Big datana kerättävien henkilötietojen arvo yksilölle

Turun yliopisto
Tietotekniikan laitos
LuK-tutkielma
Tietojenkäsittelytiede
Huhtikuu 2024
Minna Palm

TURUN YLIOPISTO
Tietotekniikan laitos

Minna Palm: Big datana kerättävien henkilötietojen arvo yksilölle

LuK-tutkielma, 24 s.
Tietojenkäsittelytiede
Huhtikuu 2024

Big datan aikakaudella datan merkitys on kasvanut. Henkilötietodataa kerätään suuria määriä, ja sillä käydään kauppaa datamarkkinoilla. Yrityksille henkilötiedot ovat arvokasta dataa, kun taas yksilöiden tietoisuus ja suhtautuminen omien tietojensa arvoon vaihtelee. Tässä tutkielmassa on tavoitteena arvioida big datana kerättävien henkilötietojen arvoa yksilönäkökulmasta. Tutkielma selvittää, millä eri tavoin henkilötietoja voidaan hinnoitella ja mitkä tekijät vaikuttavat hinta-arvioihin. Lisäksi tarkastellaan tietosuoja-asetusten asettamia rajoituksia henkilötietojen käsittelyyn. Tutkielma on toteutettu kirjallisuuskatsauksena aiheeseen liittyviin teoksiin, joista merkittävimmät ovat artikkelit Pricing Privacy – the Right to Know the Value of Your Personal Data (Malgieri ym., 2018) ja The Effect of Fair Information Practises and Data Collection Methods On Privacy-related Behaviors: A Study of Mobile Apps (Libaque-Sáenz, C.F. ym., 2021). Tutkielmassa päädytään siihen, että henkilötietojen hinnan todellinen ja tarkka arviointi on nykyisillä menetelmillä erittäin haastavaa. Arviointimenetelmillä on kullakin omat puutteensa ja itse henkilötietodata myöskin vaihtelee ominaisuuksiltaan. Epätarkat hinta-arviot voivat kuitenkin vaikuttaa yksilöiden yksityisyyden suojaamisen motivointikeinona tietosuojalakien rinnalla. Tämä tutkielma suosittaa jatkotutkimuksia yksilöiden henkilötietojen hintatietoisuuden lisäämiseen.

Asiasanat: big data, henkilötiedot, GDPR, yksityisyys

Sisällys

Kuvat

Taulukot

1 Johdanto

Data voidaan määritellä tiedoksi, jota talletetaan symbolisessa muodossa jollekin välineelle, kuten tietokoneelle [1]. Sen lähteisiin lukeutuvat muun muassa erilaiset sensorit, monitorit, ohjelmistot, puhelimet, lokitiedostot ja puettavat teknologiat [2]. Ilmiönä datalla on suunnaton määrä potentiaalia edistää innovaatioita, taloutta, tuotteliaisuutta ja tulevaa kasvua [3]. Maailmantaloudessa hyödynnetään dataa yhä enemmän, ja esimerkiksi yrityksissä pyritään perustamaan päätöksentekoa datalähteisiin. Dataa syntyy tarjolle jatkuvasti lisää, ja esimerkiksi vuoteen 2025 mennessä on määrän ennustettu kasvavan yli 180 zettatavuun [4]. Ilmiötä voidaan kutsua suurten datamäärien eli big datan vallankumoukseksi [5].

Datan ja kuluttajadatan kaupallistaminen on johtanut siihen, että yrityksillä on mahdollisuus kerätä valtavia määriä henkilödataa asiakkaistaan [6]. Yritykset ja erilaiset sosiaalisen median alustat, kuten Facebook, Twitter ja Amazon, tehostavat näin omaa liiketoimintaansa ja pyrkivät datan avulla saamaan kilpailuetua muihin yrityksiin [7][8]. Kuluttajadataa ja niihin liittyviä analysointipalveluita onkin viime vuosina ollut enenevissä määrin tarjolla datamarkkinoilla [9].

