TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

PETRI SALMINEN

EUROOPAN UNIONIN YLEISEN TIETOSUOJA-ASETUKSEN MUKAINEN DATA-ANALYTIIKKA

Kandidaatintyö

Tarkastaja: Lehtori Pasi Hellsten

TIIVISTELMÄ

PETRI SALMINEN: Euroopan unionin yleisen tietosuoja-asetuksen mukainen data-analytiikka
Tampereen teknillinen yliopisto
Kandidaatintyö, 21 sivua
Marraskuu 2017
Tietojohtamisen kandidaatin tutkinto-ohjelma
Pääaine: Tietojohtaminen
Tarkastaja: Lehtori Pasi Hellsten

Avainsanat: data-analytiikka, yleinen tietosuoja-asetus, Euroopan unioni

Tietojohtamisen kandidaatintyössä suoritetaan kirjallisuuskatsaus itseä kiinnostavaan aiheeseen. Tässä kandidaatintyössä tutkitaan Euroopan unionin yleisen tietosuoja-asetuksen mukaista data-analytiikkaa. Tutkimus suoritetaan systemaattisena kirjallisuuskatsauksena käyttäen lähteinä verkkopohjaisista tietokannoista löytyviä lähteitä.

Euroopan unionin yleistä tietosuoja-asetusta on tutkittu tieteellisessä kirjallisuudessa paljon, mutta suomenkielistä tutkimusta on erittäin vähän. Asetusta käytännössä soveltavaa tieteellistä kirjallisuutta löytyy lähinnä julkisen terveydenhuollon ja liiketoiminnan näkökulmista. Data-analytiikka on ajankohtainen aihe liiketoimintaa tukevana prosessina ja sen takia on tärkeää tutkia miten sitä voidaan suorittaa, kun yleistä tietosuoja-asetusta aletaan soveltaa jäsenvaltioissa.

Tutkimuksessa saatiin tulokseksi, että data-analytiikan käytölle ei ole estettä, mikäli sitä tukevat prosessit ja organisaation laajuiset toimintatavat on tehty asetuksen kanssa yhteensopiviksi.

ABSTRACT

PETRI SALMINEN: Data analytics in accordance with the European Union's general data protection regulation
Tampere University of Techonology
Bachelor's Thesis, 21 pages
November 2017
Master's Degree Programme in Information Technology
Degree Programme in Business and Technology Management, BSc. Information and Knowledge Management
Major: Information and Knowledge Management
Examiner: Lecturer Pasi Hellsten

Keywords: data analytics, gdpr, general data protection regulation, European Union

In the bachelor's thesis of information and knowledge management, a literature review is carried out on a topic of self's interest. In this bachelor's thesis data analytics in accordance with the European Union's general data protection regulation is researched. The research is carried out as a systematic literature review using web based databases for the source articles.

European Union's general data protection regulation has been researched in scientific literature thoroughly but there is very little of research in Finnish. Research where the regulation is applied in practice is mainly found in public healthcare and business categories. Data analytics is a current topic as a process of business and that is why it is important to research how data analytics can be carried out when the regulation will become enforceable in the member states.

The results of the research were that there are no reasons why data analytics could be used if the data analytics' supporting processes and organization-wide modes of operation have been made compatible with the regulation.

ALKUSANAT

Tämä työ on tehty kandidaatintyönä osana tietojohtamisen teknistaloudellisen kandidaatintutkintoa Tampereen teknillisellä yliopistolla syksyllä 2017. Koen aiheeni erittäin mielenkiintoiseksi sen yhdistäessä kaksi minulle mielenkiintoista ja yleisesti ottaen ajankohtaista aihetta.

Haluaisin kiittää kandidaatinkurssin vastuuopettajaa Pasi Hellsteniä, joka on tarjonnut työn aikana hyviä vinkkejä ja pitänyt yllä rentoa tunnelmaa vitseillään. Haluan myös kiittää kandidaatintyössä samassa ryhmässä kanssani olleita henkilöitä, joilta olen saanut korvaamatonta palautetta ja vertaistukea työn edetessä. Lisäksi haluan kiittää Henna Ponkalaa, joka on jaksanut kannustaa kandidaatintyön tekemisessä alusta loppuun.

Tampereella, 28.11.2017

Petri Salminen

SISÄLLYSLUETTELO

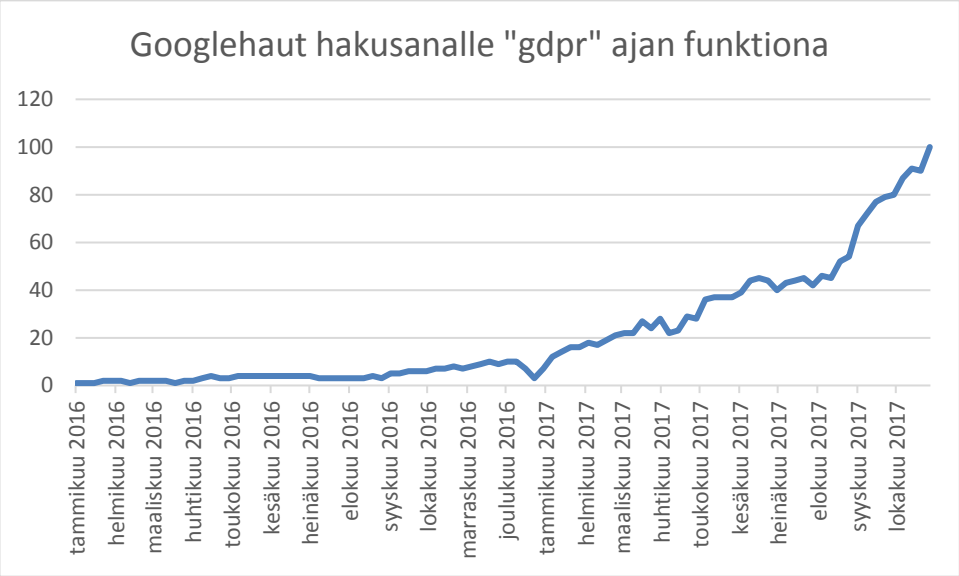
.

