TURUN YLIOPISTO

# TIIVISTELMÄ

| | |
|---|---|
| | Kandidaatintutkielma |
| x | Pro gradu -tutkielma |
| | Lisensiaatintutkielma |
| | Väitöskirja |

| | | | |
|---|---|---|---|
| Oppiaine | Yritysjuridiikka | Päivämäärä | 30.12.2019 |
| Tekijä(t) | Harri Paalanen | Sivumäärä | |
| Otsikko | Henkilötietojen suoja automaattisessa päätöksenteossa ja profiloinnissa | | |
| Ohjaaja(t) | Matti J. Sillanpää | | |


Tiivistelmä

Euroopan komission strategiaan kuuluva yhdennetty digitaalinen sisämarkkina viittaa datatalouteen, jonka raaka-aineena on muun muassa kuluttajista kerätyt tiedot – mukaan lukien henkilötiedot. Vuonna 2018 voimaan astunut yleinen tietosuoja-asetus GDPR selkiytti ja raamitti henkilötietojen käsittelyyn liittyviä teemoja sekä toi uusia velvollisuuksia rekisterinpitäjille ja uusia oikeuksia rekisteröidyille.

Henkilötietojen suoja on jatkossa perusoikeus. Yritykset ja organisaatiot tarvitsevat kuitenkin dataa päätöksenteon tueksi. Tutkin tällaisia datan perusteella tehtäviä automaattisesti tehtäviä päätöksiä etenkin tekoälyteknologioiden valossa ja käytän esimerkkinä koneoppimista. Mitä vaikutuksia yleinen tietosuoja-asetus tuo tähän asetelmaan?

Tutkimuskysymykseni liittyy automaattisesti tehtäviin päätöksiin, jotka perustuvat yrityksen keräämiin henkilötietoihin. Yritykset pyrkivät jatkuvasti automatisoimaan toimintojaan ja kehittämään uusia palveluita. Mitä vaatimuksia yleinen tietosuoja-asetus tuo tähän asetelmaan ja miten opetetun algoritmin tekemät päätökset suhteutuvat henkilötietojen suojaan?

Arvioin tutkimuksessa yleisen tietosuoja-asetuksen keskeisempien periaatteiden soveltamista nykyaikaiseen tietojenkäsittelytoimiin kuten koneoppimiseen.



| | |
|---|---|
| Asiasanat | koneoppiminen, gdpr, tietosuoja |
| Muita tietoja | |

TURUN YLIOPISTO
Kauppakorkeakoulu

# HENKILÖTIETOJEN SUOJA AUTOMAATTISESSA PÄÄTÖKSENTEOSSA JA PROFILOINNISSA

Yritysjuridiikan
pro gradu -tutkielma

Laatija(t):
Harri Paalanen

Ohjaajat:
Matti J. Sillanpää

30.12.2019
Turku

# Sisällysluettelo

## List of figures



## List of tables

# 1 JOHDANTO AUTOMAATTISEEN PÄÄTÖKSENTEKOON JA TIETOSUOJAAN

## 1.1 Alustus

Euroopan komission strategiaan kuuluu niin sanottu *digital single market* eli yhdennetyt digitaaliset sisämarkkinat. Tarkoituksena on avata digitaalisia mahdollisuuksia ihmisille ja yrityksille sekä korostaa Euroopan asemaa johtajana digitaalitaloudessa. Komission toimeksiantona toteutetun Data Market Studyn[1] mukaan datatalouden arvo on ollut vuonna 2018 jo noin 300 miljardia euroa, mikä on noin 12 prosentin kasvu edellisestä vuodesta. Datataloudessa datan toimittajien liikevaihto on kasvanut 12%, 77 miljardiin euroon vuonna 2018. Tutkimuksen mukaan myös data-ammattilaista on edelleen pulaa. On siis selvää, että datan merkitys kasvaa taloudessa yhä enemmän.

Vastaavasti kuitenkin kuluttajien asenteet datan käyttöä kohtaan ovat tietyllä tavalla ristiriitaiset. Eurobarometri 359 -tutkimuksen[2] mukaan 74% eurooppalaisista näkee henkilötietojen antamisen yhä isommaksi osaksi modernia elämää. Lisäksi 70 % eurooppalaisista on huolissaan siitä, että yritykset käyttävät annettuja henkilötietoja muihin tarkoituksiin, kuin ne on alun perin kerätty.

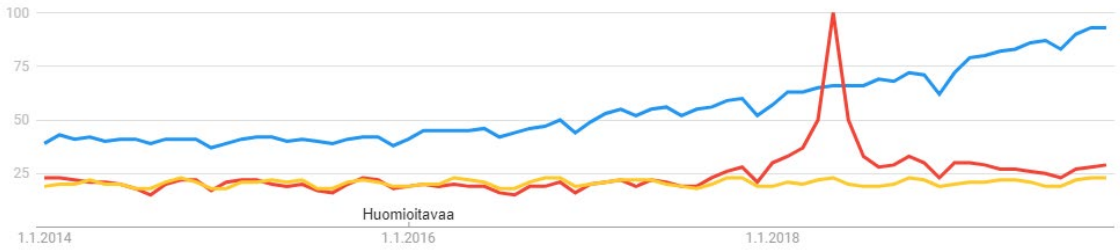
Tietojärjestelmien tulee mallintaa reaalimaailmaa riittävän hyvin, jotta ne voisivat olla älykkäämpiä. Tämän saavuttamiseksi tarvitaan valtavia määriä dataa, joka pitää varastoida strukturoidusti ja sähköisesti. Nykyisin tieto on varsin usein strukturoimatonta, mutta se tulisi voida jotenkin formalisoida ja kategorisoida, jotta eri järjestelmät voivat hyödyntää sitä. Kaiken strukturoimattoman tiedon strukturointi manuaalisesti ei ole kuitenkaan järkevää, joten monet tutkijat ovat keskittyneet oppimisalgoritmien tutkimukseen. Näitä voisi hyödyntää isojen datamäärien analysoinnissa. Vaikka isoja kehitysaskelia onkin jo tehty, vielä ei kuitenkaan voida puhua tekoälystä. [3]

Prediktiivisia eli ennustavia malleja käytetään useissa *business intelligence* tehtävissä. Mallit kehitetään esimerkiksi ennustamaan asiakaskäyttäytymistä tai vaikkapa petoksia historiadataan perustuen. Kriittinen tekijä tällaisten mallien kehittämisessä on sen datan laatu, jota käytetään mallin opettamiseen. Ennustavien mallien kehittämisessä datan laadulla tarkoitetaan esimerkiksi otoksen koostumusta, arvojen tarkkuutta sekä tuntemattomien arvojen määrää. [4]

[1] Micheletti & Pepato 2019, 10.
[2] Euroopan komissio 2011.
[3] Bengio 2009, 2.
[4] Saar-Tsechansky 2009.

Terminä *business intelligence* on tietyllä tavalla vakiintunut ja sitä käytetään useimmiten englannin kielisenä suomenkielisen *liiketoimintatiedon hallinta* termin sijaan. Sen sisältö kuitenkin vaihtelee jonkin verran ja sitä voidaan pitää tietynlaisena kattoterminä useammalle eri teknologialle ja toiminnalle. Business intelligencen määritelmää on tutkittu Suomessa avainsanojen esiintyvyyden ja kontekstin mukaan. Tutkimuksen lopputulos oli, että suurin aihealueiden prosentuaalisista osuuksista kaikissa liiketoimintatiedon hallinnan määritelmissä oli aihealueella ”päätöksenteko”. [5]

Kuvassa 1 on kuvattu aihealueeseen ”business intelligence” liittyvien hakujen määrää (sininen jana) koko maailmassa ajanjaksolla 1.1.2014–26.11.2019. Verrokkina sille on aihealueisiin "tietosuoja" punaisella janalla ja "tietoturva" keltaisella janalla liittyvien hakujen määrät samalla ajanjaksolla.



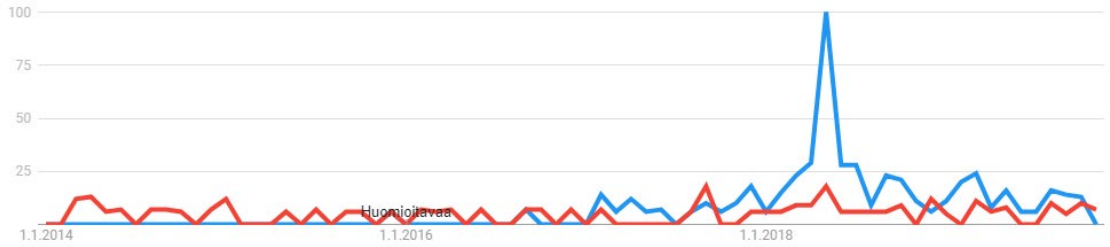

**Kuva 1. Aihealueen "business intelligence" (sininen jana) hakujen määrät. Haettu Google Trends -palvelusta 30.11.2019.**

Google Trends -palvelu paransi tietojen keräämistä 1.1.2016 alkaen, mikä on merkitty kuvaajiin ”Huomioitavaa” -huomiolla. Kuvaajien numerot esittävät haun suosiota valitulla ajanjaksolla ja alueella suhteutettuna kaavion suurimpaan arvoon. Asteikon arvo 100 on alue, jossa termi tai aihealue oli suosituin. Arvo 50 on alue, jossa hakuja tehtiin puolet vähemmän kuin 100 pisteen alueella. Vastaavasti 0 on alue, jolla ei ole hauista riittävästi tietoa.

Alkuvuodesta 2018 voidaan nähdä, että jopa koko maailmaa koskevassa aineistossa on selvä piikki aihealueeseen ”tietosuoja” liittyvissä hauissa. Samalla ajanjaksolla aihealueeseen ”business intelligence” liittyvien hakujen määrä on kasvanut tasaisesti ja siihen liittyvien hakujen määrä on ollut koko ajan suurempi kuin vertailuaihealueisiin. Huomionarvoista on myös, että ajanjakson lopulla 26.11.2019 aihealueisiin liittyvien hakujen määrien ero on suuri. Aihealueeseen ”business intelligence” liittyviä hakuja

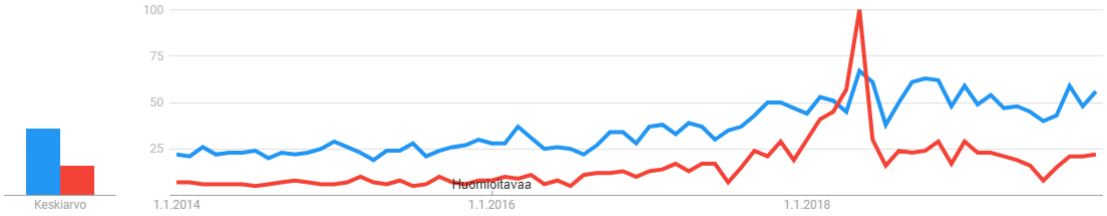
[5] Vartiainen 2014, 15-16. Tutkimuksessa käytettiin seuraavia hakutermejä: päätöksenteko, datan varastointi, analytiikka, ennustaminen, datan louhinta, OLAP (online analytical processing), liiketoiminnan analytiikka ja visualisointi. Vaikka tutkimus ei välttämättä anna täydellistä kuvaa määritelmän laajuudesta ja eri hakutermien hajonnasta, se mielestäni tarjoaa kuitenkin mielenkiintoisen ja tämän tutkielman kannalta relevantin näkökulman.

tehdään 75 – 100 pisteen alueella, kun aihealueisiin ”tietosuoja” ja ”tietoturva” liittyvien hakujen määrä on tasoittunut takaisin 25 pisteen alueelle.




**Kuva 2. Google -hakujen määrä aihealueissa "Yleinen tietosuoja-asetus" (sininen) ja "tietosuoja" (punainen). Haettu Google Trends -palvelusta 26.11.2019. Maantieteellisesti rajattu Suomeen.**

Google Trends -palvelun kautta voi tarkistella Google-hauissa käytettyjä hakusanoja ja hakumääriä. Kuvassa 2 on esitetty niiden hakujen määrät, joissa on aihealueena ollut ”tietosuoja” punaisella janalla ja verrokkina sinisellä janalla ”yleinen tietosuoja-asetus”. Siitä voidaan nähdä, että vaikka tietosuoja on ollut ajanjaksolla 1.1.2014–26.11.2019 melko tasaisesti mukana Google -hauissa, ”yleinen tietosuoja-asetus” on noussut kiinnostavaksi vasta vuoden 2018 aikana. GDPR tuli voimaan 25.5.2016 ja sitä on alettu soveltamaan EU:n jäsenvaltioissa 25.5.2018 alkaen.




**Kuva 3. Google hakujen määrät aihealueessa ”tekoäly” sinisellä janalla ja ”tietosuoja” punaisella janalla. Haettu Google Trends -palvelusta 30.11.2019. Maantieteellisesti rajattu Suomeen.**

Samalla ajanjaksolla esimerkiksi aihealue ”tekoäly” on esiintynyt yhä useammassa Google -haussa. Kuvasta 3 voidaan nähdä, että aihealueen ”tekoäly” haut ovat kasvaneet tasaisesti samalla ajanjaksolla. Aihealue ”tietosuoja” on esiintynyt useammassa Google -haussa vain juuri ennen yleisen tietosuoja-asetuksen soveltamisen alkamista.

## 1.2 Informaation epäsymmetria päätöksenteossa

Luotto- ja palvelumarkkinoille on tyypillistä informaation epäsymmetrisyys, mikä erottaa ne hyödykemarkkinoista. Epäsymmetrinen informaatio kuvaa nimenomaan sopimusosapuolten välistä informaatioeroa. [6] Kun sopimuksia tehdään puutteellisin tiedoin, niiden riskisyys kasvaa. Sopimuksen osapuolet harvoin tietävät toisistaan kaiken, mitä oikean luottopäätöksen tekemiseen tarvittaisiin. Vastaavasti hyödykemarkkinoilla on enemmän tietoa osapuolten välillä saatavilla, kun ostaja voi vaikkapa kokeilla fyysistä hyödykettä. Lisäksi kauppa on yleensä käteiskauppaa.

Haitallinen valikoituminen (*adverse selection)* on ennen sopimuksen tekoa (*ex ante)* syntyvä ongelma, joka on epäsymmetrinen informaation aiheuttama. Tällöin vaarana on, että luotollista myyntiä tapahtuu ja rahoitus kanavoituu riskipitoisempiin projekteihin ja turvalliset, varmatuottoisammat projektit jäävät ilman rahoitusta. [7]

Luottoriski kasvaa, jos asiakkaaksi valikoituu joidenkin sopimusten ehtojen vuoksi huonoja ja erityisen riskialttiita asiakkaita vähäriskisten asiakkaiden sijaan. Tällaisia ehtoja voivat olla esimerkiksi vakuusvaatimukset tai korkea korko. Korkean vakuuden hyväksyvä asiakas voi olla todennäköisemmin korkean riskin asiakas.

Epäsymmetrisesta informaatiosta sopimusosapuolen välillä voi aiheutua myös moraalikatoa (*moral hazard*), joka syntyy yleensä sopimuksenteon jälkeen (*post ante*). Moraalikato on yleensä myyjään tai luotonantajaan kohdistuva riski. Asuntolainoissa moraalikato voi aiheutua siitä, että velallinen voi käyttääkin saamansa lainan muuhun kuin asunnon hankkimiseen ja siten alentaa todennäköisyyttä lainan takaisinmaksulle.[8] Toisaalta myös pankkivalvontaviranomaisen ja pankin välillä moraalikato voi syntyä, kun pankki ottaa toiminnassaan tietoisesti liian suuria riskejä, koska talletusturva korvaa mahdolliset menetykset. [9]

## 1.3 Luottotietorekisterit päätöksenteon tukena

Luotonmyöntäjä ottaa kantaakseen luotonottajan maksukykyyn ja -halukkuuteen liittyvän riskin. Vastapuoli ei välttämättä maksa kokonaan tai oikea-aikaisesti saamaansa luottoa takaisin. Luotonantajille onkin jo varhain syntynyt tarve pyrkiä mahdollisuuksien mukaan etukäteen rajoittamaan riskejään esimerkiksi vakuuksilla ja luotonhakijoiden erilaisilla seulonnoilla. Lisäksi luotonantajat ovat hyödyntäneet myös ulkoisia tieto-

[6] Lehtiö 2004, 2; Stiglitz & Weiss 1981, ??.
[7] Anttila 1996, 29.
[8] Mishkin 1997, 36, 201, 210–211.
[9] Lehtiö 2004, 34. Esimerkki on vuodelta 2004, jolloin asuntolainaa oli mahdollista vielä käyttää eri tavalla. Nykyisin sääntely on erilainen.

lähteitä, kuten luottotietoraportteja. Etenkin automatisoiduissa, digitaalisissa luotonmyöntöprosesseissa luottotietojen arvioinnin merkitys riskien hallinnan välineenä on korostunut viime vuosina varsinkin kulutusluotoissa. [10]

Sopimusosapuolen tunnistaminen ja tämän luotettavuuden arviointia pidetäänkin tärkeänä aina, kun maksu ei tapahdu samanaikaisesti kuin tavara tai palvelu annetaan. Luottoriskien hallinta liittyy siis myös muuhunkin kuin puhtaasti luotolla tapahtuvaan myyntiin.[11] Luottoriskien minimoinnin näkökulmasta keskeinen vaihe onkin nimenomaan asiakasvalinta eli se vaihe, kun sopimus tehdään osapuolten välillä. Riskien hallitsemiseksi myyjä pyrkii keräämään tietoja eri lähteistä, joita ostajan antamien tietojen lisäksi voi olla muun muassa myyjän omat asiakasrekisterit, viranomaisten rekisterit sekä muut vastaavat julkiset rekisterit.[12]

Erilaisia tietolähteitä, joita voidaan käyttää luottokelpoisuuden arviointiin ovat pääasiassa luotonhakija itse, tulorekisteri, tilitiedot ja uusi EU-tasoinen AnaCredit -tietokanta. Velvollisuus tuntea asiakas on ollut pitkään luottoriskien hallinnan kulmakivi. Tätä samaa periaatetta on laajennettu ja periaatteellista merkitystä on korostettu entisestään rahanpesun estämistä koskevan sääntelyn yhteydessä. [13]

Luottotietorekisterit ovat olennainen osa niitä tietolähteitä, joita myyjä yleensä tarkastaa osana asiakasvalintaa. Suomessa luottotietorekisterit perustettiin vuonna 2007 voimaan tulleella luottotietolailla.[14] Luottotietorekisterejä on Euroopan ja Pohjoismaiden laajuisesti pääasiassa muutamia erilaisia. Luottotietorekisterin pitäjä voi olla yksityinen tai julkinen taho ja tallennettavat tiedot voivat olla luonteeltaan negatiivisia tai positiivisia. Esimerkiksi Suomen luottotietorekistereistä voidaan puhua negatiivisena luottotietorekisterinä, sillä rekisteriin tallennetaan vain negatiivisena pidettävät maksuhäiriötiedot. Positiivisen luottotietorekisteriin tallennetaan myös niin sanottuja positiivista maksukykyä indikoivia tietoja, kuten tulotietoja, ammatti, siviilisääty[15] sekä jo otetut lainat ja velat sekä toisinaan myös niiden maksutiedot. Positiivisella luottotiedolla ei ole kuitenkaan yksiselitteistä määritelmää. Positiivisen luottotiedon käsite ymmärretään negatiivisen luottotiedon vastakäsitteenä ja vakiintuneen näkemyksen mukaan positiivisilla luottotiedoilla tarkoitetaan muita henkilön taloudellisesta tilasta kertovia tietoja kuin maksuhäiriötietoja.[16]

Suomessa on ollut keskustelua vuosien varrella positiivisen luottotietorekisterin perustamisesta. Vuonna 2018 valmistuneen selvityksen mukaan positiivisten luottotietojen tarve on ilmeinen, koska luottojen tarjonta on kasvanut ja luotonanto on digitalisoitunut.

---

[10] Kontkanen 2018, 11.
[11] HE 241/2006 vp, 4.
[12] HE 241/2006 vp, 4.
[13] Kontkanen 2018, 36.
[14] Luottotietolaki (11.5.2007/527).
[15] Kts. esim. Riestra 2002. Avioero voi ennustaa jossain määrin kasvanutta velkaantumisriskiä.
[16] Kontkanen 2018, 9.

Selvityksessä ehdotetaan perustettavaksi muita kuin maksuhäiriötietoja koskeva positiivinen luottotietorekisteri vuoden 2019 toimintansa aloittavan tulorekisterin yhteyteen. [17]

Digitalisaation myötä monet yritykset ovat ottaneet käyttöön erilaisia automaattisia pisteytysjärjestelmiä. Järjestelmät käyttävät yleensä tilastotieteellisiä menetelmiä kohteena olevan henkilön arviointiin. Varsinkin kulutusluottojen myöntäminen on pitkälle automatisoitu tällaisiin järjestelmiin perustuen.[18] Tietosuoja-asetuksessa on säädetty, että rekisteröidyllä on oikeus olla joutumatta sellaisen päätöksen kohteeksi, joka perustuu pelkästään automaattisen tietojen käsittelyyn, kuten profilointiin, ja jolla on häntä koskevia oikeusvaikutuksia tai joka vaikuttaa häneen vastaavalla tavalla merkittävästi. Kieltoon on kuitenkin kolme poikkeustilannetta. Yksi näistä on, jos päätös on välttämätön rekisteröidyn ja rekisterinpitäjän välisen sopimuksen tekemiseksi tai sen täytäntöönpanoa varten. Poikkeus kieltoon tietyissä tilanteissa voidaan vahvistaa myös kansallisessa lainsäädännössä tai se voi perustua rekisteröidyn nimenomaiseen suostumukseen. [19]

## 1.4 Tekoälyn hyödyntäminen päätöksenteossa ja sen apuna

Soveltamalla tekoälyä liiketoiminnassa voidaan kehittää parempia liiketoimintaan liittyviä prosesseja ja toimintoja, luoda uusia lisäarvoa tuottavia palveluita datan avulla ja hyödyntää datasta saatavaa tietoa erilaisen päätöksenteon tukena.[20] Tekoälystä käytetään usein melko harhaanjohtavasti useita eri nimityksiä ja lyhenteitä. Useimmiten käytetään lyhennettä AI (*Artificial Intelligence*), joka on eräänlainen kattotermi usealle eri tekniikalle. Suomenkielisessä keskustelussa voidaan käyttää myös muita hiukan tarkempia termejä kuten keinoäly, tekoäly, koneoppiminen ja syväoppiminen. Ei ole kuitenkaan olemassa vain yhtä teknologiaa, jolla tekoäly toteutetaan. Tekoäly koostuu useammasta eri metodista ja teknologiasta, joista tulisi aina pystyä valitsemaan paras käsillä olevaan ongelmaan. [21]

Tekoälyn ydin on ohjelmointia, matematiikka ja tilastotiedettä. Näitä on hyvä ymmärtää käsitetasolla, jotta pystyy paremmin hahmottamaan, millaisia ongelmia tekoälyn avustuksella on mahdollista ratkaista. Tekoälyn keskeisimmät periaatteet voidaan kuvata suhteellisen pienellä määrällä matemaattisia kaavoja ja jopa melko yksinkertaisella matematiikallakin. [22]

---

[17] Ks. tarkemmin Kontkanen 2018. Tulorekisteri aloitti vuonna 2019 ja positiivinen luottotietorekisteri aloittaa 2023. Ks. tarkemmin Oikeusministeriö 2019.
[18] Kontkanen 2018, 11.
[19] Yleinen tietosuoja-asetus 22 artikla.
[20] Kananen & Puolitaival 2019, 199.
[21] Kananen & Puolitaival 2019, 27.
[22] Kananen & Puolitaival 2019, 28.

Tekoäly eroaa perinteisemmästä, sääntöpohjaisesta ohjelmoinnista. Sääntöpohjaisessa, deterministisessä ohjelmoinnissa ohjelmalle koodataan tietyt säännöt, joiden mukaan se pystyy käsittelemään annettua dataa. Esimerkiksi luonnollisen kielen simulointi deterministen ohjelmoinnin menetelmillä on hankalaa. Sen sijaan tekoälypohjaisessa ohjelmoinnissa käytetään algoritmeja, jotka löytävät datasta säännönmukaisuudet. Tekoälylle annetaan siis dataa ja tiedetyt vastaukset, jolloin tekoäly löytää datasta säännöt data - vastaus -parien kautta. Nämä säännöt perustuvat todennäköisyyksiin. Tästä on kyse, kun tekoälyä koulutetaan. Tällaisen kouluttamisen jälkeen, tekoälyä vielä muokataan siten, että se voi antaa täysin uudella datajoukolla ennusteita. [23]

On kuitenkin painotettava, että nykyiset tekoälyteknologiat ovat niin sanottua heikkoa tekoälyä eikä voida puhua aidosti tietoisesta tekoälystä. Ne ovat tarkoin rajattuihin käyttötarkoituksiin sopivia työkaluja, mutta eivät kykene itsenäiseen, oma-aloitteiseen päättelyyn. Vahvasta tekoälystä voidaan puhua, kun tekoäly kykenee ihmisen tasoiseen päättelyyn itsenäisesti ja on yleiskäyttöinen. [24]

## 1.5 Tutkimuskysymyksen asettelu ja rajaukset

Yritysjuridiikan lähtökohta on yrityksen tavoitteet, kun taas perinteisessä oikeustieteessä lähtökohta on normatiiviset systeemit ja niiden tutkiminen. Digitalisoituvassa maailmassa liiketoiminnan harjoittajien tulee ymmärtää vero- ja kirjanpito-oikeuden lisäksi myös muuta toimintaa säätelevää juridiikkaa. Mielestäni yksi tärkeimmistä nykyisen toimintaympäristön oikeudenaloista on tietosuojajuridiikka. Tutkin aihetta liiketaloudellis-normatiivisesti.

Tutkimuskysymykseni liittyy automaattisesti tehtäviin päätöksiin, jotka perustuvat yrityksen keräämiin henkilötietoihin. Yritykset pyrkivät jatkuvasti automatisoimaan toimintojaan ja kehittämään uusia palveluita. Mitä vaatimuksia yleinen tietosuoja-asetus tuo tähän asetelmaan ja miten opetetun algoritmin tekemät päätökset suhteutuvat henkilötietojen suojaan?

Tutkielman aiheen rajauksen kannalta on tarpeellista määritellä tietosuojan ja tietoturvan käsitteet. Tietosuoja on perusoikeus, joka turvaa luonnollisen henkilön oikeuden henkilötietojensa suojaan. Se turvaa rekisteröidyn oikeuksien ja vapauksien toteutumisen henkilötietojen käsittelyssä. Tietoturva sen sijaan on yksi tietosuojan toteuttamisen keino ja sen tarkoitus on suojata nimenomaan tietoaineistot ja tietojärjestelmät. Se tarkoittaa karkeasti kaikkia niitä organisatorisia ja teknisiä toimenpiteitä, joilla varmiste-

[23] Kananen & Puolitaival 2019, 30.
[24] Kananen & Puolitaival 2019, 38.

taan tietojen luottamuksellisuus ja eheys. [25] Karkeasti voitaneen todeta, että tietosuoja on laajempi, juridinen käsite kuin tietoturva, joka on luonteeltaan enemmän tekninen. Yksityisyyden suoja on vielä erillinen käsite tietosuojasta, vaikka tietosuoja onkin osa yksityisyyden suojaa. Yksityisyyden suoja on käsitteenä laajempi ja se on erillinen perusoikeus henkilötietojen suojasta[26].

Lähdeaineistona käytän pääasiassa EU-oikeuslähteitä kuten unionin oikeuden oikeustapauksia, säädöksiä ja niiden tulkinnasta annettuja linjauksia. Tutkimuskohteeni on melko tuore lainsäädäntö, joten varsinaista oikeuskäytäntöä ei ole käytettävissä paljoakaan. Merkittävin lähteeni on varsinaisen yleisen tietosuoja-asetuksen tekstin ja johdantokappaleiden lisäksi WP29-tietosuojatyöryhmän lausunnot ja ohjeistukset direktiivin ja asetuksen tulkinnasta.

Tutkielman kieleksi on tietoisesti valittu suomen kieli, koska tietosuojajuridiikasta ei ole käytettävissä paljoakaan ajantasaista materiaalia suomeksi. Samalla myös tutkielman yksi tarkoituksista on kääntää ja tuoda englannin kielisestä aineistosta termistöä luontevammaksi osaksi suomenkielistä keskustelua.

Tutkielma keskittyy pääasiallisesti tietosuoja-asetukseen ja suurin osa erityislainsäädännöstä johtuvat vaikutukset on rajattu ulkopuolelle. Esimerkiksi finanssialalla on erityissäädöksiä liittyen luotonantoon, jolla on vaikutuksia automaattiseen päätöksentekoon, kuten luotonantoprosessiin. Profilointiin tietojen käsittelyn yhteydessä keskityn vain siltä osin kuin se on tarpeellista automaattisen päätöksenteon tutkimiseksi. Tekoälyn luotettavuuteen ja eettisyyteen liittyen käsittelen aihetta vain siltä osin kuin se on tarpeellista automaattisen päätöksenteon ja oikeusvaikutusten kannalta. Siten esimerkiksi tekoälyn eettisyyteen liittyvät muut mahdolliset kiinnostavat näkökulmat rajautuvat tarkastelun ulkopuolelle. Lisäksi tietosuoja-asetuksessa on myös muita kiinnostavia teemoja, kuten henkilötietojen käsittely ja siirto EU:n talousalueen ulkopuolelle. Tämä on myös rajattu tutkielman aihealueen ulkopuolelle.

[25] Tietosuoja 2019.

[26] Euroopan unionin perusoikeuskirja (2012/C 326/02) 7 artikla ”yksityis- ja perhe-elämän kunnioittaminen” mukaan ”[j]okaisella on oikeus siihen, että hänen yksityis- ja perhe-elämäänsä, kotiaan ja viestejään kunnioitetaan2. 8 artiklan taas määrittelee henkilötietojen suojan perusoikeudeksi.

# 2 TIETOSUOJAN JURIDINEN VIITEKEHYS

## 2.1 Tietosuojadirektiivi 95/46/EY

Ennen yleistä tietosuoja-asetusta, voimassa oli direktiivi 95/46/EY[27] yksilöiden suojelusta henkilötietojen käsittelyssä ja näiden tietojen vapaasta liikkuvuudesta (aiempi tietosuojadirektiivi[28]). Yleisen tietosuoja-asetuksen 94 artikla mukaan direktiivi kumottiin ja viittauksia kyseiseen direktiiviin pidetään viittauksina yleiseen tietosuoja-asetukseen. Lisäksi saman artiklan mukaan viittauksia direktiivin 95/46/EY 29 artiklalla perustettuun tietosuojatyöryhmään pidetään viittauksia yleisellä tietosuoja-asetuksella perustettuun tietosuojaneuvostoon.

Aiemman direktiivin lähtökohta vuonna 1995 oli yksityisyyden suojan ja perusoikeuksien sijaan mielestäni markkinalähtöisempi. Direktiivin johdantokappaleen 1 kohdassa mainitaan unionin perustamissopimus ja viitataan muun lisäksi sen tarkoitukseen turvata yhteisellä toiminnalla taloudellinen ja sosiaalinen edistys. Saman direktiivin johdantokappaleen 4 kohdassa todetaan, että taloudellisen ja sosiaalisen toiminnan eri aloilla turvaudutaan yhä useammin henkilötietojen käsittelyyn ja että tietotekniikan kehittyminen helpottaa merkittävästi kyseisten tietojen käsittelyä ja niiden vaihtoa.

Aiemman tietosuojadirektiivin 95/46/EY soveltamisala oli melkein sama kuin voimassa olevan yleisen tietosuoja-asetuksen. Aiemman tietosuojadirektiivin soveltamisala määriteltiin 3 artiklassa, jonka mukaan sitä sovelletaan ”*osittain tai kokonaan automatisoituun tietojenkäsittelyyn sekä sellaisten henkilötietojen manuaaliseen käsittelyyn, jotka muodostavat rekisterin osan tai joiden on tarkoitus muodostaa rekisterin osa*”. Yleisen tietosuoja-asetuksen aineellinen soveltamisala on määritelty asetuksen 2 artiklassa, jonka sanamuoto on vastaavankaltainen: ”*[t]ätä asetusta sovelletaan henkilötietojen käsittelyyn, joka on osittain tai kokonaan automaattista, sekä sellaisten henkilötietojen käsittelyyn muussa kuin automaattisessa muodossa, jotka muodostavat rekisterin osan tai joiden on tarkoitus muodostaa rekisterin osa*”. Kummassakin säännöksissä mainitaan automaatio, henkilötiedot ja manuaalinen käsittely soveltamisalaan vaikuttavina tekijöinä, mutta hieman eri yhteyksissä.

Yleisen tietosuoja-asetuksen 3 artikla määrittelee alueellisen soveltamisalan seuraavasti: ”*[t]ätä asetusta sovelletaan henkilötietojen käsittelyyn, jota suoritetaan unionin alueella sijaitsevassa rekisterinpitäjän tai henkilötietojen käsittelijän toimipaikassa toiminnan yhteydessä, riippumatta siitä, suoritetaanko käsittely unionin alueella vai ei*”.

---

[27] Euroopan parlamentin ja neuvoston direktiivi 95/46/EY annettu 24 päivänä lokakuuta 1995, yksilöiden suojelusta henkilötietojen käsittelyssä ja näiden tietojen vapaasta liikkuvuudesta

[28] Ei vakiintunut termi, kirjoittajan oma.

Rekisterinpitäjän ja tietojen käsittelijän käsitteet ovat siis oleellisia alueellisen soveltamisen kannalta.