Datamarkkinoiden myötä ovat vakiintuneet datanvälitysyritykset, jotka erikoistuvat henkilötietojen keräämiseen ja myymiseen jälleenmyyjille, mainostajille, markkinoille ja valtion virastoille.[5] Tällainen kuluttajien digitaalisten identiteettien kaupallistaminen on yhä kasvava ilmiö, jossa kuitenkaan kuluttajilla ei näytä olevan tietoa omien henkilötietojensa todellisesta arvosta [10]. Henkilötietoja on luonneh-

dittu Internetin uudeksi öljyksi ja digitaaliseksi valuutaksi [8], jolla kuluttajat ovat maksaneet ”ilmaisista” palveluista Internetissä [11].

Yrityksillä on liiketoimintamalleja, joissa tarjotaan näennäisesti ilmainen palvelu kuluttajille todellisuudessa vastineeksi heidän henkilötietodatastaan [5]. Nykyisin suurin osa digitaalisista palveluista on niin ikään maksettava henkilötiedoilla. Tällaisiin palveluihin lukeutuvat noin 30 % virustentorjunta- ja navigointiohjelmistoista sekä pilvitallennuspalveluista, 77 % suoratoistotapahtumista ja yli 50 % elokuvista, TV-ohjelmista, e-kirjoista ja peleistä.[10]

Yritysten ja kuluttajien välinen tietomäärä henkilötietojen arvosta on epäsymmetrinen [10][11]. Yksilöiden tietoisuutta omista henkilötiedoistaan tulisi lisätä ja kuluttajien asemaa datamarkkinoilla vahvistaa, jotta heillä olisi paremmat edellytykset suojella omia tietojaan [10]. Tässä tutkielmassa tavoitteena on selvittää, minkälainen todellinen arvo henkilötietodatalla on ja mitä lainsäädännöllisiä asetuksia yksilöiden henkilötietojen turvaamiseksi on jo olemassa. Tutkimuskysymyksinä ovat tk1 = ”Miten henkilötietodatan arvo voidaan määritellä?” ja tk2 = ”Mitä lainsäädännöllisiä askelia on otettu yksilöiden henkilötietodatan suojaamiseksi?”

Tutkielma on toteutettu kirjallisuuskatsauksena, jossa on haettu aineistoa eri tietokannoista ja sovellettu kuhunkin tietokantaan sopivia hakulausekkeita. Suuremmissa tietokannoissa on hyödynnetty enemmän rajaavia avainsanoja, jotka on esitetty taulukossa 1.1. Haussa on aineiston rajaamisen kriteereinä käytetty englanninkielisyyttä ja julkaisuvuosia 2010–2024. Aineistoa on myös jälkikäteen rajattu jättämällä sisällön perusteella pois julkaisut, jotka viittaavat erityisaloihin, kuten lääketieteeseen tai yrityspolitiikkaan.

Taulukko 1.1: Tutkielman aineistonhaussa käytetyt tietokannat ja hakulausekkeet tuloksineen.

	tietokanta	hakusana	tulokset yhteensä	otsikon perusteella valitut	sisällön perusteella lopullisesti valitut
1	Volter	”big data” AND ”personal data”	1563	24	2
2	Volter	”big data” AND ”personal data” AND (pric* OR market)	237	26	3
3	Scopus	”big data” AND ”personal data” AND individual AND privacy AND ”data collection”	621	21	1
4	ACM	”personal data” AND ”private data” AND privacy AND ”data collection”	540	38	1

Taulukko 1.2: Tutkielman aineistot aiheineen.

	Aineisto	Datan määrittely	Datan hinnoittelu	Yksityisyys ja GDPR
1	Hung, Patrick C. K. *Big Data Applications and Use Cases.* (2016)	x		
	Pietsch, Wolfgang. *Big Data.* (2021)	x		
2	Malgieri, Gianclaudio ja Custers, Bart. *“Pricing Privacy – the Right to Know the Value of Your Personal Data”.* (2018)		x	x
	Spiekermann, Sarah, ym. *“The Challenges of Personal Data Markets and Privacy”.* (2015)		x	x
	Becerril, Anahiby Anyel. *“The Value of Our Personal Data in the Big Data and the Internet of All Things Era”.* (2018)		x	
3	Libaque-Sáenz, C.F. ym. *”The Effect of Fair Information Practices and Data Collection Methods on Privacy-related Behaviors: A Study of Mobile Apps”.* (2021)			x
4	Zainab, Syeda Sana e ja Kechadi, Tahar. *”Sensitive and Private Data Analysis: A Systematic Review”.* (2019)	x		x