LYHENTEET JA MERKINNÄT

EU	Euroopan unioni
GDPR	Katso yleinen tietosuoja-asetus
Henkilötietoryhmä	Henkilötietojen osa, esimerkiksi nimi, sukupuoli tai ammattiliiton jäsenyys
Profilointi	”Mikä tahansa henkilötietojen automaattista käsittely, jossa henkilötietoja käyttämällä arvioidaan luonnollisen henkilön tiettyjä henkilökohtaisia ominaisuuksia, erityisesti analysoidaan tai ennakoidaan piirteitä, jotka liittyvät kyseisen luonnollisen henkilön työsuoritukseen, taloudelliseen tilanteeseen, terveyteen, henkilökohtaisiin mieltymyksiin, kiinnostuksen kohteisiin, luotettavuuteen, käyttäytymiseen, sijaintiin tai liikkeisiin” (Yleinen tietosuoja-asetus 2016)
Pseudonymisointi	”Henkilötietojen käsittelemistä siten, että henkilötietoja ei voida enää yhdistää tiettyyn rekisteröityyn käyttämättä lisätietoja, edellyttäen että tällaiset lisätiedot säilytetään erillään ja niihin sovelletaan teknisiä ja organisatorisia toimenpiteitä, joilla varmistetaan, ettei henkilötietojen yhdistämistä tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön tapahdu” (Yleinen tietosuoja-asetus 2016)
Rekisteri	”Mikä tahansa jäsenneltyä henkilötietoja sisältävä tietojoukko, josta tiedot ovat saatavilla tietyin perustein, oli tietojoukko sitten keskitetty, hajautettu tai toiminnallisin tai maantieteellisin perustein jaettu” (Yleinen tietosuoja-asetus 2016)
Rekisterinpitäjä	”Luonnollinen henkilö tai oikeushenkilö, viranomainen, virasto tai muu elin, joka yksin tai yhdessä toisten kanssa määrittelee henkilötietojen käsittelyn tarkoitukset ja keinot; jos tällaisen käsittelyn tarkoitukset ja keinot määritellään unionin tai jäsenvaltioiden lainsäädännössä, rekisterinpitäjä tai tämän nimittämistä koskevat erityiset kriteerit voidaan vahvistaa unionin oikeuden tai jäsenvaltion lainsäädännön mukaisesti” (Yleinen tietosuoja-asetus 2016)
Rekisteröity	luonnollinen henkilö, johon henkilötiedot liittyvät
Yleinen tietosuoja-asetus	Euroopan unionin asetus 2016/679

1. JOHDANTO

Tietotekniikan kehityksen tuloksena datan määrä on kasvanut huomattavasti, mikä on johtanut myös henkilötietojen käytön monipuolistumiseen. Tämä on yksi syy siihen, että Euroopan Unioni (EU) teki yleisen tietosuojasetuksen 2016/679 (yleinen tietosuoja-asetus). Yleisellä tietosuoja-asetuksella pyritään nykyaikaistamaan ja yhtenäistämään EU:n jäsenvaltioiden tietosuojakäytäntöjä henkilötietojen käsittelyyn ja liikkuvuuteen liittyen ja suojelemaan luonnollisia henkilöitä vahingollisilta tilanteilta, joita voi aiheutua henkilötietojen kontrolloimattomasta keräämisestä tai käytöstä. (Yleinen tietosuoja-asetus 2016) Kuvassa 1 havainnollistetaan kasvanutta kiinnostusta yleistä tietosuoja-asetusta kohtaan googlehakujen määrän avulla. Kuvasta huomataan, että googlehaut hakusanalla ”gdpr” (General data protection regulation) ovat viisinkertaistuneet seitsemässä kuukaudessa maaliskuun ja lokakuun välillä vuonna 2017.



Kuva 1. *Googlehaut hakusanalle "gdpr" ajan funktiona. Y-akselin arvot on sovitettu niin, että suurin arvo = 100. (mukaillen gdpr - Explore - Google Trends 2017)*

Data-analytiikka on vakiinnuttanut asemansa olennaisena osana liiketoiminnan päätöksentekoa (Jain et al. 2014). On mahdollista, että yleisen tietosuoja-asetuksen myötä data-analytiikan hyödyntäminen muuttuu tai vaikeutuu. Tämä johtuu siitä, että käytettävästä teknologiasta riippumatta kaikki prosessit, joissa käytetään henkilötietoja, ovat yleisen tietosuoja-asetuksen alaisia. Asiakkaisiin liittyvät henkilötiedot ovat tärkeä tiedonlähde data-analytiikassa (Runkler 2012), joten on tarpeellista tutkia data-analytiikkaa asetuksen näkökulmasta.

1.1 Tutkimuskysymykset ja rajaus

Tavoitteena tutkimuksessa on kartoittaa Euroopan unionin yleisen tietosuoja-asetuksen vaikutuksen laajuutta, vakavuutta ja mahdollisia vaadittuja toimenpiteitä data-analytiikan näkökulmasta.

Päätutkimuskysymys on:

- Miten organisaation suorittamaa data-analytiikkaa ja siihen liittyviä prosesseja pitää muuttaa, että se on yhteensopiva Euroopan unionin yleisen tietosuoja-asetuksen kanssa?

Päätutkimuskysymykseen saadaan vastaus vastaamalla ensin alatutkimuskysymyksiin:

- Mitä on data-analytiikka?
- Mitä prosesseja data-analytiikkaan liittyy?
- Mitä ovat henkilötiedot?
- Miten data-analytiikassa voidaan hyödyntää henkilötietoja?
- Miten Euroopan unionin yleinen tietosuoja-asetus vaikuttaa henkilötietojen käsittelyyn?
- Miten Euroopan unionin yleinen tietosuoja-asetus vaikuttaa data-analytiikkaan?

Koska data-analytiikka on aiheena laaja, rajoitan tutkimustani koskemaan data-analytiikkaa vain niiltä osin, joihin asetuksella on selkeä vaikutus. Koska asetus käsittelee yksinomaan henkilötietoja, voidaan henkilötietoja käsittelemätön data-analytiikka jättää pois tutkimuksesta.

Kaikki muut syyt saada käsitellä henkilötietoja, paitsi henkilön oma suostumus, mielletään tässä tutkimuksessa yllä oleviksi poikkeuksiksi, koska muiden syiden ei nähdä vaikuttavan liiketoimintalähtöiseen data-analytiikkaan merkittävästi.

Yleisen tietosuoja-asetuksen poikkeuksia käsitellään vain, jos niillä on erittäin merkittävä rooli tutkimuksen kannalta. Esimerkiksi asetuksessa on määritelty, että rekisterinpitäjä saa käsitellä henkilötietoja sille kuuluvan julkisen vallan käyttämiseksi (Yleinen tietosuoja-asetus 2016). Vaikka kyseiseen julkisen vallan käyttöön voi liittyä data-analytiikkaa, ei se ole olennaista tämän tutkimuksen kannalta.

Yleistä tietosuoja-asetusta ei käsitellä Euroopan unionin ulkopuolisten toimijoiden osalta tutkimuksen selkeyttämiseksi.

1.2 Tutkimuksen rakenne

Tässä tutkimuksessa luvussa 2 esitellään tutkimusaineisto ja tutkimuksessa käytetyt menetelmät. Luvussa 3 esitellään ensin yleinen tietosuoja-asetus, jonka jälkeen käsitellään syvemmin tutkimuksen kannalta olennaisimpia yleisen tietosuoja-asetuksen säädöksiä. Luvussa 4 käsitellään data-analytiikkaa soveltuvin osin ja luvussa 5 yhdistetään kappaleet 3 ja 4 muotoon, yleisen tietosuoja-asetuksen periaatteiden näkökulmasta. Luvussa 5 käsitellään erityisesti ongelmia, mitä nykyisissä data-analytiikan menetelmissä ja siihen liittyvissä prosesseissa on. Luvussa 6 pohditaan mahdollisia vastauksia siihen, miten data-analytiikkaa ja siihen liittyviä prosesseja pitää suorittaa, jotta se olisi yhteensopiva yleisen tietosuoja-asetuksen kanssa. Luvussa 7 esitellään tulokset, arvioidaan niitä ja esitetään mahdollisia jatkotutkimuskohteita.