## 2.2 General Data Protection Regulation, GDPR

EU:n yleinen tietosuoja-asetus[29] (*General Data Protection Regulation,* GDPR) on tullut voimaan 25.5.2016, ja sen soveltaminen alkoi jäsenvaltioissa 25.5.2018. Sen keskeisempiä säädöksiä ovat se, että luonnollisten henkilöiden suojelu henkilötietojen käsittelyn yhteydessä on perusoikeus. Euroopan unionin perusoikeuskirjan 8 artiklan 1 kohdan ja Euroopan unionin toiminnasta tehdyn sopimuksen 16 artiklan 1 kohdan mukaan jokaisella on oikeus henkilötietojensa tietosuojaan. Tietosuoja-asetus tunnistaa myös teknologian kehittymisen ja globalisaation tuomat haasteet henkilötietojen käsittelyyn. Uudet teknologiat mahdollistavat aiempaa huomattavasti laajemman tietojen käsittelyn. [30] Lisääntyvän sääntelyn tarpeen vastapainona on kuitenkin myös yleisen taloudellisen toimeliaisuuden vaatima tietojen vaihto ja vapaa liikkuvuus.

Yleistä tietosuoja-asetusta sovelletaan sekä julkisella että yksityisellä sektorilla. Vaikka EU:n yleinen tietosuoja-asetus on kansallisesti suoraan sovellettava säädös, se jättää jäsenvaltioille direktiivinomaista kansallista liikkumavaraa.

Suomessa EU:n yleisen tietosuoja-asetuksen mukaiset muutokset toteutettiin säätämällä uusi tietosuojalaki (5.12.2018/1050), joka toimii henkilötietojen käsittelyä koskevana yleislakina. Tietosuojalailla täydennettiin ja täsmennettiin EU:n yleistä tietosuoja-asetusta. Yleinen tietosuoja-asetus koskee lähtökohtaisesti kaikenlaista henkilötietojen käsittelyä. Se sisältää säännökset rekisteröidyn oikeuksista sekä rekisterinpitäjän ja henkilötietojen käsittelijän velvollisuuksista.[31] Suomessa kansallista tietosuojalainsäädäntöä täydentää tietosuojalain lisäksi laki sähköisen viestinnän palveluista (7.11.2014/917), laki yksityisyydensuojasta työelämässä (13.8.2004/759) ja luottotietolaki (11.5.2007/527). Nämä ovat erityislakeja suhteessa tietosuojalakiin.

Yleinen tietosuoja-asetus ei aseteta luonnollisille tai oikeushenkilöille lisävelvoitteittasuhteessa sähköisen viestinnän tietosuojadirektiiviin 2002/58/EY[32] nähden, kun kyse on sellaisesta tietojenkäsittelytoimesta, joka liittyy yleisesti saatavilla olevien sähköisten viestintäpalvelujen tarjoamiseen yleisissä viestintäverkoissa unionissa.

---

[29] Euroopan parlamentin ja neuvoston asetus (EU) 2016/679, annettu 27 päivänä huhtikuuta 2016, luonnollisten henkilöiden suojelusta henkilötietojen käsittelyssä sekä näiden tietojen vapaasta liikkuvuudesta ja direktiivin 95/46/EY kumoamisesta (yleinen tietosuoja-asetus).
[30] Yleisen tietosuoja-asetuksen johdantokappaleen 3 kohta.
[31] Eduskunta 2018.
[32] Euroopan parlamentin ja neuvoston direktiivi 2002/58/EY, annettu 12 päivänä heinäkuuta 2002, henkilötietojen käsittelystä ja yksityisyyden suojasta sähköisen viestinnän alalla (sähköisen viestinnän tietosuojadirektiivi)

Yleinen tietosuoja-asetus ei myöskään vaikuta aiemmin tehtyihin sellaisiin sopimuksiin koskien henkilötietojen siirtoa kolmansiin maihin tai kansainvälisille järjestöille, jotka on tehty ennen 24.6.2016. Ne jäävät voimaan, kunnes niitä muutetaan, korvataan tai kumotaan.

Euroopan tietosuojaneuvosto perustettiin yleisellä tietosuoja-asetuksella ja se korvasi aiemman tietosuojatyöryhmän. Tietosuojatyöryhmä tunnettiin myös nimellä WP 29 tai working party 29 ja se perustettiin aiemman direktiivin 95/46/EY 29 artiklan mukaisesti. Se toimi kunnes 25.5.2018 asti, kunnes yleinen tietosuoja-asetus astui voimaan. Tietosuojaneuvoston muodostavat jäsenvaltioiden tietosuojavaltuutetut ja Euroopan tietosuojavaltuutettu. Neuvosto toimii riippumattomasti. Neuvoston tehtävät on määritelty yleisen tietosuoja-asetuksen 70 artiklassa. Tiivistäen sen tehtävä on ensisijaisesti varmistaa, että asetusta sovelletaan yhdenmukaisesti sekä antaa suosituksia, suuntaviivoja, parhaita käytänteitä ja oma-aloitteisesti tarkastella asetuksen soveltamista koskevia kysymyksiä. Lisäksi tietosuojaneuvosto antaa komissiolle lausunnon kolmannen maan tai kansainvälisen järjestön tietosuojan tason riittävyyden arvioimiseksi. Neuvosto pitää yllä myös sähköistä rekisteriä päätöksistä, joita viranomaiset ja tuomioistuimet ovat tehneet yhdenmukaisuusmekanismissa käsitellyissä asioissa.

Aiemmin toimineen tietosuojatyöryhmän (WP29) tehtäviin kuului myös huolehtia tehtävistä myös sähköisen viestinnän tietosuojadirektiivin soveltamisalaan kuuluvissa asioissa kuten perusoikeuksien ja -vapauksien sekä oikeutettujen etujen suojaamisessa sähköisen viestinnän alalla. [33]

Suomessa yleistä tietosuoja-asetusta on täydennetty ja täsmennetty tietosuojalailla (1050/2018). Verrattuna tietosuoja-asetukseen, kansallinen tietosuojalaki laajentaa asetuksen aineellista soveltamisalaa. Tietosuoja-asetuksessa aineellinen soveltamisala on rajoitettu asetuksen 2 artiklan 2 kohdassa siten, että asetusta ei sovelleta henkilötietojen käsittelyyn, joka suoritetaan sellaisen toiminnan yhteydessä, joka ei kuulu unionin lainsäädännön soveltamisalaan tai, jota suorittavat jäsenvaltiot toteuttaessaan SEU V osaston 2 luvun soveltamisalaan kuuluvaa toimintaa. Tällaista toimintaa on rajavalvonta-, turvapaikka- ja maahanmuuttopolitiikka. Lisäksi Suomessa yleistä tietosuoja-asetusta ei sovelleta eduskunnan valtiopäivätoimintaan tai sellaiseen henkilötietojen käsittelyyn, josta säädetään henkilötietojen käsittelystä rikosasioissa ja kansallisen turvallisuuden ylläpitämisen yhteydessä annetussa laissa. [34] Lisäksi tietosuojalaki määrittelee lapsien

---

[33] Sähköisen viestinnän tietosuojadirektiivi 2002/58/EY 15 artiklan 3 kohta: *Direktiivin 95/46/EY 29 artiklalla perustetun tietosuojatyöryhmän on huolehdittava direktiivin 30 artiklassa tarkoitetuista tehtävistä myös tämän direktiivin soveltamisalaan kuuluvissa asioissa eli perusoikeuksien ja -vapauksien sekä oikeutettujen etujen suojaamisessa sähköisen viestinnän alalla.* Yleisen tietosuoja-asetuksen loppusäännöksien 94 artiklan 2 kohdan mukaan viittauksia ko. direktiiviin pidetään viittauksina yleiseen tietosuoja-asetukseen.

[34] Laki henkilötietojen käsittelystä rikosasioissa ja kansallisen turvallisuuden ylläpitämisen yhteydessä (1054/2018 § 1.1):

kohdalla käsittelyn lainmukaisesti, jos lapsi on vähintään 13 vuotias. Vastaavaa tarkkaa määrittelyä yleisessä tietosuoja-asetuksessa ei ole.

Tietosuojalaissa on myös määritelty poikkeuksia tietosuoja-asetukseen nähden erityisiin henkilötietoryhmiin koskevaan tietojenkäsittelyyn. Tietosuoja-asetuksen 9 artiklan 1 kohdan mukaisten erityisten henkilötietoryhmien käsittely on kiellettyä. Tietosuojalaissa kuitenkin sallitaan näiden tietojen käsittely tietyissä tilanteissa[35]:

- vakuutuslaitoksen käsitellessä vakuutustoiminnassa saatuja tietoja
- tietojen käsittelyyn, josta säädetään laissa tai joka johtuu välittömästi rekisterinpitäjälle laissa säädetystä tehtävästä
- ammattiliittoon kuulumista koskevaan tiedon käsittelyyn, joka on tarpeen rekisterinpitäjän erityisten oikeuksien ja velvoitteiden noudattamiseksi työoikeuden alalla
- terveydenhuollon palveluntarjoaja järjestäessään tai tuottaessaan palveluja käsittelee tässä toiminnassa saamiaan rekisteröidyn hoidon kannalta välttämättömiä tietoja
- sosiaalihuollon palveluntarjoaja järjestäessään tai tuottaessaan palveluja tai myöntäessään etuuksia käsittelee tässä toiminnassa saamiaan tai tuottamiaan rekisteröidyn palvelun tai etuuden myöntämisen kannalta välttämättömiä tietoja
- terveyttä koskevien ja geneettisten tietojen käsittelyyn antidopingtyössä ja vammaisurheilun yhteydessä siltä osin kuin näiden tietojen käsittely on välttämätöntä antidopingtyön tai vammaisten ja pitkäaikaissairaiden urheilun mahdollistamiseksi
- tieteellistä tai historiallista tutkimusta taikka tilastointia varten tehtävään tietojen käsittelyyn
- tutkimus- ja kulttuuriperintöaineistojen käsittelyyn yleishyödyllisessä arkistointitarkoituksessa geneettisiä tietoja lukuun ottamatta.

Tietosuojalaissakin on kuitenkin määrätty tietyistä erityisistä ja asianmukaisista suojatoimista rekisteröidyn oikeuksien suojaamiseksi. Nämä toimet noudattelevat yleisen tietosuoja-asetuksen määräyksiä. [36]

---

”Tätä lakia sovelletaan toimivaltaisten viranomaisten käsitellessä henkilötietoja, kun kyse on:
1) rikosten ennalta estämisestä, paljastamisesta, selvittämisestä tai syyteharkintaan saattamisesta;
2) syyteharkinnasta ja muusta rikokseen liittyvästä syyttäjän toiminnasta;
3) rikosasian käsittelemisestä tuomioistuimessa;
4) rikosoikeudellisen seuraamuksen täytäntöönpanemisesta;
5) yleiseen turvallisuuteen kohdistuvilta uhkilta suojelemisesta tai tällaisten uhkien ehkäisemisestä 1–4 kohdassa tarkoitetun toiminnan yhteydessä.”

[35] Ks täydellinen listaus tietosuojalaki (1054/2018) 6.1 §

[36] Ks. täydellinen listaus tietosuojalaki (1054/2018) 6.2 §

## 2.3 Tietosuojajuridiikan käsitteet

Alla on luettelo tämän työn kannalta keskeisimmistä käsitteistä, joka on kirjoitettu mukaillen yleisen tietosuoja-asetuksen 4 artiklan määritelmiä. Aiheen kannalta olennaisimpia tutkitaan myöhimmässä luvuissa tarkemmin.

- *Henkilötiedot* tarkoittavat kaikkia tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön liittyviä tietoja. Tunnistettavissa olevana pidetään luonnollista henkilöä, joka voidaan suoraan tai epäsuorasti tunnistaa erityisesti tunnistetietojen, kuten nimen, henkilötunnuksen, sijaintitiedon, verkkotunnistetietojen taikka yhden tai useamman hänelle tunnusomaisen tekijän perusteella. Tällaisia tunnusomaisia tekijöitä ovat tietosuoja-asetuksessa fyysiset, fysiologiset, geneettiset, psyykkiset, taloudelliset, kulttuuriset ja sosiaaliset tekijät. [37]
- *Rekisteröity* viittaa tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön (katso myös edellä henkilötiedot).
- *Käsittely* viittaa siihen toimintoon tai niihin toimintoihin, joita kohdistetaan henkilötietoihin tai henkilötietoja sisältäviin tietojoukkoihin joko automaattista tietojenkäsittelyä käyttäen tai manuaalisesti. Tällaisia toimintoja ovat muun muassa tietojen kerääminen, tallentaminen, järjestäminen, jäsentäminen, säilyttämien, muokkaaminen ja muuttaminen. Tietojen luovuttaminen on myös tällaista käsittelyä tai tietojen saataville asettaminen. Myös tietojen yhteensovittaminen ja yhdistely, tietojen rajoittaminen, poistaminen sekä tuhoaminen ovat myös tietojen käsittelyä.
- *Käsittelyn rajoittamisella* tarkoitetaan tallennettujen henkilötietojen merkitsemistä tarkoituksena rajoittaa niiden myöhempää käsittelyä.
- *Profiloinnilla* tarkoitetaan mitä tahansa henkilötietojen automaattista käsittelyä, jossa henkilötietoja käyttämällä arvioidaan luonnollisen henkilön tiettyjä henkilökohtaisia ominaisuuksia. Profilointia on myös, mikäli tietoja analysoidaan tai ennakoidaan henkilön piirteitä, jotka liittyvät kyseisen luonnollisen henkilön työsuoritukseen, taloudelliseen tilanteeseen, terveyteen, henkilökohtaisiin mieltymyksiin, kiinnostuksen kohteisiin, luotettavuuteen, käyttäytymiseen, sijaintiin tai liikkeisiin.
- *Pseudonymisoinnilla* tarkoitetaan henkilötietojen käsittelemistä siten, että henkilötietoja ei voida enää yhdistää tiettyyn rekisteröityyn käyttämättä lisätietoja. Pseudonymisoinnin edellytys on, että tällaiset lisätiedot säilytetään erillään ja niihin sovelletaan teknisiä ja organisatorisia toimenpiteitä, joilla varmistetaan,

[37] Ks. myös asia C-582/14 Patrick Breyer v. Saksan liittotasavalta, tuomio. Sähköinen oikeustapauskokoelma. Kohta 49. Dynaaminen IP-osoite on palveluntarjoajaan nähden direktiivissä 95/46/EY tarkoitettu henkilötieto.

ettei henkilötietojen yhdistämistä tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön tapahdu.

- *Rekisteri* tarkoittaa mitä tahansa jäsenneltyä henkilötietoja sisältävää tietojoukkoa, josta tiedot ovat saatavilla tietyin perustein. Sillä ei ole merkitystä onko tietojoukko keskitetty, hajautettu tai toiminnallisin tai maantieteellisin perustein jaettu.
- *Rekisterinpitäjä* tarkoittaa luonnollista henkilöä tai oikeushenkilöä, viranomaista, virastoa tai muuta elintä, joka yksin tai yhdessä toisten kanssa määrittelee henkilötietojen käsittelyn tarkoitukset ja keinot.
- *Henkilötietojen käsittelijä* tarkoittaa luonnollista henkilöä tai oikeushenkilöä, viranomaista, virastoa tai muuta elintä, joka käsittelee henkilötietoja rekisterinpitäjän lukuun.
- *Kolmas osapuoli* tarkoittaa luonnollista henkilöä tai oikeushenkilöä, viranomaista, virastoa tai muuta toimielintä kuin rekisteröityä, rekisterinpitäjää, henkilötietojen käsittelijää ja henkilöä, joilla on oikeus käsitellä henkilötietoja suoraan rekisterinpitäjän tai henkilötietojen käsittelijän välittömän vastuun alaisena.
- Rekisteröidyn *suostumuksella* tarkoitetaan mitä tahansa vapaaehtoista, yksilöityä, tietoista ja yksiselitteistä tahdonilmaisua, jolla rekisteröity hyväksyy henkilötietojensa käsittelyn antamalla suostumusta ilmaisevan lausuman tai toteuttamalla selkeästi suostumusta ilmaisevan toimen.
- *Yrityksellä* tarkoitetaan taloudellista toimintaa harjoittavaa luonnollista henkilöä tai oikeushenkilöä sen oikeudellisesta muodosta riippumatta, mukaan lukien kumppanuudet tai yhdistykset, jotka säännöllisesti harjoittavat taloudellista toimintaa.
- *Valvontaviranomaisella* tarkoitetaan jäsenvaltion perustamaa riippumatonta viranomaista. Suomessa valvontaviranomainen on tietosuojavaltuutettu.

## 2.4 Tietojenkäsittelyn lainmukaisuus

Aiemman tietosuojadirektiivin (95/46/EY) säädökset tietojenkäsittelyn laillisuutta koskevista periaatteista (7 artikla) ovat samanlaisia kuin voimassa olevan yleisen tietosuoja-asetuksen vastaavat säädökset (6 artikla). Alla taulukossa 1 vertaillaan asetuksen ja direktiivin säädöksiä henkilötietojen laillisen käsittelyn perusteista.

| Säädöksen kohta | Yleinen tietosuoja-asetus (2016/679/EU) | Direktiivi 95/46/EY |
|---|---|---|
| a) | rekisteröity on antanut suostumuksensa henkilötietojensa käsittelyyn **yhtä tai useampaa erityistä tarkoitusta varten**; | jos rekisteröity on **yksiselitteisesti** antanut suostumuksensa, tai |
| b) | käsittely on tarpeen sellaisen sopimuksen täytäntöön panemiseksi, jossa rekisteröity on osapuolena, tai sopimuksen tekemistä edeltävien toimenpiteiden toteuttamiseksi rekisteröidyn pyynnöstä; | jos käsittely on tarpeen sellaisen sopimuksen täytäntöön panemiseksi, jossa rekisteröity on osallisena, tai sopimusta edeltävien toimenpiteiden toteuttamiseksi rekisteröidyn pyynnöstä, tai |
| c) | käsittely on tarpeen rekisterinpitäjän lakisääteisen velvoitteen noudattamiseksi; | jos käsittely on tarpeen rekisterinpitäjän laillisen velvoitteen noudattamiseksi, tai |
| d) | käsittely on tarpeen rekisteröidyn **tai toisen luonnollisen henkilön** elintärkeiden etujen suojaamiseksi; | jos käsittely on tarpeen rekisteröidyn elintärkeän edun suojaamiseksi, tai |
| e) | käsittely on tarpeen yleistä etua koskevan tehtävän suorittamiseksi tai rekisterinpitäjälle kuuluvan julkisen vallan käyttämiseksi; | jos käsittely on tarpeen yleistä etua koskevan tehtävän suorittamiseksi tai sellaisen julkisen vallan käyttämiseksi, joka kuuluu rekisterinpitäjälle **tai sivulliselle, jolle tiedot luovutetaan**, tai |
| f) | käsittely on tarpeen rekisterinpitäjän tai kolmannen osapuolen oikeutettujen etujen toteuttamiseksi, paitsi milloin henkilötietojen suojaa edellyttävät rekisteröidyn edut tai perusoikeudet ja -vapaudet syrjäyttävät tällaiset edut, **erityisesti jos rekisteröity on lapsi**. | jos käsittely on tarpeen rekisterinpitäjän **tai tiedot saavan sivullisen** oikeutetun intressin toteuttamiseksi, paitsi milloin tämän intressin syrjäyttävät rekisteröidyn 1 artiklan 1 kohdan perusteella suojaa tarvitsevat intressit ja perusoikeudet ja -vapaudet. |

Taulukko 1. Tietosuojadirektiivin ja tietosuoja-asetuksen vertailu laillisen käsittelyn perusteista.

Kuten taulukosta 1 nähdään, säädökset on muotoiltu hyvin samankaltaisesti. Olennaisimmat erot on tummennettu. Kohdassa a suostumuksen osalta tietosuoja-asetuksessa on muutettu hieman suostumuksen kohdetta siten, että tietosuoja-asetuksessa nimenomaisessa mainitaan, että suostumuksen voi antaa yhtä tai useampaa tarkoitusta varten. Kohdassa d tietosuoja-asetuksessa on laajennettu elintärkeiden etujen suojaamista rekisteröidyn lisäksi myös toiseen luonnolliseen henkilöön. Kohdassa e puolestaan säädöstä

on kavennettu siten, että se ei koske enää sivullisia. Samoin on toimittu myös kohdassa f sekä lisätty erityismaininta lapsista.

## 2.5 Hallinnollinen seuraamusmaksu

Yleisen tietosuoja-asetuksen luvussa VIII määritellään oikeussuojakeinot, vastuut ja seuraamukset. Rekisteröidyn oikeussuojakeino on oikeus tehdä valitus valvontaviranomaiselle. Tietosuoja-asetuksen artiklat 78 ja 79 edelleen vahvistavat tätä oikeutta määrittelemällä muun lisäksi oikeuden tehokkaisiin oikeussuojakeinoihin valvontaviranomaista sekä rekisterinpitäjää tai henkilötietojen käsittelijää vastaan.

Luonnollinen henkilö on oikeutettu korvaukseen, mikäli tietosuoja-asetuksen rikkomisesta seuraa aineellista tai aineetonta vahinkoa. Henkilötietojen käsittelijä on vastuussa vain, jos se ei ole noudattanut nimenomaisesti käsittelijöille osoitettuja velvoitteita tai jos se on toiminut rekisterinpitäjän ohjeistuksen ulkopuolella tai sen vastaisesti. [38]

Valvontaviranomaisen tutkintavaltuudet on lueteltu tyhjentävästi tietosuoja-asetuksen 58 artiklan 1 kohdassa. Lisäksi 2 kohdassa luetellaan toimivaltuudet korjaaviin toimenpiteisiin, joita ovat muun muassa varoituksien ja huomautuksien antaminen. Varsinaiset seuraamukset määritellään kuitenkin tietosuoja-asetuksen 83 artiklassa, jossa määritellään hallinnollisten sakkojen määräämisen yleiset edellytykset. Varoituksen tai huomautuksen antaminen rekisterinpitäjälle ei ole edellytys, että seuraamusmaksu voitaisiin langettaa rekisterinpitäjälle. [39] Hallinnollisen sakon määräämisestä ja sen suuruudesta päätettyä tulisi ottaa seuraavat tekijät huomioon kussakin yksittäistapauksessa:

- ”rikkomisen luonne, vakavuus ja kesto, kyseisen tietojenkäsittelyn luonne, laajuus tai tarkoitus huomioon ottaen, sekä niiden rekisteröityjen lukumäärä, joihin rikkominen vaikuttaa, ja heille aiheutuneen vahingon suuruus
- rikkomisen tahallisuus tai tuottamuksellisuus
- rekisterinpitäjän tai henkilötietojen käsittelijän toteuttamat toimet rekisteröidyille aiheutuneen vahingon lieventämiseksi
- rekisterinpitäjän tai henkilötietojen käsittelijän vastuun aste, ottaen huomioon heidän 25 (*Sisäänrakennettu ja oletusarvoinen tietosuoja*) ja 32 artiklan (*Käsittelyn turvallisuus*) nojalla toteuttamansa tekniset ja organisatoriset toimenpiteet

[38] Yleinen tietosuoja-asetus 86 artikla.
[39] Yleinen tietosuojasetus 83 artikla 2 kohdan tarkka sanamuoto: ” Hallinnolliset sakot määrätään kunkin yksittäisen tapauksen olosuhteiden mukaisesti 58 artiklan 2 kohdan a–h ja j alakohdassa tarkoitettujen *toimenpiteiden lisäksi tai niiden sijasta*.”

- rekisterinpitäjän tai henkilötietojen käsittelijän mahdolliset aiemmat vastaavat rikkomiset
- yhteistyön aste valvontaviranomaisen kanssa rikkomisen korjaamiseksi ja sen mahdollisten haittavaikutusten lieventämiseksi
- henkilötietoryhmät, joihin rikkominen vaikuttaa
- tapa, jolla rikkominen tuli valvontaviranomaisen tietoon, erityisesti se, ilmoittiko rekisterinpitäjä tai henkilötietojen käsittelijä rikkomisesta ja missä laajuudessa
- jos kyseiselle rekisterinpitäjälle tai henkilötietojen käsittelijälle on aikaisemmin määrätty samasta asiasta 58 artiklan 2 kohdassa tarkoitettuja toimenpiteitä, näiden toimenpiteiden noudattaminen
- 40 artiklan mukaisten hyväksyttyjen käytännesääntöjen tai 42 artiklan mukaisten hyväksyttyjen sertifiointimekanismien noudattaminen
- mahdolliset muut tapaukseen sovellettavat raskauttavat tai lieventävät tekijät, kuten rikkomisesta suoraan tai välillisesti saadut mahdolliset taloudelliset edut tai rikkomisella vältetyt tappiot”[40]

Valvontaviranomaisen määräyksen noudattamatta jättämisestä voidaan määrätä hallinnollinen sakko. Sakko voi olla enintään 20 000 000 € tai, jos kyseessä on yritys, 4 % edeltävän tilikauden vuotuisesta maailman laajuisesta kokonaisliikevaihdosta sen mukaan kumpi on suurempi. [41] Suomessa hallinnollisen seuraamusmaksun määrää tietosuojavaltuutetun ja apulaistietosuojavaltuutettujen yhdessä muodostama seuraamuskollegio. Se on päätösvaltainen kolmijäsenisenä. Lisäksi tietosuojalaki lisää vanhentumissäännöksen, jonka mukaan maksua ei saa määrätä, jos on kulunut yli 10 vuotta siitä, kun rikkomus on tapahtunut. Tietosuojalaki myös rajaa tiettyjä julkisia toimijoita seuraamusmaksun ulkopuolelle. Näitä ovat muun muassa valtion viranomaiset, valtion liikelaitokset, Suomen evankelis-luterilainen kirkko ja Suomen ortodoksinen kirkko. [42] On huomattava, että tietosuojaan liittyviä rangaistussäännöksiä löytyy myös esimerkiksi rikoslaista, jossa säädetään viestintäsalaisuuden loukkauksesta.

[40] Yleinen tietosuoja-asetus 83 artikla 2 kohta.
[41] Yleinen tietosuoja-asetus 83 artikla 6 kohta.
[42] Tietosuojalaki 24 §.

# 3 HENKILÖTIEDOT

## 3.1 Mikä on henkilötietoa?

Yleinen tietosuoja-asetus koskee nimenomaan henkilötietoja, joilla tarkoitetaan tietoja, jotka voidaan liittää tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön. Henkilö voi olla tunnistettavissa suorasti tai epäsuorasti hänelle tunnusomaisen fyysisen, fysiologisen, geneettisen, psyykkisen, taloudellisen, kulttuurisen tai sosiaalisen tekijän perusteella. [43] Tietosuoja-asetuksen henkilötietojen määritelmä on siis melko laaja. On myös olennaista huomioida, että yritys tai yhteisö kerää tietoja useista eri lähteistä ja näitä eri tietoja yhdistelemällä henkilöitä voidaan tunnistaa.

Tietosuojatyöryhmän[44] mukaan henkilötietojen käsite koostuu neljästä komponentista, jotka seuraavat suoraan säädöksen sanamuodoista: ”kaikenlaisia tietoja”, ”luonnollista henkilöä koskevia tietoja”, ”tunnistettavissa tai tunnistettavissa olevaa” ja ”luonnollinen henkilö”. Kaikenlaisilla tiedoilla viitataan tiedon laajaan käsitteeseen: se voi olla joko objektiivista tai subjektiivista. Tiedon ei myöskään välttämättä tarvitse olla oikeaksi todistettua, jotta se voitaisiin tulkita henkilötiedoksi. [45]

Luonnollista henkilöä koskevilla tiedoilla viitataan puolestaan siihen, että käsiteltävällä tiedolla on jokin relaatio luonnolliseen henkilöön. Käsiteltävällä tiedolla voi olla suora relaatio luonnolliseen henkilöön, kuten sairauskertomuksissa. Toisaalta tieto voi olla epäsuorasti yhdistettävissä jonkin objektin kautta luonnolliseen henkilöön. Esimerkiksi talo voisi olla tällainen objekti, kun sille tehdään hinta-arvio. Tällaisen objektin kautta tieto voi olla yhdistettävissä luonnolliseen henkilöön joissakin tilanteissa ja olla sitä kautta tietosuoja-asetuksen tarkoittamaa henkilötietojen käsittelyä. [46]

Kolmas henkilötiedon komponentti on ”tunnistettavissa tai tunnistettavissa oleva”. Tunnistetulla viitataan tilanteeseen, jossa yksittäinen luonnollinen henkilö on mahdollista tunnistaa joukosta. Tunnistettavissa oleva puolestaan viittaa tilanteeseen, jossa yksittäinen luonnollinen henkilö voisi olla mahdollista tunnistaa joukosta, mutta ei ole vielä tunnistettu. Tässä siis pitää arvioida asiayhteys, jossa yksilöinti on mahdollista tehdä. Vaikkapa tietyn luokan opiskelijoista voi olla mahdollista tunnistaa yleisellä sukunimellä kuten Virtanen yksi tietty oppilas, mutta vastaavasta kaupungin kaikista opiskelijoista ei. [47]

---

[43] Yleinen tietosuoja-asetus 4 artikla 1 kohta.
[44] WP136 2007, 6. Tietosuojatyöryhmän lausunto ennen voimassa olevaa yleistä tietosuoja-asetusta. Tietosuojadirektiivin 95/46/EY 2 artiklan a kohdan mukainen henkilötietojen määritelmä vastaa yleisen tietosuoja-asetuksen 4 artiklan 1 kohdan mukaista henkilötietojen määritelmää olennaisilta osin.
[45] WP136 2006, 6-9.
[46] WP136 2007, 9.
[47] WP136 2007, 12.

Yleisen tietosuoja-asetuksen johdantokappale 26 tarjoaa lisätietoja koskien henkilötietoja ja henkilön tunnistettavuutta: ”*[j]otta voidaan määrittää, onko luonnollinen henkilö tunnistettavissa, olisi otettava huomioon kaikki keinot, joita joko rekisterinpitäjä tai muu henkilö voi kohtuullisen todennäköisesti käyttää mainitun luonnollisen henkilön tunnistamiseen suoraan tai välillisesti, kuten kyseisen henkilön erottaminen muista*”. Rekisterinpitäjän tulisi siis arvioida, että mitkä kaikki henkilöä koskevat tunnisteet ovat sellaisia, joita voitaisiin käyttää yksittäisen luonnollisen henkilön tunnistamiseen. Nimenomaista listaa tällaisista tekijöistä ei ole, mutta asetuksen 4 artiklan 1 kohdan määritelmissä listataan tällaiset tunnisteluokat: ”- - *kuten nimen, henkilötunnuksen, sijaintitiedon, verkkotunnistetietojen taikka yhden tai useamman hänelle tunnusomaisen fyysisen, fysiologisen, geneettisen, psyykkisen, taloudellisen, kulttuurillisen tai sosiaalisen tekijän perusteella*”.

Tietojen pseudonymisoinnilla viitataan menetelmiin, joilla voidaan peittää yksittäisten luonnollisten henkilöiden identiteetti. Pseudonymisointi on hyödyllistä silloin, kun halutaan kerätä tietoa yksilöstä, mutta yksilön identiteettiä ei ole tarpeellista tietää. Tämä tulee usein kysymykseen etenkin tilastollisessa tutkimuksessa. Pseudonymisointi voidaan tehdä niin, että pseudonimisoitua tietoa ei voida enää uudelleen palauttaa identiteetteihin. Tällöin voidaan sanoa, että data on anonymisoitua. Jos identiteetit ovat palautettavissa, tietoja voidaan pitää henkilötietoina, joista yksittäinen luonnollinen henkilö on tunnistettavissa. Tällöin kuitenkin tietosuojariskejä luonnollisille henkilöille voidaan pitää pienempinä ja se voi auttaa rekisterinpitäjiä noudattamaan tietosuojavelvoitteitaan. [48] On tärkeää kuitenkin huomata, että pseudonymisointi ei ole vain yksi anonymisoinnin metodi vaan tapa vähentää relaatioita alkuperäisen rekisteröidyn identiteettiin.[49] Kun arvioidaan, onko pseudonymisoitua dataa mahdollista palauttaa yksittäisen henkilön identiteettiin, tulee yleisen tietosuoja-asetuksen johdantokappaleen 26 mukaan ottaa huomioon ”*kaikki objektiiviset tekijät, kuten tunnistamisesta aiheutuvat kulut ja tunnistamiseen tarvittava aika sekä käsittelyajankohtana käytettävissä oleva teknologia ja tekninen kehitys*”.

Anonyymit tiedot ovat tietoja, jotka eivät liity tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön tai sellaisia tietoja, joista tunnistetiedot on poistettu siten, ettei yksittäisen henkilön tunnistaminen ole enää mahdollista. Yleinen tietosuoja-asetus ei koske tällaisten tietojen käsittelyä. [50]

[48] WP136 2007, 18.
[49] WP216 2014, 3.
[50] Yleinen tietosuoja-asetus johdantokappale 26.

## 3.2 Henkilötietojen anonymisointi

Henkilötietojen (datan) pseudonymisointi tavalla, joka estää yksilöiden identiteetin palauttamisen, on datan anynomisointia. Kun arvioidaan, että voiko identiteettiä palauttaa datasta, rekisterinpitäjän tulee arvioida mitä kohtuudella voidaan päätellä datasta. Anonymisointitekniikoita tulisi arvioida sekä juridisesta, mutta myös teknisestä näkökulmasta, jotta voidaan päätellä, mikä on kohtuudella mahdollista.