Tutkielma rakentuu siten, että ensin luvussa 2 alustetaan big datan ja henkilötietojen sekä niihin liittyvien käsitteiden määritelmät. Seuraavaksi luvussa 3 tarkastellaan tietosuojalakeja, joiden määräämissä raameissa henkilötietodatan käsittely ja arvotus tulee tehdä. Luku 4 käsittelee tarkemmin datan hinnoitteluun vaikuttavia

tekijöitä, erilaisia hinnoittelumetodeja ja niihin liittyviä ongelmia. Lopuksi luvussa 5 pohditaan, voiko yksilöitä kannustaa omien tietojensa suojaamiseen tietosuojalakien ja henkilötietojen hinnoittelusta tiedottamisen avulla.

2 Big datan ja henkilötietojen määrittely

Big data eli massadata on käsite, jolle ei ole olemassa täysin yksikäsitteistä määritelmää. Alasta ja organisaatiosta riippuen se voidaan määritellä eri tavoin ja tarkentuu siihen liittyvän teknologian kehittyessä [12]. Yleistajuisesti big datalla voidaan kuitenkin viitata sellaiseen dataan, jota on kertynyt niin paljon, ettei sitä pystytä käsittelemään perinteisillä datan hallintaan tarkoitetuilla työkaluilla [13].

Alun perin big datan määrittelemisessä käytettiin apuna ns. kolmea V:tä (V's of Big Data), jotka keksi Doug Laney vuonna 2001. Nämä v-kirjaimella alkavat termit, volume, variety ja velocity, ovat ominaisuuksia, jotka tekevät datasta juuri big dataa. Näistä ominaisuuksista volume viittaa datan määrään, variety datan erityyppisiin rakenteisiin ja velocity siihen nopeuteen, jolla dataa tuotetaan.[14] Myöhemmin v-kirjaimella alkavia termejä on lähteestä riippuen lisätty alkuperäisten rinnalle. Tällaisia ovat esimerkiksi variability (datan vaihtelu), veracity (todenmukaisuus), value (arvo) tai visualization (visualisointi).[1]

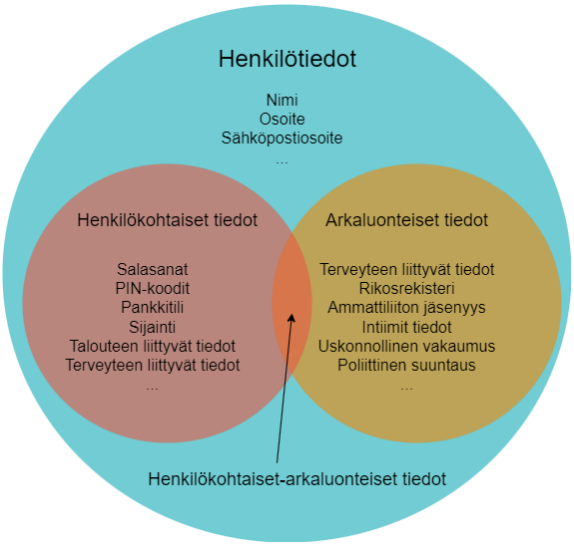
Tässä tutkielmassa big datan määrityksessä keskitytään alkuperäiseen kolmeen V:hen. Tällöin big dataksi luokiteltava data on jatkuvasti kasvavaa ja voi saavuttaa jopa kvadriljoonan tavun koon. Lisäksi sen rakenne voi olla niin strukturoitua kuin ei-strukturoituakin. Toisin sanoen data joko on jossain ennalta määrätyssä muodossa, kuten tietokannassa, Excel-taulukossa tai videona, tai sitä ei ole järjestetty

tietynlaiseen formaattiin, kuten käsinkirjoitetuissa muistiinpanoissa. Tällaista, big dataksi luokiteltavaa dataa tuotetaan ja päivitetään nopealla tahdilla.[3]

Big data-ilmiön tilalla tai yhteydessä käytetään yleisesti myös termejä digital exhaust, digital footprints ja personal information tai personal data. Digital exhaust (digitaalinen pakokaasu) viittaa big dataan sivutuotteena, jota syntyy ihmisen ja digitaalisen teknologian välisestä vuorovaikutuksesta [2][15]. Digital footprints (digitaaliset jalanjäljet) on dataa, jota käyttäjät tuottavat ja jättävät jälkeensä itsestään digitaalisilla alustoilla käydessään [7]. Termit eroavat siinä, että yksittäisillä ”digitaalisen pakokaasun partikkeleilla” ei ole merkitystä, mutta yhdessä ne muodostavat mielekkäitä kokonaisuuksia, joita kutsutaan digitaalisiksi jalanjäljiksi [16].