2. TUTKIMUSAINEISTO JA -MENETELMÄT

Tämä tutkimus suoritetaan kirjallisuustutkimuksena, jossa lähdemateriaalina käytetään verkkopohjaisista tietokannoista löydettyjä kirjoja, sanakirjoja ja artikkeleita.

2.1 Tutkimusmenetelmä

Fink, E. 2014. Conduction research literature reviews From the Internet to Paper. University of California. Los Angeles.

Tässä tutkimuksessa hyödynnetään Finkin kirjallisuuskatsauksen prosessimallia, jotta varmistutaan, että tutkimuksen lähestymistapa on systemaattinen ja toistettavissa oleva. Finkin (2005) mukaan kirjallisuuskatsauksen voi jakaa seitsemään tehtävään:

1. Tutkimuskysymysten valinta
2. Tietokantojen valinta
3. Hakutermien valinta
4. Hakutulosten seulonta käytännöllisillä seuloilla (esim. Ajanjakso, kieli, vertaisarvioidut)
5. Hakutulosten seulonta metodologisilla seuloilla (esim. Subjektiivinen arvio tieteellisestä laadusta)
6. Kirjallisuuskatsauksen kirjoittaminen
7. Tulosten syntetisointi.

Tehtävä 1, tutkimuskysymysten valinta, on esitetty tutkimuskysymyksineen luvussa 1.1 Tutkimuskysymykset ja rajaus.

Tutkimuksessa käytetään lähdemateriaalin etsimisessä tietokantoina Andoria ja Scopusta. Andorin vahvuus lähdemateriaalin etsimisessä on sen laajuus, sillä se etsii aineistoja kaikista Tampereen Teknillisen Yliopiston kirjastoon hankituista aineistoista. Scopus on hyödyllinen käyttäessä hakutermejä, joilla Andor palauttaa paljon epäoleellisia aineistoja.

Taulukossa 1 on esitelty tutkimuksen tiedonhaussa käytetyt hakutermit. Kaikkien hakutermien osumiin ei vaikuttanut, vaikka rajoitin hakua kattamaan vain vuonna 2015 julkaistut tai uudemmat julkaisut. Tämä johtuu todennäköisesti siitä, että yleinen tietosuoja-asetus on tullut julkiseksi ja astunut voimaan vasta vuonna 2016.

Taulukko 1. *Hakutermeillä saadut tulokset eri tietokannoista*

Hakutermi	**Lisäseula**	**Aineistojen määrä (Andor)**	**Aineistojen määrä (Scopus)**
”gdpr” OR ”general data protection regulation”		5853	217
("data analysis" OR "data analytics" OR "data mining")		2212366	392943
("data analysis" OR "data analytics" OR "data mining")	Vuosi >= 2015	555841	88733
("data analysis" OR "data analytics" OR "data mining")	Vuosi >= 2015, Aihe-alue = “computer science”	24239	36339
("gdpr" OR "general data protection regulation") AND ("data analysis" OR "data analytics" OR "data mining")		469	11

3. EUROOPAN UNIONIN YLEINEN TIETOSUOJA-ASETUS

EU:n yleinen tietosuoja-asetus astui voimaan toukokuussa 2016 ja sitä on sovellettava EU:n jäsenvaltioissa kahden vuoden siirtymäajan jälkeen toukokuussa vuonna 2018. Asetus korvaa vanhan, vuoden 1995 direktiivin 95/46/EC. (Nyrén et al. 2014; Yleinen tietosuoja-asetus 2016) Yleinen tietosuoja-asetus eroaa aiemmasta direktiivistä henkilötietojen määritelmän, rekisterinpitäjän velvollisuuksien, rekisteröidyn oikeuksien, sekä täytäntöönpanon käytäntöjen kannalta (van der Sloot 2014).

Eräs suurimmista yleisen tietosuoja-asetuksen merkittävyyteen liittyvistä seikoista on rikkomisista annettavat sanktiot. Sakkojen suuruus vaihtelee erittäin monen seikan vuorovaikutuksesta. Esimerkiksi sakon suuruuteen vaikuttaa: rikkomisen luonne, laajuus, vakavuus, kesto ja tahallisuus, toimet jotka on tehty vahinkojen pienentämiseksi ja tapa, jolla rikkomus on tullut valvontaviranomaisen tietoon. (Lähde) On siis selvää, että omalla toiminnalla voi vaikuttaa saataviin sakkoihin pienentävästi esimerkiksi ilmoittamalla rikkomisesta välittömästi valvontaviranomaiselle. Suurin mahdollinen sakko on 20 miljoonan euron tai viime vuoden liikevaihdosta neljän prosentin suuruinen. (Yleinen tietosuoja-asetus 2016; ITGP Privacy Team 2016) Maksimimäärä sakoissa määräytyy sen mukaan, kumpi edellä mainituista sakoista on suurempi. (Yleinen tietosuoja-asetus 2016; ITGP Privacy Team 2016)

3.1 Henkilötietojen määritelmä asetuksessa

Henkilötietojen määritelmä on laajentunut direktiivistä 95/46/EC yleiseen tietosuoja-asetukseen (Rumbold & Pierscionek 2017). Asetus ottaa huomioon laajemmin tunnistetietoja ja tekijöitä, jotka saattavat suoraan tai epäsuorasti johtaa henkilön (rekisteröity) tunnistamiseen (van der Sloot 2014, s. 311). Tunnistetiedot ovat tietoja, joita pidetään erityisesti henkilötietoina. Yleisessä tietosuoja-asetuksessa rekisteröidyn erityisen merkittäviksi tunnistetiedoiksi luetellaan nimi, henkilötunnus, sijaintitieto ja verkkotunnistetieto, kuten esimerkiksi IP-osoite. (Yleinen tietosuoja-asetus 2016) Direktiivissä 95/46/EC tunnistetiedoksi lueteltiin vain henkilötunnus (van der Sloot 2014, s. 311).

Tunnistetietojen lisäksi yleisessä tietosuoja-asetuksessa mainitaan ryhmä tekijöitä, jotka myös vaikuttavat siihen, onko kyseessä henkilötiedoksi luokiteltava tieto vai ei. Näitä tekijöitä on yleisessä tietosuoja-asetuksessa monipuolistettu. Yleisessä tietosuoja-asetuksessa mainitut tekijät ovat henkilölle tunnusomaiset fyysiset, fysiologiset, geneettiset, psyykkiset, taloudelliset, kulttuurilliset ja sosiaaliset tekijät, joiden perusteella on mahdollista tunnistaa henkilö suorasti tai epäsuorasti (Yleinen tietosuoja-asetus 2016). Käytännössä tämä tarkoittaa sitä, että vaikka henkilöistä ei olisi kerätty rekisteriin esimerkiksi

nimeä, mutta heistä oltaisiin kerätty huomattavan paljon muita tietoja, voidaan rekisteri laskea henkilötietorekisteriksi.