Yleisen tietosuoja-asetuksen lisäksi myös sähköisen viestinnän tietosuojadirektiivissä[51] on viittauksia tietojen anonymisointiin. Sähköisen viestinnän tietosuojadirektiivin johdantokappaleessa 26 mainitaan myös anonymisointi: ”Viestintäpalvelujen markkinointiin tai lisäarvopalvelujen tarjoamiseen käytettävät liikennetiedot olisi myös poistettava tai tehtävä nimettömiksi palvelun tarjoamisen jälkeen”. Vastaava direktiivin 6 artiklassa 1 kohdassa mainitaan verkkoliikennedatan anonymisointi ja saman direktiivin 9 artiklan 1 kohdassa mainitaan sijaintidatan anonymisointi.

Kansainvälinen standardointijärjestö ISO määrittelee anonymisoinnin vapaasti käännettynä prosessiksi, jonka tuloksena henkilötietoja sisältävä data on muunnettu peruuttamattomasti niin, että henkilöä ei voida enää suorasti tai epäsuorasti tunnistaa rekisterinpitäjän yksin tai yhdessä kolmansien osapuolien kanssa.[52]

Tietosuojatyöryhmä[53] jäsentää ”kohtuudella mahdollista” seuraavien kysymysten kautta:

1. Onko edelleen mahdollista tunnistaa yksilöä joukosta?
2. Voiko yhdistää tietoa niin, että yksilö on mahdollista tunnistaa relaatioiden kautta?
3. Voiko käsiteltävistä tiedoista tehdä päätelmiä yksilöä koskien?[54]

Tietosuojatyöryhmä[55] korostaa myös, että tietojen anonymisointiin liittyy aina riski siitä, että anonymisoidut tiedot voitaisiinkin linkittää yksittäisiin luonnollisiin henkilöihin. Toisaalta kuitenkaan kaikkia tietoja ei tule anonymisoida, jotta tietyt rekisteröidyn ja rekisterinpitäjän oikeudet voidaan turvata.

Tietosuojatyöryhmän mukaan rekisterinpitäjän tulisi ensisijaisesti keskittyä konkreettisiin tapoihin anonymisoida tiedot etenkin kustannuksiin ja vaadittuun tietotaitoon pei-

---

[51] Euroopan parlamentin ja neuvoston direktiivi 2002/58/EY, annettu 12 päivänä heinäkuuta 2002, henkilötietojen käsittelystä ja yksityisyyden suojasta sähköisen viestinnän alalla (sähköisen viestinnän tietosuojadirektiivi).

[52] ISO/IEC 29100:2011(E) alkuperäinen englanninkielinen sanamuoto: ”process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party”

[53] WP216 2014, 3.

[54] WP216 2014, 3. Alkuperäinen englanninkielinen sanamuoto:” (i) is it still possible to single out an individual, (ii) is it still possible to link records relating to an individual, and (iii) can information be inferred concerning an individual?”

[55] WP216 2014, 6.

laten. Anonymisointiin käytetty aika ja kustannukset tulisi olla tasapainossa muiden tiedonhankintaan liittyvien tekijöiden kanssa. Tällaisia tekijöitä ovat muun muassa kustannusten alentuminen yksilöiden tunnistamiseen isojen tietomassojen tietojen käsittelyssä, yhä laajemmin saatavilla oleva julkinen data (niin sanottu open data) sekä mahdolliset esiin tulleet esimerkit epäonnistuneita anonymisoinneista. [56]

Toiseksi rekisterinpitäjän tulisi myös huomioida ”*tunnistettu tai tunnistettavissa oleva*” -komponentti henkilötiedon käsitteessä. On siis tärkeää ymmärtää, että vaikka rekisterinpitäjä pseudonymisoisi erään tietojoukon, joka sisältää henkilötietoja, ja antaa tämän kolmannelle osapuolelle, tuloksena on silti henkilötietoja sisältävä tietojoukko. Vain jos rekisterinpitäjä koostaisi tiedot tietojoukosta tarpeeksi yleiselle tasolle, jolloin henkilöitä ei voida tunnistaa, tietojoukkoa voitaisiin pitää anonyymina. Tietosuojatyöryhmä käyttää esimerkkinä organisaatiota, joka kerää matkustustietoja yksilötasolla. Yksittäiset matkustusreitit ja -ajat muodostuvat henkilötietoja sisältävän rekisterin. Jos tämä organisaatio piilottaa varsinaiset henkilön tunnisteet kuten sosiaaliturvatunnukset tiedoista ja luovuttaa sen eteenpäin, kyseistä tietojoukkoa voidaan silti pitää henkilötietoina, koska niistä voidaan tunnistaa yksittäiset henkilöt. Jos organisaatio poistaisi tällaisen tapahtumatason datan ja sen sijaan tarjoaisi tietoa, kuten ”maanantaisin reitillä X on 160% enemmän matkustajia kuin tiistaisin”, sitä ei pidettäisi enää henkilötietoina. Tehokas anonymisointitekniikka siis estää sekä rekisterinpitäjää, että kolmatta osapuolta tunnistamista yksilöitä tietojoukosta. [57]

Tietosuojatyöryhmä tunnistaa kolme erilaista riskiä anonymisointiin liittyen: yksilöitävyys (*singling out*), yhdisteltävyys (*linkability)* ja pääteltävyys (*inference*). Yksilöitävyys viittaa mahdollisuuteen löytää yksilö tai yksilöitä tietojoukosta. Yhdisteltävyys viittaa mahdollisuuteen löytää yhdistäviä tekijöitä, joiden avulla yksilö voidaan tunnistaa. Jos tietojoukosta on löydettävissä tekijöitä, joiden perusteella voidaan päätellä, että kaksi eri tietuetta kuuluvat samaan yksilöiden joukkoon, mutta yksittäisiä yksilöitä ei voida tunnistaa, kyseinen anonymisointitekniikka tarjoaa suojaa yksilöitävyyttä vastaan, mutta ei yhdisteltävyyttä vastaan. Pääteltävyys puolestaan viittaa mahdollisuuteen päätellä huomattavalla todennäköisyydellä tietyn tietojoukon attribuutin[58] arvo isommasta joukosta attribuutteja. [59] Kaikkia kolme riskiä liittyvät luonnollisen henkilön paljastumiseen tietojoukosta. Voidaan siis puhua paljastumisriskistä.

Työryhmän lausunnon mukaan anonymisointia voidaan lähestyä yleensä kahdella eri tavalla: satunnaistamismenetelmä tai yleistäminen. Lisäksi työryhmä nostaa esiin myös seuraavat: pseudonymisointi, *differential privacy* -tekniikka, l-diversiteetti ja *t-closeness*

[56] WP216 2014, 8-9.
[57] WP216 2014, 9.
[58] Attribuutilla viitataan yleisesti tietotekniikassa entiteetin ominaisuuteen.
[59] WP216 2014, 11-12.

-tekniikka. Satunnaistamismenetelmät ovat joukko menetelmiä, jotka muokkaavat tietojen totuudenmukaisuutta häivyttämällä datan ja yksilön välisiä relaatioita. Satunnaistaminen ei itsessään poista datasta tietoja rivi- tai tapahtumatasolla, koska jokainen tietue on edelleen johdettu yksilöistä. Satunnaistaminen vähentää kyllä pääteltävyyteen liittyviä riskejä. Lisämenetelmiä voidaan tarvita, jotta yksilöä ei voida tunnistaa satunnaistetusta tietojoukosta. [60]

Satunnaistamisen eräs tekniikka on niin sanottu melun lisäys (*noise addition*). Tekniikkaa hyödyntämällä tietojoukon arvoja muutetaan niin, etteivät ne ole enää niin tarkkoja ja siten yksilöiviä. Esimerkiksi yksilön pituus on alun perin mittausdatassa senttimetrin tarkkuudella, mutta anonymisoidussa tietojoukossa, joka on käsitelty melun lisäys -tekniikalla, pituus voikin olla 10 senttimetrin tarkkuudella. Tekniikka yksinään ei ole täysin riittävä vaan vaatii lisäkseen esimerkiksi tiettyjen tunnisteiden poistamisen. Satunnaistamisen toinen tekniikka ovat permutaatiot.[61] Permutaatio -tekniikassa tietojoukkoa käsitellään niin, että joukon yksittäisten havaintojen arvoja sekoitetaan keskenään. Näin joukon sisällä arvojen jakaumat säilyvät, mutta ne linkitetään eri havaintoihin. Tällöin korrelaatiot yksilöiden ja arvojen välillä tietysti häviävät, mutta jakauma ja havaintovälit eivät. Joissain tapauksissa voi olla tärkeää arvojoukkoja sekoitetaan keskenään yhden arvon sijaan. [62]

Differential privacy -tekniikkaa voidaan myös pitää eräänä satunnaistamisen tekniikkana. Tässä tekniikassa rekisterinpitäjä tuottaa anonymisoidun näkymän dataan, mutta pitää itsellään alkuperäisen kopion tiedoista. Anonymisoitu näkymä dataan voi olla alkuperäisen tietojoukon alajoukko, johon on lisätty niin sanottua melua. Tässä tekniikassa kolmas osapuoli ei saa siis suoraan pääsyä tietojoukkoon tai aineistoon vaan osapuoli voi eri kyselyn perusteella saada kyselyn tulokset, joka on siis anonymisoitua. [63]

Yleistäminen on toinen anonymisointitekniikoiden ryhmä, joka koostaa menetelmistä, joilla voidaan vaimentaa yksilöivien tietojen relaatioita tai niiden yleistämisestä. Vaikka yleistäminen on yleisesti ottaen tehokas menetelmä yksilöitävyyden riskien pienentämiseen, se ei tarjoa tehokasta anonymisointia kaikissa tapauksissa. Aggregointi ja k-anonymiteetti ovat tekniikoita, joiden tavoitteena on estää rekisteröidyn yksilöiminen ryhmittelemällä heitä ainakin *k* muun yksilön kanssa samaan ryhmään. Tämän saavuttamiseksi havainnot yleistetään riittävälle tasolle niin, että jokaisella henkilöllä on sama arvo. Esimerkiksi kaupungin sijaan yleistämällä havainnot maan tarkkuudelle saadaan samaan aineistoon suurempi määrä henkilöitä. Vastaavasti vaikkapa syntymäajat voidaan yleistää päivämääräväleihin tai ryhmitellä esimerkiksi kuukausiin. K-

[60] WP216 2014, 12.
[61] WP216 2014, 12-13
[62] WP216 2014, 13-15.
[63] WP216 2014, 15-16. Differential privacy -tekniikalla ei ole vakiintunutta suomenkielistä vastinetta. Mielestäni yksi mahdollinen termi voisi olla eriytetty yksityisyys.

anonymiteettiin perustuvan tekniikan suurin puute onkin se, että se ei poista pääteltävyyteen liittyviä riskejä. Esimerkiksi jos hyökkääjä tietää mihin ryhmään tietty henkilö kuuluu, on helppoa hakea tällaisesta aineistosta kyseiseen henkilöön liittyvät arvot. [64]

L-diversiteetti laajentaa k-anonymiteettia, jotta voidaan varmistua siitä, että aineistossa ei ole yhdisteltäviä tekijöitä, joiden perusteella mahdollinen hyökkääjä voisi yksilöidä aineistosta henkilöitä. Tällä tekniikalla varmistetaan, että jokaisessa kategoriassa tai luokassa attribuuteilla eli havaintojen ominaisuuksilla on vähintään *l* määrä eri arvoja. Tavoitteena on siis vähentää sellaisia havaintojen luokkia, joissa on vähän havaintojen eri arvoja. Näin hyökkääjälle jää aina tietty epävarmuus yhdisteltävyydestä. L-diversiteetti -tekniikka on hyödyllinen yhdisteltävyyteen liittyvien riskien torjumiseen, kun tietojoukon arvot ovat jakautuneet hyvin. T-läheisyys tarkentaa l-diversiteettia edelleen. Se pyrkii luomaan yhteismitallisia luokkia, jotka muistuttavat alkuperäisen datan jakaumaa. Tätä tekniikka käytetään, kun anynomisoitu data halutaan pitää mahdollisimman alkuperäisen kaltaisena. Sen lisäksi *l* määrä eri arvoja pitää esiintyä yhteismitallisissa luokissa, jokaisen arvon täytyy esiintyä niin usein kuin tarpeen, jotta anonymisoidun datan vastaavien arvojen jakauma vastaa alkuperäisen datan vastaavien arvojen jakaumaa.[65]

## 3.3 Henkilötietojen pseudonymisointi

Pseudonymisointi tarkoittaa yleensä yhden tietueen korvaamista toisella. Tämä tietue on tyypillisesti jokin uniikki tai yksilöivä id. Luonnollinen henkilö voidaan pseudonymisoinnin jälkeen tunnistaa epäsuorasti, joten pseudonymisointi ei yksinään tee tietojoukosta anonyymia. Tietosuojatyöryhmä ottaa kuitenkin lausunnossaan kantaa siihen, koska se usein sekoitetaan anonymisointiin. [66]

Pseudonymisointi vähentää tietojen yhdisteltävyyteen liittyviä riskejä ja siten riskiä siitä, että luonnollinen henkilö olisi mahdollista tunnistaa tietojoukosta. Se on siis kelvollinen tietoturvan keino ja osoitus rekisterinpitäjän tietojenkäsittelyn velvollisuuksien hoitamisesta. Pseudonymisoinnin tuloksena voi olla alkuperäisestä tietojoukosta itsenäisiä arvoja, kuten silloin kun tietojoukossa korvataan esimerkiksi henkilötunnus satunnaisella numerolla. Pseudonymisoinnin tuloksena voi syntyä myös uusia arvoja, jotka perustuvat alkuperäisiin tietojoukon arvoihin. Esimerkiksi niin sanotun tiivistefunktion

[64] WP216 2014, 16-17.
[65] WP216 2014, 18.
[66] WP216 2014, 20.

(*hash*) tai muun salaustoiminnon avulla tuotettu arvo perustuu aina johonkin alkuperäisen tietojoukon arvoihin. [67]

Tietosuojatyöryhmän lausunnon[68] mukaan yleisemmin käytetyt pseudonymisointitekniikat ovat: tietojen salaus salausavaimella, tiivistefunktiolla ja näiden yhdistelmä salausavaimella suojattu tarkistussumma. Salausavaimen käyttö tietojen salaamiseen eli kryptaamiseen tarkoittaa sitä, että tietoja ei pääse tarkastelemaan ilman salausavainta. Henkilötiedot ovat siis edelleen tietojoukossa, mutta salatussa muodossa ja vain salausavaimella tietoja voi lukea ja käsitellä. Tiivistefunktion tarkoitus on, että funktio palauttaa mistä tahansa otteesta määrämittaisen palautusarvon ja tätä arvoa ei voi sellaisenaan lukea tai käsitellä ihmisen toimesta. Tiivistefunktio on kuitenkin altis monenlaisille hyökkäyksille. Jos hyökkääjä tietää riittävästi alkuperäisestä tietojoukosta, on mahdollista, että muodostetaan kaikille mahdollisille arvoille uudet tiivistefunktion arvot. Sitten näitä arvoja voi verrata käsiteltävään aineistoon ja löytää sieltä säännönmukaisuudet. Lisäksi työryhmä nostaa esiin myös deterministinen salauksen ja tokenisoinnin. [69]

Tietosuojatyöryhmän lausunnon[70] mukaan kaikkien pseudonymisointitekniikoiden heikkoutena on edelleen yksilöitävyys. Vaikka itse yksilöivä tieto kuten henkilöturvatunnus muutettaisiin jonkinlaiseksi numerosarjaksi (pseudonymisoitu tieto), josta ei suoraan voida henkilötunnusta palauttaa, se on silti yksilöivä tieto. Samoin yhdisteltävyys on riski pseudonymisointiakin käytettäessä. Samaa pseudonymisoitua tietoa käyttäen eri tietoja yhdistelemällä voidaan yksilöitä tunnistaa. Vain jos tietojoukossa ei ole arvoja, jotka voitaisiin yhdistelemällä palauttaa yksilöön, ja pseudonymisoitujen arvojen ja niiden alkuperäisten arvojen relaatiot on poistettu, voidaan sanoa, että yhdisteltävyys ei ole riski. Myös pääteltävyyteen liittyvät riskit ovat pseudonymisoinninkin jälkeen edelleen olemassa.

Tietosuojatyöryhmä korostaa, että pseudonymisoitu aineisto ei ole tietosuoja-asetuksen tarkoittamalla tavalla anonymisoitua ja siten sitä tulee kohdella kuin henkilötietoja. Pseudonymisointi vähentää tiettyjä riskejä huomattavasti ja on selvää, että pseudonymisoitu tietojoukko ei ole juridisesti täysin sama asia kuin suojaamattomat henkilötiedot. Pseudonymisointi sekä anonymisointitekniikat eivät ole aukottomia ja niistä jää aina jäännösriski. On kuitenkin rekisterinpitäjän vastuulla arvioida, onko näiden tekniikoiden jäännösriskiä riittävästi lievennetty. Arviointi eri tekniikoiden käyttämisestä tulee tehdä aina tapauskohtaisesti. [71] Valitettavasti ei ole olemassa yleisesti hyväksyttyä menetelmää arvioida eri tekniikoiden turvallisuutta eikä ole myöskään olemassa vastaa-

[67] WP216 2014, 20.
[68] WP216 2014, 20-21.
[69] WP216 2014, 20. Alkuperäiset englanninkieliset termit ovat *encryption with secret key, hash function*, *keyed-hash function with stored key, deterministic enrcryption or keyed-has function with deletion of the key* ja *tokenization.*
[70] WP216 2014, 21.
[71] WP216 2014, 23.

vasti menetelmää arvioida, mikä on ”kohtuullinen” aika ja työmäärä, joka vaaditaan anonymisoidun datan käsittelemiseksi niin, että luonnollinen henkilö voitaisiin taas tunnistaa. [72]

## 3.4 Tilastolliset tietosuojamenetelmät

Tietosuojatyöryhmä[73] määritteli luonnollisen henkilön paljastumisriskin kolmen riskikomponentin kautta, jotka olivat yksilöitävyys, yhdisteltävyys ja pääteltävyys. Paljastmisriskiä voidaan lieventää erilaisilla pseudonymisointi- ja anonymisointitekniikoilla. Nämä tekniikat aiheuttavat aineistoon tietynlaista informaatiokatoa ja vähentävät siten sen käyttökelposuutta. Paljastumisriskin ja informaatiokadon optimointi ovatkin olennaisia näkökulmia, kun tietoa halutaan antaa ulkopuolisille.

Avoimen datan vaatimukset[74] toisaalta mahdollistavat uusia asioita, kun enemmän tietoa on julkisesti saatavilla. Toisaalta avoin data mahdollistaa aiempaa enemmän erilaisten tietojen yhdistelemistä ja siihen perustuen päättelyiden tekemistä, mikä lisää luonnollisen henkilön paljastumisriskiä. Etenkin niin sanotun mikrodatan julkistaminen lisää tällaista riskiä. Mikrodatalla viitataan tietoon, joka koostuu yksittäisistä henkilöistä. Rivitasolla tiedot voidaan peittää tai vaikkapa henkilöturvatunnukset voidaan muuntaa erilaiseksi, mutta tiedot ovat kuitenkin tiettyjä yksilöitä koskevia. Jos tietojoukossa on riittävästi erilaisia attribuutteja tai ominaisuuksia, pseudonymisoitukin tietojoukko voidaan palauttaa riittävällä tilastollisella todennäköisyydellä yksittäisiin henkilöihin. [75]

Mikrodataan liittyvät riskit realisoituivat, kun vuonna 2006 Netflix, silloin maailman suurin DVD vuokraaja, ilmoitti kilpailusta, jossa oli mahdollisuus voittaa miljoona dollaria henkilölle, joka onnistuisi parantamaan parhaiten Netflixin suosittelupalvelua. Tätä varten Netflix julkisti tietojoukon, joka sisälsi yli 100 miljoonaa yksittäistä elokuva-arvostelua yhteensä yli 480 000 käyttäjältä, jotka olivat käyttäneet palvelua vuosien 1999 ja 2005 välisenä aikana. Julkaistut tiedot olivat Netflixin mukaan anonymisoitu. Netflix oli ottanut myös kantaa arvostelujen tietosuojaan kertomalla, että varsinaisten arvostelujen sisältö ja niiden ajankohdat eivät ole riittäviä tunnistamaan henkilöitä koko tietojoukosta. Lisäksi Netflix oli sisällyttänyt vain murto-osan keräämistään tiedoista julkiseen tietojoukkoon ja lisäksi tietoja oli käsitelty siten, että siihen oli lisätty niin sanottua häiriötä (*noise addition*) tunnistamisen hankaloittamiseksi. Narayanan & Shama-

[72] WP216 2014, 27.
[73] WP216 2014.
[74] Avoimella datalla tarkoitetaan aineistoa, joka on avointa ja sen jakelutapa täyttää seuraavat ehdot: saavutettavuus, uudelleenjakelu, uudelleenkäyttö ja vapaa teknisistä rajoitteista. Kts. tarkemmin Open Knowledge Foundationin määritelmä, saatavilla http://opendefinition.org/od/1.1/fi/ . Viitattu 18.12.2019.
[75] Narayanan & Shmatikov 2008, 1.

tikov toteuttivat tapaustutkimuksen vuonna 2008, jossa he arvioivat, että voisiko Netflixin julkistamista tiedoista paljastaa yksittäisiä henkilöitä, vaikka tietyt yksilöivät tiedot olikin pseudonymisoitu. Heidän tutkimuskysymyksensä oli, kuinka paljon taustatietoa tarvittaisiin, jotta yhden käyttäjän voisi tunnistaa julkisista tiedoista. [76]

Tutkijat käyttivät kehittämäänsä *Scoreboard-RH* -algoritmia[77] ja yrittivät deanonymisoida Netflixin julkistaman tietojoukon. Tutkijat käyttivät lisätietoina IMDb:n (Internet Movie Database)[78] julkisia käyttäjätietoja. Lopputuloksena he totesivat, että jo pienellä määrällä taustatietoja voitiin tunnistaa keskimääräinen käyttäjä Netflixin tietojoukosta. Tutkijat pystyivät yksilöimään 99% julkistetusta tietojoukosta yhdistämällä Netflixin julkistamat anonymisoidut tiedot IMDb:n julkisiin tietoihin. 99 % kattavuuteen päästiin vertailemassa kahdeksan eri elokuvan arvosteluja ja arvostelujen ajankohtia. Jopa kahdella elokuva-arvostelulla ja niiden ajankohdilla voitiin yksilöidä 68 % tietojoukosta. Etenkin, jos elokuva ei ollut laaja hitti, sen arvostelulla voitiin tunnistaa käyttäjä huomattavasti tarkemmin. Tutkijat huomattavat myös, että tässä tapauksessa kyse oli laajasta tietojoukosta. Mikäli hyökkääjä haluaisi tunnistaa vain jonkin tietyn henkilön laajasta tietomassasta, hän voisi tehdä myös helposti pienemmillä taustatiedoilla. [79]

Tutkimuksessa keskustellaan myös siitä, että mikä merkitys elokuva-arvosteluilla on yksityisyyden suojan näkökulmasta. Nykyisen lainsäädännön näkökulmasta ensisijainen kysymys on, että ovatko tällaiset tiedot henkilötietoja ja kuka on rekisterinpitäjä. Yleisen tietosuoja-asetuksen näkökulmasta se voisi olla asetuksen tarkoittamalla tavalla henkilötietoa ja siten sen käsittelyä koskisi myös asetuksen periaatteet ja säännökset. Henkilötieto on kuitenkin käsitteenä asetuksessa melko laaja ja tietyissä määrin jopa tulkinnanvarainenkin. Erilaiset tietosuojaan liittyvät velvollisuudet koskettavat vain rekisterinpitäjää, jonka tulee arvioida, voiko tiettyjä tietoja julkistaa ylipäätään. Se, että kolmas osapuoli tutkii ja käsittelee niitä, on siitä erillinen asia.

Olisi myös syytä arvioida, olisiko vastaavan tutkimuksen toteuttaminen ylipäätään mahdollista tietosuojan näkökulmasta. Vaikka kyseessä onkin akateeminen tutkimus, sen todennäköisesti tulisi noudattaa asetuksen mukaisia tietosuojasäännöksiä. Netflixin ja IMDb:n tietojen yhdistäminen voisi olla kuluttajan näkökulmasta asetuksen tarkoittamalla tavalla yllättävää. Toisaalta kun kyseessä on tilastollinen tutkimus, jonka tarkoitus on osoittaa mahdollisia haavoittuvuuksia, se voisi olla perusteltu.

---

[76] Narayanan & Shmatikov 2008, 8.
[77] Algoritmi tuottaa scoren eli pisteytyksen, kuinka hyvin syötteenä annetulla tiedolla löytyy annetuista lisätiedoista vastineita. Pisteytys kuvaa siis sitä, miten paljon vastineita löytyy raakadatan ja annettun lisätietojen välillä. Ks. tarkempi kuvaus algoritmista Narayanan & Shmatikov 2008, 5-6.
[78] IMDb eli Internet Movie Database on julkisesti saatavilla oleva, hyvinkin tunnettu tietokanta elokuvista, niiden arvosteluista ja esimerkiksi elokuvien lyhyistä esittelyvideoista (trailerit). Ks. tarkemmin www.imdb.com.
[79] Narayanan & Shmatikov 2008, 10-11.

Pseudonymisoinnin ja anonymisoinnin tekniikoiden käyttämistä tulisi harkita ja arvioida tarkoin tapauskohtaisesti. Silloin, kun aineistoa muutetaan, on myös riski, että yksilö voidaan tunnistaa ja hyökkääjälle paljastuu yksilöstä väärää tietoa.

# 4 HENKILÖTIETOJEN KERÄÄMISESTÄ

## 4.1 Tietosuojaa koskeva vaikutustenarviointi

Yleisen tietosuoja-asetuksen 35 artiklassa esitellään tietosuojaa koskevan vaikutustenarvioinnin käsite. Tällaisella vaikutustenarvioinnilla on tarkoitus kuvata henkilötietojen käsittelyä, arvioida sen tarpeellisuutta ja oikeasuhteisuutta sekä tietenkin tukea luonnollisten henkilöiden oikeuksiin ja vapauksiin kohdistuvaa riskienhallintaa. Siinä arvioidaan riskit ja määritellään toimenpiteet, joilla niihin puututaan. [80]

Vaikutustenarviointi on toisaalta menettely, jolla parannetaan vaatimusten noudattamista ja toisaalta tapa, jolla osoitetaan vaatimustenmukaisuus.[81] Osoitusvelvollisuudella tarkoitetaan sitä, että rekisterinpitäjän pitää osoittaa sen noudattavan yleisen tietosuoja-asetuksen vaatimuksia. Tietosuojaa koskeva vaikutusten arviointi on siis suhteessa tärkeä osa osoitusvelvollisuutta. [82] Yleisen tietosuoja-asetuksen jakso 3 kuvaa rekisterinpitäjän velvollisuutta arvioida tietosuojaa koskevia vaikutuksia:

> Jos tietyntyyppinen käsittely etenkin uutta teknologiaa käytettäessä todennäköisesti aiheuttaa – käsittelyn luonne, laajuus, asiayhteys ja tarkoitukset huomioon ottaen – luonnollisen henkilön oikeuksien ja vapauksien kannalta korkean riskin, rekisterinpitäjän on ennen käsittelyä toteutettava arviointi suunniteltujen käsittelytoimien vaikutuksista henkilötietojen suojalle.[83]

Riskillä tarkoitetaan skenaariota, jolla kuvataan tietojen käsittelyyn liittyvää tapahtumaa ja sen seurauksia sekä arvioidaan niiden todennäköisyyttä ja vakavuutta. Luonnollisen henkilön oikeuksilla ja vapauksilla viitataan puolestaan oikeutta tietosuojaan ja oikeutta yksityisyyteen, mutta se vois myös toisaalta pitää sisällään myös muitakin perusoikeuksia kuten sananvapauden tai syrjintäkiellon. Joka tapauksessa vaikutustenarvioinnin (DPIA) tekeminen on riskiperusteista ja siten arviointi ei ole pakollista kaikkien käsittelytoimien osalta. Säännöksen[84] tekstissä mainittu ”*todennäköisesti aiheuttaa*” viittaa rekisterinpitäjän yleiseen velvollisuuteen toteuttaa toimenpiteitä, joilla voidaan hallita rekisteröityjen oikeuksiin ja vapauksiin kohdistuvia riskejä. Käytännössä tämä tarkoittaa jatkuvaa käsittelytoimien riskiperusteista valvontaa. [85]

---

[80] WP248 2017, 4.
[81] WP248 2017, 4. Kts. myös yleisen tietosuoja-asetuksen johdantokappale 84: ”Arvioinnin tulos olisi otettava huomioon määriteltäessä asianmukaisia toimia, jotka on toteutettava, jotta voidaan osoittaa, että henkilötietojen käsittely on tämän asetuksen säännösten mukaista”.
[82] WP251 2018, 31.
[83] Yleinen tietosuoja-asetus artikla 35.
[84] Yleinen tietosuoja-asetus artikla 35 kohta 1.
[85] WP248 2017, 7.

Vaikutustenarviointi voi koskea lähtökohtaisesti vain yhtä henkilötietojen käsittelytoimea. Asetuksen 35 artiklan 1 kohdassa todetaan kuitenkin, että ”*[y]htä arviota voidaan käyttää samankaltaisiin vastaavia korkeita riskejä aiheuttaviin käsittelytoimiin*”. Lisäksi asetuksen johdanto-osan kappaleessa 92 todetaan, että joissain tilanteissa voi olla järkevää ja taloudellista tehdä vaikutustenarviointi, jossa tarkastellaan asioita laajemmin kuin vain yhden käsittelytoimen kannalta. Yhtä yksittäistä vaikutustenarviointia voidaan käyttää useiden käsittelytoimien arviointiin, mikäli niiden luonne, asiayhteys, tarkoitus ja riskit ovat samankaltaisia. Vaikutustenarviointia ei siis tarvitse tehdä, jos sellainen on jo tehty samankaltaiseen käsittelytoimeen aiemmin. [86]

Säännöksessä korostetaan velvollisuutta etenkin uuden teknologian käyttöönoton yhteydessä. Useissa yhteyksissä vaikutustenarvioinnista käytetään myös lyhennettä DPIA (*Data Protection Impact Assesment*). DPIA:n yhteydessä rekisterinpitäjän on konsultoitava tietosuojavastaavaa, jos sellainen on organisaatioon nimetty[87]. DPIA vaaditaan 35 artiklan 3 kohdan mukaan erityisesti seuraavissa tapauksissa:

a) luonnollisten henkilöiden henkilökohtaisten ominaisuuksien järjestelmällinen[88] ja kattava arviointi, joka perustuu automaattiseen käsittelyyn, kuten profilointiin, ja johtaa päätöksiin, joilla on luonnollista henkilöä koskevia oikeusvaikutuksia tai jotka vaikuttavat luonnolliseen henkilöön vastaavalla tavalla merkittävästi;
b) laajamittainen käsittely, joka kohdistuu 9 artiklan 1 kohdassa tarkoitettuihin erityisiin henkilötietoryhmiin tai 10 artiklassa tarkoitettuihin rikostuomioita tai rikkomuksia koskeviin tietoihin; tai
c) yleisölle avoimen alueen järjestelmällinen valvonta laajamittaisesti.

A kohta viittaa nimenomaan siis luonnollisten henkilöiden henkilökohtaisten ominaisuuksien järjestelmälliseen ja kattavaan arviointiin, joka johtaa päätöksiin, joilla on oikeusvaikutuksia tai jotka vaikuttavat muuten merkittävästi. B kohta viittaa puolestaan erityisiin henkilötietoryhmiin kuten rekisteröidyn poliittiseen mielipiteeseen. C kohdalla viitataan taas erilaisiin valvontajärjestelmiin.

Yleisen tietosuoja-asetuksen 35 artiklan kohtien 5 ja 6 mukaan valvontaviranomainen voi laatia luettelon henkilötietojen käsittelytoimien tyypeistä, joista ainakin vaaditaan DPIA tai vastaavasti luettelon, joista ei nimenomaan vaadita DPIA:ta. Nämä luette-

[86] WP248 2017, 8.
[87] Kts. yleinen tietosuoja-asetus 4 jakso, artiklat 37-39. Rekisterinpitäjän tulee nimittää tietosuojavastaava mm. kun ” rekisterinpitäjän tai henkilötietojen käsittelijän ydintehtävät muodostuvat käsittelytoimista, jotka luonteensa, laajuutensa ja/tai tarkoitustensa vuoksi edellyttävät laajamittaista rekisteröityjen säännöllistä ja järjestelmällistä seurantaa”.
[88] Tietosuojatyöryhmän (WP243 2016, 10) tulkinnan mukaan ”järjestelmällinen” viittaa yhteen tai useampaan seuraavista:

- Jotain järjestelmää noudattaen tapahtuva
- ennalta järjestetty, organisoitu tai menetelmällinen
- osana tietojen keruuta koskevaa yleissuunnitelmaa tapahtuva
- osana strategiaa noudatettava

lot on toimitettava tietosuojaneuvostolle[89]. Jos ei ole selvää, että pitääkö arviointi tehdä vai ei, tietosuojatyöryhmä suosittelee, että se tehdään kuitenkin. Arviointi auttaa rekisterinpitäjiä noudattamaan tietosuojalainsäädäntöä. [90]

Yleisen tietosuoja-asetuksen 35 artiklan kohdan 7 mukaan DPIA:n tulee sisältää ainakin:

| | |
|---|---|
| a) | järjestelmällinen kuvaus suunnitelluista käsittelytoimista, ja käsittelyn tarkoituksista, mukaan lukien tarvittaessa rekisterinpitäjän oikeutetut edut; |
| b) | arvio käsittelytoimien tarpeellisuudesta ja oikeasuhteisuudesta tarkoituksiin nähden; |
| c) | arvio 1 kohdassa tarkoitetuista rekisteröityjen oikeuksia ja vapauksia koskevista riskeistä; ja |
| d) | suunnitellut toimenpiteet riskeihin puuttumiseksi, mukaan lukien suoja- ja turvallisuustoimet ja mekanismit, joilla varmistetaan henkilötietojen suoja ja osoitetaan, että tätä asetusta on noudatettu ottaen huomioon rekisteröityjen ja muiden asianomaisten oikeudet ja oikeutetut edut. |

Vaikutustenarviointia varten on myös annettu ohjeellisia pohjia, joita voi käyttää vaikutustenarvioinnin analysoinnissa[91].

