettu 11.1.2018 osoitteesta <http://vm.fi/documents/10623/1107406/Avoimen+tiedon+vaikuttavuus/cd515174-dfeb-4959-a193-246ef2fbefc7?version=1.0>

Valtiovarainministeriö (2015b). Avoimesta datasta innovatiiviseen tiedon hyödyntämiseen, Avoimen tiedon ohjelman 2013 – 2015 loppuraportti, Valtiovarainministeriön raportti – 31/2015. Haettu 11.1.2018 osoitteesta <http://vm.fi/documents/10623/1107406/Avoimen+tiedon+ohjelman+loppuraportti/8eaaee68-6f3b-4a48-8b57-c5866315bf13?version=1.0>

Vilka, H. & Airaksinen, T. (2003). *Toiminnallinen opinnäytetyö*. Helsinki: Tammi.

Virtuaaliammattikorkeakoulu (n.d.). Käytettävyydestä. Haettu 13.5.2018 osoitteesta <http://www2.amk.fi/digma.fi/www.amk.fi/opintojak-sot/030308/1111676348138/1111677021119/1111677206424/1111677569162.html>

World Wide Web Foundation (2017). The Open Data Barometer. Haettu 9.4.2018 osoitteesta https://opendatabarometer.org/?_year=2016&indicator=ODB

AvoHILMO: Käyntien odotusajat perusterveydenhuollossa (käynti toteutunut), Kanta-Häme ja Forssan seudun terveyskeskus

https://sampo.thl.fi/pivot/prod/fi/avo/hpaasyph01/fact_ahil_hpaasyph01_kaikki.csv?row=palveluntuottaja-31127&column=odotusaika-25575

https://sampo.thl.fi/pivot/prod/fi/avo/hpaasyph01/fact_ahil_hpaasyph01_kaikki.csv?row=palveluntuottaja-30690&column=odotusaika-25575

Rokotusrekisteri: Lasten rokotuskattavuus, Espoo, syntymävuodet 2009-2011

https://sampo.thl.fi/pivot/prod/fi/vaccreg/coverage/fact_coverage_infant.csv?row=area-13044&column=vaccine-172606.&column=time-133277.133264.133270

Erikoissairaanhoidon jonotilanne, neurokirurgia, Helsingin ja Uudenmaan sairaanhoitopiiri

https://sampo.thl.fi/pivot/prod/fi/eshjono/eshj/fact_eshj.csv?row=tuottaja-255447.&column=time-430&filter=eala-255530#

Erikoissairaanhoidon jonotilanne, lastentaudit, Helsingin ja Uudenmaan sairaanhoitopiiri, jäsenkuntien asukasluku, mediaani vrk, lähete käsitelty yli 21 vrk, lähete käsitelty 4-21 vrk:ssa, lähete käsitelty 3 vrk:ssa, lähetteet yhteensä, vuodet 2013-2017 (30.4., 31.8. ja 31.12.)

https://sampo.thl.fi/pivot/prod/fi/eshjono/eshj/fact_eshj.csv?row=tuottaja-255447.&column=time-430&filter=eala-255520&filter=measure-18798#

https://sampo.thl.fi/pivot/prod/fi/eshjono/eshj/fact_eshj.csv?row=tuottaja-255447.&column=time-430&filter=eala-255520&filter=measure-18646#

https://sampo.thl.fi/pivot/prod/fi/eshjono/eshj/fact_eshj.csv?row=tuottaja-255447.&column=time-430&filter=eala-255520&filter=measure-18790#

https://sampo.thl.fi/pivot/prod/fi/eshjono/eshj/fact_eshj.csv?row=tuottaja-255447.&column=time-430&filter=eala-255520&filter=measure-18811#

https://sampo.thl.fi/pivot/prod/fi/eshjono/eshj/fact_eshj.csv?row=tuottaja-255447.&column=time-430&filter=eala-255520&filter=measure-18659#

https://sampo.thl.fi/pivot/prod/fi/eshiono/eshj/fact_eshj.csv?row=tuottaja-255447.&column=time-430&filter=eala-255520&filter=measure-18653#

Avohilmo: Perusterveydenhuollon asiakkaat, Helsinki, 0 v., 1-6 v. ja 7-14 v.