3.2 Rekisteröidyn oikeudet asetuksessa

Yleisessä tietosuoja-asetuksessa (2016) rekisteröidyn oikeudet ovat:

- Pääsy henkilötietoihin
- Oikeus tietojen oikaisemiseen
- Oikeus tietojen poistamiseen (”oikeus tulla unohdetuksi”)
- Oikeus tietojen siirtämiseen
- Vastustamisoikeus
- Oikeus olla joutumatta automatisoidun päätöksenteon kohteeksi, mukaan lukien profilointi
- Muut oikeudet

”Pääsy henkilötietoihin” viittaa yleisen tietosuoja-asetuksen 15 artiklaan. 15 artiklassa määritellään tiedot, joihin rekisteröidyllä on oikeus saada pääsy tai jotka on pyydettäessä toimitettava rekisteröidylle. Artiklassa määritellään, että rekisteröidyllä on oikeus saada tieto seuraavista asioista: käsitelläänkö hänen henkilötietojansa vai ei, henkilötietojen käsittelyn tarkoitukset, käytetyt henkilötietoryhmät, tahot, joille tietoja ollaan luovutettu tai saatetaan luovuttaa, henkilötietojen suunniteltu säilytysaika, henkilötietojen alkuperä, automaattisen päätöksenteon olemassaolo ja logiikka. (Yleinen tietosuoja-asetus 2016)

Seuraavat oikeudet, ”Oikeus tietojen oikaisemiseen”, ”Oikeus tietojen poistamiseen” ja ”Oikeus tietojen siirtämiseen”, ovat lähes itsestään selviä. Oikaiseminen tarkoittaa virheellisen tiedon muuttamista, poistaminen tarkoittaa tietojen hävittämistä ja siirtäminen tietojen siirtämistä rekisteröidyn haluamaan paikkaan. Ottaen huomioon tämän tutkimuksen laajuuden ja rajauksen, ei niiden tarkempaa sisältöä ole tarpeellista tarkastella. Oikeuksien sisältö löytyy kokonaisuudessaan yleisen tietosuoja-asetuksen III luvun 3 jaksossa (Yleinen tietosuoja-asetus 2016).

”Vastustamisoikeus” viittaa yleisen tietosuoja-asetuksen 21 artiklaan. Artiklassa eritellään, millaisissa tilanteissa rekisteröidyllä on oikeus vastustaa henkilötietojensa käsittelyä.

”Oikeus olla joutumatta automatisoidun päätöksenteon kohteeksi, mukaan lukien profilointi” viittaa yleisen tietosuoja-asetuksen 22 artiklaan. Tästä oikeudesta esimerkki on, että rekisteröidylle ei saa tarjota henkilötietojen perusteella eri tuotteita tai hintoja kuin muille ilman hänen lupaansa (Zuiderveen Borgesius & Poort 2017).

Yleisen tietosuoja-asetuksen III luvun 5 jaksossa käsitellään tilanteita, joissa rekisteröidyn oikeudet eivät päde. Nämä tilanteet ovat tämän tutkimuksen kannalta epäolennaisia.

3.3 Asetuksen muut säännökset

Aiemmin lueteltujen määrittelyjen lisäksi yleisessä tietosuoja-asetuksessa on säädetty paljon muitakin asioita. Luonnollisesti kaikkiin rekisteröidyn oikeuksiin liittyy säädös siitä, että rekisterinpitäjällä on velvollisuus toimia oikeuksien mukaisesti ja siitä, että rekisterinpitäjän pitää parhaimman mukaansa noudattaa asetusta (Yleinen tietosuoja-asetus 2016).

Yleisen tietosuoja-asetuksen mukaan rekisterinpitäjän on mahdollisuuksien mukaan pyrittävä sisäänrakennettuun ja oletusarvoiseen tietosuojaan. Sisäänrakennetulla tietosuojalla tarkoitetaan sitä, että rekisterinpitäjän on toteutettava tehokkaasti tietosuojaperiaatteita henkilötietojen käsittelytapojen määrittämisessä ja itse henkilötietojen käsittelyn yhteydessä. Rekisterinpitäjän on siten toteutettava vaaditut tekniset ja organisatoriset toimenpiteet tietosuojan parantamiseksi. Esimerkiksi tietojen minimointi ja pseudonymisointi ovat tällaisia toimenpiteitä. Pseudonymisointi tarkoittaa henkilötietojen muuttamista niin, ettei niitä voida yhdistää luonnolliseen henkilöön ilman ulkopuolista tietoa. Pseudonymisointi voi tapahtua esimerkiksi siirtämällä erityisen merkittävät henkilötiedot erilliseen paikkaan, jonne pääsyä tulee rajata. Oletusarvoinen tietosuoja tarkoittaa tietotarpeiden määrittelyä, jonka avulla pystytään määrittelemään henkilötiedoille kohtuullinen määrä, käsittelyn laajuus, säilytysaika ja saatavilla olo. Etenkin se on tärkeää, että henkilötietoja julkisteta ihmisille, keiden ei pidä niitä nähdä. (Yleinen tietosuoja-asetus 2016)

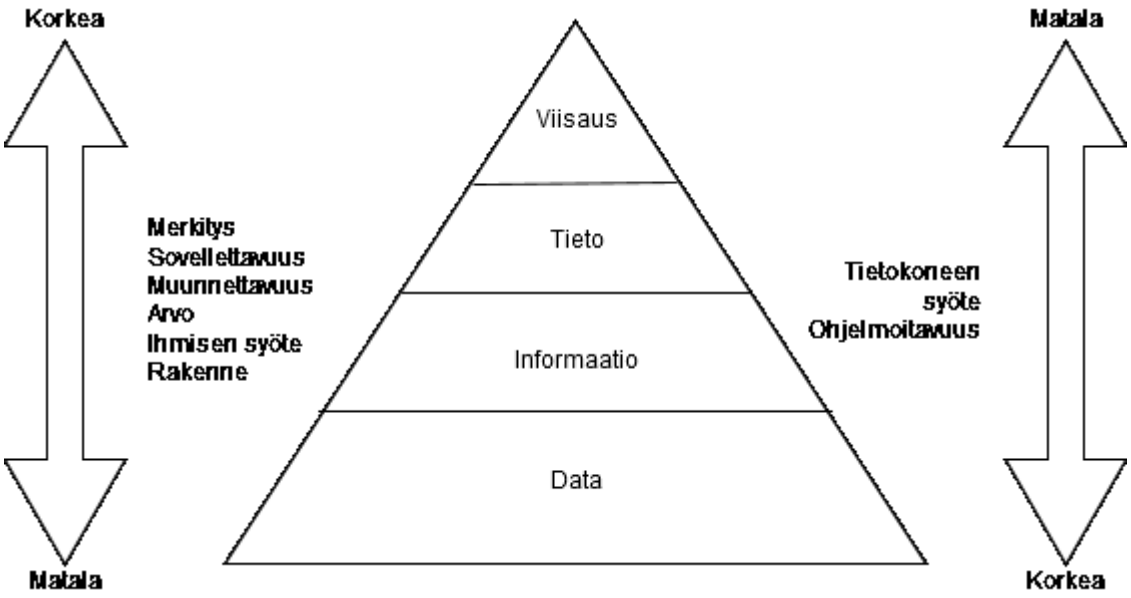
Yleisessä tietosuoja-asetuksessa määritellään henkilötietojen käsittelijä, joka on taho, joka käsittelee henkilötietoja rekisterinpitäjän puolesta. Myös henkilötietojen käsittelijän on toteutettava riittävät suojatoimet henkilötietojen suojaamiseksi. (Yleinen tietosuoja-asetus 2016)

4. DATA-ANALYTIIKKA

Data-analytiikka on merkityksellisten rakenteiden etsimistä ja löytämistä suuresta määrästä dataa ja niiden hyödyntämistä tulevaisuuden ennakoimisessa erityisesti liiketoimintakontekstissa (Data analytics 2017). Useimmiten data-analytiikka perustuu tilastoihin, jakaumiin ja todennäköisyyksiin (Piegorsch 2015).