Henkilötiedot (engl. personal information/data) ovat myös dataa, jota syntyy päivittäisessä vuorovaikutuksessa digitaalisen teknologian kanssa [15]. Vaikka termille ei ole yleismaailmallisesti vakiintunutta määritelmää, viittaa se yleisesti mihin tahansa dataan, josta henkilö voidaan tunnistaa. Henkilötietojen termiin liittyy monitulkintaisuutta, sillä se usein liitetään henkilökohtaisiin tietoihin (engl. private information) ja arkaluonteisiin tietoihin (engl. sensitive information).[17]

Suomen Tietosuojavaltuutetun toimisto määrittelee henkilötiedoiksi esimerkiksi yhteystiedot, joita ovat muun muassa nimi, osoite, puhelinnumero ja sähköpostiosoite, henkilökortin numeron, auton rekisterinumeron, paikannustiedot, IP-osoitteen ja terveystiedot [18]. Henkilökohtaiset ja arkaluonteiset tiedot ovat kaksi, osittain päällekkäistä kategoriaa henkilötieto-käsitteen sisällä. Zainabin ja Kechadin (2019) mukaan henkilökohtaiset tiedot viittaavat sellaisiin henkilötietoihin, joita henkilö ei halua paljastaa julkisesti. Arkaluonteinen data on taas tietoa, joka liittyy mm. yksilön hyvinvointiin, seksuaalisuuteen, etnisyyteen ja biometriseen tunnistamiseen. Yhdessä ne viittaavat tietoon, joka ei välttämättä liity tiettyyn tunnettuun henkilöön, mutta jota ei sovi julkisesti paljastaa.[17]



Kuva 2.1: Kaavio henkilötietojen, henkilökohtaisten tietojen ja arkaluonteisten tietojen suhteista. Mukaillen lähdettä [17].

Arviolta 98 % matkusteluun ja liikkumiseen liittyvistä iOS-sovelluksista kerätystä datasta voidaan yhdistää käyttäjän identiteettiin [19]. Henkilötietodatan kerääminen onkin helpompaa kuin aiempina vuosina [2]. Arkipäiväisessä elämässä henkilötietodataa voi muodostua esimerkiksi hakukoneissa, sosiaalisessa mediassa ja verkkokaupoissa vierailun yhteydessä.[15]

Henkilötietodataa voi kerätä joko pitkäaikaisesti yhdeltä yksilöltä tai lyhyemmällä aikavälillä useilta, sadoilta miljoonilta ihmisiltä. Dataa kerääviä teknologioita on puhelimissa, ladattavissa ohjelmistoissa, puettavissa kameroissa ja monessa muussa. Nämä tekniikat jäljittävät ja tallettavat yksityishenkilöiden sijaintia, tapoja, ostoksia, rutiineja, sosiaalista vuorovaikutusta ja mielipiteitä.[2] Datan keräämisen ja käsittelyn lainmukaisuutta ja ihmisten tietosuojaoikeuksien toteutumista valvoo Suomessa Tietosuojavaltuutetun toimisto. Henkilötietojen tietosuojaa käsitellään tarkemmin luvussa 3 ja arvoa luvussa 4.

3 Yksityisyys ja tietosuojakäytänteet

Ennen kuin on mahdollista tarkastella datalle asetettavaa arvoa, on otettava huomioon datan keräämiseen ja käsittelyyn vaikuttavat lait ja yksilöiden oikeudet yksityisyyteen. **Yksityisyys** viittaa henkilöiden vapauteen päättää ja kontrolloida, mitä henkilötietodataa he haluavat itse jakaa Internetissä [7]. Yksityisyyden käsite voidaan jakaa neljään kategoriaan: henkilötietojen käsittelyä koskevaan yksityisyyteen, fyysiseen yksityisyyteen eli ruumiilliseen koskemattomuuteen, sosiaaliseen yksityisyyteen ja alueelliseen yksityisyyteen [17]. Tässä tutkielmassa keskitytään rajatusti yksityisyyteen, joka koskee henkilötietojen käsittelyä.