Mikäli DPIA osoittaa, että henkilötietojen käsittely kuvatulla tavalla aiheuttaisi korkean riskin, jos rekisterinpitäjä ei ole toteuttanut toimenpiteitä riskin pienentämiseksi, tulee rekisterinpitäjän kuultava ennakkoon valvontaviranomaista. Jos valvontaviranomainen katsoo, että suunniteltu käsittely rikkoisi yleistä tietosuoja-asetusta, erityisesti jos rekisterinpitäjä ei ole tunnistanut riittävästi riskejä tai pienentänyt niitä, on valvontaviranomaisen annettava ohjeet rekisterinpitäjälle kirjallisesti. [92]

Sen lisäksi, että DPIA voidaan osoittaa vaatimustenmukaisuus, se on myös hyvä väline rekisterinpitäjälle. Rekisterinpitäjä voi määrittää, mitä toimenpiteitä se toteuttaa käsittelyyn liittyvien tietosuojariskien torjumiseksi. [93]

### *4.1.1 Korkeariskiset käsittelytoimet*

Yleisen tietosuoja-asetuksen 35 artiklan 3 kohdassa annettujen, edellä kuvattujen esimerkkien lisäksi tietosuojatyöryhmä on tunnistanut muitakin tapauksia, joissa DPIA olisi välttämätön.

Tietosuojatyöryhmän ohjeistuksessa luetellaan eri tilanteita esimerkinomaisesti, joissa ainakin vaaditaan vaikutusten arviointia. Ensimmäinen esimerkki kuuluu: ”*erityisesti ’rekisteröidyn työsuorituksen, taloudellisen tilanteen, terveyden, henkilökohtaisten mieltymysten tai kiinnostuksen kohteiden, luotettavuuden tai käyttäytymisen, sijainnin tai liikkumisen’ arviointi tai pisteytys, mukaan lukien profilointi ja ennakointi*”. Käytännön

[89] Kts. tietosuojaneuvoston jäsenistö https://edpb.europa.eu/about-edpb/board/members_fi, viitattu 25.10.2019.
[90] WP248 2017, 9.
[91] Kts. https://gdpr.eu/data-protection-impact-assessment-template/ , viitattu 25.10.2019.
[92] Yleisen tietosuoja-asetuksen 36 artikla. Suomessa valvontaviranomainen on tietosuojavaltuutettu.
[93] WP251 2018, 32.

sovellus voisi olla rahoitusalan laitos, joka arvioi asiakkaitaan luottotoimintaan liittyvän tietokannan, terrorismin rahoituksen torjumiseen liittyvän tietokannan ja petoksia koskevan tietokannan valossa. [94]

Toinen tietosuojatyöryhmän antama esimerkki on ”*automaattinen päätöksenteko, jolla on oikeusvaikutuksia tai vastaavia merkittäviä vaikutuksia: käsittely, jonka tavoitteena on tehdä rekisteröidyistä päätöksiä, joilla on 'luonnollista henkilöä koskevia oikeusvaikutuksia' tai jotka 'vaikuttavat luonnolliseen henkilöön vastaavalla tavalla merkittävästi' ”*. Tällaisia vaikutuksia voivat olla henkilön ulkopuolelle jättäminen tai syrjintä. Käsittely, joka vaikuttaa henkilöön vain vähän tai ei ollenkaan, ei täytä tätä kriteeriä. [95]

Kolmas esimerkki on ”*järjestelmällinen valvonta: rekisteröityjen tarkkailuun, seurantaan tai valvontaan käytettävä tietojenkäsittely sekä tietojen kerääminen verkkojen välityksellä tai 'yleisölle avoimen alueen järjestelmällinen valvonta laajamittaisesti'*”. Tässä viitataan tilanteisiin, joissa rekisteröidyt eivät välttämättä ole täysin tietoisia siitä, että kuka kerää heidän tietojaan ja mihin tarkoituksiin. Toisaalta myös yksittäisen henkilön voi olla mahdotonta olla joutumatta tällaisen tietojenkäsittelyn kohteeksi esimerkiksi julkisissa tiloissa. [96]

Tietosuojatyöryhmä jatkaa, että DPIA tulisi tehdä myös aina, kun käsitellään arkaluontoisia tietoja tai luonteeltaan hyvin henkilökohtaisia tietoja, joihin kuuluvat yleisen tietosuoja-asetuksen 9 artiklassa määritellyt *erityiset henkilötietoryhmät* (kuten tiedot henkilöiden poliittisista mielipiteistä) sekä *rikostuomioihin ja rikkomuksiin liittyvät henkilötiedot*, jotka määritellään puolestaan 10 artiklassa. Esimerkiksi potilastietojen ylläpitämiseksi yleisessä sairaalassa tai vaikkapa rikoksentekijätietoja etsivän yksityisetsivän tulisi tehdä DPIA. Tällaiset tiedot katsotaan arkaluontoisiksi, koska ne liittyvät yksityistä kotitaloutta koskevaan ja yksityiseen toimintaan, tai koska ne vaikuttavat perusoikeuden käyttämiseen tai koska niiden loukkaamiseen liittyy selvästi rekisteröidyn arkeen kohdistuvia vakavia vaikutuksia. [97]

Arviointia edellyttää myös aina asetuksen 35 artiklan 3 kohdan b alakohdan mukaan ”*tietojen laajamittainen käsittely*”. Tätä ei säännöksessä määritellä tarkemmin, mutta asetuksen 91 johdantokappaleessa annetaan joitakin ohjeita siihen liittyen. Johdantokappaleessa todetaan, että asetusta tulisi soveltaa erityisesti ”- - *laajoihin käsittelytoimiin, joissa on tarkoitus käsitellä huomattavia määriä henkilötietoja alueellisella, kansallisella tai ylikansallisella tasolla, jotka voivat vaikuttaa suureen määrään rekisteröityjä ja joihin todennäköisesti liittyy korkea suuri riski esimerkiksi tietojen arkaluonteisuuden vuoksi* - -”. Laaja käsittely voi siis viitata maantieteelliseen laajuuteen, mutta

[94] WP248 2016, 10.
[95] WP248 2016, 10. Kts myös kappale 4 profilointiin liittyen.
[96] WP248 2016, 10.
[97] WP248 2016, 10.

siihen liittyy kuitenkin aina myös käsiteltävien henkilötietojen määrä sekä rekisteröityjen määrä, joita tietojenkäsittely koskee. [98]

Tietosuojatyöryhmä toteaa, että on käytännössä mahdotonta tarkkoja lukumääriä tai rajoja, joilla määritellään, milloin tietojenkäsittely on ”*laajamittaista*”. On kuitenkin mahdollista, että ajan kuluessa muodostuu vakiintuneet käytännöt, joiden pohjalta voidaan jatkossa määritellä tarkemmin mitä tällä tarkoitetaan tietyn tyyppisten käsittelytoimien yhteydessä. [99]

Kuitenkin tietosuojatyöryhmä[100] suosittelee, että ”laajamittaisen käsittelyn” arvioinnissa otetaan aina huomioon seuraavat tekijät:

| | |
|---|---|
| a) | asianomaisten rekisteröityjen lukumäärä, joko täsmällisenä lukuna tai osuutena kyseisestä väestöstä |
| b) | käsiteltävien tietojen määrä ja/tai käsiteltävien erillisten tietoyksikköjen määrä |
| c) | tietojenkäsittelytoimen kesto tai pysyvyys |
| d) | käsittelytoimen maantieteellinen ulottuvuus |

Korkeariskisenä tulisi myös tietosuojatyöryhmän mukaan pitää käsittelytoimea, joka sovittaa yhteen tietokokonaisuuksia tai yhdistää niitä ”- - *rekisteröidyn kohtuulliset odotukset ylittävällä tavalla**,** kun kyseessä ovat esimerkiksi kahdesta tai useammasta eri tarkoituksiin suoritetusta ja/tai eri rekisterinpitäjien suorittamasta tiedonkäsittelytoimesta peräisin olevat tietokokonaisuudet*”. [101] Tietosuojatyöryhmä rinnastaa kohtuulliset odotukset käyttötarkoitussidonnaisuuden arviointiin. [102] Arvioinnissa tulisi ottaa huomioon, mikä olisi yleisesti hyväksyttävää ja tavanomaista kyseisessä kontekstissa sekä rekisteröidyn ja rekisterinpitäjän välisessä suhteessa. Kun arvioidaan tätä suhdetta, tulisi tarkastella myös sitä, että oliko tiedot julkisia ja oliko vaikkapa jokin kolmas osapuoli pakottavan lainsäädännön takia pakotettu toimittamaan tietoja. Yleisesti voidaan kuitenkin todeta, että mitä yllättävämpää tietojen käyttö muihin tarkoituksiin olisi, sitä todennäköisemmin tietojen käyttötarkoitus ei vastaa sitä tarkoitusta, jota varten ne on kerätty. [103]

Arviointi tulee myös tehdä aina, kun käsittelytoimen kohteena on heikossa asemassa olevat rekisteröityjen tiedot. Heikolla asemalla viitataan rekisterinpitäjän ja rekiste-

[98] Yleinen tietosuoja-asetus 37 artikla 1 kohta. Kun tietojenkäsittely on ’laajamittaista’, rekisterinpitäjän tulee nimetä tietosuojavastaava.
[99] WP243 2016, 9.
[100] WP248 2016, 11.
[101] WP248 2016, 12.
[102] Asetuksen 5 artiklan 1 kohdan alakohdan b mukaan henkilötietoja on kerättävä vain tiettyä, nimenomaista ja laillista tarkoitusta eikä niitä saa käsitellä myöhemmin näiden vastaisesti (käyttötarkoitussidonnaisuus).
[103] WP203 2013, 24.

röidyn väliseen valtasuhteeseen. Tällöin rekisteröidyn ei ole helppo antaa suostumusta, vastustaa tietojensa käsittelyä tai käyttää oikeuksiaan. [104]

Arviointi tulee tehdä myös aina, kun sovelletaan uusia tai käytetään innovatiivisella tavalla teknisiä tai organisatorisia ratkaisuja. Tällaiseen uudenlaiseen käyttöön tai soveltamiseen liittyy usein uudenlaisia tapoja kerätä tietoa ja käyttää tietoa, joista voi johtua korkea riski. Esimerkiksi tietyillä sovelluksilla voi olla merkittävä vaikutus yksityisten henkilöiden arkielämään ja yksityisyyteen, joten tällaiset sovellukset edellyttävät DPIA:n tekemistä. [105]

Arviointia edellyttää myös aina käsittelytoimet, jotka ”- - *valvontaviranomainen katsoo, että käsittelyyn todennäköisesti liittyy korkea riski rekisteröityjen oikeuksien ja vapauksien kannalta erityisesti siitä syystä, että ne estävät rekisteröityjä käyttämästä oikeutta tai palvelua tai sopimusta.*” [106] Tällä tarkoitetaan toimia, joiden tavoitteena on sallia tai evätä rekisteröidyn oikeus käyttää palvelua, tehdä sopimus tai muuttaa kyseisestä oikeutta. Esimerkiksi pankin tekemä luottopäätös lainatarjouksen yhteydessä voisi olla tällainen tilanne. [107]

Mitä useampi edellä kuvatuista kriteereistä täyttyy, sitä todennäköisemmin käsittelytoimi aiheuttaa korkean riskin rekisteröityjen oikeuksille ja vapauksille ja tämän vuoksi DPIA tulisi tehdä. Rekisterinpitäjän tulisi huomioida kuitenkin, että jo vähintään kahden kriteerin täyttyminen edellyttää DPIA:ta. [108] Vaikka käsittelytoimi täyttäisikin tietyt kriteerit, rekisterinpitäjä voi silti katsoa, että se ei aiheuta ”korkeaa riskiä”. Tällöin rekisterinpitäjän tulisi kuitenkin dokumentoida syyt, miksi tähän on päädytty. Lisäksi perusteluihin tulisi kirjata myös tietosuojavastaavan näkemykset. [109]

### *4.1.2 Käyttötarkoitussidonnaisuus*

Yleisen tietosuoja-asetuksen 5 artiklan kohdan 1 alakohdan b määrittelee henkilötietojen käsittelyä koskevan käyttötarkoitussidonnaisuuden. Käyttötarkoitussidonnaisuudella[110] viitataan siihen, että rekisteröity ymmärtää, että tiedot on kerätty tiettyä tarkoitusta

[104] WP248 2016, 12. Kts myös yleisen tietosuoja-asetuksen johdantokappale 75. Erityisesti lapset luetaan tällaiseen heikompaan asemaan kuuluvaan rekisteröityjen ryhmään.
[105] WP248 2016, 12.
[106] Yleisen tietosuoja-asetuksen johdantokappale 91.
[107] WP248 2016, 12.
[108] WP248 2016, 12.
[109] WP248 2016, 14.
[110] Englanniksi *purpose limitation* viittaa mielestäni enemmän siihen, että käyttötarkoitus rajoittaa tietojen käsittelyä. Suomeksi sama termi *käyttötarkoitussidonnaisuus* antaa mielestäni enemmän liikkumavaraa eikä nimenomaisesti rajoita vaan sitoo tietojen keräämisen tiettyyn käyttötarkoitukseen niiden myöhemmän käsittelyn kanssa.

varten. Tämän tarkoituksen on oltava laillinen ja nimenomainen eikä tietoja saa käsitellä myöhemmin näiden tarkoitusten kanssa yhteensopimattomalla tavalla.

Tietosuojatyöryhmän[111] mukaan käyttötarkoituksen määritteleminen on olennaisin ja tärkein askel tietosuojatoimien käyttöönotossa. Tietojen käsittelyn käyttötarkoituksen määritteleminen vaikuttaa olennaisesti myös muihin tietojenkäsittelytoimien vaatimuksiin, kuten täsmällisyyteen ja säilytyksen rajoittamiseen. Tällä periaatteella on pyritty antamaan raameja, miten tiettyä käyttötarkoitusta varten kerättyjä tietoja voidaan käsitellä ja mihin sitä voidaan käyttää. Periaate voidaan jakaa kahteen komponenttiin:

- rekisterinpitäjä voi kerätä tietoja vain tiettyä, nimenomaista ja laillista tarkoitusta varten,
- datan keräämisen jälkeen, sitä ei voida käyttää tavalla, joka olisi epäyhteensopiva mainitun käyttötarkoituksen kanssa

Ensinnäkin käyttötarkoituksen tulee olla jokin *tietty* eli riittävän hyvin määritelty, jotta voidaan määritellä tarvittavat suojaustoimet ja jotta käsittelytoimi voidaan rajata. Toiseksi jotta käyttötarkoitus olisi *nimenomainen*, sen tulee riittävän selvästi ilmaistu ja yksiselitteinen. Kolmanneksi tietojenkäsittelyn käyttötarkoituksen tulle olla *laillinen*. Tämä viittaa sekä asetuksen 6 artiklan vaatimukseen käsittelyn lainmukaisuudesta, mutta myös muuhun mahdollisesti sovellettavaan lainsäädäntöön. On syytä huomata, että sekä uuden, että vanhan käsittelytoimen käyttötarkoituksen tulee täyttää nämä vaatimukset. [112] Käyttötarkoitus tulee olla myös määriteltynä ennen tietojenkäsittelytoimien aloittamista tai ainakin viimeistään, kun se alkaa. [113]

Käyttötarkoitussidonnaisuuden periaate suojaa myös kohtuullisia odotuksia ja luottamusta tietojenkäsittelytoimia kohtaan. Kun kuluttaja jakaa henkilötietojaan, hän voi kohtuudella luottaa siihen, että tietoja käsitellään tietyllä tavalla. Tämän tärkeän kulmakiviperiaatteen onkin tarkoitus myös estää niin sanottu *mission creep* [114] eli alkuperäisen tarkoituksen tai tavoitteen vähittäiseen muuttumiseen toiseksi. [115]

Toisaalta tietosuojatyöryhmä näkee arvoa myös siinä, että sallitaan tietojen käsittely – tietyissä rajoissa – myös alkuperäisestä tarkoituksesta poikkeavaan tarkoitukseen. Tiettyä tarkoitusta varten kerätyt tiedot voivat olla aidosti hyödyllisiä myös muissa tarkoituksissa, joita ei aiemmin ymmärretty. Asetuksessa mainittu ”*yhteensopimaton tapa*” ei kategorisesti rajaa kaikkia muita nimenomaisia käsittelyn käyttötarkoituksia pois,

---

[111] WP203 2013, 4.
[112] WP203 2013, 12.
[113] WP203 2013, 21.
[114] *Mission creep* on mainittu terminä tietosuojatyöryhmän englanninkielisessä analyysissa, mutta sille ei tarjota enempää selostusta. Merriam-Webster online-sanakirja määrittelee sen seuraavasti: ” *the gradual broadening of the original objectives of a mission or organization*”. Lähde: https://www.merriam-webster.com/dictionary/mission%20creep, viitattu 7.11.2019. Tästä voitaneen vetää johtopäätös, että termillä viitataan tilanteeseen, jossa tietojenkäsittelytoimen alkuperäinen käyttötarkoitus vähitellen muuttuu toisenlaiseksi kuin se oli alun perin tarkoitettu. Ei suomenkielistä vakiintunutta käännöstä.
[115] WP203 2013, 4.

kunhan uusi käsittelytapa on ”*yhteensopiva*” vanhan kanssa. Tietosuojatyöryhmän mukaan yhteensopivuuden analysoinnissa pitäisi pystyä määrittelemään käsittelytoimen sekä yhteensopiva, että yhteensopimattomat tavat käsitellä tietoa. Analyysissa tulisi ottaa huomioon myös ennustettavuus ja läpinäkyvyys rekisteröidyn näkökulmasta. [116]

Tietojen myöhemmällä käsittelyllä viitataan siis tietojenkäsittelytoimiin, joita suoritetaan tietojen keräämisen jälkeen, jolloin käsittelytoimien käyttötarkoitus on (ensimmäisen kerran) määritelty. Liittyen yhteensopivuuteen, lainsäätäjä ei ole antanut nimenomaisia yhteensopivuusvaatimuksia vaan säätänyt kiellon epäyhteensopivuudesta. Näyttäisi siis siltä, että lainsäätäjä on halunnutkin jättää hieman liikkumavaraa säädöksen tulkintaan. Vaikka tietojen myöhempi käsittely tehtäisiinkin eri käyttötarkoitusta varten, se ei ole automaattisesti kiellettyä. Sen sijaan tulee arvioida, onko tämä eri käyttötarkoitus yhteensopiva aiemmin määritellyn käyttötarkoituksen kanssa. [117]

Tällainen arviointi tai analyysi voidaan karkeasti jakaa kahteen eri tyyppiin: muodollinen arviointi ja aineellinen arviointi. Muodollinen arviointi tarkoittaisi kirjallisena toimitettujen sekä uuden että vanhan käyttötarkoituksen arviointia. Aineellinen puolestaan tarkoittaisi menisi muodollista arviointia pidemmälle ja siinä arvioitaisiin käyttötarkoituksia ottaen huomioon sen millaisena ne on ymmärretty (tai olisi pitänyt ymmärtää) riippuen kontekstista ja muista tekijöistä. [118]

Joitakin tiettyjä tilanteita voidaan tunnistaa, missä tällainen arviointi tulisi suorittaa:

- Yhteensopivuus on ilmeistä (*prima facie*)
- Yhteensopivuus ei ole ilmeistä ja tilanne edellyttää tarkempaa arviointia
- Yhteensopimattomuus on ilmeistä

Eri tilanteiden tunnistamista varten tietosuojatyöryhmä on määritellyt joitakin avaintekijöitä: [119]

a) Käyttötarkoitus, jota varten tiedot on kerätty ja käyttötarkoitus, jota varten tietoja tultaisiin myöhemmin käsittelemään[120] (alkuperäinen ja myöhemmän käsittelyn käyttötarkoitus)
b) Asiayhteys, jossa tiedot on kerätty ja rekisteröidyn kohtuulliset odotukset sen mahdolliseen myöhempään käsittelyyn
c) Kerättyjen tietojen ominaisuudet ja tietojen myöhemmän käsittelyn vaikutukset rekisteröidylle

---

[116] WP203 2013, 4.
[117] WP203 2013, 21.
[118] WP203 2013, 21.
[119] WP203 2013, 22-23.
[120] WP203 2013, 23. Tietosuojatyöryhmä korostaa, että arvioinnissa ei annettaisi niin paljoa arvoa sille, miten käyttötarkoitus on kirjallisesti määritelty vaan nimenomaan sen aineelliseen arviointiin, mikä on näiden kahden käyttötarkoituksen suhde. Mitä kauempana alkuperäinen käyttötarkoitus ja suunniteltu, tuleva käyttötarkoitus ovat toisistaan, sitä todennäköisempää on, että käsittelytoimien käyttötarkoitukset eivät ole säädöksessä tarkoitetulla tavalla yhteensopivia.

d) Rekisterinpitäjän suojatoimet, joilla varmistetaan tietojen reilu käsittely ja joilla varmistetaan tarpeettomat ja kohtuuttomat vaikutukset rekisteröityihin

Mikäli rekisterinpitäjän tietojenkäsittelytoimi ei täytä edellä mainittuja vaatimuksia ja myöhemmän käsittelytoimen käyttötarkoitus ei ole yhteensopiva alkuperäisen käyttötarkoituksen kanssa, käsittelytoimi voidaan tulkita laittomaksi. Rekisterinpitäjä ei voi siis erotella myöhempiä samojen tietojen käsittelytoimia irrallisiksi alkuperäisestä ja siten kiertää yhteensopivuuden vaatimusta. [121]

## 4.2 Eri roolit henkilötietojen käsittelyssä, 4 luku

### *4.2.1 Rekisterinpitäjä*

Yleisen tietosuoja-asetuksen 4 artiklassa määritellään tietojenkäsittelyn roolit. Rekisterinpitäjällä tarkoitetaan ”- - *luonnollista henkilöä tai oikeushenkilöä, viranomaista tai muuta elintä, joka yksin tai yhdessä toisten kanssa määrittelee henkilötietojen käsittelyn tarkoitukset ja keinot* - -” ja henkilötietojen käsittelijällä puolestaan tarkoitetaan ”*luonnollista henkilöä tai oikeushenkilöä, viranomaista, virastoa tai muuta elintä, joka käsittelee henkilötietoja rekisterinpitäjän lukuun*”. Rekisterinpitäjän käsite on olennainen, koska se määrittää, kuka vastaa tietosuojasäännösten noudattamisesta ja miten rekisteröidyt voivat käytännössä käyttää oikeuksiaan [122].

Yleisen tietosuoja-asetuksen 4 artiklan 7 kohta jatkuu: ”- - *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- -*”. Tietosuojavaltuutettu on ottanut asiaan kantaa ratkaisussaan TSV 02.09.2019 (7713/163/2018) henkilötietojen rekisterinpitäjästä ja käsittelijästä käsitteli juuri tätä kysymystä. Kuntayhtymä oli saanut kunnilta tiedusteluja rekisterien omistajuudesta, jotka liittyvät sosiaalihuollon asiakasrekisteriin. Kuntayhtymä ei ollut löytänyt selkeää vastausta siihen, kuka on tietojen omistaja. Kyseessä on kuntayhtymä, mutta kuntalaisten sosiaalihuolto on kunnan vastuulla. Tietosuojavaltuutettu totesi, että ”se viranomainen, jolla on jonkin palvelun järjestämisvastuu, on myös tämän palvelun antamisen yhteydessä kerättyjen asiakastietojen rekisterinpitäjä. Palvelunjärjestäjä rekisterinpitäjänä voi joko itse tuottaa kyseessä olevan palvelun tai ostaa sen joltain ulkopuoliselta, kuten esim. kuntayhtymältä”.

[121] WP203 2013, 40.
[122] WP169 2010, 4.

Rekisterinpitäjän vastuu on määritelty asetuksen 24 artiklassa 1 kohdassa:

> Ottaen huomioon käsittelyn luonne, laajuus, asiayhteys ja tarkoitukset sekä luonnollisten henkilöiden oikeuksiin ja vapauksiin kohdistuvat, todennäköisyydeltään ja vakavuudeltaan vaihtelevat riskit rekisterinpitäjän on toteutettava *tarvittavat tekniset ja organisatoriset toimenpiteet*, joilla voidaan *varmistaa ja osoittaa*, että käsittelyssä noudatetaan tätä asetusta. Näitä toimenpiteitä on tarkistettava ja päivitettävä tarvittaessa.

On huomattava, että jos yleinen tietosuoja-asetus ei tule sovellettavaksi asetuksen 2 artiklan 2 kohdan mukaisen luettelon tarkoittamissa tapauksissa, myöskään edellä kuvattuja vastuista ei ole rekisterinpitäjällä. [123] Asiassa C 319/07 T*ietosuojavaltuutettu v. Jehovan todistajat* [124] unionin tuomioistuin on ottanut kantaa rekisterinpitäjän käsitteeseen. Jehovan todistajien ovelta ovelle -saarnaamisen yhteydessä kerääntyi luonnollisten henkilöiden tietoja, joita myös käsiteltiin myöhemmin. Tietoja kerättiin muistiinpanoihin, mutta niin, että niihin oli myöhemmin helppo palata. Organisaatio myös kannusti saarnaamistyöhön ja vaikka saarnaajia ei erikseen ohjeistettukaan tietojen keräämisestä, Jehovan todistajien katsottiin olevan yleisen tietosuoja-asetuksen tarkoittamia rekisterinpitäjiä. Tuomiossa todettiin myös, että käsitettä voidaan soveltaa, ei ole tarpeen, että tiedot tallennetaan kortistoon tai luetteloihin ja että olisi tietojen hakemista palvelevia järjestelyjä.

Rekisterinpitäjän käsite on tärkeä myös, kun arvioidaan mitä laki sovelletaan tiettyihin käsittelytoimiin. Asetuksen 3 artiklan 1 kohdan mukaan asetusta sovelletaan henkilötietojen käsittelyyn, jota suoritetaan unionin alueella sijaitsevassa rekisterinpitäjän tai henkilötietojen käsittelijän toimipaikassa toiminnan yhteydessä. Tästä seuraa, että rekisterinpitäjän määrittäminen ja rekisterinpitäjän toimipaikka ovat ratkaisevassa asemassa, mitä lakia sovelletaan. Lähtökohta on kuitenkin, että kukin jäsenvaltio soveltaa henkilötietojen käsittelyyn kansallisia säädöksiä. Jos sama rekisterinpitäjä on sijoittunut useamman jäsenvaltion alueelle, sen tulee varmista, että kussakin noudatetaan säännöksiä. [125]

---

[123] Yleisen tietosuoja-asetuksen 2 artiklan 2 kohdan luettelo:

2. *Tätä asetusta ei sovelleta henkilötietojen käsittelyyn,*
   a) *jota suoritetaan sellaisen toiminnan yhteydessä, joka ei kuulu unionin lainsäädännön soveltamisalaan;*
   b) *jota suorittavat jäsenvaltiot toteuttaessaan SEU V osaston 2 luvun soveltamisalaan kuuluvaa toimintaa;*
   c) *jonka luonnollinen henkilö suorittaa yksinomaan henkilökohtaisessa tai kotitalouttaan koskevassa toiminnassa;*
   d) *jota toimivaltaiset viranomaiset suorittavat rikosten ennalta estämistä, tutkintaa, paljastamista tai rikoksiin liittyviä syytetoimia varten tai rikosoikeudellisten seuraamusten täytäntöönpanoa varten, mukaan lukien yleiseen turvallisuuteen kohdistuvilta uhkilta suojelua ja tällaisten uhkien ehkäisyä varten.*

[124] EUVL N:o C 319, 10.09.2018, C 319/07.

[125] WP169 2010, 5.

Rekisterinpitäjän käsitteen tarkempi analyysi osoittaa, että asetuksen 4 artiklan 7 kohdan mukaisessa määritelmässä on kolme komponenttia:

- ”luonnollinen oikeushenkilö, julkinen viranomainen, virasto tai muu toimielin”
- ”joka, yksin tai yhdessä toisten kanssa”
- ”määrittelee henkilötietojen käsittelyn tarkoituksen ja keinot”

Ensimmäinen komponentti liittyy määritelmän henkilökohtaiseen osuuteen. Toinen komponentti laventaa määritelmää yhden yksittäisen rekisterinpitäjän sijasta useampaan rekisterinpitäjään. Kolmas komponentti puolestaan sisältää olennaiset osatekijät, joiden perusteella rekisterinpitäjä voidaan tunnistaa. [126]

Kolmannen komponentin tarkastelu on käytännön syistä tehdä ensin. Tässä ensimmäinen osatekijä ”määrittelee” on näkyvissä myös säädännön historiallisessa kehityksessä, jossa korostuu kaksi eri seikkaa. Ensinnäkin rekisterinpitäjän on mahdollista toimia rekisterinpitäjänä erityisestä toimivallasta riippumatta, joka annetaan laissa. Toiseksi rekisterinpitäjän määritteleminen on yhteisön käsite, jolla on oma erityinen asema yhteisön säännöstössä. Tällä turvataan yhtenäistä ja itsenäistä tulkintaa rekisterinpitäjän käsitteelle. Tällainen tulkinta varmistaa asetuksen tehokkaan soveltamisen. [127]

Rekisterinpitäjän käsitteen viimeinen määritys on riippumattomuus. Riippumattomuudella tarkoitetaan riippumattomuutta muusta lainsäädännöstä, kun rekisterinpitäjää määritellään. Vaikka ulkoisista oikeuslähteistä voikin olla hyötyä rekisterinpitäjän tunnistamiseksi, toimea tulisi tulkita kuitenkin pääasialliseesti vain tietosuojalainsäädännön kautta. Esimerkiksi teollis- ja tekijänoikeuksien todellisena haltijana toimiminen ei sulje pois mahdollisuutta toimia myös rekisterinpitäjänä. [128]

Rekisterinpitäjän käsitettä voidaan pitää toiminnallisena käsitteenä, jonka tarkoituksena on jakaa vastuuta sinne, missä sen todellinen vaikutus on. Tämän tarkoitus on siis nimenomaan tosiasiallinen eikä muodollinen analyysi. Tietosuojatyöryhmä toteaa, että tällainen todellinen vaikutus voidaan päätellä olosuhteiden perusteella. Olosuhteita voidaan puolestaan analysoida ja jakaa kolmeen ryhmään: 1) eksplisiittinen oikeustoimikelpoisuuteen perustuva vastuu, 2) implisiittiseen toimivaltaan perustuva vastuu ja 3) tosiasialliseen vaikuttamiseen perustuva vastuu. Näistä kaksi ensimmäistä osoittavat tietosuojatyöryhmän mukaan varmemmin määrittelevän yhteisön ja voivat kattaa jopa 80 % käytännössä relevanteista tilanteista. Jälkimmäinen kolmas ryhmä vaatii kuitenkin tarkempaa analyysia. [129]

Tosiasialliseen vaikuttamiseen tietojen käsittelyn tarkoituksesta ja keinoista perustuva vastuu perustuu siis arvioon tosiasiallisista olosuhteista. Osapuolten välinen sopimus

[126] WP169 2010, 8.
[127] WP169 2010, 8.
[128] WP169 2010, 9.
[129] WP169 2010, 9.

voi usein selventää tätä tilannetta. Arviosta voidaan myös tehdä päätelmä, jossa rekisterinpitäjän asema ja vastuu annetaan usealle osapuolelle. Työryhmän mukaan tällainen päätelmä voisi olla hyödyllinen erityisesti mutkikkaassa ympäristössä, jossa usein käytetään myös uutta tietotekniikkaa ja jossa relevantit toimijat eivät itse katso olevansa rekisterinpitäjinä vastuullisia vaan enemmänkin ”yhteyshenkilöitä”. Osapuolten välisessä sopimuksessa voidaan antaa aineksia osoittaa rekisterinpitäjän vastuu ilmeisen hallitsevalle osapuolelle. [130]

Myös muut seikat kuin osapuolten välisen sopimuksen määräykset voivat olla hyödyllisiä rekisterinpitäjän aseman ja vastuun tulkinnassa. Tällaisia seikkoja voivat olla osapuolen tosiasiallisesti harjoittaman vastuun määrä, rekisteröidyille annettu kuva ja rekisteröidyn kohtuulliset odotukset näkyvyyden suhteen. [131]

Kolmannen komponentin ”määrittelee tietojenkäsittelyn tarkoituksen ja keinot” toinen osa ”tarkoitus ja keinot” on tietosuojatyöryhmän mukaan olennaisin osa: mitä osapuolen olisi määriteltävä rekisterinpitäjän ominaisuuksiksi. Asetuksen mukaan henkilötietoja voi kerätä vain nimenomaista ja laillista tarkoitusta varten, eikä tietoja saa käsitellä myöhemmin näiden tarkoitusten kanssa yhteensopimattomalla tavalla. Tarkoituksen ja keinojen määrittelemisen voidaan katsoa olevan sen määrittelyä, että ”miksi” ja ”miten” tietyt käsittelytoimet suoritetaan. Epäselvissä tilanteissa rekisterinpitäjän tunnistaminen voi edellyttää siis pitkäaikaista selvitystä ja toiminnallista analyysiä. Tietojen käsittelyn tarkoituksen määritteleminen on joka tapauksessa rekisterinpitäjän vastuulla. [132]

### *4.2.2 Yhteisrekisterinpitäjä*

Alati mutkistuvassa toimintaympäristössä on yhä tärkeämpää, että rekisterinpitäjän asema ja vastuu ovat helposti osoitettavissa. Näin varmistetaan, että yhteisvastuun mutkikkuus ei johda toimimattomaan vastuun jakoon ja näin estäisi tietosuojalainsäädännön tehokkuuden. Tietosuojatyöryhmän lausunnossa vuodelta 2010 korostetaan, että mahdollisten järjestelyn moninaisuuden vuoksi tyhjentävää luetteloa erilaisista yhteisvastuun ilmenemismuodoista on mahdotonta laatia. Sen sijaan tällaisen yhteisvastuun arvioinnissa tulisi tehdä arvio yksinkertaisesta vastuusta ja käytettävä olennaista ja toimivaa lähestymistapaa. Tärkeintä olisi keskittyä siihen, onko tarkoituksen ja keinojen määrittelijöinä ollut yksi vai useampi taho. [133]

---

[130] WP169 2010, 11.
[131] WP169 2010, 11.
[132] WP169 2010, 13.
[133] WP169 2010, 18. Lausunto perustuu vuonna 2010 voimassa olleeseen lainsäädäntöön, mutta sen tulkintaa voidaan pitää relevanttina myös kirjoitushetken voimassa olevan lainsäädännön tulkinnassa.