4.1 Tiedon ja datan välinen yhteys

Jotta datasta saataisiin analytiikan avulla liiketoiminnassa hyödyntävää tietoa, on datalla ja tiedolla oltava yhteys. Zelenyn (2005) mukaan data on strukturoimatonta informaatiota, jota puolestaan käytetään tiedon luomisprosesseissa ”ainesosana”. Kuvassa 2 on esitelty data-informaatio-tieto-tietämys -hierarkia. Kuvassa on valittu yhteensä 8 mittaria, jotka vaikuttavat tiedon tason määrittämiseen.



Kuva 2. *Data-informaatio-tieto-tietämys -hierarkia (mukaillen Rowley 2007 s. 167)*

Data-analytiikassa siis otetaan syötteenä koneluettavaa dataa ja käsitellään sitä koneellisin ja kognitiivisin menetelmin kohti tietoa ja viisautta. Tämä on erittäin tärkeää, koska tiedolla ja tietämyksellä on merkittävästi enemmän merkitystä esimerkiksi liiketoiminnassa.

4.2 Data-analytiikan hyödyntämiskohteet

Sosiaalista mediaa hyödyntämällä voidaan data-analytiikan avulla kartoittaa asiakkaiden profiileja, jonka jälkeen voidaan parantaa omaa tuotetta ja markkinointia kohtaamaan asiakkaiden tarpeet ja kiinnostuksenkohteet. Esimerkiksi tietokonevalmistaja Dell on tehnyt näin hyödyntämällä tietoa LinkedIn-palvelusta. (Larose & Larose 2015)

Salehan & Kim (2016) esittelevät tutkimuksessaan menetelmän, jonka avulla pystyy analysoimaan verkkotuotteisiin liittyvien arvostelujen ja asiakkaan kokeman mielenkiinnon suhdetta. Tutkimuksessa selvisi esimerkiksi, että pitkillä otsikoilla oli negatiivinen vaikutus asiakkaan kiinnostukseen, kun taas leipätekstin pituudella vaikutus oli positiivinen.

Data-analytiikan avulla voidaan personoida palveluita. Esimerkiksi web-palvelussa voidaan näyttää käyttäjille eri mainoksia perustuen heidän metadataansa, kuten evästeisiin, paikkatietoihin ja verkko-ostoskäyttäytymiseen. (Hofman et al. 2017) Useat suuret web-yritykset, kuten Google, Facebook ja Amazon, hyödyntävät käyttäjältä kerättyä dataa palvelun personoimiseen (Garcia-Rivadulla 2016).

4.3 Data-analytiikan tietolähteet ja -varastot

Data-analytiikassa voidaan käyttää tietolähteinä organisaation sisäisiä tietokantoja ja ulkoisia rajapintoja. Tietokannat voivat olla esimerkiksi muodoissa: JSON, CSV, SQL, Excel tai XML (Uddin & Lee 2017).

```
sukunimi,etunimi,sähköposti,puhelinnumero
Meikäläinen,Matti,matti.meikalainen@example.com,0701234567
Leidi,Liisa,liisa@example.com,0701234568
Teekkari,Teemu,teemu.teekkari@example.com,0701234569
```

Kuva 3. *Kuva 1 Esimerkki CSV-tiedoston sisällöstä*

Kuvassa 3 on esitelty yksinkertaisen CSV-tiedoston sisältö, joka sisältää henkilötietoja. CSV on lyhenne sanoista Comma Separated Values. Kuvasta 3 on nähtävissä, että tiedoston rivit on eroteltu rivinvaihdolla ja sarakkeet on eroteltu pilkulla.

4.4 Data-analytiikan menetelmät

Tässä luvussa esitellään data-analytiikan yleisesti käytettyjä menetelmiä. Menetelmät on jaettu yksinkertaisiin ja monimutkaisiin menetelmiin menetelmien oletetun monimutkaisuuden mukaan.

4.4.1 Yksinkertaiset menetelmät

Yksinkertaisimmillaan data-analytiikalla kuvaillaan joukkoa alkioita pelkistetyillä tunnusluvuilla, jotka kuvaavat koko joukkoa. Esimerkiksi summa, mediaani, keskiarvo ja moodi ovat tällaisia menetelmiä. (Piegorsch 2015) Nämä menetelmät ovat hyviä tilanteissa, joissa alkuperäinen tietojoukko on liian suuri ihmisen arvioitavaksi manuaalisesti.

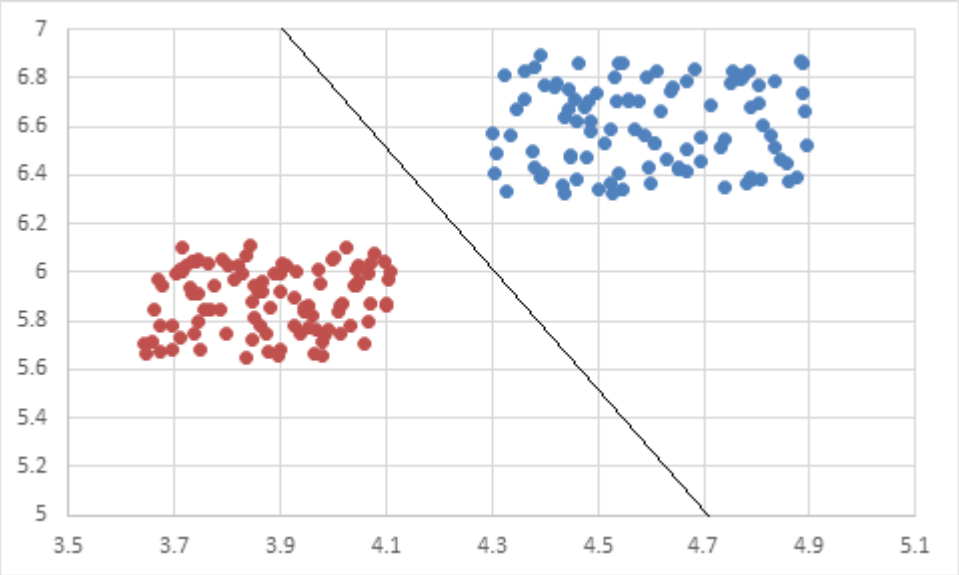
Yllä olevat menetelmät eivät sovellu tilanteisiin, joissa etsitään tietojoukon äärimmäisiä alkioita. Tähän tarkoitukseen minimi ja maksimi sopivat hyvin. (Piegorsch 2015)

Data-analytiikaksi voidaan laskea myös arkiset toimenpiteet, kuten kaavioiden piirtäminen ja muu datan visualisointi. Yleisesti käytetyt visualisointimenetelmät, kuten esimerkiksi pylväs- ja hajontakaaviot perustuvat yksinkertaisiin matemaattisiin menetelmiin, kuten esimerkiksi erilaisten alkioiden ryhmittelyyn tietyn attribuutin mukaan.