Termejä yksityisyys ja turvallisuus on usein käytetty synonyymeinä eri tilanteissa, mutta ne tarkoittavat eri asioita [17]. Yksityisyyttä voi kuvata käyttäjien kokemana riskinottona omiin tietoihinsa liittyen, kun taas turvallisuus viittaa siihen, että pääsyä tietoihin suojataan luvattomilta toimijoilta. Kun yksilöt luovat profiilejaan, jakavat kiinnostuksen kohteitaan ja henkilötietodataansa sosiaalisessa mediassa, vaarantavat he potentiaalisesti yksityisyyttään ja turvallisuuttaan [7].

Internetin käyttäjät saavat vapaasti jakaa tietojaan, mikä on johtanut siihen, että yhä useammat nykypäivän organisaatiot harjoittavat kuluttajatietojen kauppaa. Tällaiset organisaatiot voivat toimia lain harmailla vyöhykkeillä henkilötietojen käsittelyssä. Datamarkkinoilla ei välttämättä välitetä yksityisyyttä lisäävistä tekniikoista ja yksityisyyden kokemisesta ihmisoikeutena.[8]

Tutkimuksissa on väitettävästi todettu, että kuluttajien henkilötietoihin liittyvien riskien tuntemus vaikuttaa merkittävästi heidän tapoihinsa käyttää sosiaalista mediaa ja että tämä vaikutus näkyy myös heidän digitaalisten jalanjälkiensä määrässä [7]. Kuluttajien turvaksi useissa valtioissa henkilötietojen käyttöä säännellään tiukasti [8]. Esimerkiksi vuonna 2019 Euroopassa arviolta 98 % yrityksistä noudatti Euroopan unionin tietosuoja-asetuksia jonkin verran, suurilta osin tai täydellisesti asteikolla ei ollenkaan - täydellisesti. Vastaava luku oli Yhdysvalloissa toimiville yrityksille noin 73 %.[20] Seuraavaksi käsitellään tarkemmin Yhdysvaltojen ja Euroopan tietosuojakäytänteitä sekä tuodaan esille lyhyesti käytänteiden alkuperän historiaa.

3.1 FIP

Yhdysvalloissa kuluttajien tietosuoja nousi esille alun perin Yhdysvaltain ministeriön automatisoituja henkilötietojärjestelmiä käsittelevän neuvoa-antavan komitean (the US Secretary's Advisory Committee on Automated Personal Data Systems) vuoden 1973 raportissa. Siinä käsiteltiin ns. reilun tietokäytännön periaatteita (engl. Fair Information Practises, FIP), jotka kuvaavat henkilötietojen yksityisyyden käsittelemiseen liittyviä kansainvälisesti tunnustettuja käytänteitä.[6]

Henkilötietojen yksityisyyttä suojaamaan kehitetty FIP määrää, että kaikkien henkilötietoja keräävien yritysten tulisi noudattaa neljää yleisesti tunnustettua periaatetta: ilmoitusta, pääsyä, valintaa ja turvallisuutta. Toisin sanoen dataa keräävien yritysten tulisi 1) julkaista tietokäytäntönsä eli ilmoittaa kuluttajille henkilötietojen keräämisestä, 2) sallia kuluttajille pääsy heidän omiin tietoihinsa, 3) hankkia suostumus kuluttajilta heidän tietojensa käyttämiseen ja 4) suojata kerätty data luvattomilta osapuolilta.[6][21] Viides periaate lisättiin vuonna 2010 [6], ja sen mukaan datan kerääjien tulisi ottaa käyttöön toimeenpanomekanismit tietosuojakäytänteiden ylläpitämiseen [21].