Yhteisvastuussa osapuolten osallistuminen tietojenkäsittelytoimien tarkoitusten ja keinojen määrittämiseen voi kuitenkin saada erilaisia muotoja ja vastuu määrittämisestä ei välttämättä jakaudu tasapuolisesti. Pelkästään se, että eri osapuolet ovat yhteistyössä henkilötietojen käsittelyssä, ei automaattisesti tarkoita sitä, että ne pitäisivät rekisteriä yhdessä. Jos osapuolet kuitenkin muodostavat yhteisen perusrakenteen kunkin omia yksittäisiä tarkoituksia varten, tämä arvio voi muuttua. Jos tässä tapauksessa osapuolet määrittelevät tällaisen perusrakenteet perustamisen yhteydessä keinot tietojen käsittelyyn tai niiden olennaiset osat, osapuolet katsotaan todennäköisesti yhteisrekisterinpitäjiksi – vaikka niillä ei olisikaan täysin yhteisiä tarkoituksia. Työryhmä[134] antaa esimerkin tällaisesta tilanteesta: [135]

| |
|---|
| Matkatoimisto, hotelliketju ja lentoyhtiö päättävät perustaa yhteisen internetvarausjärjestelmän parantaakseen yhteistyötään matkavarausten hallinnassa. Ne sopivat käytettävien keinojen tärkeistä osista, kuten mitä tietoja säilytetään, miten varauksia otetaan vastaan ja vahvistetaan ja kenellä on oikeus tutustua säilytettyihin tietoihin. Lisäksi ne päättävät pitää yhteiset asiakastiedot voidakseen toteuttaa yhdessä markkinointitoimia. |

Tässä tapauksessa matkatoimisto, lentoyhtiö ja hotelliketju vastaavat yhdessä asiakkaidensa henkilötietojen käsittelytavasta, joten ne ovat yhteisvastuussa tietojenkäsittelytoiminnoista, jotka perustuvat niiden yhteiseen internetvarausjärjestelmään. Kukin niistä on kuitenkin edelleen yksin vastuussa muista tietojenkäsittelytoiminnoista, esim. kunkin henkilöresurssien hoidosta.

Tässä siis työryhmä katsoo tietojenkäsittelytoimien olennaisiksi osiksi ainakin sen, miten tietoja säilytetään, miten varauksia otetaan vastaan ja kenellä on oikeus tutustua säilytettyihin tietoihin sekä se, että asiakastiedot ovat yhteisiä. Työryhmä toteaa myös, että samoja henkilötietoja voi käsitellä useampi taho peräkkäin. Näissä tapauksissa on mahdollista tietysti, että läheltä katsottuna, mikrotasolla eri toiminnot ovat toisistaan hyvinkin erillään, koska niillä on eri tarkoitus ainakin näennäisesti. Työryhmä kuitenkin korostaa, että eri toimintoja tulisi katsella ”toimintojen kokonaisuutena”: onko niillä sama tarkoitus ja käytetäänkö yhteisesti määriteltyjä keinoja. Työryhmä[136] antaa kaksi esimerkkiä tällaisista tilanteista: [137]

| |
|---|
| XYZ-niminen yritys kerää ja käsittelee työntekijöidensä henkilötietoja palkkojenmaksua, työmatkoja, sairausvakuutuksia jne. varten. Laissa kuitenkin edellytetään, että yritys lähettää kaikki palkkatiedot veroviranomaisille verovalvonnan tehostamiseksi. |

Tässä esimerkissä sekä ZYZ että veroviranomainen käsittelevät samoja palkkatietoja, mutta koska niillä ei ole kuitenkaan yhteistä tarkoitusta ja yhteisesti määriteltyjä keinoja käsittelyssä, ne on katsottava erillisiksi rekisterinpitäjäksi.[138]

---

[134] WP169 2010, 19, esimerkki 5.
[135] WP169 2010, 19.
[136] WP169 2010, 20. Esimerkki 9.
[137] WP169 2010, 19.
[138] WP169 2010, 19.

Seuraavassa esimerkissä[139] tutkitaan tapausta, jossa kaksi tahoa voidaan katsoa yhteisiksi rekisterinpitäjiksi:

> Esimerkkinä on pankki, joka käyttää rahoitustietojen välittäjää liiketoimiensa toteuttamisessa. Pankki ja tietojen välittäjä sopivat rahoitustietojen käsittelyssä käytettävistä keinoista. Rahoitustoimiin liittyviä henkilötietoja käsittelee ensin rahoituslaitos ja sen jälkeen rahoitustietojen välittäjä.

Vaikka kummallakin osapuolella on läheltä katsottuna, mikrotasolla oma tarkoituksensa, kauempaa katsottuna, makrotasolla tietojen käsittelyn eri vaiheet, tarkoitus ja keinot liittyvät tiivisti toisiinsa. Tässä tapauksessa työryhmä katsoisi pankin ja välittäjän olevan yhteisrekisterinpitäjiä. [140] Käsittelytoimet muodostavat siis tosiasiallisesti jatkumon tai kokonaisuuden, josta ei voida erottaa yhtä komponenttia itsenäiseksi toiminnoksi.

Työryhmä haluaa korostaa lausunnossaan, että rekisterinpitäjän vastuu säilyy niissäkin tapauksissa, joissa rekisterinpitäjän velvoitteita ei voida suoraan täyttää (esimerkiksi tietojen varmista ja tiedonsaantioikeutta), rekisterinpitäjän vastuu kuitenkin säilyy. Toiset osapuolet voivat käytännössä täyttää tällaisia vastuita, joita rekisterinpitäjä itse ei voi. Rekisterinpitäjä on kuitenkin vastuussa velvoitteistaan ja kantaa myös vastuun niiden mahdollisesti rikkomisesta. [141]

Mutkikkaissakin järjestelyissä olisi pidettävänä aina lähtökohtana sitä, että määritetään selvästi tietosuojasääntöjen noudattaminen ja vastuu näiden sääntöjen mahdollisesta rikkomisesta silloinkin, kun eri rekisterinpitäjillä on erilainen asema käsittelytoimissa. Tällainen määrittäminen tulisi tehdä, jotta henkilötietosuojaa ei kavennettaisi tai että ei syntyisi kielteinen toimivaltaristiriita eikä myöskään tilanteita, joissa yksikään osapuoli ei olisi vastuussa. [142] Lisäksi saman artiklan 2 kohdassa tarkennetaan vielä, että rekisterinpitäjän panee täytäntöön asianmukaiset tietosuojaa koskevat toimintaperiaatteet.

### *4.2.3 Tietojen käsittelijä*

Henkilötietojen käsittelijä 28 artiklan mukaan käsittelee henkilötietoja rekisterinpitäjän lukuun. Rekisterinpitäjä määrittää tietojen käsittelijän käsittelytoimia sopimuksella. Käsittelytoimia voidaan määrittää myös unionin oikeudella tai jäsenvaltion lainsäädännön mukaisella asiakirjalla, joka sitoo henkilötietojen käsittelijää suhteessa rekisterinpitäjään. Tällaisessa tietojenkäsittelysopimuksessa (DPA, *Data Processing Agreement*)

[139] WP169 2010, 20. Esimerkki 10.
[140] WP169 2010, 20.
[141] WP169 2010, 21.
[142] WP169 2010, 22.

tulee 28 artiklan 3 kohdan mukaan määritellä tietojenkäsittelijälle tietyt säännöt käsittelytoimien suhteen. Tällaisia sääntöjä ovat:

- Tietojen käsittelijä käsittelee tietoja vain rekisterinpitäjän antamien dokumentoitujen ohjeiden mukaisesti
- Käsittelijä varmistaa, että henkilöt, joilla on oikeus käsitellä henkilötietoja, on salassapitovelvollisuus
- Käsittelijä toteuttaa kaikki 32 artiklan mukaiset käsittelyn turvallisuuteen liittyvät toimet
- käsittelijä ei saa käyttää toisen henkilötietojen käsittelijän palveluksia ilman rekisterinpitäjän erityistä tai yleistä kirjallista ennakkolupaa
- Jos käsittelijä käyttää toisen henkilötietojen käsittelijän palveluksia, tähän toiseen käsittelijään sovelletaan samoja tietosuojavelvoitteita
- Käsittelijän tulee saattaa rekisterinpitäjän saataville kaikki ne tiedot, jotka ovat tarpeen säädettyjen velvollisuuksien osoittamista varten, sekä sallii rekisterinpitäjän tai valtuutetun auditoijan suorittamaan auditoinnin sekä osallistuu niihin

Asetuksen 28 artiklan 10 kohdan mukaan, jos tietojen käsittelijä määrittää itsenäisesti henkilötietojen käsittelyn tarkoitukset ja keinot, kyseistä tietojen käsittelijää pidetään rekisterinpitäjänä tietojen käsittelijän sijaan.

Yleisen tietosuoja-asetuksen 30 artiklan 1 kohdan mukaan rekisterinpitäjän on ylläpidettävä selostetta vastuullaan olevista käsittelytoimista. Selosteen on pidettävä sisällään kaikki seuraavat tiedot:

a) rekisterinpitäjän ja mahdollisen yhteisrekisterinpitäjän, rekisterinpitäjän edustajan ja tietosuojavastaavan nimi ja yhteystiedot;
b) käsittelyn tarkoitukset;
c) kuvaus rekisteröityjen ryhmistä ja henkilötietoryhmistä;
d) henkilötietojen vastaanottajien ryhmät, joille henkilötietoja on luovutettu tai luovutetaan, mukaan lukien kolmansissa maissa tai kansainvälisissä järjestöissä olevat vastaanottajat;
e) tarvittaessa tiedot henkilötietojen siirtämisestä kolmanteen maahan tai kansainväliselle järjestölle, mukaan lukien tieto siitä, mikä kolmas maa tai kansainvälinen järjestö on kyseessä, sekä asianmukaisia suojatoimia koskevat asiakirjat, jos kyseessä on 49 artiklan 1 kohdan toisessa alakohdassa tarkoitettu siirto;
f) mahdollisuuksien mukaan eri tietoryhmien poistamisen suunnitellut määräajat;
g) mahdollisuuksien mukaan yleinen kuvaus 32 artiklan 1 kohdassa tarkoitetuista teknisistä ja organisatorisista turvatoimista

Lisäksi 30 artiklan 2 kohdan mukaan henkilötietojen käsittelijän tulee ylläpitää selostetta kaikista rekisterinpitäjän lukuun suoritettavista käsittelytoimista. Kyseisen selosteen pitää sisältää:

a) henkilötietojen käsittelijän tai käsittelijöiden ja kunkin rekisterinpitäjän, jonka lukuun henkilötietojen käsittelijä toimii, sekä rekisterinpitäjän tai tarvittaessa henkilötietojen käsittelijän edustajan ja tietosuojavastaavan nimi ja yhteystiedot;
b) kunkin rekisterinpitäjän lukuun suoritettujen käsittelyiden ryhmät;

c) tarvittaessa tiedot henkilötietojen siirtämisestä kolmanteen maahan tai kansainväliselle järjestölle, mukaan lukien tieto siitä, mikä kolmas maa tai kansainvälinen järjestö on kyseessä, sekä asianmukaisia suojatoimia koskevat asiakirjat, jos kyseessä on 49 artiklan 1 kohdan toisessa alakohdassa tarkoitettu siirto;
d) mahdollisuuksien mukaan yleinen kuvaus 32 artiklan 1 kohdassa tarkoitetuista teknisistä ja organisatorisista turvatoimista.

Näiden selosteiden tulee olla kirjallisessa muodossa (mukaan lukien sähköinen muoto). Lisäksi selosteita ei tarvitse tehdä tietyin rajoituksin, jos rekisterinpitäjänä on yritys tai järjestö, jossa on alle 250 työntekijää. [143]

[143] Yleinen tietosuoja-asetus 30 artikla 4 ja 5 kohdat.

# 5 PROFILOINTI JA AUTOMAATTINEN PÄÄTÖKSENTEKO

## 5.1 Profilointi

Yleisessä tietosuoja-asetuksessa ei keskitytä vain automaattisen käsittelyn tai profiloinnin tuloksena tehtyihin päätöksiin. Sitä sovelletaan tietojen keräämiseen profiilien luomista varten sekä näiden profiilien soveltamiseen yksittäisiin henkilöihin.[144] Yleisen tietosuoja-asetuksen mukaan profiloinnilla tarkoitetaan

> Mitä tahansa henkilötietojen automaattista käsittelyä, 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.[145]

Profilointi koostuu siis seuraavista osatekijöistä:

- sen on oltava *automatisoitua* käsittelyä
- se on suoritettava *henkilötiedoilla* ja
- profiloinnin tavoitteena on oltava luonnollisen henkilön *henkilökohtaisten ominaisuuksien arvioiminen*.[146]

Laajasti katsottuna profilointi tarkoittaa tietojen keräämistä henkilöstä (tai henkilöiden ryhmästä) ja henkilön ominaisuuksien tai käyttäytymismallien arvioimista tämän sijoittamiseksi tiettyyn luokkaan tai ryhmään, jotta voidaan analysoida ja/tai ennustaa esimerkiksi henkilön kykyä suorittaa jokin tehtävä, hänen kiinnostuksen kohteita tai hänen todennäköistä käyttäytymistä. [147] Lisäksi on huomattava, että profiili ja profilointi voivat tarkoittaa eri asiayhteyksissä erilaisia asioita. Esimerkiksi tilastotieteessä profiili voidaan määritellä yksilön identifioivien muuttujien arvojen yhdistelmäksi. Yksilön identifioivat muuttujat puolestaan ovat tilastotieteessä muuttujia, joiden avulla yksiköt ovat yhdistettävissä toisiin yksiköihin.[148]

Juridisesta näkökulmasta profilointia voidaan lähtökohtaisesti käyttää kolmella eri tavalla[149]:

a) yleinen profilointi
b) profilointiin perustuva päätöksenteko ja

[144] WP251rev.01 2018, 6.
[145] WP251rev.01 2018, 6.
[146] WP251rev.01 2018, 8.
[147] WP251rev.01 2018, 8.
[148] Konnu 2006, 102.
[149] WP251rev.01 2018, 9.

c) pelkästään automatisoitu päätöksenteko, profilointi mukaan luettuna, jolla on rekisteröityä koskevia oikeusvaikutuksia tai joka vaikuttaa häneen vastaavalla tavalla merkittävästi (22 artiklan 1 kohta).

Kohtien b ja c tärkein ero on, missä määrin ihminen osallistuu päätöksentekoon. Mikäli esimerkiksi algoritmi päättää lainan myöntämisestä ja päätös toimitetaan henkilölle tiedoksi ilman ihmisen vuorovaikutusta, kyse on c kohdassa kuvatusta tavasta käyttää profilointia ja sitä arvioidaan kriittisemmin. Rekisterinpitäjät voivat siis tehdä profilointia tietojenkäsittelytoimissaan ja käyttää automaattista päätöksentekoa, mikäli ne voivat noudattaa kaikkia periaatteita ja käsittelytoimilla on laillinen peruste. Jos kyseessä aiemmin kohdassa c kuvattu automaattinen päätöksenteko ilman ihmisen vuorovaikutusta, profilointi mukaan luettuna, tällöin sovelletaan täydentäviä suojatoimia ja rajoituksia. [150]

Profilointiin liittyy myös komponentti ”oikeusvaikutukset” tai ”vastaavalla tavalla merkittävät” vaikutukset. Vain vakavat vaikutukset kuuluvat yleisen tietosuoja-asetuksen 22 artiklan soveltamisalaan. Oikeusvaikutuksilla automaattisen päätöksen vaikutuksia henkilön laillisiin oikeuksiin, kuten yhdistymisvapauteen, äänioikeuteen tai oikeuteen ryhtyä oikeustoimiin. Oikeusvaikutus voi olla myös seikka, joka vaikuttaa henkilön oikeudelliseen asemaan tai sopimusperusteisiin oikeuksiin. Tietosuojatyöryhmä listaa esimerkkejä, joita ovat sopimuksen peruuttaminen, tietyn lakisääteisen etuuteen kuten lapsilisän myöntäminen tai epääminen sekä maahantulokielto. Oikeusvaikutuksien lisäksi automaattisella päätöksellä voi olla vastaavalla tavalla merkittäviä vaikutuksia. Jotta vaikutus olisi ”merkittävä”, sen tulisi mahdollisesti vaikuttaa merkittävästi kyseisen henkilön olosuhteisiin, käyttäytymiseen tai valintoihin. Se voisi myös vaikuttaa rekisteröityyn pitkäaikaisesti tai pysyvästi. Lisäksi kysymyksen voisi ääritapauksessa tulla myös päätökset, jotka johtavat henkilöiden syrjäytymiseen tai syrjintään. Tietosuojaryhmä korostaa, että merkittävyyden kynnyksen ylittymisen määrittäminen on vaikeaa, mutta esimerkiksi seuraavat päätökset voisivat ylittää tämän kynnyksen: luottokelpoisuuspäätökset, päätös mahdollisuudesta saada terveydenhuoltopalveluja, työllistymismahdollisuuksiin liittyvät päätökset ja päätökset, jotka vaikuttavat koulutusmahdollisuuksiin. [151]

Yleiseen kieltoon on säädetty poikkeukset yleisen tietosuoja-asetuksen 22 artiklan 2 kohdassa. Kieltoa ei sovelleta, jos päätös

[150] WP251rev.01 2018, 9.
[151] WP251rev.01 2018, 22.

a) on välttämätön sopimuksen tekemistä tai täytäntöönpanoa varten
b) on hyväksytty rekisterinpitäjään sovellettavassa unionin oikeudessa tai jäsenvaltion lainsäädännössä, jossa vahvistetaan myös asianmukaiset toimenpiteet rekisteröidyn oikeuksien ja vapauksien sekä oikeutettujen etujen suojaamiseksi tai
c) perustuu rekisteröidyn nimenomaiseen suostumukseen.

Sopimuksen tekemistä tai sen täytäntöönpanoa automaattisen päätöksenteon tulee olla ”välttämätöntä”. Tietosuojatyöryhmän[152] mukaan rekisterinpitäjän on voitava osoittaa, että automaattinen päätöksenteko on välttämätöntä, kun otetaan huomioon, voisiko se käyttää vähemmän yksityisyyteen puuttuvaa menetelmää. Jos siis on olemassa muita tehokkaita ja vähemmän yksityisyyteen puuttuvia keinoja saman päämäärän saavuttamiseksi, tällainen käsittely ei ole yleisen tietosuoja-asetuksen mukaan ”välttämätöntä”.[153]

Automatisoituja päätöksiä, profilointi mukaan luettuna, voidaan tehdä 22 artiklan 2 kohdan b alakohdan nojalla, jos sen käyttö on hyväksytty unionin oikeudessa tai kyseisen jäsenvaltion lainsäädännössä. Tiettävästi Suomen tietosuojavaltuutettu ei ole kirjoitushetkellä ottanut kantaa profilointiin liittyviin poikkeuslupiin eikä niistä ole säädetty myöskään Suomen lainsäädännössä. Automatisoituja päätöksiä voidaan tehdä 22 artiklan 2 kohdan c alakohdan nojalla, kun rekisteröity on antanut nimenomaisen suostumuksensa.

Profilointia voidaan siis kuvata tietojen keräämisenä niin sanottua yleistä profiilia varten. Tällaisia tietoja voisivat olla vaikkapa henkilön nimi, sukupuoli, ikä ja osoite. Näihin tietoihin perustuen voidaan tehdä päätöksiä joko ihmisen tai automaattisen käsittelytoimien tuloksena. Automaattisen käsittelytoimen tuloksena voi syntyä myös jonkinlainen profiilissa oleviin tietoihin perustuva pisteytys, jota käytetään mahdollisissa muissa jatkokäsittelytoimissa ja automaattisissa päätöksissä. Tällaisia käsittelytoimia voisi olla esimerkiksi asiakkaan pisteytys maksukykyyn perustuen ja tähän pisteytykseen tukeutuva päätös vaikkapa lainahakemuksesta.

## 5.2 Automaattinen päätöksenteko

Automaattisilla päätöksillä on erilainen soveltamisala, mutta ne voivat olla kuitenkin päällekkäisiä profiloinnin kanssa tai olla sen tuloksia. Itsenäisesti automaattinen päätöksenteko viittaa kykyyn tehdä päätöksiä tekniikan avulla ilman ihmisen osallistumista. Tällaiset päätökset voivat perustua mihin tahansa tietoihin, kuten

- kyseisten henkilöiden suoraan antamat tiedot (esimerkiksi vastaukset kyselylomakkeeseen)

[152] WP251rev.01 2018, 25.
[153] Ks. tarkemmin ”välttämättömyydestä” EDPS 2017.

- henkilöistä havainnoidut tiedot (kuten sovelluksen avulla kerätyt paikannustiedot)
- johdetut tai päätellyt tiedot, kuten henkilöstä jo luotu profiili (esimerkiksi luottopisteytys).[154]

Vaikka automatisoituja päätöksiä ja profilointia sovelletaankin erikseen, ne ovat kuitenkin selvästi yhteydessä toisiinsa. Se, mikä alkaa yksinkertaisena automatisoituna päätöksentekoprosessina, voi muuttua profiloinniksi sen mukaan, miten tietoja käytetään. Esimerkiksi ylinopeussakkojen määrääminen pelkästään nopeuskameroiden antamien tietojen perusteella ei ole profilointia, mutta on automatisoitua päätöksentekoa. Siitä tulisi sen sijaan profilointiin perustuva automaattinen päätöksentekoprosessi, jos nopeuskameroiden antamiin tietoihin yhdistettäisiin edistyneempää analytiikkaa, kuten henkilön ajotapoja pidemmältä ajalta ja sakon määrä määräytyisi myös muidenkin tekijöiden summana, kuten onko ylinopeus toistuvaa tai onko kuljettaja syyllistynyt muihin rikkomuksiin. [155]

## 5.3 Profilointia ja automatisoituja päätöksiä koskevat periaatteet

Seuraavia yleisen tietosuoja-asetuksen 5 artiklan 1 kohdan mukaisia tietosuojaperiaatteita sovelletaan kaikkeen profilointiin ja automatisoituihin päätöksiin, joihin liittyy henkilötietoja: [156]

- Lainmukaisuus, kohtuullisuus ja läpinäkyvyys[157]
- Myöhempi käsittely ja käyttötarkoituksenmukaisuus
- Tietojen minimointi
- Täsmällisyys
- Säilytyksen rajoittaminen

Tekoälyä käsittelevän korkean tason asiantuntijaryhmän[158] mukaan laajamittainen kansalaisten pisteyttäminen eli niin kutsuttu *scoring* on erityinen huolenaihe tekoälyn eettisen kehityksen kannalta. Yleisesti yhteiskuntien tulisi pystyä suojelemaan kaikkien kansalaisten vapautta ja itsemääräämisoikeutta. Kaikenlainen pisteyttäminen voi mahdollisesti johtaa itsemääräämisoikeuden menettämiseen ja vaarantaa syrjimättömyyden periaatteen. Pisteytystä tulisi käyttää vain, jos sille on selvä peruste ja jos toimenpiteet ovat oikeasuhteisia ja oikeudenmukaisia. Ohjeistuksessa korostetaan, että huoli koskee

[154] WP251rev.01 2018, 8.
[155] WP251rev.01 2018, 8.
[156] WP251rev.01 2018, 10.
[157] Läpinäkyvyyttä käsitellään tarkemmin luvussa 7.
[158] Tekoälyä käsittelevän korkean tason asiantuntijaryhmä 2019b, 43.

erityisesti kansalaisten normatiivisten pisteyttämistä eli tietynlaisen ”moraalisen persoonallisuuden” tai ”eettisen integriteetin” yleistä arviointia. Asiantuntijaryhmä tuo esiin, että pisteytyksen käyttämisen edellytys on sen läpinäkyvyys, mutta se ei ole kuitenkaan yleislääke pisteytykseen liittyviin ongelmiin. Heidän mukaansa on erityisen tärkeää huolellisesti arvioida missä mittakaavassa pisteytystä käytetään, ihannetilanteessa annettaisiin mahdollisuus jättäytyä sen ulkopuolelle ja tarjottaisiin myös mekanismeja pistemäärien kyseenalaistamiseksi. Erityisen tärkeää tällaisten suojamekanismien käyttö olisi, kun osapuolten välillä vallitsee vallan epätasapaino.

### *5.3.1 Lainmukaisuus, kohtuullisuus ja läpinäkyvyys*

Yleisen tietosuoja-asetuksen 5 artiklan alakohta a sisältää tietosuoja-asetuksen yhden olennaisimmista periaatteista, joka on tietojen käsittelyn läpinäkyvyys. Profilointiprosessi on usein automaattinen ja rekisteröidyn kannalta näkymätön. Prosessin yhteydessä luodaan usein käsiteltävistä henkilötiedoista tietyllä tavalla johdettuja tai pääteltyjä lisätietoja henkilöistä. Näitä tietoja voitaneen pitää ikään kuin uusina henkilötietoina, joita rekisteröityneet itse eivät ole toimittaneet. Rekisteröityneen voi olla hankalaa ymmärtää profilointiin ja automatisoituun päätöksentekoon liittyviä tekniikoita. [159]

Esimerkiksi vakuutusyhtiöt ovat alkaneet tarjota Suomessakin vakuutuksia, jonka maksut perustuvat henkilön ajokäyttäytymiseen. [160] Tällaisessa palvelussa käsiteltäviä tietoja voivat olla ajettu matka, käytetty reitti sekä muut eri antureista saadut tiedot. Kerättyjä tietoja käytetään profilointiin huonon ajokäyttäytymisen tunnistamiseksi (kuten nopeat kiihdytykset, äkilliset jarrutukset ja ylinopeus). Näitä tietoja voidaan vielä vaikkapa vertailla ristiin ja muihin tietolähteisiin, kuten sää- ja liikennetietoihin. Tässä tapauksessa rekisterinpitäjän on varmistettava, että ensinnäkin tämän tyyppiselle käsittelylle on laillinen peruste. Lisäksi rekisteröidylle on annettava tiedot kerätyistä tiedoista ja tarvittaessa tieto automatisoidun päätöksenteon olemassaolosta, käsittelyyn liittyvästä logiikasta ja käsittelyn merkittävyydestä sekä mahdollisista seurauksista. [161]

Tietojen käsittelyn on oltava myös kohtuullista ja läpinäkyvää. Profilointi voi olla kohtuutonta, jos se aiheuttaa syrjintää esimerkiksi estämällä henkilöiden työmahdollisuuksien, lainojen tai vakuutusten saannin tai kohdentamalla heille liian riskialttiita tai kalliita rahoitustuotteita. [162]

---

[159] WP251rev.01 2018, 10.
[160] Esim. ALD Automotive tarjoaa sekä autovakuutuksia, että ajoneuvotietopalvelua. Kts. https://www.aldautomotive.fi/asiakkaalle/tarjoamme/ald-ajotietopalvelu . Viitattu 9.4.2019.
[161] Ks. tarkemmin yleisen tietosuoja-asetuksen artiklan 22 artiklan 1 ja 4 kohdat.
[162] WP251rev.01 2018, 11.

### *5.3.2 Myöhempi käsittely ja käyttötarkoituksenmukaisuus*

Profilointiin voi liittyä tietoja, jotka on alun perin kerätty toista tarkoitusta varten. Se, onko myöhempi käsittely yhteensopivaa sen tarkoituksen kanssa, johon ne alun perin oli kerätty, riippuu monista tekijöistä, kuten siitä, mitä rekisterinpitäjä tietoja antoi alun perin rekisteröidylle. Erilaiset tekijät on otettu huomioon yleisessä tietosuoja-asetuksessa[163]:

- tietojen keräämisen alkuperäisten tarkoitusten ja myöhemmän käsittelyn tarkoitusten välinen suhde
- asiayhteys, jossa tiedot kerättiin, ja rekisteröityjen kohtuulliset odotukset niiden myöhemmän käytön suhteen
- tietojen luonne
- myöhemmän käsittelyn vaikutus rekisteröityihin ja

rekisterinpitäjän soveltamat suojatoimet, joilla varmistetaan asianmukainen käsittely ja estetään tarpeettomat vaikutukset rekisteröityihin.

### *5.3.3 Kerättyjen tietojen minimointi, täsmällisyys ja säilytyksen rajoittaminen*

Tallennuskustannuksien alentuminen ja toisaalta suurien tietomäärien kasvavat käsittelyvalmiudet luovat jatkuvasti uusia liiketoimintamahdollisuuksia. Tämä voi johtaa siihen, että organisaatiot keräävät enemmän henkilötietoja kuin ne todellisuudessa tarvitsevat siltä varalta, että ne osoittautuvat myöhemmin tarpeellisiksi. [164]

Rekisterinpitäjien tulisi pystyä aina perustelemaan ja selittämään selvästi tarve kerätä ja säilyttää henkilötietoja. Samoin rekisterinpitäjien tulisi harkita myös yhdistelmätietojen, anonymisoitujen tietojen tai pseudonymisoitujen tietojen käyttämistä esimerkiksi profilointiin.[165]

Profilointi ja dataan perustuva automaattinen päätöksenteko voi olla vain yhtä tarkkaa kuin data, johon se perustuu. Tästä syystä on tärkeää ottaa huomioon tietojen täsmällisyys kaikissa profilointiprosessin vaiheissa ja etenkin:

- tietoja kerätessä
- tietoja analysoidessa
- kun luodaan henkilöstä profiilia tai
- kun profiilia käytetään henkilöön vaikuttavan päätöksen tekemiseen[166]

---

[163] Yleisen tietosuoja-asetuksen 6 artiklan 4 kohta.
[164] WP251rev.01 2018,12.
[165] WP251rev.01 2018,12.
[166] WP251rev.01 2018,12.

Vaikka niin sanottu raakadata kirjataankin täsmällisesti, tietue ei ole välttämättä täysin edustava. Päätöksiä voidaan myös tehdä esimerkiksi vanhentuneiden tietojen perusteella tai ulkoisten tietojen perusteella, joita on tulkittu väärin. Tietojen analyysissa voi olla myös piilevä vinouma[167]. Tällaiset epätarkkuudet voivat johtaa epäasianmukaisiin ennusteisiin tai vaikkapa lausuntoihin luotto- tai vakuutusriskeistä. [168]

Etenkin koneoppimisen algoritmit on suunniteltu käyttämään suuria tietomääriä ja luomaan assosiaatioita, joiden avulla yritykset voivat luoda henkilöistäkin hyvin kattavia ja henkilökohtaisia profiileja. Tällaisten tietojen säilyttämisestä olisi yrityksen näkökulmasta hyötyä, mutta rekisterinpitäjien on kuitenkin noudatettava tietojen minimoinnin periaatetta kerätessään ja käsitellään henkilötietoja. Samalla on varmistettava myös, että kerätyt tiedot säilytetään vain niin kauan kuin on tarpeen ja tarkoituksenmukaista. [169]

## 5.4 Tietojen käsittelyn oikeudelliset perusteet

Automatisoidut päätökset ovat sallittuja vain, jos yleisessä tietosuoja-asetuksessa nimenomaisesti kuvattuja poikkeuksia sovelletaan. [170] Yleisen tietosuoja-asetuksen[171] mukaan rekisteröidyllä on oikeus olla joutumatta sellaisen päätöksen kohteeksi, joka perustuu pelkästään automaattisen käsittelyyn ja jolla on häntä koskevia oikeusvaikutuksia, tai joka vaikuttaa häneen vastaavalla tavalla merkittävästi. Asetus määrittelee kuitenkin myös poikkeukset tähän kieltoon. Yksi näistä poikkeuksista on, jos automaattinen päätös on *välttämätön* rekisteröidyn ja rekisterinpitäjän välisen sopimuksen tekemistä tai täytäntöönpanoa varten.

Henkilötietojen suoja on määritelty perusoikeudeksi ja sitä voidaan rajoittaa vain tietyin edellytyksin. Oikeuksien ja vapauksien käyttämistä voidaan rajoittaa ainoastaan lailla sekä kyseisten oikeuksien ja vapauksien keskeistä sisältöä kunnioittaen. Rajoituksia voidaan säätää ainoastaan, jos ne ovat *välttämättömiä*. [172] Tästä voidaan vetää tiettyjä johtopäätöksiä myös tietosuoja-asetuksen profilointikiellon poikkeuksen edellytyksille.

---

[167] Vinoumasta tarkemmin kappaleessa 7.
[168] WP251rev.01 2018,12.
[169] WP251rev.01 2018,13.
[170] Yleinen tietosuoja-asetus 22 artikla 1 ja 2 kohdat.
[171] Yleinen tietosuoja-asetus 22 artiklan 1 ja 2 kohdat.
[172] Rajoittaminen on määritelty Euroopan unionin perusoikeuskirjan (2012/C 326/02) 52 artiklan 1 kohdassa.