4.4.2 Edistyneet menetelmät

Lineaarinen regressio on monille tuttu menetelmä taulukkokäsittelyohjelmistojen sovitesuorasta. Piegorschin (2015) mukaan lineaarisella regressiolla saadaan approksimoitu suora, joka kuvastaa kahden muuttujan välistä suhdetta mahdollisimman hyvin. Lineaarisessa regressiossa käytetään minimoidun neliösumman menetelmää, jotta paljon poikkeavat arvot saavat vähemmän painoarvoa. Lineaarisen regression tuloksien oikeellisuutta voidaan tarkastella residuaalianalyysin avulla. Residuaalianalyysissä verrataan alkioiden etäisyyksien (sovitesuorasta) suhteita. (Piegorsch 2015)

Siinä missä esimerkiksi keskiarvo kuvaa tietojoukon yhtä muuttujaa kerrallaan, voidaan korrelaatiolla selvittää tietojoukon muuttujien korrelaatiota (Piegorsch 2015). Korrelaatiolla tarkoitetaan muuttujien välistä ”jos x, niin y”-suhdetta.



Kuva 4. *Keinotekoinen havainnollisus kahden muuttujan avulla suoritetusta klusteroinnista, jonka avulla ollaan huomattu kaksi selvää ryhmää*

Eräs data-analytiikan menetelmistä on klusterointi. Klusteroinnin avulla jaetaan alkioita ryhmiin, jotka ovat ominaisuuksiltaan samankaltaisia. (Piegorsch 2015) Kuvassa 4 on esitetty esimerkki keinotekoisesti tuotetusta kaksiulotteisesta tietojoukosta. Kuvasta ilmenee, että tietojoukosta on eroteltavissa kaksi selvää erinäistä ryhmää.

5. EUROOPAN UNIONIN YLEISEN TIETOSUOJA-ASETUKSEN PERIAATTEIDEN TOTEUTUMINEN DATA-ANALYTIIKASSA

Yleinen tietosuoja-asetus voidaan jakaa kuuteen yksityisyyden periaatteeseen. Niitä ovat:

1. Lainmukaisuus, kohtuullisuus, läpinäkyvyys
2. Käyttötarkoitussidonnaisuus
3. Tietojen minimointi
4. Täsmällisyys
5. Säilytyksen rajoittaminen
6. Eheys ja luottamuksellisuus. (ITGP Privacy Team 2016; Yleinen tietosuoja-asetus 2016)

Tässä luvussa käydään läpi yllä olevat periaatteet ja tarkastellaan, toteutuvatko ne jo nyt tyypillisessä data-analytiikassa.

5.1 Lainmukaisuus, kohtuullisuus, läpinäkyvyys

Henkilötietojen käsittely täyttää tämän periaatteen, jos henkilö on antanut suostumuksensa siihen, tiedon käsittely vastaa sitä, mihin henkilö on antanut suostumuksensa ja että henkilölle ollaan läpinäkyviä tiedon käsittelyn prosesseista. (ITGP Privacy Team 2016) Data-analytiikassa tämä saattaa olla ongelma, sillä usein käytetään liiketoiminnan sivutuotteena saatuja tietoja.

5.2 Käyttötarkoitussidonnaisuus

Käyttötarkoitussidonnaisuus liittyy edellisen periaatteen ”kohtuullisuus” ja ”lainmukaisuus” -kohtiin, sillä sillä tarkoitetaan sitä, että henkilötietoja saa käyttää vain siihen tarkoitukseen, johon henkilö on antanut luvan. Niin kuin luvussa 5.1, ongelmat liittyvät liiketoiminnan sivutuotteena saadun tiedon uudelleenkäyttämiseen.

5.3 Tietojen minimointi

Yleisen tietosuoja-asetuksen (2016) mukaan käsiteltävien tietojen on koostuttava vain tarkoitukseen tarpeellisista tietoryhmistä. Esimerkiksi toimistotyöhön hakevalta ei tule hankkia kyselyllä tietoa erityisistä terveyteen liittyvistä asioista, elleivät ne ole työn kannalta merkittäviä (ITGP Privacy Team 2016). Koneoppimisen menetelmissä voidaan hyötyä siitä, että käsitellään ihmiselle merkityksettömän oloista tietoa. On epäselvää, miten

yleisen tietosuoja-asetuksen näkökulmasta koneoppimisalgoritmeihin suhtaudutaan tiedon minimoinnin näkökulmasta.

5.4 Täsmällisyys

Tietojen täsmällisyys liittyy rekisteröidyn oikeuteen oikaista häntä koskevat väärät tiedot. Täsmällisyys siis tarkoittaa sitä, että henkilötiedot ovat rekisterinpitäjän hallussa olevan tiedon mukaisesti oikein. (ITGP Privacy Team 2016) Täsmällisyyden periaate ei aseta haasteita data-analytiikalle – päinvastoin. Toteuttamalla rekisteröidylle mahdollisuuden muuttaa tietojaan oikeiksi ja ajantasaisiksi, varmistetaan myös data-analytiikassa hyödynnettävän tiedon oikeellisuus.

5.5 Säilytyksen rajoittaminen

Säilytyksen rajoittaminen viittaa sekä säilytyksen ajalliseen rajoittamiseen että säilytyksenaikaisen saatavuuden rajoittamiseen. Ajallisesti säilytystä pitää pyrkiä rajoittamaan niin, että kun tieto ei ole enää hyödyllistä alkuperäiseen tarkoitukseensa, se tulee poistaa. Säilytyksenaikaista saatavuutta pitää rajoittaa esimerkiksi salauksella, anonymisoimalla tieto tai pseudonymisoimalla tieto. Myös eri tietoryhmien säilyttäminen eri paikoissa on tapa rajoittaa tiedon saatavuutta säilytyksen aikana. Varmin tapa on tietysti poistaa tieto heti, kun mahdollista. (ITGP Privacy Team 2016)

On yleinen käytäntö, että kaikki tieto säilötään, koska ei voi ikinä tietää milloin sitä tarvitaan. Data-analytiikassa tiedon ”varmuuden vuoksi” -säilöminen ei ole yleisen tietosuoja-asetuksen jälkeen helppoa.

5.6 Eheys ja luottamuksellisuus

Eheydellä ja luottamuksellisuudella viitataan siihen, että vain ne (henkilöt, tahot tai prosessit) millä on lupa käsitellä henkilötietoja, pystyvät käsittelemään niitä (ITGP Privacy Team 2016). Tämä saattaa vaikeuttaa data-analytiikkaa, koska tietojen näyttäminen esimerkiksi kollegalle on kiellettyä, ellei toisin ole sovittu.