FIP:n periaatteet toimivat vain suosituksina, sillä niiden lainvalvonta on rajattua. Kuitenkin monet yritykset voivat hyödyntää näitä periaatteita omien, sisäisten yksityisyyskäytänteidensä rakentamiseen. FIP on ollut vaikuttamassa oleellisesti myös useiden maiden ja osavaltioiden nykyisten käytänteiden taustalla. Yksi merkittävimmistä FIP:n pohjalta luoduista tietosuoja-asetuksista on Euroopan unionin GDPR.[6][21]

3.2 GDPR

Yhdysvaltojen FIP-periaatteista inspiroituneet Euroopan unionin tietosuojakäytännöt saivat alkunsa 1980-luvulla. Vuodesta 2018 lähtien on sovellettu kenties kaikista tunnetuinta käytäntöä, **GDPR**:ää eli yleistä tietosuoja-asetusta (engl. General Data Protection Regulation).[6] GDPR on henkilötietojen käsittelyyn kantaa ottava laki, joka alun perin astui voimaan 24.5.2016 Euroopan unionin maissa ja joka siirtymäajan jälkeen otettiin käyttöön 25.5.2018 [22].

Asetus määrittelee artiklassa 4 henkilötiedoiksi kaikki tunnistettuun tai tunnistettavissa olevaan luonnolliseen henkilöön liittyvät tiedot. Tällaisia tietoja ovat esimerkiksi nimi, henkilötunnus, sijaintitiedot ja verkkotunnistetiedot. Laajemmin tunnistetiedoiksi lasketaan kaikki fyysiset, fysiologiset, geneettiset, psyykkiset, taloudelliset, kulttuuriset ja sosiaaliset tekijät, joista henkilö voidaan suoraan tai epäsuorasti tunnistaa.[22]

Artiklan 3 mukaan asetus koskee niin Euroopan unionin sisäisiä kuin ulkopuolisia toimijoita, jotka käsittelevät henkilötietoja. Tarkemmin sanottuna sitä sovelletaan Euroopan unionin alueella sijaitseviin henkilötietoja käsitteleviin toimijoihin, vaikka itse käsittely tapahtuisi muualla. Lisäksi asetusta sovelletaan EU:n ulkopuolisiin toimijoihin silloin, kun ne käsittelevät unionin alueella sijaitsevien henkilöiden henkilötietoja. Tilanteet, joissa asetusta ei sovelleta, koskevat henkilökohtaista tai kotitalouteen liittyvää toimintaa, kuten kirjeenvaihtoa ja sosiaalista verkostoitumis-

ta. Myöskin siinä tapauksessa, että henkilö on kuollut, hänen henkilötietoihinsa ei asetusta sovelleta.[22]

Kokonaisuudessaan GDPR on yksi maailman tiukimmista henkilötietoja suojaavista laeista ja se on vaikuttanut merkittävästi seuraavaan neljään osa-alueeseen: EU:n kansalaisten oikeudet, henkilötietojen määritys, organisaatioiden vastuu datan tietoturvallisesta käsittelystä ja hallinnasta sekä tietosuoja-asetusten noudattaminen [6]. Seuraavaksi esitellään lyhyesti näitä osa-alueita perustellen ne asetuksen asianmukaisilla kohdilla.

GDPR on parantanut alueensa kansalaisten oikeuksia henkilötietojen suojaamiseen. Sen mukaan datan kerääjien tulee selkeällä tavalla ilmoittaa rekisteröidylle henkilötietojen keräämisestä (artiklat 10–14) ja myöskin taata tälle pääsy omiin tietoihinsa (artikla 15). Lisäksi rekisteröidyllä on oikeudet tietojensa oikaisuun (artikla 16) ja poistamiseen (artikla 17) sekä siirtämiseen toiseen järjestelmään (artikla 20). Rekisteröity on oikeutettu myös vastustamaan henkilötietojensa käsittelyä, ellei datan kerääjä pysty osoittamaan erityisen painavia syitä käsittelylle (artikla 21).[22]

GDPR-asetuksen määritelmä henkilötiedoille on laaja ja kattaa kaikki niin suorat kuin epäsuoratkin tunnisteet sekä muun muassa biometrisen tunnistamisen, geneettisen datan ja tietokoneen evästeet. Asetus määrittelee erikseen tilanteet, joissa henkilötietojen kerääminen on perusteltua tietosuoja-asetuksesta huolimatta.[6] Tällaisia tilanteita ovat esimerkiksi kansalliseen ja yleiseen turvallisuuteen sekä puolustukseen liittyvät toiminnot (artikla 23).[22]