Euroopan tietosuojavaltuutettu[173] on laatinut ohjeistuksia siitä, miten välttämättömyyttä tulisi tulkita tietosuojan ja yksityisyyden suojan yhteydessä. Ohjeistukseen kuulu neljän kohdan tarkistuslista uusien lainsäädäntötoimien arvioimiseksi. Ensimmäinen askel on toimen kuvaus faktapohjaisesti. Toinen on niiden perusoikeuksien tunnistaminen, joita tietojenkäsittely rajoittaa. Kolmanneksi uuden lainsäädännön tavoite tulisi määritellä ja lopuksi vielä ohjeistus kehottaa valitsemaan sen vaihtoehdon, joka on tehokkain, mutta vähiten rajoittava. Näistä erityisesti on kiinnostava henkilötietojen käsittelyn välttämättömyyden tarkastelun kannalta.

Euroopan tietosuojavaltuutetun kanta on, että henkilötietojen käsittely on lähtökohtaisesti aina henkilötietojen suojan perusoikeuden rajoittamista ja sen välttämättömyyttä tulee tarkastella. Sillä ei ole merkitystä, ovatko käsiteltävät henkilötiedot arkaluontoisia tai onko niiden käsittely aiheuttanut haittaa rekisteröidylle. [174]

### *5.4.1 Rekisteröidyn nimenomainen suostumus*

Henkilölle esitettäviin tietojenkäsittelyn hyväksymispyyntöihin tulee soveltaa tiukkoja vaatimuksia, koska kyse on yksilön perusoikeuksista ja koska rekisterinpitäjä haluaa toteuttaa tietojenkäsittelytoimen, joka olisi laiton ilman rekisteröidyn suostumusta. [175] Yleisen tietosuoja-asetuksen 4 artiklan 11 alakohdassa mainitulla suostumuksella tarkoitetaan ”*mitä tahansa vapaaehtoista, yksilöityä, tietoista ja yksiselitteistä tahdonilmaisua, jolla rekisteröity hyväksyy henkilötietojensa käsittelyn antamalla suostumusta ilmaisevan lausuman tai toteuttamalla selkeästi suostumusta ilmaisevan toimen*”.

Hieman yleistäen suostumus voi olla laillinen peruste tietojenkäsittelylle vain, jos rekisteröidylle tarjotaan aidosti mahdollisuus valvoa tietojensa käyttöä sekä aito tilaisuus valita vapaasti, hyväksyykö hän tarjotut ehdot vai hylkääkö hän ne, eikä hylkäämisestä koidu hänelle haittaa. [176] Pätevän suostumuksen osatekijät koostuvat tahdonilmaisusta, joka on vapaaehtoinen, yksilöity, tietoinen ja yksiselitteinen. Tahdonilmaisulla rekisteröity hyväksyy henkilötietojensa käsittelyn antamalla suostumusta ilmaisevan lausuman tai toteuttamalla selkeästi suostumusta ilmaisevan toimen. [177]

Pätevän suostumuksen osatekijä ”vapaaehtoinen” edellyttää rekisteröidyn todellista vapaan valinnan ja valvonnan mahdollisuutta. Silloin, kun rekisteröity tuntee olevansa pakotettu antamaan suostumuksensa tai jos hänelle aiheutuu kielteisiä seurauksia siitä, ettei anna suostumustaan tietojensa käsittelyyn, annettua suostumusta ei voida pitää

[173] EDPS 2017.
[174] EDPS 2017, 11.
[175] WP259 rev.01 2018, 3.
[176] WP259 rev.01 2018, 3.
[177] WP259 rev.01 2018, 5.

vapaaehtoisena ja siten pätevänä. [178] Euroopan tuomioistuimen tuomiossa Planet49 -tapauksessa[179] oli ratkaistavana, että voiko valmiiksi rastitetulla ruudulla käyttäjä antaa suostumuksensa tietojenkäsittelyyn. Tapaukseen liittyvässä julkisasiamiehen ratkaisuehdotuksessa[180] korostettiin sitä, että käyttäjältä edellytetään aktiivisia toimenpiteitä, jotta suostumus voidaan antaa pätevästi. Myös tuomioistuin päätyi siihen, että suostumusta tietojenkäsittelyyn ei voida antaa pätevästi valmiiksi rastitetulla ruudulla vaan edellytetään aktiivisia toimia. [181]

Pätevän suostumuksen osatekijät ’yksilöity’ ja ’tietoinen’ tuli esille myös esille edellä mainitun Planet49 -tapauksen käsittelyn yhteydessä. Julkisasiamiehen ratkaisuehdotuksen kohdassa 72 annetaan pätevästä suostumuksesta esimerkkejä; käyttäjä voi verkkosivuilla vieraillessaan rastittaa ruudun, jolla antaa suostumuksensa tietojen käsittelyyn, tai toimii tavalla, joka selkeästi osoittaa, että hän hyväksyy henkilötietojensa käsittelyä koskevan ehdotuksen.

Lisäksi sama tapaus osoitti, että pätevän suostumuksen eri osatekijät tulee täyttyä yhdessä eikä erikseen. Julkisasiamies tulkitsi[182], että jotta suostumus annetaan ”vapaaehtoisesti” ja ”tietoisesti”, suostumuksen tulee olla paitsi aktiivinen, myös ”erillinen”. Erillisyydellä viitataan siihen, että pätevän suostumuksen antaminen ei voi olla osa käyttäjän muuta toimintaa vaan nimenomaan siitä erillinen. Käyttäjä ei voi siis esimerkiksi videota katsoessaan antaa pätevästi suostumusta tietojen käsittelyyn.

### *5.4.2 Tarpeellisuus ja oikeutettu etu*

Yritykset saattavat haluta käyttää profilointia tai automatisoitua päätöksentekoprosessia osana esimerkiksi luotonhallinnan prosessejaan, koska

- Ne voivat lisätä prosessin johdonmukaisuutta tai oikeudenmukaisuutta, kun esimerkiksi inhimillisten virheiden, syrjinnän ja vallan väärinkäytön mahdollisuudet pienenevät
- Ne vähentävät luottoriskiä
- Ne lisäävät tehokkuutta nopeuttamalla prosesseja[183]

---

[178] WP259 rev.01 2018, 5-6.
[179] EUTI C-673-/17 Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband eV v. Planet49 GmbH, ennakkoratkaisu 1.10.2019.
[180] Julkisasiamiehen ratkaisuehdotus 21.3.2019 asiassa C-673/17, Planet49 GmbH v. Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband e.V. kohdat 59, 60, 61 ja 122.
[181] EUTI C-673-/17 Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband eV v. Planet49 GmbH, ennakkoratkaisu 1.10.2019, ratkaisun kohta 1.
[182] Julkisasiamiehen ratkaisuehdotus 21.3.2019 asiassa C-673/17, Planet49 GmbH v. Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband e.V. , kohta 66.
[183] WP251rev.01 2018, 14.

Yksinään nämä näkökohdat eivät kuitenkaan riitä osoittamaan, että tällainen tietojen käsittely on tietosuoja-asetuksen tarkoittamalla tavalla[184] tarpeellista.

Tietosuoja-asetuksen 6 artiklan 1 kohdan mukaan:

> 1. Käsittely on lainmukaista ainoastaan jos ja vain siltä osin kuin vähintään yksi seuraavista edellytyksistä täyttyy:
> [- -]
> f) käsittely on tarpeen **rekisterinpitäjän** tai **kolmannen osapuolen** ***oikeutettujen etujen*** toteuttamiseksi, paitsi milloin henkilötietojen suojaa edellyttävät rekisteröidyn edut tai perusoikeudet ja -vapaudet syrjäyttävät tällaiset edut, erityisesti jos rekisteröity on lapsi.[185]

Oikeutettuja etuja[186] ovat ainakin käsittely suoramarkkinointitarkoituksessa ja ehdottoman välttämätön käsittely petosten estämistarkoituksessa. Tätä kohtaa ei kuitenkaan voi soveltaa vain siksi, että rekisterinpitäjällä tai kolmannella osapuolella on oikeutettu etu. Rekisterinpitäjän on suoritettava niin sanottu tasapainotesti, jossa arvioidaan, menevätkö rekisteröidyn edut tai perusoikeudet ja -vapaudet sen etujen edelle. [187]

Ainakin seuraavat seikat ovat tasapainotestin kannalta merkityksellisiä:

- Profiilin yksityiskohtaisuus
- Profiilin kattavuus
- Profiloinnin vaikutukset (seuraukset rekisteröidylle)
- Suojaavat toimet, joilla pyritään varmistamaan profilointiprosessin kohtuullisuus, syrjimättömyys ja täsmällisyys[188]

Jotta tietojenkäsittelyn toimeen liittyvä etu olisi oikeutettu, edun tulee olla laillinen[189], riittävän tarkasti määritelty ja koskea aitoa etua. Pelkästään se, että oikeutettu etu olisi laillinen, ei yksinään riitä vielä täyttämään tasapainotestin kriteereitä. Esimerkiksi rekisterinpitäjän edun mukaista voisi olla suorittaa profilointia ja personoida tarjouksia asiakkailleen heidän mieltymystensä mukaisesti. 6 artiklan kohdan 1 alakohdan f mukaista oikeutettua etua voisi mahdollisesti käyttää profiloinnin perusteena, etenkin jos asianmukaiset varotoimet on pantu täytäntöön. Tällaisia varotoimia ovat esimerkiksi rekisteröidyn mahdollisuus vastustaa automaattista käsittelyä. Rekisterinpitäjä ei voi kuitenkaan automaattisesti aina tukeutua tähän säännökseen ja esimerkiksi yhdistellä laajamittaisesti dataa eri lähteistä kootakseen hyvin yksityiskohtaisen profiilin rekisteröidystä. Jos näin tapahtuisi, rekisteröidyn oikeudet ohittaisivat todennäköisesti rekisterinpitäjän oikeutetun edun. [190]

---

[184] Tietosuoja-asetuksen 6 artiklan 1 kohdan b alakohta.
[185] Tätä kohtaa ei sovelleta tietojenkäsittelyyn, jota viranomaiset suorittavat tehtäviensä yhteydessä.
[186] Englanniksi käytetään sanaa *'legitimate interest'* (Ks. esim. WP217 2014, 25). Sana *interest* viittaa mielestäni enemmän mielenkiinnon kohteeseen, kuin etuun sinänsä.
[187] WP251 2018, 15.
[188] WP251 2018, 15.
[189] WP217 2014, 25 alaviite 48. Laki laajassa merkityksessä.
[190] WP217 2014, 25.

Tietosuojatyöryhmän kannanotossa SWIFT -tapaukseen[191] arvioitavaksi tuli, että tulisiko SWIFT -organisaation luovuttaa pankkisanomiin liittyviä tietoja terrorismin vastaista työtä varten vai tulisiko rekisteröityjen etuja suojella, koska heille ei ollut etukäteen informoitu tällaisesta tietojen käsittelystä. Tapauksessa ei keskitytty terrorismin ehkäisyyn sinänsä arvona vaan analysoitiin tilannetta SWIFT:n kannalta välittömien vaikutuksien kautta. SWIFT joutuisi uhkasakon uhalla luovuttamaan tiedot. Tietosuojatyöryhmä kuitenkin päätteli, että vaikka etu olisikin oikeutettu, se ei olisi riittävän painava syrjäyttämään rekisteröityjen oikeuksia. Työryhmä kuvaili tietojenkäsittelyä ”piileväksi ja laajamittaiseksi” eikä käsittelyyn liittyen rekisteröidyillä ollut ennakkotietoa siitä ja siten mahdollista vastustaa käsittelyä.

Unionin tuomioistuimen ennakkoratkaisussa koskien ”Rīgas satiksme” -tapausta[192] todettiin, että oikeutettua etua on tulkittava siten, ettei siinä aseteta velvollisuutta luovuttaa henkilötietoja sivulliselle, jotta tämä voi nostaa siviilioikeudellisen vahingonkorvauskanteen vaatiakseen henkilön, jota henkilötietosuoja koskee, aiheuttaman vahingon korvaamista. Julkisasiamiehen ratkaisuehdotuksessa[193] käsitellään tarkemmin oikeutettua etua. Kyseisen säännöksen soveltaminen edellyttää ensinnäkin, että joko rekisterinpitäjällä tai kolmannella osapuolella on oikeutettu etu. Julkisasiamies pitää selvänä, että tällaisena oikeutettuna etuna voidaan pitää sivullisen intressiä saada hänen omaisuuttaan vahingoittaneen henkilön henkilötiedot, jotta hän voi vaatia tältä vahingonkorvauksia. Intressivertailussa otettiin huomioon rekisteröidyn ikä (alaikäinen), kanteen nostamiseksi vaadittavat tiedot, näiden tietojen luonne ja arkaluonteisuus, niiden saatavuus yleisistä lähteistä sekä rikkomuksen vakavuus. Myös kansallisella lainsäädännöllä on luonnollisesti merkitystä asian arvioinnissa. Tässä tapauksessa tuomiossa todettiin lopuksi vielä, että direktiivin 95/46/EY artiklan 7 alakohtaa f (oikeutettu etu) ei ole kuitenkaan esteenä tietojen luovuttamiselle. Vaikka siis oikeutetun edun todettiin olevan olemassa, se ei kuitenkaan läpäissyt tasapainotestiä.

---

[191] WP128 2006, executive summary. SWIFT (Society for Worldwide Interbank Financial Telecommunication) fasilitoi kansainvälisiä rahasiirtoja pankkien välillä. Se säilyttää operaatioihin liittyvät sähköiset viestit 124 päivän ajan kahdessa keskuksessa, joista toinen on EU:n sisällä ja toinen USA:ssa. Keskukset peilaavat tietosisällön välillään eli kummassakin on samat tiedot kaikkina ajanhetkinä. Sanomatiedoissa on henkilötietoja kuten maksajan nimi ja maksunsaajan nimi. Vuoden 2001 terroristihyökkäyksien jälkeen USA vaati tuolloin kanteella, että SWIFT luovuttaa USA:n keskuksen tiedot käyttöönsä. Asia tuli julkiseksi vasta 2006, kun asiasta kirjoitettiin lehdistössä. Tietosuojatyöryhmän mietintö arvioi tapahtumia EU-lainsäädännön näkökulmasta.

[192] EUVL 2017/C 213/10. C-13/16. Valsts policijas Rīgas reģiona pārvaldes Kārtības policijas pārvalde vs. Rīgas pašvaldības SIA ”Rīgas satiksme”. Tuomioistuin on tehnyt ratkaisunsa direktiivin 95/46/EY perusteella ja viittaa tuomiossaan sen artiklan 7 artiklan f alakohtaan. Sen tarkka sanamuoto on ”*jos käsittely on tarpeen rekisterinpitäjän tai tiedot saavan sivullisen oikeutetun intressin toteuttamiseksi, paitsi milloin tämän intressin syrjäyttävät rekisteröidyn 1 artiklan 1 kohdan perusteella suojaa tarvitsevat intressit ja perusoikeudet ja -vapaudet*”.

[193]

Profilointi voi olla sallittua oikeutetun edun perusteella myös silloin, kun kyse on kolmannesta osapuolesta, jolle tiedot on annettu. Tietosuojatyöryhmä[194] tarjoaa muutamia esimerkkitilanteita, joissa voisi tulla kysymykseen tällainen kolmannen osapuolen oikeutettu etu. "*Tietojen julkistaminen avoimuuden ja vastuuvelvollisuuden takia* voisi olla yksi tällainen tilanne. Tässä tilanteessa tulisi arvioida nimenomaan kenen edun mukaista tietojen julkistaminen olisi. Vain rekisterinpitäjän edun mukaista se ei saisi olla vaan nimenomaan muiden, ulkopuolisten sidosryhmien kuten journalistien tai suuren yleisön. Joka tapauksessa tasapainotesti osoittaa lopulta, että voiko oikeutettua etua käyttää perusteena. [195] Toinen tapaus, jossa kolmannen osapuolen oikeutettu etu voisi tulla kyseeseen on *historiallinen ja muunlainen tieteellinen tutkimus*. Tämä peruste voisi tulla kyseeseen etenkin, kun tutkimuksen tekeminen edellyttää pääsyä tiettyihin tietokantoihin. Tässäkin tapauksessa tulee huolehtia tarvittavista suojatoimista. [196]

Kolmannen osapuolen oikeutettu etu voi tulla kysymykseen tietojen käsittelyn oikeudellisena perusteena myös muussakin tapauksessa. Silloin kun kyse on *suuren yleisön tai muun kolmannen osapuolen oikeutetusta edusta*, joka voi olla kysymyksessä, kun rekisterinpitäjä ajaa omaa etuaan, joka korreloi myös toisen tahon intressien kanssa. Esimerkiksi rekisterinpitäjä saattaa mahdollisesti viranomaisen kehotuksesta avustaa lainvalvontaa vaikkapa rahanpesun estämiseen tähtäävissä toimenpiteissä. [197]

Tietosuojatyöryhmän lausunto oikeutetusta intressistä perustuu vanhempaan tietosuojadirektiiviin 95/46/EY, sen sisältämät esimerkit ovat kuitenkin edelleen valideja profilointia suorittaville rekisterinpitäjille. [198]

### *5.4.3 Lakisääteiset velvoitteet, elintärkeät ja yleiset edut*

Tietyissä tapauksissa voi olla lakisääteinen velvoite tehdä profilointia esimerkiksi petosten tai rahanpesun ennaltaehkäisyyn. Yleisen tietosuoja-asetuksen johdantokappale 41 viittaa tällaiseen lakisääteiseen velvoitteeseen:

---

[194] WP217 2014, 27.
[195] Tietosuojatyöryhmä korostaa myös, että asetuksen 5 artiklan mukainen avoimuus -periaate voi olla myös tietojen käsittelyn oikeudellinen peruste ja tietojen julkaisun peruste sinänsä.
[196] WP217 2014, 28. Ks. myös WP203 2013 käyttötarkoitussidonnaisuudesta: tietojen sekundääriselle käytölle tulisi tehdä toinen tasapainotesti. Ensin tulisi varmistaa käsittelytoimen käyttötarkoituksenmukaisuus ja sitten sen oikeudellinen peruste.
[197] WP217 2014, 28.
[198] WP251 2018, 15.

> "Aina kun tässä asetuksessa viitataan käsittelyn oikeusperusteeseen tai lainsäädäntötoimeen, siinä ei välttämättä edellytetä parlamentissa hyväksyttyä säädöstä, sanotun kuitenkaan rajoittamatta asianomaisen jäsenvaltion perustuslaillisen järjestyksen edellyttämien vaatimusten soveltamista. Kyseisen käsittelyn oikeusperusteen tai lainsäädäntötoimen olisi kuitenkin oltava selkeä ja täsmällinen ja sen soveltamisen henkilöiden kannalta ennakoitavissa olevaa Euroopan unionin tuomioistuimen, jäljempänä 'unionin tuomioistuin', ja Euroopan ihmisoikeustuomioistuimen oikeuskäytännön mukaisesti."[199]

Tietojen käsittelyn oikeusperusteen ei tarvitse olla siis laki tai säädös, mutta sen tulee olla selkeä, täsmällinen ja sen soveltamisen tulee olla henkilöiden kannalta ennakoitavissa. Yleisen tietosuoja-asetuksen johdantokappale 45 laajentaa ja täsmentää tietojenkäsittelyn oikeusperustetta lisää:

> "Kun käsittely tapahtuu rekisterinpitäjää koskevan lakisääteisen velvoitteen mukaisesti tai kun se on tarpeen yleisen edun vuoksi toteutettavan tehtävän tai rekisterinpitäjälle kuuluvan julkisen vallan käyttämiseksi, käsittelyllä olisi oltava perusta unionin oikeudessa tai jäsenvaltion lainsäädännössä. Tässä asetuksessa ei edellytetä, että kaikkia yksittäisiä tiedonkäsittelytilanteita varten olisi olemassa erityislaki. Useiden käsittelytoimien perustana oleva yksi laki voi olla riittävä käsittelyn perustuessa rekisterinpitäjän lakisääteisen velvoitteeseen tai jos käsittely on tarpeen yleisen edun vuoksi toteutettavan suorittamiseksi tai julkisen vallan käyttämiseksi - -." [200]

Kaikkia yksittäisiä tiedonkäsittelytilanteita varten ei tarvitse siis olla aina omaa erityislakia vaan useiden käsittelytoimien perustana oleva yksi laki voi olla riittävä. Tätä käsittelyn tarkoitusta tulisi tosin määriteltävä unionin oikeudessa tai jäsenvaltion lainsäädännössä. Lisäksi todetaan, että kyseisessä oikeudessa voidaan myös täsmentää yleisen tietosuoja-asetuksen säätämiä yleisiä edellytyksiä.

Profilointi ja automaattinen päätöksenteko ovat sallittuja, kun se on tarpeellista elintärkeiden etujen suojaamiseksi. Tässä viitataan nimenomaan luonnollisen henkilön hengen kannalta elintärkeisiin etuihin. Tietynlainen tietojenkäsittely voi palvella sekä tärkeitä yleisiä etuja että rekisteröidyn elintärkeitä etuja.[201] Tietosuojatyöryhmä esittää esimerkkinä profiloinnin, joka on tarpeen hengenvaarallisten sairauksien leviämistä ennustavien mallien laatimiseksi tai humanitäärisissä tilanteissa. Näissä tapauksissa rekisterinpitäjä voi käyttää elintärkeitä etuja perusteena vain, jos käsittelylle ei ole muuta perustetta. [202]

Tietosuojatyöryhmä toteaa, että "[y]leisen tietosuoja-asetuksen 6 artiklan 1 kohdan e alakohta saattaa olla asianmukainen peruste julkisella sektorilla suoritettavalle profiloinnille tietyissä olosuhteissa" sekä, että "[t]ehtävällä tai toiminnolla on oltava selvä lakisääteinen perusta".

---

[199] Yleisen tietosuoja-asetuksen johdanto-osan 41 ja 45 kappale.
[200] Yleisen tietosuoja-asetuksen johdanto-osan 45 kappale.
[201] WP251rev.01 2018, 15.
[202] Yleisen tietosuoja-asetuksen johdanto-osan 46 kappale.

## 5.5 Rekisteröidyn oikeudet

### *5.5.1 Oikeus saada tietoja*

Yleisen tietosuoja-asetuksen artiklat 13 ja 14 käsittelevät avoimuusperiaatetta, jonka mukaan rekisterinpitäjien on varmistettava, että ne selittävät rekisteröidyille selkeästi ja yksinkertaisesti, miten profilointi ja automatisoitu päätöksentekoprosessi toimivat.[203] Etenkin kun profiloinnin perusteella tehdään päätöksiä[204], rekisteröidylle on tehtävä selväksi, tietojen käsittelyn tarkoituksena sekä profilointi, että päätöksen tekeminen luodun profiilin perusteella. Rekisteröidylle on myös oikeus vastustaa profilointia tietyissä tilanteissa riippumatta siitä, että tehdäänkö sen perusteella päätöksiä. [205]

Rekisteröidyllä on oikeus saada tietoja, kun rekisterinpitäjä tekee tietosuoja-asetuksessa tarkoitettuja automatisoituja päätöksiä, rekisterinpitäjän on 1) ilmoitettava rekisteröidylle harjoittavansa tällaista toimintaa, 2) annettava merkityksellisiä tietoja käsittelyyn liittyvästä logiikasta ja 3) selitettävä käsittelyn merkittävyys ja mahdolliset seuraukset. Vaikka tietojen käsittelytoimet eivät täyttäisikään automatisoitujen päätöksien tai profiloinnin tunnusmerkkejä tietosuoja-asetuksen tarkoittamalla tavalla, on kuitenkin hyvän käytännön mukaista antaa nämä tiedot joka tapauksessa. Tietosuojatyöryhmä käyttää esimerkkinä koneoppimista ja kertoo sen yleistyvän sekä monimutkaistuvan. Tämä haastaa rekisteröidyn oikeuksien toteutumista. [206]

Tietosuoja-asetus edellyttää, että rekisterinpitäjä toimittaa merkityksellisiä tietoja rekisteröidylle tietojen käsittelyyn liittyvästä logiikasta, mutta monimutkaista selitystä siinä käytettävistä algoritmeista tai koko algoritmin paljastamista ei välttämättä edellytetä. Toisaalta algoritmin monimutkaisuus ei myöskään ole peruste olla toimittamatta tietoja. Joka tapauksessa toimitettujen tietojen tulisi olla kuitenkin riittävän kattavat, jotta rekisteröity ymmärtää päätöksen perusteet. Toimitettujen tietojen tulee olla merkityksellisiä ja sen tueksi tietosuojatyöryhmä mainitseekin, että konkreettisia esimerkkejä olisi toimitettava näiden tietojen mukana. [207]

---

[203] Ks. myös WP251 2018, 17.
[204] Tässä viitataan myös päätöksiin, joilla ei ole yleisen tietosuoja-asetuksen artiklan 22 mukaisia oikeudellisia vaikutuksia.
[205] WP251 2018, 18.
[206] WP251 2018, 26-27. Ks. myös yleisen tietosuoja-asetuksen johdantokappale 60.
[207] WP251 2018, 27. Ks. myös yleisen tietosuoja-asetuksen johdantokappale 58.

### *5.5.2 Oikeus saada pääsy tietoihin*

Yleisen tietosuoja-asetuksen artiklan 15 mukaan rekisteröidyllä on oikeus saada tietoja profilointiin käytetyistä henkilötiedoista mukaan lukien niin sanotut erityiset henkilötietoryhmät. [208] Rekisteröidyllä on oikeus saada vahvistus siitä, että häntä koskevia tietoja käsitellään tai ei käsitellä. Jos tietoja käsitellään, hänellä on oikeus saada pääsy kyseisiin tietoihin sekä lisäksi yleisessä tietosuoja-asetuksessa 15 artiklan kohdan 1 alakohdissa määritellyt tiedot. Rekisteröidyllä on näiden alakohtien mukaan oikeus saada tiedot muun muassa seuraavista asioista:

- tietojen käsittelyn tarkoitus
- suunniteltu käsittelyaika tai ainakin tämän ajan määrittämisen kriteerit
- jos tietoja ei kerätä rekisteröidyltä, tietojen kaikki alkuperästä käytettävissä olevat tiedot
- automaattisen päätöksenteon (mukaan lukien profiloinnin) olemassaolo sekä ainakin merkitykselliset tiedot tällaiseen käsittelyyn liittyvästä logiikasta, käsittelyn merkittävyys ja mahdolliset seuraukset rekisteröidylle

Rekisterinpitäjä on lisäksi velvollinen asettamaan saataville profiilin luomisessa syöttötietoina käytetyt tiedot sekä antamaan pääsyn tietoihin profiilista ja niistä segmenteistä, joihin rekisteröity on sijoitettu. Tämä velvollisuus eroaa 20 artiklassa säädetystä rekisteröidyn oikeudesta siirtää tietoja järjestelmästä toiseen. Sen mukaan rekisterinpitäjän on ilmoitettava vain rekisteröidyn toimittamat tai rekisterinpitäjän havainnoimat tiedot, mutta ei itse profiilia. [209]

Yleisen tietosuoja-asetuksen johdantokappaleessa 63 annetaan täsmennyksiä sekä jonkin verran suojaa myös rekisterinpitäjälle, joka voi olla huolissaan immateriaalioikeuksistaan:

> ”[- -] Tämä oikeus ei saisi vaikuttaa epäedullisesti muiden oikeuksiin ja vapauksiin, joita ovat esimerkiksi liikesalaisuudet tai henkinen omaisuus ja erityisesti ohjelmistojen tekijänoikeudet. Näiden seikkojen huomioon ottaminen ei kuitenkaan saisi johtaa siihen, että rekisteröidylle ei anneta minkäänlaista tietoa. Jos rekisterinpitäjä käsittelee merkittäviä määriä rekisteröityä koskevia tietoja, rekisterinpitäjä olisi voitava pyytää rekisteröityä täsmentämään riittävällä tavalla ennen tietojen luovuttamista, mitä tietoja tai mitä käsittelytoimia rekisteröidyn pyyntö koskee.”

[208] Erityiset henkilötietoryhmät kts. tarkemmin yleinen tietosuoja-asetus 9 artikla. Erityisiä henkilötietoryhmien tietoja ei saa käsitellä. Erityisinä henkilötietoryhminä pidetään ainakin rotua, etnistä alkuperää, uskonnollista ja filosofista vakaumusta, poliittista mielipidettä, ammattiliiton jäsenyyttä, seksuaalista suuntautumista ja esimerkiksi terveyttä koskevia tietoja. Tietoja voi kuitenkin käsitellä artiklan 2 kohdassa säädetyin poikkeuksin, kuten rekisteröidyn suostumuksella tai jos rekisteröity on jo muuten saattanut tiedon julkiseksi. Jäsenvaltiot voivat ottaa käyttöön lisäehtoja ja rajoituksia, jotka koskevat geneettisten tietojen, biometristen tietojen tai terveystietojen käsittelyä.

[209] WP251 2018, 18.

Siitä huolimatta, että kappaleessa mainitaan ”*ei saisi vaikuttaa*”, se ei anna täsmennyksiä kuitenkaan siihen, että miten rekisterinpitäjä voisi lieventää riskiä omien immateriaalioikeuksien loukkaamiseen.

### *5.5.3 Oikeus tietojen poistamiseen, käsittelyn rajoittamiseen ja tietojen oikaisemiseen*

Yleisen tietosuoja-asetuksen artiklan 16 mukaan rekisteröidyllä on oikeus tietojensa oikaisemiseen. Hänellä on oikeus vaatia, että rekisterinpitäjä oikaisee ilman aiheetonta viivytystä häntä koskevat epätarkat ja puutteelliset tiedot. Artiklan 17 mukaan rekisteröidyllä on niin sanottu ”oikeus tulla unohdetuksi” eli rekisterinpitäjä tulee vaatimuksesta poistaa ilman aiheetonta viivytystä rekisteröityä koskevat tiedot sillä edellytyksellä, että tietty perustevaatimus täyttyy. Vastaavasti rekisterinpitäjällä on myös edellä kuvatusta riippumaton velvollisuus itsenäisesti poistaa rekisteröidyn tiedot, mikäli jokin perustevaatimuksista täyttyy.

Tällaisia perusteita voivat olla muun muassa seuraavat tilanteet:

- Henkilötietoja ei tarvita enää niihin tarkoituksiin, joihin ne on kerätty
- Rekisteröity peruuttaa suostumuksensa tietojen käsittelyyn
- Henkilötietoja on käsitelty lainvastaisesti

Jos rekisterinpitäjä on julkistanut rekisteröidyn tiedot ja sillä on artiklan 17 kohdan 1 mukainen velvollisuus poistaa ne, sen on käytettävissä oleva teknologia ja toteuttamiskustannukset huomioon ottaen toteutettava kohtuulliset toimenpiteet, ilmoittaakseen tietoja käsitteleville tahoille, että rekisteröity on pyytänyt rekisterinpitäjiä poistamaan näihin henkilötietoihin liittyvät linkit tai tietoihin liittyvät jäljennökset ja kopiot. Tähän on kuitenkin säädetty poikkeukset artiklan 17 kohdassa 3, jolloin rekisterinpitäjällä ei ole velvollisuutta poistaa tietoja. Tällaisia tilanteita ovat esimerkiksi sanavapautta ja vapaata tiedonvälitystä koskevan oikeuden käyttäminen.

Rekisteröidyllä on myös oikeus rajoittaa tietojen käsittelyä artiklan 18 mukaan silloin, kun

- Rekisteröity kiistää tietojensa paikkansapitävyyden
- tietojen käsittely on lainvastaista ja rekisteröity vastustaa niiden poistamista
- rekisterinpitäjä ei enää tarvitse kyseisiä tietoja käsittelyn tarkoituksiin, mutta rekisteröity tarvitsee niitä oikeudellisen vaateen tekemiseksi
- rekisteröity on vastustanut tietojen käsittelemistä 21 artiklan 1 kohdan nojalla odottaessa, että syrjäyttääkö rekisterinpitäjän oikeutetut perusteet rekisteröidyn oikeudet

Rekisterinpitäjällä on lisäksi velvollisuus 19 artiklan mukaan ilmoittaa kaikista artiklan 16, 17 ja 18 mukaisista henkilötietojen oikaisuista, poistoista tai käsittelyn rajoituksista jokaiselle vastaanottajalle, jolle henkilötietoja on luovutettu ottaen huomioon kohtuullisuuden.

### *5.5.4 Vastustamisoikeus*

Rekisteröidyllä on oikeus vastustaa häntä koskevaa henkilötietojen käsittelyä[210] milloin tahansa hänen henkilökohtaiseen erityiseen tilanteeseensa liittyvällä perusteella. Tällöin rekisterinpitäjä ei saa enää käsitellä rekisteröidyn henkilötietoja, ellei rekisterinpitäjä voi osoittaa, että käsittelyyn on olemassa huomattavan tärkeä ja perusteltu syy, joka syrjäyttää rekisteröidyn edut, oikeudet ja vapaudet tai jos se on tarpeen oikeusvaateen laatimiseksi. Rekisteröidyllä on kuitenkin aina oikeus vastustaa suoramarkkinointiin liittyvää henkilötietojen käsittelyä. Tämä vastustamisoikeus on nimenomaisesti saatettava rekisteröidyn tietoon ja se on esitettävä selkeästi ja muusta tiedotuksesta erillään.

[210] Viittaa nimenomaan yleisen tietosuoja-asetuksen 6 artiklan 1 kohdan alakohtaan e tai f, jotka liittyvät oikeutettuun etuun ja yleistä etua koskevan tehtävän suorittamiseen tai julkisen vallan tehtävän suorittamiseen.