6. YLEISEN TIETOSUOJA-ASETUKSEN MUKAINEN DATA-ANALYTIIKKA

Yleisen tietosuoja-asetuksen yhteensovittamiseen omiin liiketoimintaprosesseihin ei ole vain yhtä oikeaa ratkaisua, koska asetuksen tulkinta on vielä epäselvä. Esimerkiksi EU:n henkilötietodirektiiviin 95/46/EC on tullut tulkintalinjauksia vielä yli 20 vuotta direktiivin säätämisen jälkeenkin (Talus et al. 2017).

6.1 Organisaation laajuiset käytännöt

Hyvä geneerinen ohje on luoda viitekehys, jonka mukaan muutokset yleisen tietosuoja-asetuksen noudattamiseksi tehdään. Tämä tarkoittaa sitä, että määritellään koko organisaation laajuiset ohjeet ja käytännöt henkilötietojen käsittelylle. Tämä saattaa organisaatiosta riippuen johtaa suuriinkin muutoksiin. (ITGP Privacy Team 2016) Viitekehyksellä pyritään määrittämään yleisesti hyväksi koettuja käytäntöjä, joita kannattaa seurata (Tankard 2016). Ennen data-analytiikkakäytäntöjen miettimistä, on tärkeää muodostaa koko organisaation väliset säännöt ja käytännöt henkilötietojen suojelemiseksi. Tämä pitää tehdä riskiperusteista lähestymistapaa käyttäen, eli arvioida henkilötiettojen käsittelyyn liittyvät riskit ja tehdä toimenpiteet, joilla minimoidaan riskit (Talus et al. 2017).

Yleinen tietosuoja-asetus on tehty erityisesti kunnioittamaan yksityisen henkilön oikeuksia, eikä käytettävällä teknologialla ole vaikutusta asetuksen sovellettavuuteen (Yleinen tietosuoja-asetus 2016). Näiden syiden takia on tärkeää implementoida sisäänrakennettu ja oletusarvoinen tietosuoja kaikkiin organisaation prosesseihin (Tankard 2016).

6.2 Henkilötietojen kerääminen ja säilöminen

Ennen data-analytiikan suorittamista on kiinnitettävä erityistä huomiota siihen, että henkilötiedot on luvallisesti kerätty. Tämä tarkoittaa käytännössä sitä, että jokaisen henkilön henkilötietojen käsittelemiseen on saatu lupa ja rekisteröityä on informoitu henkilötietojen käsittelyyn liittyvän myös data-analytiikkaa (ITGP Privacy Team 2016). Ei ole määritelty, kuinka tarkasti henkilötietojen käsittelyyn liittyvät menetelmät on kerrottava rekisteröidylle. On mahdollista, että informointi data-analytiikasta yleisesti on riittävää. Toinen mahdollisuus on se, että rekisteröityä on informoitava yksityiskohtaisesti data-analytiikan tarkoituksista ja prosesseista.

Ellei kyse ole reaaliajassa tehdystä tietojen käsittelystä, on henkilötiedot säilöttävä tietovarastoon, josta niitä voidaan käyttää myöhemmin. Tietojen säilömisessä on erittäin olennaista tietoturvallisuus ja sen takia riskiperusteinen lähestymistapa on erittäin kriittistä.

Tietojen säilömisessä on pyrittävä joko anonymisoimaan peruuttamattomasti henkilötiedot tai vähintään pseudonymisoitava ne. Lisäksi tietojen bittitason salauksesta ja fyysisistä turvatoimista on pidettävä huolta fyysisen varkauden vahinkojen ja riskin vähentämiseksi. Nämä seikat saattavat vaikeuttaa data-analytiikkaa huomattavasti, mutta ne ovat rekisteröidyn oikeuksien suojaamiseksi välttämättömiä.

6.3 Käytön- ja pääsynvalvonta ja lokitiedot

Koska rekisteröidylle on pystyttävä informoimaan selkeässä muodossa, milloin, miksi ja kenen toimesta hänen henkilötietojaan on käsitelty, on jonkinasteinen käytönvalvonta ja käyttölokitus välttämätöntä. Tämä tarkoittaa sitä, että kuka tahansa organisaation jäsen ei voi hyödyntää analytiikkaosaamistaan käsittelemällä henkilötietoja, vaikka häneltä löytyisi tarvittava osaaminen henkilötietojen käsittelyyn. Mahdollisuutena ongelman ratkaisemiseksi on ilmaista selkeästi henkilötietojen käsittelyn lupapyynnössä, ketkä henkilötietoja tulee käsittelemään. Esimerkiksi käyttöehdoissa voi ilmaista, että henkilötietoja käytetään organisaation data-analytiikkaosaston toimesta, jolloin on mahdollista, että kaikki osaston työntekijät saavat luvan käyttää tietoja. Käytön- ja pääsynvalvontaan on kiinnitettävä silti huomiota myös riskiperusteisesta näkökulmasta, jolloin jokainen ylimääräinen henkilö, kenellä on oikeudet nähdä tai muokata henkilötietoja, kasvattaa riskiä henkilötietojen väärästä käytöstä.

Lokitiedoista on tultava selkeästi ilmi henkilötietojen käsittelemisen tiedot. Viestintäviraston (2016) mukaan käyttökelpoisessa lokitiedossa pitäisi olla vähintään:

- Aikaleima
- Tapahtuma
- Toimija
- Käyttöoikeus
- Tapahtumalähde
- Tapahtuman kohde
- Tapahtuman tila

Näin kattavan lokitiedon avulla voidaan vastata yleisen tietoturva-asetuksen tarpeisiin, mikäli lokitietoa kerätään jokaisesta merkittävästä tapahtumasta ja tiedon muuttamiseen ei ole (ainakaan ihannetilanteessa) kenelläkään muutosoikeuksia. Viestintävirasto (2016) kertoo, että 6-24 kuukautta on yleensä riittävä lokien säilytysajaksi.

6.4 Henkilötietojen analysoiminen

Mikäli kaikki muut henkilötietojen käsittelyyn liittyvät aspektit ovat organisaation laajuisesti toteutettuja, ei henkilötietojen analysoinnissa data-analytiikan menetelmin ole ylei-

sen tietosuoja-asetuksen näkökulmasta ongelmia. On kuitenkin noudatettava erityistä varovaisuutta siinä, mitä henkilötietoja käytetään osana analytiikkaa ja mitä ei. Syksyllä 2017 Terveyden ja hyvinvoinnin laitoksen julkaisemassa materiaalissa oli vahingossa mukana lähes 6000 henkilön nimi ja henkilötunnus johtuen inhimillisestä virheestä henkilötietojen käsittelyssä (THL). On mahdollista, että tämän tyylinen tietovuoto olisi voitu välttää minimoimalla käytettävät henkilötiedot.

7. YHTEENVETO

Euroopan yleisen tietosuoja-asetuksen vaikutukset henkilötietojen käsittelyyn ovat laajat. Tässä kirjallisuuskatsauksessa tutkittiin erityisesti asetuksen asettamia velvoitteita suhteessa data-analytiikkaan ja siihen liittyviin prosesseihin.