Tietosuoja-asetus määrää, että organisaatioiden on noudatettava tietosuojaperiaatteita, kuten tietojen minimointia ja pseudonymisointia (artikla 25). Tietojen minimoinnilla tarkoitetaan käytännössä sitä, että kerätään pienin mahdollinen määrä dataa tavoiteltuun tarkoitukseen (artikla 5). Pseudonymisointi viittaa käsittelyyn, jonka jälkeen henkilötietoja ei voi enää yhdistää henkilöön ilman lisätietoja (artikla 4). Lisäksi asetus määrää, ettei henkilötietoja voi oletusarvoisesti saattaa

rajoittamattoman henkilömäärän saataville ilman, että siihen on suostumus rekisteröidyltä.[22] Tämä viittaa siihen, että datan kerääjien on pyydettävä suostumus kuluttajalta silloin, kun henkilötietoja myydään tai käytetään uudelleen [6].

GDPR asettaa myös seuraamuksia tietosuoja-asetuksen noudattamatta jättämisestä [6]. Artiklan 82 mukaan henkilöllä on oikeus korvauksiin, jos hänelle on asetuksen rikkomisesta aiheutunut aineellista tai aineetonta vahinkoa [22]. Esimerkiksi vuonna 2023 määrättiin Euroopan unionissa korvauksia noin 2,1 miljardin euron edestä, mikä oli enemmän kuin edeltävinä vuosina 2019–2021 yhteensä. Pääsyy ilmiölle oli ennätyksellisen suuri 1,2 miljardin euron sakko Facebookin emoyhtiö Metalle.[23]

FIP:n ja GDPR:n välillä ollut asetelma on nykyisin kääntynyt toisinpäin. On arveltu, että toisin kuin aiemmin, nyt Euroopan GDPR:llä on vaikutusvaltaa Yhdysvaltojen osavaltioiden yksityisyydensuojalakeihin. Myös monet muut maat ovat ottaneet käyttöön tietosuoja-asetuksia GDPR:n voimaan astumisen jälkeen. Tällaisia maita ovat esimerkiksi Brasilia, Intia, Australia ja Brexitin jälkeinen Iso-Britannia.[6]

Tietosuoja-asetukset voidaan kokea yksilöiden passiivisena puolustuskeinona datan keräämistä, käyttöä ja uudelleenkäyttöä vastaan. Malgierin ja Custersin (2018) mukaan tämä passiivinen tapa toimii vain tietojen tunnearvon suojelemiseen ja tehokkaampi, sekä realistisempi puolustuskeino olisi yksilöiden kannustaminen omien henkilötietojensa hallintaan. Yksityisyyden suojaamiseen voi motivoida tieto omien henkilötietojen arvosta.[10] Seuraavassa luvussa syvennytään tarkemmin henkilötietojen arvoon, hinnoittelutapoihin ja hinnoitteluun liittyviin ongelmiin.

4 Datan hinta

Henkilötietomarkkinoita on ennakoitu jo 1990-luvulta lähtien tutkimuspiireissä [8]. Henkilötiedot ovat nykyään yksi Euroopan tärkeimmistä taloudellisista voimavaroista, mutta niiden potentiaalia ei hyödynnetä tarpeeksi: vain kaksi 20:stä henkilötiedoista rahaa saavista suurimmista yrityksistä ovat eurooppalaisia [15]. Henkilötietodataa on kutsuttu ”Internetin uudeksi öljyksi” ja ”digitaalisen maailman valuutaksi”, koska sillä katsotaan olevan rahallista potentiaalia niin yrityksille kuin kuluttajillekin [8].