# 6 DATASTA SAATAVAT HYÖDYT

## 6.1 Datan merkitys

Data on tiedon perusyksikkö ja sitä voi olla monessakin eri muodossa. Data on niin sanottua raakadataa, joka on yleisemmin numeroita, tekstiä, kuvia ja vaikkapa lukuja. Kun dataan liitetään jokin merkitys, siitä muodostuu tietoa. Kun tietoa edelleen sidotaan kontekstiin, saadaan merkityksellistä tietämystä. Vasta tällä on merkitystä ja niin sanotulla raakadatalla ei yksinään. Tekoälyn kouluttamiseen tarvitaan dataa, mutta sen pitää olla laadukasta eli yhdenmukaista ja yksiselitteistä. Tekoäly ei voi saada huonolaatuisesta datasta parempaa. [211]

Datan lähteitä voi olla useita, mutta yrityksen on helpointa lähteä tutkimaan omien rekisteriensä dataa. Tällä tarkoitetaan vaikkapa asiakasrekisteriä, taloustietoja ja esimerkiksi verkkokäyttäytymistä. [212] Datan hyödyntäminen riippuu muun muassa siitä, onko data hyödyntäjän omaisuutta (rekisterinpitäjä) ja mihin tarkoitukseen se on kerätty.

Kun tekoälyllä tehdään käytännön sovelluksia, on erittäin tärkeää tunnistaa ratkaistavaan ongelmaan vaikuttavat asianhaarat tai ominaisuudet ja löytää nämä asiat sovelluksen pohjana käytettävästä datasta. Tässä tarvitaan ihmistä, jolla on vahva kyseisen asian tuntemus. Tästä vaiheesta käytetään myös puhekielessä nimitystä muuttujien valintana. Muuttujien valinta voidaan tehdä manuaalisesti asiantuntijan toimesta, käyttäen soveltuvaa matemaattista tapaa[213] tai näitä kahta tapaa yhdistelemällä.[214]

Dataa koskevat kysymykset liittyvät rekisteröityjen tietosuojan lisäksi myös yleisesti dataa koskeviin vääristymiin. Monet tekoälyjärjestelmät tarvitsevat valtavia määriä dataa toimiakseen hyvin, on hyvin tärkeää ymmärtää, miten data vaikuttaa tekoälyjärjestelmän käyttäytymiseen. Jos tekoälyn opettamisessa käytettävässä datassa on vaikkapa rasistisia vääristymiä, tällaisella datalla opetettu tekoäly ei pysty tekemään yleistyksiä asiallisella tavalla ja sen käyttäminen voi johtaa epäoikeudenmukaisiin päätöksiin. [215]

Yleisesti ottaen Euroopan alueella on viime vuosina järjestäydytty unionin tasolla tekoälyn tuomien haasteiden ympärille. 25 Euroopan maata allekirjoittivat 10.4.2018 tekoälyn kehittämisen yhteistyöhön tähtäävän julistuksen, jolla allekirjoittajamaat sitoutuivat tekemään yhteistyötä tekoälyyn liittyvien oleellisempien kysymyksien osalta.

---

[211] Kananen & Puolitaival 2019, 71.
[212] Kananen & Puolitaival 2019, 83.
[213] Käyttämällä matemaattisia tapoja, voidaan löytää datasta merkittävimmät asiaan vaikuttavat ominaisuudet. Kananen & Puolitaival 2019,96.
[214] Kananen & Puolitavail 2019,96.
[215] Tekoälyä käsittelevä korkean tason asiantuntijaryhmä 2019, 4-5.

Tällaisia ovat muun lisäksi EU:n kilpailukyky tekoälyteknologioiden kehittämisessä ja soveltamisessa sekä siihen liittyvissä sosiaalisissa ja eettisissä kysymyksissä. [216]

Euroopan komissio on esittänyt 25.4.2018 ja 7.12.2018 antamissaan tiedonannoissa tekoälyä koskevan visionsa. Visiossa tuetaan ”eettistä, turvallista ja alan viimeisintä kehitystä edustavaa Euroopassa tuotettua tekoälyä”. Visio perustuu kolmen pilarin varaan: tekoälyyn tehtävien julkisten ja yksityisten investointien lisääminen tekoälyn käyttöönoton lisäämiseksi, valmistautuminen sosioekonomiseen muutoksiin ja asianmukaisen eettisen ja oikeudellisen kehityksen varmistaminen. [217]

Vuonna 2018 komissio myös nimitti osana tekoälystrategiaansa AI HLEG *(AI High-Level Expert Group on Artificial Intelligence)* -työryhmän, johon kuuluu 52 asiantuntijaa eri aloilta. Työryhmä on tehnyt perustamisensa kesäkuun 2018 jälkeen eettiset suuntaviivat tekoälyn kehittämiselle. Suuntaviivoissa esitellään ihmiskeskeinen lähestymistapa tekoälyyn ja se listaa myös seitsemän avainvaatimusta, joilla tekoälyjärjestelmistä voidaan saada luotettavampia. Vaatimuksia pilotoidaan ja pilotin odotetaan päättyvän vuonna 2020. [218]

## 6.2 Luotettava tekoäly

Eräs tapa jäsentää tekoälyäteknologioita on katsoa sitä opettamisen näkökulmasta. Ohjelma voidaan opettaa kolmella eri tavalla: ohjatun oppimisen, ohjaamattoman oppimisen tai vahvistetun oppimisen kautta. Ohjatussa oppimisessa käytetään usein neuroverkkoja ja koneoppimisen menetelmiä. Ohjaamaton oppiminen viittaa yleensä koneoppimiseen. Kummatkin tavat edellyttävät kuitenkin isoja määriä (laadukasta) dataa kouluttamiseen. Vahvistusoppimisessa ohjelma oppii ikään kuin yrityksen ja erehdyksen kautta, mutta sen toimintaympäristö on kuitenkin mallinnettu. [219]

Tekoälyä käsittelevän korkean tason asiantuntijaryhmä [220] määrittelee tekoälyn hieman laajemmin ja jäsentää tekoälyn eri tieteenalojen kautta: päättely ja päätöksenteko sekä oppiminen. Päättely ja päätöksenteko on ryhmä tekoälyn tekniikoita, joilla raakadatasta voidaan esittää tietämystä ja päättelyitä, sekä tehdä suunnittelua, hakuja ja optimointeja. Tietämyksen esittämisen jälkeen sen perusteella voidaan tehdä päätelmiä. Päätelmiä voidaan tehdä erilaisten symbolisten sääntöjen avulla. Tietämyksen esittämisen jälkeen voidaan tehdä myös suunnittelua koskevia toimia tai hakuja suuresta määrästä erilaisia ratkaisuja. Lopullinen vaihe on päätelmien perusteella tehtävä päätös to-

[216] Digibyte 2018.
[217] Kts. tarkemmin COM(2018) 237 ja COM(2018) 795.
[218] Digital Single Market 2019.
[219] Kananen & Puolitaival 2019, 43.
[220] Tekoälyä käsittelevä korkean tason asiantuntijaryhmä 2019a, 3-4.

teuttavasta toimesta. Oppiminen tekoälyn tekniikkana on ryhmä tekniikoita, jotka sisältävät muun muassa koneoppimisen, neuroverkot, syväoppimisen ja päätöspuun. Näiden tekniikoiden avulla tekoäly voi oppia ratkaisemaan ongelmia, joita ei voida esittää täsmällisesti tai ongelmia, joiden ratkaisua ei voida kuvata symbolisilla päättelysäännöillä.

Tekoälyä käsittelevän korkean tason asiantuntijaryhmä on tehnyt myös luotettavaa tekoälyä koskevat eettiset ohjeet. Sen mukaan luotettavalla tekoälyllä on kolme edellytystä, joiden olisi täytyttävä tekoälyjärjestelmän kaikissa elinkaaren vaiheissa. Tekoälyjärjestelmän tulisi olla lainmukaista ja noudatettava kaikkia sovellettavia lakeja ja määräyksiä. Tekoälyjärjestelmän tulisi olla myös eettinen ja sen olisi varmistettava eettisten periaatteiden ja arvojen noudattaminen. Lopuksi tekoälyjärjestelmän tulisi olla teknisesti ja sosiaalisesti luotettavaa, sillä tekoälyjärjestelmillä voidaan aiheuttaa tahatonta haittaa. Asiantuntijaryhmä korostaa, että kukin näistä edellytyksistä on välttämätön, jotta tekoäly olisi luotettava, mutta ne eivät sinänsä ole riittäviä luotettavan tekoälyn tuottamiseksi. Eettisissä ohjeissa asiantuntijaryhmä ei käsittele lainkaan ensimmäistä vaatimusta eli lainmukaista tekoälyä eikä ohjeet ole tarkoitettu oikeudellisiksi neuvoiksi tai ohjeiksi lainsäädännön soveltamisesta. Ohjeistuksessa annetaan kuitenkin lopuksi luotettavan tekoälyn arviointilista. [221]

Asiantuntijaryhmä listaa myös neljä eettistä periaatetta, jotka perustuvat perusarvoihin. Niitä täsmennetään vielä eettisinä vaatimuksina, koska niitä tulisi aina noudattaa, kun kehitetään tekoälyä. Nämä periaatteet ovat 1) ihmisen itsemääräämisoikeuden kunnioittaminen, 2) vahinkojen välttäminen, 3) oikeudenmukaisuus ja 4) selitettävyys. Näitä periaatteita on myös laajennettu ohjeistuksessa. [222]

## 6.3 Tekoälyn kouluttaminen

### *6.3.1 Koneoppiminen esimerkkinä*

Koulutusdatan avulla luodaan matemaattinen malli, jota voidaan kouluttamisen jälkeen käyttää uudelle datalle. Malli kuvaa siis tietojoukon (datasetin) eri tietuiden väliset riippuvuudet ja pystyy tekemään näistä yleistyksiä todennäköisyyksiin perustuen tietyin rajoituksin. Kouluttamisessa on kyse siis siitä, että luotu matemaattinen malli muotoutuu koulutusdatan mukaan. Koulutettu malli pystyy siis tunnistamaan koulutusdatassa olleet riippuvuudet uudesta datasta ja ennustamaan etsittyjä arvoja. [223]

---

[221] Tekoälyä käsittelevä korkean tason asiantuntijaryhmä 2019b, 2.
[222] Tekoälyä käsittelevä korkean tason asiantuntijaryhmä 2019b, 14.
[223] Kananen & Puolitaival 2019, 46.

Tekoälyn kouluttamiseen tarvitaan siis paljon dataa. Kaikilla yrityksillä on paljon dataa, mutta sen laatu ja ominaisuudet vaihtelevat. Dataa voi olla esimerkiksi asiakkaista ja heidän transaktioistaan ja myynnistä. Data voi olla muodoltaan tekstejä, kuvia, sensorin mittaustuloksia, käyntitietoja ja niin edelleen. Data voi olla joko rakenteellista eli strukturoitua tai ei-rakenteellista eli strukturoimatonta. Jotta tekoälyä voidaan alkaa kouluttamaan, data pitää valmistella. Tarkoitus on luoda yhtenäinen tietokanta, jota voidaan käyttää eri tarkoituksiin: mallin kouluttamiseen sekä mallin testaukseen ja validointiin. Testaus ja validointi tehdään, jotta voidaan varmistua siitä, että malli toimii myös täysin uudella datalla, jota se ei ole aiemmin käsitellyt. [224]

Koneoppimisen teknologiat siis voivat olla erilaisia, mutta yleisemmin koneoppimiseksi ajatellaan neuroverkot, tiedon louhinta ja geneettiset algoritmit. Näiden avainominaisuudet ovat:

- Ne vaativat paljon historiadataa. Osa tästä datasta tarvitaan mallin harjoittamiseen ja osa sen validointiin.
- Kun tietyt parametrit on asetettu, malli voi oppia itsenäisesti assosiaatioita ja lisäohjelmointia ei tarvita.
- Assosiaatiot voidaan esittää esimerkiksi linkkeinä eri tietojen välillä tai klustereina.
- Tällainen järjestelmä ei pysty itsenäisesti antamaan muita perusteluita, kuin tilastolliset korrelaatiot eri datapisteiden välillä. [225]

Euroopan ihmisoikeustuomioistuin teki analyysin tarkoituksenaan ennustaa eri tapauksien lopputulemia käyttämällä koneoppimisen ja luonnollisen kielen prosessoinnin tekniikoita. Tutkijat onnistuivat kehittämään mallin, joka pystyi 79 % tarkkuudella ennustamaan käsiteltyjen tapauksien lopputuloksen. Tutkijat huomasivat, että käsiteltävän tapauksien formaalit faktat ennustivat parhaiten käsittelyn lopputulosta, mutta myös suullisen ja kirjallisen käsittelyn teemoilla oli vaikutusta. Kun esimerkiksi keskusteltiin, onko artiklaa 3[226] rikottu, tietyt käytetyt sanat liittyen valtion velvollisuuksiin (esimerkiksi ”suojelu”, ”kompensaatio”, jne.), pidätysolosuhteisiin (esimerkiksi ”pääsy” tai ”tilausahtaus”) tai viranomaisten kohteluun (esimerkiksi ”altistettu” tai ”voima”) liittyen indikoivat sitä, että artiklan rikkomus todettiin. Lisäksi tutkimuksessa tunnistettiin sanaryhmiä, jotka indikoivat sitä, että rikkomusta ei todettu. [227]

Koneoppiminen on siis toisaalta hyvinkin tehokas, mutta toisaalta epätarkka teknologia. Se pystyy tekemään assosiaatioita harjoitusdatan perusteella siitä riippumatta ovat-

[224] Kananen & Puolitaival 2019, 46.
[225] Kingston 2017.
[226] Ihmisoikeustuomioistuin artikla 3 viittaus
[227] Aletras et al 2016.

ko assosiaatiot eksplisiittisiä, relevantteja tai perusteltuja. Viitaten edelliseen esimerkkiin, sanaryhmä, joka ennusti sitä, ettei artiklan rikkomusta todettu, olivat tiettyjen maiden nimet. On mahdollista, että tuomioistuin on käsittelemissään tapauksissa ollut tilastollisesti taipuvaisempi ratkaisemaan tapauksen tietyllä tavalla maasta riippuen, mutta on kyseenalaista voisiko jatkossa koneoppimisen malli käyttää tätä tekijänä ratkaistessaan tulevia tapauksia.[228]

### *6.3.2 Koneoppimisen mallien mittarit*

Yhtä yksittäistä tai edes muutamaa tunnistettua tunnuslukua ei ole olemassa, joilla voitaisiin riippumattomasti arvioida koneoppimismallien ennustetarkkuutta tai soveltuvuutta. Sen sijaan tilastotieteissä ja sovelletun matematiikan haarassa informaatioteoriassa on esitetty useita mahdollisia tunnuslukuja.

Mallissa käytettävien ominaisuuksien hyvyydelle on olemassa tiettyjä tilastotieteellisiä mittareita tai tunnuslukuja. Tällaisia ovat gini -kerroin, *information gain* tai oddsratio. Mallin hyvyyttä mitataan tällöin sillä, miten hyvin valitut ominaisuudet pystyvät mallintamaan ilmiön. [229] Näillä mittareilla pyritään kuvaamaan mallinnetun ilmiön ja varsinaisen mallin tuottamien tulosten välisiä eroja ja siten mallin hyvyyttä.

Gini -kertoimella kuvataan yleensä tuloerojen jakautumista, mutta yleisesti se kuvaa siis todennäköisyysjakaumaa. Sillä voidaan kuvata tietyn jakauman epätasaisuutta. *Information gain* on synonyymi Kullback-Leibler -informaatiolle, joka mittaa kahden jakauman etäisyyttä toisistaan. Sitä voidaan siis käyttää arvioimaan mallin rakenteelle ja ominaisuuksien määrälle sopiva estimaatti, joka on mahdollisimman lähellä kuvattavaa ilmiötä. [230] *Odds -ratio* eli vetosuhde kuvaa sitä kuinka etäällä kaksi todennäköisyyttä tai suhteellista osuutta ovat toisistaan. [231]

Tekoälyteknologioiden käyttämiseen liittyy samoja haasteita kuin tilastolliseen tutkimukseen yleisestikin. Ylisovittaminen ja alisovittaminen ovat kuitenkin tärkeitä käsitteitä etenkin koneoppimisen tekniikoilla tuotettujen ennustemallien arvioimisessa. Ylisovittaminen tarkoittaa sitä, että koneoppimisen tekniikalla tuotettu malli ennustaa sen luomisessa käytetyn opetusdatan arvot liian hyvin. Teoriassa malli voidaan toteuttaa niin, että se ennustaa opetusdatan arvot täydellisesti. Tämä on kuitenkin ongelma, kun mallia haluttaisiin yleistää ja käyttää datalla, jota malli ei ole aiemmin käyttänyt. Aineistoissa on aina satunnaisvaihtelua eli kohinaa. Jos mallista luodaan siis liian hyvin ope-

[228] Kingston 2017.
[229] Kananen & Puolitavail 2019, 95 – 100.
[230] Kullback & Leibler 1951.
[231] Rita 2004.

tusdataan sovitettu eli ylisovitettu, se alkaa ennustamaan tätä satunnaisvaihtelua muun lisäksi. Ylisovittamisen seurauksena malli ei yleistä ja kun mallia käytetään uuden aineiston avulla ennustamiseen, ennusteet ovat huonoja. Tällaista huonoa ennustetta voidaan kuvata ennustevirheellä, joka on opetusdatan ja uuden datan arvojen välinen ero. [232] Käyttökelpoinen koneoppiva malli tunnistaa ilmiöitä, kuten trendejä ja soveltaa oppejaan yleiseen dataan.

## 6.4 Automaattisen päätöksenteon tunnetut haasteet

### *6.4.1 Mitä enemmän dataa, sitä oikeudenmukaisemmat päätökset?*

Dataa voidaan käyttää erilaisten ennustemallien tekemiseen. Malli oppii siis esimerkin avulla. Tämä tarkoittaa sitä, että se mitä malli oppii, riippuu datasta, jota sille syötetään. Tästä datasta käytetään usein nimitystä *training data* tai opetusdata. Jos halutaan esimerkiksi mallille opettaa oikeuskäytäntöä, jonka perusteella malli pystyisi tekemään tapauksien ratkaisusuosituksia, tulisi sille syöttää paljon esimerkkitapauksia. Näissä tapauksissa voi mahdollisesti olla ihmisen päätöksentekoon perustuvaa epätarkkuutta ja virheellisyyttäkin. Malli oppii vain sen, mitä tässä opetusdatassa on. Jos esimerkeissä on päätöksiä, joihin vaikkapa ennakkoluulot ovat vaikuttaneet, ne siirtyvät myös malliin opetusdatan kautta. Siten mallista voi tulla jopa syrjivä.

Yleinen oletus on, että kun arkaluontoisen tiedon kuten henkilön etnistä taustaa koskevien tietojen käsittelyä rajoitetaan ennustemallien rakentamisessa, pienennetään mahdollisuutta siihen, että malli toisintaisi ennusteisiin syrjintää. Ajatus on siis, että kun syrjintään liitettyjä ominaisuuksia tai tietoja ei oteta huomioon mallin opettamisessa käytetyssä datassa, koulutetun mallin tulokset muuttuvat automaattisesti syrjimättömäksi.[233] On kuitenkin löydetty todisteita siitä, että algoritmien tai ennustemallien opettaminen sopimattomalla tavalla voi vääristää algoritmeja ja siten aiheuttaa syrjintää. [234] Mallien opettamisessa käytettävä data voi sisältää säännönmukaisuuksia ja assosisaatioita, joihin on vaikuttanut vaikkapa rotuun perustava syrjintä.

---

[232] Jamet et al 2017, 31-39.
[233] Žliobaitė & Custers 2016, 3.
[234] Esimerkiksi Edelman & Luca 2014 osoittivat, että Airbnb.com -palvelussa vuokraisännän entinen tausta vaikutti airbnb -asunnon vuokran hintaan. Kay et al 2015 osoittivat, että sukupuoleen ja eri ammatteihin liittyy puolueellisuutta, joka näkyy kuvahaun tuloksissa.

Barocas & Selbst [235] ovat tutkineet ja systematisoineet eri tekijöitä mallien rakentamisessa ja sen eri vaiheissa, joilla voi olla syrjivä vaikutus mallin tuottamiin tuloksiin. Mallin rakentaminen alkaa yleensä tavoitemuuttujien ja luokittelutekijöiden määrittelyllä. Tavoitemuuttuja kuvaa sitä, mitä halutaan ennustaa ja luokittelutekijä jakaa tavoitemuuttujan arvot eri luokkiin. Etenkin tavoitemuuttujan määrittely on ratkaisevaa ja se tarkoittaa yleensä ratkaistavan ongelman määrittelyä tai sellaisen kysymyksen esittämistä, johon tietokone voi löytää annetusta datasta vastauksen. Joskus voidaan luoda uusia luokkia kuten ”luottokelpoisuus”, jolla tarkoitetaan yleensä asiakkaan todennäköisyyttä maksaa luotto takaisin. Ongelman ja tavoitemuuttujien määrittely vaikuttaa siis olennaisesti siihen, mitä asioita malli tulisi ennustamaan.[236] Luottokelpoisuus voi olla vaikkapa tietyillä postinumeroalueilla parempi kuin toisilla, koska siellä on sosioekonomisesti paremmassa asemassa olevia ihmisiä. Tulisikin arvioida, onko postinumero syrjimätön tekijä mallin rakentamisessa.

Opetusdataan liittyy useita haasteita, joista esimerkkien väärät tai jopa sopimattomat merkinnät sekä puolueellisia tiedonkeruumenetelmiä pidetään oleellisimpina. Esimerkkien merkinnällä tarkoitetaan datan jakamista luokkiin merkintöjen avulla. Esimerkiksi sähköpostin merkitseminen roskapostiksi manuaalisesti olisi tässä yhteydessä datan luokittelua merkintöjen avulla. Joissain tapauksissa opetusdataa ei ole kuitenkaan luokiteltu. Tällöin mallin kehittäjien tulee luokitella data sopivien merkintöjen avulla. Tämä voi olla varsin työläs prosessi ja altis virheille. Esimerkiksi kuluttaja voidaan luokitella luottokelpoiseksi tai -kelvottomaksi sen perusteella, onko hänellä maksattomia luoton eriä 4 tai 5. Ero voi olla joidenkin kuluttajien kohdalla hyvinkin merkitsevä ja vastaus kysymykseen ei ole aina yksiselitteinen. Tietyt tapaukset voivat edustaa joitakin, mutta eivät kaikkia kriteereitä, joiden perusteella dataa tulisi luokitella puolueettomasti. Kun dataa luokitellaan manuaalisesti ihmisen toimesta, se sisältää aina tiettyä puolueellisuutta ja voi edustaa jopa ennakkoluuloja. [237]

Datan keräämiseen liittyy toinen olennainen riski, joka voi johtaa puutteelliseen päätöksentekoon. Automaattiset, algoritmin tekemät päätökset, jotka perustuvat virheelliseen, puutteellisen tai puolueellisen opetusdatan perusteella opetettuihin malleihin, voivat syrjiä suojattuja luokkia tai tiettyjä erityisryhmiä. Datan laatuun vaikuttaa tapa, jolla se on kerätty. Tiettyihin ryhmiin kuuluvista henkilöistä voidaan kerätä dataa, kuten asiakastietoja, eri tavoilla riippuen siitä, mihin ryhmään heidät luokitellaan. Yrityksen asiakkaaksi voi olla valikoitunut vain tiettyjen ryhmien edustajia ja siten tällaista dataa ei voisi käyttää sellaisen mallin opettamiseen, jonka olisi tarkoitus tehdä päätöksiä

[235] Barocas & Selbst 2016, 677.
[236] Barocas & Selbst 2016, 677-680.
[237] Barocas & Selbst 2016, 681 - 683.

kaikkia ryhmiä koskien. Yritys voi myös itse kerätä tietoa joistain asiakasryhmistä tarkemmin kuin toisista. [238]

Jotta opetusdata edustaisi niitä asioita, joita halutaan käyttää päätöksenteon perustana, tulisi datan keräämiseen kiinnittää erityistä huomiota. Datassa voi olla joitakin tiettyjä havaintoja yliedustettuna ja toisia taas aliedustettuina. Usein kuitenkin oletetaan, että kerätty data edustaa populaatiota ja kaikkia tietoluokkia. Lisäksi menetelmät, joita käytetään datan keräämiseen liiketoiminnan tarpeisiin, jättävät toivomisen varaa verrattuna vaikkapa sosiaalitieteiden menetelmiin verrattuna. [239]

Tilastotieteessä tunnetaan yleisesti otantavirhe ja siihen liittyvät virheiden lähteet. Otantavirheistä erityisesti kehikkovirhe liittyy läheisesti aiemmin keskusteltuun opetusdatan keräämiseen liittyviin haasteisiin. Kehikkovirheellä viitataan otannan kehikkoon eli tapaan tehdä otanta. Silloin, kun perusjoukko ei ole käytettävissä otosta valittaessa, otannasta voi jäädä havaintoyksiköitä pois. Tätä kutsutaan kehikkovirheeksi. [240]

Kun opetusdata on saatu valikoitua mallin opettamista varten, seuraavaksi pitää määritellä ne muuttujat tai ominaisuudet, joita mallin halutaan datasta tarkastelevan. Tätä prosessia kutsutaan *feature selection* tai vapaasti kääntäen muuttujien valinnaksi. Tällä voi olla hyvinkin merkittävä vaikutus siihen, miten hyvin malli pystyy tekemään päätöksiä koskien tiettyjä henkilötietoja tai henkilöryhmiä. Valittujen muuttujien tulisi edustaa niitä henkilöryhmiä, joita koskien mallin halutaan tekevän päätöksiä. On kuitenkin mahdollista, että kaikki valitut muuttujat eivät edusta kaikkia ryhmiä yhtä hyvin. Vaikka valittaisiinkin lisää muuttujia, ei voida usein kuitenkaan kaikkia reaalimaailman ilmiöitä mallintaa valitsemalla vain yhä enemmän muuttujia. Lisäksi mitä enemmän muuttujia valitaan, sitä enemmän kustannuksia mallin kehittämisessä voi muodostua. [241] Muuttujien etukäteisvalinnan vaihtoehtona on *feature extraction* eli vapaasti kääntäen muuttujien eristäminen, jossa havainnoista luokittelumenetelmien avulla eristetään uusia muuttujia. Näitä tekniikoita kutsutaan *dimensional reduction* -tekniikoiksi eli aineiston dimensioiden vähentämisen tekniikoiksi. Niillä pyritään siihen, että aineiston dimensioita tai muuttujia olisi käytännöllinen määrä. [242]

Tietosuoja-asetuksen johdantokappaleessa 71 keskustellaan rekisteröityä koskevasta asianmukaisesta ja läpinäkyvästä tietojenkäsittelystä. Rekisterinpitäjän tulisi ottaa huomioon henkilötietojen erityiset käsittelyolosuhteet ja asiayhteys sekä lisäksi olisi käytettävä profiloinnissa asianmukaisia matemaattisia tai tilastollisia menetelmiä sen varmistamiseksi, että muun lisäksi rekisteröidyn etuihin ja oikeuksiin kohdistuvat mahdolliset riskit otetaan huomioon ja estetään. Tällaisia riskejä ovat muun muassa luonnollisten

[238] Barocas & Selbst 2016, 684 - 686.
[239] Barocas & Selbst 2016, 684.
[240] Ks. tarkemmin Tilastokeskus 2007.
[241] Barocas & Selbst 2016, 686-689.
[242] Vellido et al 2012, 163-169.

henkilöiden syrjintä rodun tai etnisen alkuperän, poliittisten mielipiteiden, uskonnon tai vakaumuksen tai muiden vastaavien tekijöiden perusteella. Lisäksi johdantokappaleessa myös todetaan, että ”- - *taikka vaikutukset, joiden johdosta toteutetaan toimenpiteitä, joilla on tällaisia seurauksia*”. Vaikutuksilla viitataan siis edellä mainittuihin tekijöiden, joiden perusteella luonnolliselle henkilölle muodostuu riski joutua syrjinnän kohteeksi.

### *6.4.2 Mallin läpinäkyvyys, tarkkuus ja perusteltavuus*

Tarkkuus ja perusteltavuus ovat kaksi eri vaatimusta, jotka ovat hankalia tasapainottaa erilaisten pisteytysmallien kehittämisessä. Jos pisteytysmalli on perusteltavissa, mallin tarkkuus yleensä pienenee. Tutkijat kehittävät jatkuvasti uusia malleja, jotka ennustavat tarkemmin, mutta ovat monimutkaisempia perustella ja ymmärtää. Rahoitusala tarvitsisi nimenomaan malleja, jotka ovat tarkkoja ja perusteltavissa. [243]

Esimerkiksi luottopäätöksen pohjana olevan pisteytysmallin pitäisi tarkkuuden lisäksi, sen päätöksien tulisi olla perusteltavissa [244]. Mallin tekemät päätökset pitää siis pystyä perusteleman ihmisen toimesta. Lisäksi läpinäkyvä, ymmärrettävä ja perusteltavissa oleva malli on helpompi ottaa käyttöön, kun mallin ymmärtäminen ei vaadi tilastotieteiden syvää osaamista. Lopulta myös mallia käyttävät yritykset oppivat ymmärtämään paremmin eri tekijöitä, jotka vaikuttavat luottopäätöksen. [245] Luottopisteytyksen tuottamiseen käytetyistä metodeista suosituin on perinteisesti ollut logistinen regressio. Muut metodit kuten lineaarinen regressio ja päätöspuut ovat myös suosittuja. Ne ovat helposti tulkittavissa, niiden suorituskyky on hyväksyttävällä tasolla ja ennen kaikkea ne ja niiden tulokset ovat helpoiten selitettävissä. Viime aikoina myös esimerkiksi koneoppimisen menetelmiä kuten neuroverkkoja on käytetty petosten tunnistamiseen ja muihin luottoihin liittyviin taustaprosesseihin, koska niissä ei ole yhtä korkeat selitettävyysvaatimukset kuin luottopäätöksissä, joilla on suurempi vaikutus henkilöön. [246]

Tekoälyä käsittelevän korkean tason asiantuntijaryhmä[247] on luotettavaa tekoälyä koskevassa eettisessä ohjeistuksessaan listannut neljän eettistä periaatetta, joista luotettava tekoäly ainakin koostuu. Näihin kuuluvat oikeudenmukaisuus ja selitettävyyden periaatteet.

Oikeudenmukaisuuden periaatteella viitataan muun lisäksi myös siihen, että yksilöihin tai ryhmiin ei kohdistu epäoikeudenmukaista puolueellisuutta, syrjintää tai leimaa-

[243] Hayashi et al 2018,1.
[244] Englanninkielisissä lähteissä käytetään sanaa *interpretability*, jonka suora käännös olisi tulkittavuus. Tällä viitataan siihen, että luottopäätösmallien päätöksen pitää olla perusteltavissa myös ihmisen toimesta. Mielestäni perusteltavuus kuvaa tätä ominaisuutta paremmin kuin tulkittavuus.
[245] Chen & Cheng 2013.
[246] Finlay 2011.
[247] Tekoälyä käsittelevä korkean tason asiantuntijaryhmä 2019b, 14.

mista. Lisäksi asiantuntijaryhmän mukaan oikeudenmukaisuuden periaatteen noudattaminen edellyttää, että tekoälyalan toimijoiden olisi kunnioitettava keinojen ja päämäärien oikeasuhteisuuden periaatetta ja erityisesti harkittava, miten kilpailevat edut ja tavoitteet tasapainotetaan.[248] Nämä periaatteet vastaavat yleisen tietosuoja-asetuksen periaatteita erityisten henkilöryhmien suojelemisesta tietojenkäsittelyssä sekä tietojenkäsittelyn vaikutustenarviointi- eli DPIA-prosessia, jossa tulee tehdä tasapainotesti rekisterinpitäjän oikeutettujen etujen ja rekisteröidyn henkilötietojen suojan välillä.

Selitettävyyden periaate on asiantuntijaryhmän mukaan keskeinen seikka, jolla voidaan luoda tekoälyjärjestelmien käyttäjien tunteman luottamuksen luomiseksi ja ylläpitämiseksi. Erityisesti prosessien tulisi olla avoimia, tekoälyjärjestelmän kapasiteetti ja tarkoitus on ilmoitettava avoimesta. Lisäksi päätökset tulisi mahdollisuuksien mukaan pystyä selittämään niille, joihin ne suoraan tai välillisesti vaikuttavat. Vaatimus selitettävyyden tasosta riippuu asiantuntijaryhmän mukaan paljolti kontekstista sekä virheellisten tai muutoin epätarkkojen tulosten aiheuttamien seurausten vakavuudesta. [249] Selitettävyyden periaate tekoälyjärjestelmien kehitystä koskevassa eettisessä ohjeistuksessa vastaa hyvin yleisen tietosuoja-asetuksen vaatimuksia avoimuudesta etenkin rekisteröityjen oikeuksiin saada tietoja ja oikeuteen vastustaa tietojen käsittelyä liittyen.