https://sampo.thl.fi/pivot/prod/fi/avo/perus03/fact_ahil_perus03.csv?row=alue-94845.&column=aika-87596&filter=ikaluokat-95161#

https://sampo.thl.fi/pivot/prod/fi/avo/perus03/fact_ahil_perus03.csv?row=alue-94845.&column=aika-87596&filter=ikaluokat-95145#

https://sampo.thl.fi/pivot/prod/fi/avo/perus03/fact_ahil_perus03.csv?row=alue-94845.&column=aika-87596&filter=ikaluokat-95146#

Lastensuojelun määräaikojen seuranta, Helsinki, aineistot vuosilta 2014-2018, 1.4.-30.9. ja 1.10.-31.3.

https://sampo.thl.fi/pivot/prod/fi/lasu/seurantaj/fact_lasu_seurantaj.csv?row=area-287619&column=time-287372&column=measure-13345.55139.13347.55138.13226.55140.55141.13144.13475.13376.13358.13264.13283.13495.13430.13214.13284.13295.13309.

https://sampo.thl.fi/pivot/prod/fi/lasu/seurantaj/fact_lasu_seurantaj.csv?row=area-287619&column=measure-13345.55139.13347.55138.13226.55140.55141.13144.13475.13376.13358.13264.13283.13495.13430.13214.13284.13295.13309.&filter=time-224550

https://sampo.thl.fi/pivot/prod/fi/lasu/seurantaj/fact_lasu_seurantaj.csv?row=area-287619&column=measure-13345.55139.13347.55138.13226.55140.55141.13144.13475.13376.13358.13264.13283.13495.13430.13214.13284.13295.13309.&filter=time-175999

https://sampo.thl.fi/pivot/prod/fi/lasu/seurantaj/fact_lasu_seurantaj.csv?row=area-287619&column=measure-13345.55139.13347.55138.13226.55140.55141.13144.13475.13376.13358.13264.13283.13495.13430.13214.13284.13295.13309.&filter=time-20538

https://sampo.thl.fi/pivot/prod/fi/lasu/seurantaj/fact_lasu_seurantaj.csv?row=area-287619&column=measure-13345.55139.13347.55138.13226.55140.55141.13144.13475.13376.13358.13264.13283.13495.13430.13214.13284.13295.13309.&filter=time-21110

https://sampo.thl.fi/pivot/prod/fi/lasu/seurantaj/fact_lasu_seurantaj.csv?row=area-287619&column=measure-13345.55139.13347.55138.13226.55140.55141.13144.13475.13376.13358.13264.13283.13495.13430.13214.13284.13295.13309.&filter=time-13253

https://sampo.thl.fi/pivot/prod/fi/lasu/seurantaj/fact_lasu_seurantaj.csv?row=area-287619&column=measure-13345.55139.13347.55138.13226.55140.55141.13144.13475.13376.13358.13264.13283.13495.13430.13214.13284.13295.13309.&filter=time-13180

https://sampo.thl.fi/pivot/prod/fi/lasu/seurantaj/fact_lasu_seurantaj.csv?row=area-287619&column=measure-13345.55139.13347.55138.13226.55140.55141.13144.13475.13376.13358.13264.13283.13495.13430.13214.13284.13295.13309.&filter=time-13163

VÄESTÖAINEISTOJEN LATAUSOSOITTEET

Liite 2

Kuntajaot ja asukasluvut

<https://www.kuntaliitto.fi/asiantuntijapalvelut/vaestotietoja-kunnittain>

Pääkaupunkiseudun väestöennuste 2012-2021

http://www.hel2.fi/tietokeskus/data/helsinki/Vaestoennusteet/PKS_Vaestoennusteet2012_2021.csv

Helsingin väestö viisivuotiskäryhmittäin, siviilisäädyn ja sukupuolen mukaan 2004 alkaen

http://www.hel2.fi/tietokeskus/data/helsinki/Vaestotilastot/2017/Hki_vaesto_taulu14a.xls

HAVAINTOLOMAKE

Liite 3

lähde	0=ei onnistunut, 1=onnistui	kommentit
tavoite		
datan lataus CSV-muodossa		
datan lataus Excel-muodossa		
rajanpintakutsun käyttö BI-välineessä datan latauksessa		
rajanpintakutsun käyttö parametrien kanssa BI-välineessä datan latauksessa		
tietomallin muokkaus ja tarkastelu		
grafiikan tekeminen		
taulukon tekeminen		
tuotoksen jakaminen pilvessä BI-välineellä		
muuta:		
BI-välineen soveltuvuus datan käsittelyyn ja tiedon visualisointiin		
rajapinnan standardinmukaisuus		