7.1 Tulokset

Kirjallisuuskatsauksessa huomattiin, että data-analytiikkaa ei voida yleisen tietosuoja-asetuksen näkökulmasta lähestyä yksittäisenä työkaluja, vaan on otettava huomioon myös kaikki siihen liittyvät prosessit. Näiden prosessien tekeminen yleisen tietosuoja-asetuksen kanssa yhteensopivaksi on koko organisaation laajuinen suuri projekti, jossa yhdenkään alaprojektin merkityksellisyyttä ei voi väheksyä. Jos henkilötietojen käsittelyssä yksikin prosessi on toteutettu huonosti, on henkilötietojen väärinkäyttö, tietovuoto tai tietomurto mahdollinen.

Data-analytiikan prosessit voidaan jakaa keräämiseen, säilömiseen ja prosessointiin. Jokaisella osa-alueella yhteisiä yleisen tietosuoja-asetuksen asettamia vaatimuksia ovat läpinäkyvä informointi rekisteröidylle, henkilötietojen turvaaminen, käyttötietojen lokitus ja lainmukaiset toimintatavat.

Hyvänä ohjesääntönä voidaan pitää, että henkilötietoja käsitellessä pitää aina toimia mahdollisimman avoimesti ja tehdä parhaansa henkilötietojen ja rekisteröidyn yksityisyyden turvaamiseksi. Kun tällainen toiminta yhdistetään organisaation selkeään, rekisteröityä suojaavaan tietoturvastrategiaan, on epätodennäköistä, että rikotaan yleistä tietosuoja-asetusta vakavasti. Tämän tutkimuksen nojalla voidaan sanoa, että yleisen tietosuoja-asetuksen tarkoitus on rekisteröityjen oikeuksien ja tietojen suojaamisen lisäksi lisätä läpinäkyvyyttä, avoimuutta ja suunnitelmallisuutta henkilötietojen käsittelyssä.

7.2 Tulosten arviointi

Tutkimuksen tuloksia voidaan pitää pääpiirteittäin onnistuneina. Kuitenkin yleisen tietosuoja-asetuksen ollessa vielä hyvin tulkinnanvarainen, on mahdollista, että sitä jatkossa tulkitaan eri tavalla, kuin miten tämän hetken kirjallisuudessa sitä tulkitaan.

Työn otsikon näkökulmasta tulokset ovat painottuneet hyvin paljon data-analytiikkaa tukeviin prosesseihin, eivätkä data-analytiikkaan työkaluna. Tämä johtuu mahdollisesti yleisen tietosuoja-asetuksen suunnitelmallisuutta vaativasta luonteesta, jota toteuttamalla päädytään siihen, että tietojen prosessointia ei tarvitse muuttaa, mikäli kaikki muu on tehty hyvin.

7.3 Jatkotutkimusmahdollisuudet

Yleisen tietosuoja-asetuksen mukaisesta data-analytiikasta on mahdollista tehdä jatkotutkimusta etenkin niillä osa-alueilla, jotka rajattiin tästä tutkimuksesta ulkopuolelle. Esimerkiksi Euroopan unionin ulkopuolisten tahojen näkökulma täydentäisi tätä tutkimusta erittäin hyvin.

Yleistä tietosuoja-asetusta koskien on mahdollista tehdä jatkossa tutkimusta etenkin sen vaikutuksista sen jälkeen, kun sitä on sovellettu Euroopan unionin jäsenvaltiossa tarpeeksi pitkän aikaa. Tällaisessa tutkimuksessa olisi myös mahdollista peilata silloista tietoa asetuksesta ja sen tulkinnasta vuosien 2014-2017 kirjallisuuteen, jolloin täyttä tietoa asetuksen tulkinnasta ei ole ollut saatavilla.

LÄHTEET

Data analytics (2017). Oxford University Press,

Garcia-Rivadulla, S. (2016). Personalization vs. privacy: An inevitable trade-off? IFLA Journal, Vol. 42(3), pp. 227-238.

Hofman, D., Duranti, L. & How, E. (2017). Trust in the Balance: Data Protection Laws as Tools for Privacy and Security in the Cloud, Algorithms, Vol. 10(2), pp. 47.

Jain, P., Sharma, P. & Jayaraman, L. (2014). Behind Every Good Decision : How Anyone Can Use Business Analytics to Turn Data into Profitable Insight, AMACOM, Saranac Lake,

Larose, D.T. & Larose, C.D. (2015). Data Mining and Predictive Analytics, Second; 2; 2nd ed. John Wiley & Sons Inc, US,

Nyrén, O., Stenbeck, M. & Grönberg, H. (2014). The European Parliament proposal for the new EU General Data Protection Regulation may severely restrict European epidemiological research, European journal of epidemiology, Vol. 29(4), pp. 227-230.

Piegorsch, W.W. (2015). Statistical Data Analytics : Foundations for Data Mining, Informatics, and Knowledge Discovery, 1st ed. John Wiley & Sons, Incorporated, New York,

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy, Journal of Information Science, Vol. 33(2), pp. 163-180.

Rumbold, J. & Pierscionek, B.K. (2017). A critique of the regulation of data science in healthcare research in the European Union, BMC MEDICAL ETHICS, Vol. 18

Runkler, T.A. (2012). Data Analytics: Models and Algorithms for Intelligent Data Analysis, 2012th ed. Vieweg+Teubner Verlag, Wiesbaden,

Salehan, M. & Kim, D.J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics, Decision Support Systems, Vol. 81 pp. 30-40.

Talus, A., Autio, E., Hänninen, A., Pihamaa, H. & Kantonen, S. (2017). Miten valmistautua EU:n tietosuoja-asetukseen? Oikeusministeriö,

Tankard, C. (2016). What the GDPR means for businesses, Network Security, Vol. 2016(6), pp. 5-8.

ITGP Privacy Team. (2016). EU General Data Protection Regulation (GDPR) : An Implementation and Compliance Guide, ITGP, Ely,

THL Luottamuksellisia henkilötietoja ollut löydettävissä verkosta - Tiedote - THLhttp://www.thl.fi/fi/-/luottamuksellisia-henkilotietoja-ollut-loydettavissa-verkosta.

Uddin, M. & Lee, J. (2017). Proposing stochastic probability-based math model and algorithms utilizing social networking and academic data for good fit students prediction, Social Network Analysis and Mining, Vol. 7(1), pp. 1-21. https://link-springer-com.libproxy.tut.fi/article/10.1007/s13278-017-0448-z.

van der Sloot, B. (2014). Do data protection rules protect the individual and should they? An assessment of the proposed General Data Protection Regulation, International Data Privacy Law, Vol. 4(4), pp. 307-325.

Viestintävirasto [Teema] Loki on ylläpidon tärkein turvallisuustyökaluhttps://www.viestintavirasto.fi/kyberturvallisuus/tietoturvanyt/2016/03/ttn201603091742.html.

Yleinen tietosuoja-asetus (2016). 2016/679. Available: http://eur-lex.europa.eu/legal-content/FI/ALL/?uri=CELEX%3A32016R0679.

Zeleny, M. (2005). Human systems management: Integrating knowledge, management and systems, World Scientific Pub Co Pte, GB,

Zuiderveen Borgesius, F. & Poort, J. (2017). Online Price Discrimination and EU Data Privacy Law, Journal of Consumer Policy, Vol. 40(3), pp. 347-366.

Žliobaitė, I. & Custers, B. (2016). Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models, Artificial Intelligence and Law, Vol. 24(2), pp. 183-201.