Tarkkuus – perusteltavuus -dilemmaa voidaan tarkastella myös läpinäkyvyysperiaatteen kautta. Läpinäkyvyys on jo vakiintunut periaate EU:n lainsäädännössä. [250] Yleisessä tietosuoja-asetuksessa[251] on myös määrätty henkilötietojen käsittelyä koskevasta läpinäkyvyyden velvollisuudesta. Velvollisuus kattaa asetuksen mukaan kolme osa-aluetta: 1) tietojen asianmukaista käsittelyä koskevan tiedon antaminen rekisteröidylle, 2) rekisterinpitäjien tapa tiedottaa rekisteröidyille näiden tietosuoja-asetukseen perustuvista oikeuksista ja 3) rekisterinpitäjien keinot auttaa rekisteröityjä käyttämään oikeuksiaan. [252]

Läpinäkyvyyttä ei määritellä nimenomaisesti tietosuoja-asetuksessa. Johdantokappaleesta 39 voi kuitenkin saada viitteitä läpinäkyvyyden periaatteen merkityksestä ja vaikutuksesta tietojen käsittelyssä:

---

[248] Tekoälyä käsittelevä korkean tason asiantuntijaryhmä 2019b, 15.
[249] Tekoälyä käsittelevä korkean tason asiantuntijaryhmä 2019b, 15.
[250] Euroopan unionista tehdyssä sopimuksessa viitataan monesti 'avoimuuteen'. Esimerkiksi sopimuksen 1 artiklan mukaan päätöksen tulee tehdä "mahdollisimman avoimesta ja mahdollisimman lähellä kansalaista".
[251] Yleisen tietosuoja-asetuksen johdantokappaleet 58 ja 60 erityisesti käsittelevät läpinäkyvyyttä rekisteröidylle. Lisäksi asetuksen 5 artiklan 1 kohdan a alakohdan mukaan henkilötietoja on käsiteltävä "*lainmukaisesti, asianmukaisesti ja rekisteröidyn kannalta läpinäkyvästi*".
[252] WP260 2018, 4.

> Luonnollisille henkilöille olisi oltava läpinäkyvää, miten heitä koskevia henkilötietoja kerätään ja käytetään ja niihin tutustutaan tai niitä käsitellään muulla tavoin sekä selvillä siitä, missä määrin henkilötietoja käsitellään tai on määrä käsitellä. Läpinäkyvyyden periaatteen mukaisesti kyseisten henkilötietojen käsittelyyn liittyvien tietojen ja viestinnän on oltava helposti saatavilla ja ymmärrettävissä ja niissä on käytettävä selkeää ja yksinkertaista kieltä. Tämä periaate koskee erityisesti rekisteröityjen tietoja rekisterinpitäjän identiteetistä ja käsittelyn tarkoituksista sekä lisätietoja, joilla varmistetaan kyseisiä luonnollisia henkilöitä koskevan käsittelyn asianmukaisuus ja läpinäkyvyys, sekä heidän oikeuttaan saada vahvistus ja ilmoitus heitä koskevien henkilötietojen käsittelystä.

Johdantokappaleessa luetellut asiat kytkeytyvät asetuksen lukuun III, joka käsittelee rekisteröityjen oikeuksia, joita ovat muun muassa oikeus saada tietoja. Tietosuojatyöryhmä käsittelee ohjeistuksessaan[253] erityisesti rekisteröidylle toimitettavia tietoja ja niiden ymmärrettävyyttä.

Tietosuoja-asetuksessa edellytetään, että silloin kun käytetään automaattista päätöksentekoa yhdessä tai erikseen profiloinnin kanssa, rekisterinpitäjän tulee toimittaa rekisteröidylle tiedot ainakin näissä tapauksissa *merkitykselliset tiedot käsittelyyn liittyvästä logiikasta* samoin kuin käsittelyn *merkittävyydestä* ja *mahdollisista seurauksista*. [254] Tietosuojatyöryhmä toteaa profilointiin liittyvässä ohjeistuksessaan, että koneoppimisen yleistymisen ja monimutkaisuuden vuoksi voi olla haastavaa ymmärtää, miten automaattinen päätöksenteko tai profilointi toimivat käytännössä. Työryhmä kuitenkin korostaa, että rekisterinpitäjän tulisi löytää yksinkertaisia tapoja kertoa päätöksen taustalla olevista syistä tai sen tekemisessä käytetyistä perusteista. Rekisterinpitäjän ei siis tarvitse toimittaa monimutkaista selitystä käytettävästä algoritmista eikä algoritmia tarvitse paljastaa välttämättä. Rekisteröidylle toimitetut tiedot tulisi olla kuitenkin riittävän kattavia, jotta rekisteröity ymmärtää päätöksen perusteet. [255]

Työryhmä käyttää luottopisteytystä esimerkkinä, mitä merkityksellisiä tietoja rekisteröidylle tulisi toimittaa ”käsittelyyn liittyvästä logiikasta”. Lähtökohta on se, että rekisterinpitäjän tulee pystyä selittämään pisteytys ja sen perusteet rekisteröidylle sekä tietysti pisteytyksen lähteenä olevat tiedot. Tässä esimerkkitapauksessa rekisterinpitäjän tulisi selittää, että prosessi auttaa tekemään oikeudenmukaisia ja vastuullisia lainapäätöksiä. Lisäksi sen tulisi kertoa myös tärkeimmistä päätöksenteossa huomioon otetuista ominaisuuksista, näiden tietojen lähteistä ja niiden merkityksestä. Tällaisia voivat olla esimerkiksi rekisteröidyn lainahakemuksessa antamat tiedot, aiemmat tilitiedot, aiemmat maksujen myöhästymiset, virallisen julkisten rekisterien tiedot kuten luottotietorekisterien tiedot. Lopuksi rekisterinpitäjän tulisi kertoa myös, että luottopisteytysmenetelmiä testataan säännöllisesti niiden asianmukaisuuden, tehokkuuden ja puolueettomuuden varmistamiseksi. [256]

---

[253] WP260 2018.
[254] Yleinen tietosuoja-asetus 13 artikla 2 kohta alakohta f.
[255] WP251 rev01 2018, 26-27.
[256] WP251 rev01 2018, 27-28.

Profilointiin ja automaattiseen päätökseen liittyen rekisteröidylle tulee toimittaa tiedot tietojenkäsittelyn merkittävyydestä ja sen mahdollisista seurauksista. Tietosuojatyöryhmän mukaan termit viittaavat suunniteltuun tai tulevaan tietojen käsittelyyn sekä automaattisten päätöksien vaikutuksista rekisteröityyn. Rekisterinpitäjän tulisi antaa konkreettisia esimerkkejä mahdollisista vaikutuksista ja rekisterinpitäjän tulisi myös havainnollistaa digitaalisessa toimintaympäristössä käsittelyn vaikutuksia. [257]

### *6.4.3 Oikeus saada tietoja vai oikeus saada selitys?*

Yleinen tietosuoja-asetus antaa rekisteröidylle monia oikeuksia ja yksi niistä on oikeus saada tietoja häntä koskevista automaattisesta päätöksenteosta ja profiloinnista. Ennen sen voimaantuloa, tehtiin jonkin verran arvioita myös siitä, että miten se vaikuttaa algoritmeilla tuotettuihin automaattisen päätöksentekoprosesseihin. Goodman & Flaxman 2016 ehdottivat, että tietosuoja-asetuksen myötä rekisteröidylle muodostuu myös ”oikeus selitykseen”. Väite rakentuu muun lisäksi sen varaan, että tietosuoja-asetuksessa korostetaan ihmisen merkitystä päätöksenteossa. Ihmisen vaikutus päätöksentekoon katsotaan siis ikään kuin lieventävänä asianhaarana ja silloin päätöksentekoa ei katsota asetuksessa tarkoitetulla tavalla ankarammin. Lisäksi yleinen syrjimättömyyden vaatimus on läsnä myös tietosuoja-asetuksessa, jonka johdantokappaleessa 71 korostetaan erityisten henkilöryhmien tietojen suojaa ja se on mukana itse säädöstekstissä tietosuoja-asetuksen 9 artiklassa. Näistä johdetaan päätelmä, että yleinen tietosuoja-asetus tuo myös ”oikeuden selitykseen”, joka toimisi samalla myös suojakeinona rekisteröidyn oikeuksien toteutumiselle. [258]

Selitys suhteessa automaattiseen päätöksentekoon voidaan jakaa kahteen eri luokkaan liittyen joko järjestelmään, joka tekee päätöksen tai liittyen tiettyyn järjestelmän tekemään päätökseen. Lisäksi selitykseen voidaan liittää myös ajoitukseen liittyviä ominaisuuksia. Selitys voidaan antaa *ex ante* eli selitys ennen automaattista päätöksentekoa tai *ex post* eli automaattisen päätöksenteon jälkeinen selitys. Wacther et al[259] tunnistivat kolme mahdollista ”oikeutta selitykseen” muodostavaa tekijää yleisestä tietosuoja-asetuksesta:

- suoja automaattisten päätöksentekoa vastaan (artikla 22, johdantokappale 71)

[257] WP251 rev01 2018, 28. Tietosuojatyöryhmä viittaa myös Euroopan neuvoston dokumenttiin *Draft Explanatory Report on the modernised version of CoE Convention 108* (luonnos selittäväksi muistioksi Euroopan neuvoston yleissopimuksen N:o 108 nykyaikaistetusta versiosta), kohtaan 75. Muistiossa todetaan, että rekisteröidyllä on oikeus saada tietoja henkilötietojensa käsittelyssä käytettävässä kielteiseen tai myönteiseen päätökseen johtavasta logiikasta eikä pelkästään itse päätöksestä. Muistiossa korostetaan, että jos näitä seikkoja ymmärretä, suojakeinoja ei voi käyttää tehokkaasti.
[258] Ks tarkemmin Flaxman & Goodman 2016. ”Right to explanation” käännetty “oikeus selitykseen”.
[259] Wachter et al 2017, 78.

- rekisterinpitäjän tiedonantovelvollisuudet (artiklat 13 ja 14 sekä johdantokappaleet 60-62)
- rekisteröidyn oikeus saada tietoja (artikla 15 ja johdantokappale 63)

Wacther et al[260] tutkivat muodostuuko yleisen tietosuoja-asetuksen perusteella edellä lueteltujen säädöksien perusteella rekisteröidylle ”oikeus selitykseen” *ex post*, joka pätisi kaikkiin heitä koskeviin automaattisiin, algoritmien tekemiin päätöksiin. He päätyivät siihen, että tietosuoja-asetuksen perusteella rekisteröidylle ei muodostu tällaista oikeutta. Ensinnäkin asetuksessa ei ole nimenomaisesti mainittu ”oikeutta selitykseen”. Tiettyjen ehtojen täyttävien automaattisten päätöksien osalta rekisteröidyllä on tiettyjä suojakeinoja käytettävissään, mutta suojakeinot eivät liity siihen, että hänellä olisi oikeus saada selitys. Oikeus saada selitys oletetaan siis laajemmaksi ja kattavammaksi, kuin mitä rekisteröidyn suojakeinot ovat. Tietosuoja-asetuksessa on vain johdantokappaleessa 71 viittaus oikeuteen saada jonkinlainen selitys:

> ”- - Tällaiseen käsittelyyn olisi kuitenkin aina sovellettava asianmukaisia suojatoimia, joihin olisi kuuluttava käsittelystä ilmoittaminen rekisteröidylle ja oikeus vaatia ihmisen osallistumista tietojen käsittelemiseen, rekisteröidyn oikeus esittää kantansa, *saada selvitys kyseisen arvioinnin jälkeen tehdystä päätöksestä ja riitauttaa päätös*.” [261]

Wacther et al mukaan asetuksen johdantokappaleet antavat taustatietoa ja viitekehystä varsinaisen asetuksen säädöksille, mutta eivät ole itsessään laillisesti sitovia. Artikla 22 mukaiset suojakeinot automaattista päätöksentekoa vastaan eivät itsenäisesti myöskään tuota tällaista oikeutta selitykseen. Vaikka ”oikeus selitykseen” olisi mahdollista tulkita tietyissä tapauksissa olevan suojakeino itsessään, jonka rekisterinpitäjä voisi toteuttaa, jotta sen automaattinen päätöksenteko olisi lainmukainen, oikeuksien pitää olla vahvistettu ennen niiden täytäntöönpanoa. Wacther et al mukaan ”oikeus selitykseen” on jätetty myös tietoisesti pois säädöksien teksteistä. [262]

Seuraavaksi Wacther et al tutkivat mahdollisuutta, että rekisterinpitäjän tiedonantovelvollisuudet muodostaisivat rekisteröidylle oikeuden selitykseen. Mainitut tiedonantovelvollisuudet ovat *ex ante* ja liittyvät päätöksentekojärjestelmään. Tästä syystä tutkijat argumentoivat, että artiklojen 13 ja 14 säädökset eivät tuota oikeutta selitykseen *ex post.* [263]

Lopuksi tutkijat analysoivat vielä rekisteröidyn oikeutta saada pääsy tietoihin artiklan 15 perusteella ja voisiko muodostaa rekisteröidylle myös oikeuden selitykseen. Artiklan

---

[260] Wachter et al 2017, 79.
[261] Yleisen tietosuoja-asetuksen johdantokappale 71. Englanninkielisessä versiossa korostettu osa: ” to obtain an explanation of the decision reached after such assessment and to challenge the decision”. Mielestäni englanninkielinen ilmaus “to obtain an explanation” viittaa enemmän oikeuteen saada selvitys kuin suomenkielisessä versiossa ”saada selvitys”.
[262] Ks tarkemmin Wacther et al 2017, 81.
[263] Wacter et al 2017, 81.

mukaan rekisterinpitäjän tulee antaa pääsy rekisteröidyn henkilötietoihin sekä muun lisäksi tieto automaattisen päätöksenteon ja profiloinnin olemassaolosta ja merkityksel-liset tiedot tällaiseen käsittelyyn liittyvästä logiikasta. Heidän mukaansa siis artiklan sanamuoto ei viittaa yksittäiseen päätökseen liittyvään selostukseen vaan enemmänkin yleisemmän tason suunnitelmaan mihin automaattista päätöksentekoa ja profilointia käytetään. [264]

Tutkijoiden mukaan artiklan 22 sanamuotoa voidaan tulkita joko nimenomaisena kieltona tai rekisteröidyn oikeutena vastustaa. Tarkka sanamuoto kuuluu: ” *Rekiste-röidyllä on oikeus olla joutumatta sellaisen päätöksen kohteeksi, joka* - -”. Oikeus olla joutumatta voidaan siis tulkita myös kielloksi. Tulkinta kielloksi antaisi rekisteröidylle tehokkaamman suojan. [265]

[264] Wacther et al 2017, 83.
[265] Wacther et al 2017, 94.

# 7 JOHTOPÄÄTÖKSET

## 7.1 Automaattinen päätöksenteko ja käyttötarkoitussidonnaisuus

Yleinen tietosuoja-asetus on mielestäni lähtökohtaisesti hyvä tietosuojan viitekehys. Se toi selkeyttä ja vakiinnutti monia tietojen käsittelyn periaatteita. Se toi myös henkilöille lisää oikeuksia ja rekisterinpitäjille lisää velvollisuuksia.

Sen lähtökohdat ovat selvästi hyvät, mutta lopulliset säädökset eivät ole mielestäni täysin onnistuneita. Sen soveltaminen on hankalaa. Vaikka asetus pyrkiikin hyvänä pidettäviin tavoitteisiin kuten tietojenkäsittelytoimien läpinäkyvyyteen ja avoimuuteen, se ei riittävässä määrin määrittele mitä nämä periaatteet tarkoittavat ja siksi niiden soveltaminen voi olla monitulkintaista. Toisaalta asetus on kuitenkin onnistunut jättämään sen verran tulkinnan varaa, että yrityselämän toiminta ja palveluiden kehittäminen ei ole täysin rajoitettua.

Tietosuoja-asetuksessa pyritään turvaamaan sellainen henkilötietojen käsittely, jonka vaikutuksena ei synny syrjivää toimintaa kuten syrjiviä tai puolueellisia automaattisia yksittäispäätöksiä. On kuitenkin kyseenalaista, että päästäänkö puolueettomuuteen tai syrjimättömyyteen vain sillä, että niitä henkilön ominaisuuksia, joita ihminen pitää syrjivinä, ei käsiteltäisi lainkaan tai niiden käsittelyä rajoitettaisiin. Ne luonnolliseen henkilöön liittyvät ominaisuudet, jotka yleisesti yhdistetään syrjintään, kuten etninen alkuperä, eivät samalla tavalla vaikuta koneelliseen, automaattiseen päätöksentekoon kuin ihmisen tekemään päätöksentekoon.

Kun algoritmisia malleja opetetaan, siirretään sille samalla opetukseen käytettävän eli historiallisen datan mukana kaikkia niitä piileviä lainalaisuuksia, joita on käytetty myös kyseisten päätöksen tekemisessä. Data ei ole siis puolueetonta ja objektiivista vaan kuvastaa aikaansa ja tekijäänsä. Onkohan mahdollista, että lainsäätäjä on olettanut, että automaattisesti tehdyt päätökset ovat lähtökohtaisesti heikompia kuin ihmisen tekemät?

Yleinen tietosuoja-asetus rajoittaa henkilötietojen keräämistä ja sen käsittelyä. Sinänsä se onkin toivottavaa yksittäisen henkilön näkökulmasta. Henkilötietoja kuitenkin pitäisi pystyä käyttämään niitä keräävän yrityksen näkökulmasta melko moneen käyttötarkoitukseen. Nämä käyttötarkoitukset tulisi kertoa selvästi myös rekisteröidylle, mutta miten voidaan innovoida uusia palveluita, kun käyttötarkoitussidonnaisuuden periaate sitoo tietojen käyttöä? Yhteen tarkoitukseen kerättyjä tietoja voidaan vain tietyin ehdoin ja rajoituksin käyttää muuta tarkoitusta varten. Näitä kaikkia sekundäärisiä käyttötarkoituksia ei tiedetä, kun tietoja kerätään.

## 7.2 Automaattinen päätöksenteko, profilointi ja selitettävyys

Automaattinen päätöksenteko ja profilointi voivat olla tarpeellisia, jossain määrin jopa välttämättömiä osia yritysten liiketoimintaprosesseissa. Yrityksien eli rekisterinpitäjien olisi hyvä kuitenkin pitää mielessä päätösten selitettävyys. Vaikka tietosuoja-asetus ei nimenomaisesti velvoita selitettävyyteen sinänsä, mielestäni johtopäätöksenä voidaan kuitenkin todeta, että se olisi vähintään rekisterinpitäjän osoittamisvelvollisuuden mukaista.

Yksittäispäätösten selitettävyys on kuitenkin haastava toteuttava. Mielestäni selitettävyys lähtee siitä, että rekisterinpitäjä itse ymmärtää monilla organisaationsa tasoilla, miten automaattisia päätöksiä tehdään. Vaikka automaattisessa päätöksessä ei käytettäisikään uusimpia tekoälyn teknologioita vaan determinististä, sääntöihin perustuvaa logiikkaa, se on kuitenkin tietyissä olosuhteissa automaattinen päätös. Etenkin kun tekoälyä kuten koneoppimista sovelletaan enemmän ja laajemmin, tulisi yhä useampien ymmärtää sen peruslogiikkaa, kuten miten malleja opetetaan ja millaisia asioita ne voivat ennustaa.

Tietojenkäsittelytieteessä on jo tutkimusta siitä, miten vaikkapa erilaisia koneoppimisen malleja voidaan arvioida. Arviointi liittyy tosin yleensä niiden ennustetarkkuuteen tai virheiden määrään. Toisaalta nämä eivät ole kaukana siitä, kun mallin tuottamaan päätöstä pitää pystyä perustelemaan.

## 7.3 Tietosuojajuridiikka ja Business Intelligence

Monilla tieteenaloilla vaikuttaa olevan samankaltaisia käsitteitä ja niiden tutkimuksissa päädytään usein samojen teemojen äärelle. Kiinnostavaa oli kuitenkin, että usein tutkijat pidättäytyvät etsimästä vastauksia oman alueensa ulkopuolelta. Esimerkiksi tilastotieteessä on jo useita tietyllä tavalla valmiita käsitteitä jäsentämään ”datan keruuseen” liittyviä teemoja. Tällaisia ovat muun muassa otantavirhe ja kehikkovirhe. Juridisessa keskustelussa keskustellaan käsitteistä eri nimillä, mutta niiden sisältö vastaa kuitenkin melko hyvin vaikkapa juuri tilastotieteen käsitteitä. Vastaavasti tietojenkäsittelytieteiden puolelta löytyy käsitteitä, joiden paremmasta ymmärryksestä olisi hyötyä myös juridiikan puolella.

Eri alojen keskusteluista on helppo löytää tiettyjä teemoja, käsitteitä ja asioita, joista puhutaan paljon. Business intelligence tai liiketoimintatiedon hallinta on yksi tällaisista asioista. Jotta siihen liittyviä menetelmiä voidaan soveltaa, tulisi ensin pystyä jäsentämään kyseiset menetelmät ja analysoida, miten ne liittyvät tietosuojajuridiikkaan. Esimerkiksi onko yleisen tietosuoja-asetuksen tarkoittama automaattinen päätös sama asia kuin vaikkapa luottopäätös tai verkkokaupan näyttämä suosittelu?

## 7.4 Lopuksi

Käsitelty aihe on ajankohtainen ja relevantti, mutta se antaa myös viitteitä mahdollista jatkotutkimusta varten. Yksityisyyden suojan tutkiminen avaisi varmasti laajemmin tätä aihealuetta, koska sitä voidaan pitää laajempana käsitteenä kuin pelkkää henkilötietojen suojaa. Silloin tarkasteluun voisi ottaa myös sähköistä viestintää laajemmin ja vaikkapa digitaalista jalanjälkeä ottamatta kantaa siihen, onko kyseessä henkilötieto vai ei.

Automaattinen päätöksenteko ja profilointi olisivat myös mielenkiintoisia jatkotutkimuksen kohteita etenkin tekoälyn valossa. Tähän liittyen päätösten perusteltavuus ja läpinäkyvyys päätöksen kohteena olevalle on uskoakseni teema, joka tulee yhä ajankohtaisemmaksi. Lisäksi päätösten vaikutus tai oikeusvaikutukset voisi olla mukana tarkastelussa.

# 8 LYHENNELUETTELO

| | |
|---|---|
| GDPR | General Data Protection Regulation, yleinen tietosuoja-asetus |
| EUVL | Euroopan unionin virallinen lehti |
| DPIA | Data Protection Impact Assessment, tietojen käsittelyn vaikutustenarviointi |
| TSV | Tietosuojavaltuutettu |

# 9 LÄHTEET

Aletras N., Tsarapatsanis D., Preotiuc-Pietro D., Lampos V. 2016. Predicting judicial decisions of the European Court of Human Rights: a natural language processing pespective. Peer J Comput Sci. doi:10.7717/peerj-cs.93

Barocas, S., Selbst, A., 2016. Big Data's Disparate Impact. 104 California Law Review 671, 2016. Saatavilla: https://ssrn.com/abstract=2477899 tai http://dx.doi.org/10.2139/ssrn.2477899 . Viitattu 1.12.2019.

Bengio, Y. 2009. Learning Deep Architectures for AI. Foundations and Trends® in Machine Learning 2(1):1-55. Saatavilla https://www.researchgate.net/publication/215991023_Learning_Deep_Architectures_for_AI , vaatii kirjautumisen. Viitattu 16.12.2019.

COM(2018) 237. Komission tiedonanto Euroopan parlamentille, neuvostolle, Euroopan talous- ja sosiaalikomitealle ja alueiden komitealle – tekoäly Euroopassa. Euroopan komissio.Saatavilla: https://eur-lex.europa.eu/legal-content/FI/TXT/HTML/?uri=CELEX:52018DC0237&from=EN . Viitattu 19.12.2019.

COM(2018) 795. Komission tiedonanto Euroopan parlamentille, Eurooppa - neuvostolle, neuvostolle, Euroopan talous- ja sosiaalikomitealle ja alueiden komitealle – koordinoitu tekoälysuunnitelma. Saatavilla: https://eur-lex.europa.eu/legal-content/FI/TXT/HTML/?uri=CELEX:52018DC0795&from=EN . Viitattu 19.12.2019.

Chen, Y.-S., Cheng, C.-H. 2013. Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry. Knowledge-Based Systems. 39, 224–239.

Digibyte. 2018. EU Member States sign up to cooperate on Artificial Intelligence. 10.4.2018. Euroopan komission www-julkaisu. Saatavilla https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence . Viitattu 7.12.2019.

Digital Single Market. 2019. High-Level Expert Group on Artificial Intelligence. Policy. Saatavilla: https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence . Viitattu 7.12.2019.

Edelman, B., Luca, M. 2014. Digital discrimination: the case of Airbnb.com. Harvard business school working paper. Saatavilla: https://www.hbs.edu/faculty/Publication%20Files/Airbnb_92dd6086-6e46-4eaf-9cea-60fe5ba3c596.pdf . Viitattu 1.12.2019.

Eduskunta. 2018. EU:n yleisen tietosuoja-asetuksen (GDPR) täytäntöönpano – Uusi tietosuojalaki. Lakihankkeen tietopaketti. Saatavilla: https://www.eduskunta.fi/FI/tietoaeduskunnasta/kirjasto/aineistot/kotimainen_oikeus/LATI/EUn-tietosuojauudistus/sivut/eun-yleinen-tietosuoja-asetus.aspx. Viitattu 30.12.2019.

European Data Protection Supervisor. 2017. Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit. Saatavilla : https://edps.europa.eu/sites/edp/files/publication/17-04-11_necessity_toolkit_en_0.pdf . Viitattu 7.12.2019.

Euroopan komission. 2011. Special eurobarometer 359. Attitudes on Data Protection and Electronic Identity in the European Union. Saatavilla: https://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_359_en.pdf . Viitattu 1.12.2019.

Finlay, S. 2011. Multiple classifier architectures and their application to credit risk assessment. European Journal of Operational Research 210, 368-378.

Goodman, B., & Flaxman, S. 2016. European Union regulations on algorithmic decision-making and a “right to explanation”. AI Magazine 38.3 (2017): 50–57. Saatavilla: https://arxiv.org/abs/1606.08813 . Viitattu 8.12.2019.

Hayashi, Y. and Oishi, T. 2018. High Accuracy-priority Rule Extraction for Reconciling Accuracy and Interpretability in Credit Scoring. New Generation Computing, 36(4), pp. 393–418. doi: 10.1007/s00354-018-0043-5.

James, G., Witten, D., Hastie, T., & Tibshirani, R. 2017. An introduction to Statistical Learning with Application in R. Springer New York Heidelberg Dordrecht London. ISBN 978-1-4614-7137-0. Saatavilla myös : http://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf . Viitattu 8.12.2019.

Kontkanen, E. 2018. Selvitys positiivisia luottotietoja koskevan järjestelmän edellytyksistä. Oikeusministeriön selvityksiä ja ohjeita 26/2018. Saatavilla

http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161002/OMSO_26_2018_Positiivinen_luottotietojarjestelma.pdf?sequence=1&isAllowed=y . Viitattu 6.5.2019.

Konnu, J. 2006. Mikroaineistojen tilastolliset tietosuojamenetelmät henkilötilastoissa. Tilastotieteen pro gradu -tutkielma, Jyväskylän yliopisto. Saatavilla: https://jyx.jyu.fi/bitstream/handle/123456789/12534/URN_NBN_fi_jyu-20079.pdf?sequence=1 . Viitattu 24.11.2019.

Kontkanen, E. 2018. Selvitys positiivisia luottotietoja koskevan järjestelmän edellytyksistä. Selvityksiä ja ohjeita 26/2018. Oikeusministeriö. Saatavilla: http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161002/OMSO_26_2018_Positiivinen_luottotietojarjestelma.pdf?sequence=1&isAllowed=y . Viitattu 16.12.2019.

Kullback, S., & Leibler, R. 1951. On Information and Sufficiency. The Annals of Mathematical Statistics, 22(1), 79-86.

Kingston, J., 2017. Using artificial intelligence to support compliance with the general data protection regulation. Artificial Intelligence and Law, 25(4), pp. 429-443.

Kananen, H., Puolitaival, H. 2019. TEKOÄLY - Bisneksen uudet tyokalut. Alma Talent verkkokirja. Luettu 24.10.2019. Vaatii käyttöoikeuden, saatavilla: https://bisneskirjasto-almatalent-fi.ezproxy.utu.fi/teos/BAXBBXATCBIED#kohta:TEKO((c4)LY((20)-((20)Bisneksen((20)uudet((20)tyokalut/piste:tV

Kay, M., Matuszek, C., Munson, S. 2015. Unequal representation and gender stereotypes in image search results for occupations. Teoksessa Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Saatavilla: https://www.csee.umbc.edu/~cmat/Pubs/KayMatuszekMunsonCHI2015GenderImageSearch.pdf . Viitattu 1.12.2019.

Micheletti, G. & Pepato, C. 2019. D2.6 Second Interim Report. The European datamarket tool: key facts & figures, first policy conclusions, data landscape and quantified stories. Saatavilla: http://datalandscape.eu/study-reports/second-interim-report-european-data-market-monitoring-tool-key-facts-figures-policy . Viitattu 1.12.2019.

Narayanan, A. & Shmatikov, V. 2008. Robust de-anonymization of large sparse datasets. In Security and Privacy, 2008. SP 2008. IEEE Symposium on (pp. 111-125).

IEEE. Saatavilla myös: https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf . Viitattu 24.11.2019.

Oikeusministeriö. 2019. Ministerityöryhmä: vauhtia ylivelkaantumisen torjuntaan. Oikeusministeriön tiedote 24.10.2019. Saatavilla: https://oikeusministerio.fi/artikkeli/-/asset_publisher/ministerityoryhma-vauhtia-ylivelkaantumisen-torjuntaan . Viitattu 30.12.2019.

Tietosuoja. 2019. Tietosuojavaltuutetun toimiston www-sivusto. Saatavilla https://tietosuoja.fi/tietosuoja . Viitattu 30.11.2019.

Tekoälyä käsittelevä korkean tason asiantuntijaryhmä. 2019a. Tekoälyn määritelmä: tärkeimmät valmiudet ja tieteenalat. Saatavilla (suomenkielinen versio): https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines. Viitattu 7.12.2019.

Tekoälyä käsittelevä korkean tason asiantuntijaryhmä. 2019b. Luotettavaa tekoälyä koskevat eettiset ohjeet. Saatavilla (suomenkielinen versio): https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines. Viitattu 7.12.2019.

Wachter, S., Mittelstadt, B. & Floridi, L. 2017. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. Artikkeli. International Data Privacy Law, 2017, Vol. 7, No. 2. Oxford University Press.

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, Issue 2. 2016. s. 183–201.

Riestra, A. 2002. Credit bureaus in today's credit markets. ECRI research report no. 4. September 2002. Saatavilla: http://aei.pitt.edu/9431/2/9431.pdf . Viitattu 16.12.2019.

Rita, H. 2004. Vetosuhde (odds ratio) ei ole todennäköisyyksien suhde. Metsätieteen aikakauskirja 2/2004. Saatavilla: http://www.metla.fi/aikakauskirja/full/ff04/ff042207.pdf . Viitattu 20.12.2019.

Saar-Tsechansky, M., Melville, P. & Provost, F. 2009. Active Feature-Value Acquisition. Management Science, vol. 55, no. 4, pp. 664-684.

Tilastokeskus. 2007. Laatua tilastoissa. Käsikirjoja 43. 2. uudistettu painos. Yliopistopaino. Helsinki.

Vartiainen, M. 2014. Liiketoimintatiedon hallinta ja analytiikka. Pro gradu -tutkielma. Tietojenkäsittelytieteen laitos. Itä-Suomen Yliopisto. Saatavilla: http://epublications.uef.fi/pub/urn_nbn_fi_uef-20141439/urn_nbn_fi_uef-20141439.pdf . Viitattu 16.12.2019.

Vellido, A., Martín-Guerrero, J., Lisboa, P. 2012. Making machine learning models interpretable. ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Saatavilla: https://pdfs.semanticscholar.org/ce0b/8b6fca7dc089548cc2e9aaac3bae82bb19da.pdf. Viitattu: 20.12.2019.

WP128. 2006. Opinion 10/2006 on the processing of personal data by the Society for Worldwide Interbank Financial Telecommunication (SWIFT). Saatavilla: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2006/wp128_en.pdf .Viitattu 27.10.2019

WP136. 2007. Opinion 4/2007 on the concept of personal data. Saatavilla: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf . Viitattu 17.11.2019.

WP169. 2010. Lausunto 1/2010 rekisterinpitäjän ja henkilötietojen käsittelijän käsitteistä. Saatavilla: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp169_fi.pdf . Viitattu 7.11.2019.

WP203. 2013. Opinion 03/2013 on purpose limitation. Saatavilla: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf . Viitattu 26.10.2019.

WP216. 2014. Opinion 05/2014 on Anonymisation Techniques. Saatavilla: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf. Viitattu 17.11.2019.

WP217. 2014. Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC. 844/14/EN. Saatavilla https://fia.org/sites/default/files/uploaded/Excerpts%20-%20Opinion%2006-2014%20on%20the%20notion%20of%20legitimate%20interests%20of%20the%20....pdf . Viitattu 13.4.2019.

WP243. 2016. Tietosuojavastaavia koskevat ohjeet. Saatavilla https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612048 . Viitattu 26.10.2019.

WP248. 2017. Ohjeet tietosuojaa koskevasta vaikutustenarvioinnista ja keinoista selvittää ”liittyykö käsittelyyn todennäköisesti” asetuksessa (EU) 2016/679 tarkoitettu "korkea riski". Saatavilla https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236 . Viitattu 25.10.2019.

WP251rev.01. 2018. Suuntaviivat automatisoiduista yksittäispäätöksistä ja profiloinnista asetuksen (EU) 2016/679 täytäntöön panemiseksi. 29 artiklan mukainen tietosuojaryhmä. Saatavilla https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 . Viitattu 9.4.2019.

WP259 rev.01. 2018. Asetuksen 2016/679 mukaista suostumusta koskevat suuntaviivat. 29 artiklan mukainen työryhmä. Saatavilla https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051 . Viitattu 11.4.2019.

WP260. 2018. Asetuksen 2016/679 mukaista läpinäkyvyyttä koskevat suuntaviivat. Saatavilla: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227 . Viitattu 18.12.2019.