Yritykset hyödyntävät henkilötietoja eri tarkoituksiin, kuten personoituihin mainoksiin ja tuotehakuun, kuluttajien riskianalyyseihin, sekä taloustoimikustannusten alentamiseen. Yrityksille henkilötiedot voivat olla strategista pääomaa, jolla ne saavat kilpailuetua markkinoilla tai jolla ne kehittävät toimintojaan.[8] Yritykset pitävät kuluttajien henkilötietoja ja niihin liittyviä algoritmeja omaisuutenaan ja varjelevat niitä liikesalaisuutena.[10] Henkilötieto on tuote itsessään, etenkin sosiaalisessa mediassa käyttäjien luoman sisällön yhteydessä.[8]

Henkilötiedoilla on näin todettu olevan rahallista arvoa datamarkkinoilla. Kuluttajat ovat saaneet vastineeksi henkilötiedoistaan ”ilmaisia” palveluita tai alennusta tuotteista ja palveluista Internetissä.[10][11] Vasta 11.6.2019 astui Euroopan unionissa voimaan direktiivi, joka rinnastaa henkilötiedot kauppahintaan, jonka kuluttaja maksaa digitaalisista palveluista tai sisällöstä. Direktiivissä digitaaliset palvelut ja sisällöt kattavat mm. pilvipalvelut, sosiaalisen median, tietokoneohjelmat, mo-

biilisovellukset, sekä digitaaliset video- ja äänitiedostot. Direktiiviä on ollut määrä soveltaa EU:n jäsenvaltioissa 1.1.2022 lähtien.[24]

Toisin kuin yritykset, yksilöt eivät ole yhtä tietoisia henkilötietojensa rahallisesta arvosta ja usein tuntuvat hyväksyvän digitaalisen identiteettinsä kaupallistamisen [10]. Tästä johtuen instituutioiden ja yksilöiden välinen valta henkilötietodataan on epäsymmetrinen ja epätasapainossa [11]. Yksilöiden asemaa henkilötietomarkkinoilla voi vahvistaa tarjoamalla enemmän tietoa henkilötietojen rahallisesta merkityksestä [10]. Seuraavaksi tässä tutkielmassa käsitellään eri tapoja määritellä henkilötietojen arvoa.

4.1 Henkilötietojen hinnoittelu ja siihen vaikuttavat tekijät

Henkilötietojen arvolle ei ole yleismaailmallista, yleisesti hyväksyttyä mittaria [11]. Henkilötietodatan arvoa voi kuitenkin arvioida eri tavoilla ja eri näkökulmista. Arvioinnissa on otettava huomioon, että tietyn tyyppiset tiedot ovat rahallisesti arvokkaampia kuin toiset.

Financial Times (2013) arvioi, että henkilötiedoista esimerkiksi ikä, sukupuoli ja sijainti ovat arvoltaan noin 0,05 Yhdysvaltain senttiä henkilöä kohden. Vastaava hinta tuhannen henkilön kohdalla olisi noin 50 senttiä. Kaikista henkilökohtaisimmat tiedot ovat kuitenkin arvokkaimpia. Esimerkiksi terveystietoja myydään 26 sentin hintaan henkilöä kohden.[25] Lisäksi on arvioitu, että autoa tai lomamatkaa ostavien henkilöiden tiedot ovat arvoltaan 0,21 senttiä henkilöltä. Tietyt elämäntapahtumat, kuten vanhemmaksi tuleminen, muutto uuteen kotiin, kihlaukset tai ero, ovat myös arvokasta dataa. Raskaana olevien naisten tiedot ovat arvoltaan noin 11 senttiä.[10]

Carrascal ym. (2013) toteuttivat tutkimuksen, jossa 168 espanjalaista osanottajaa saivat kahden viikon ajan arvioida omien henkilötietojensa arvoa. Tutkimuk-

sen tuloksista käy ilmi, että käyttäjät pitivät ns. offline-tietoja eli Internetin ulkopuolisia tietoja arvokkaampina kuin online- eli verkkotietojaan. He arvioivat online-tiedoilleen, kuten selaushistorialleen arvoksi noin 7 euroa ja offline-tiedoilleen, kuten iälle, sukupuolelle, osoitteelle ja palkkatiedoille noin 25 euroa.[26]

Tosiasiassa yksittäisten henkilötietojen, kuten iän tai sukupuolen arvottaminen ei ole järkevää, sillä henkilötiedot ovat arvokkaita vain kokonaisuudessaan. Toisin sanoen ikä ja sukupuoli ovat yhdessä arvokkaampaa tietoa kuin erikseen. Voidaan sanoa, että henkilötietojen arvottaminen on itse asiassa digitaalisten identiteettien arvottamista. Tämä hinnoittelutapa ottaa huomioon myös sen, että usein yritykset ostavat suuria määriä datasettejä yksittäisten tietoattribuuttien sijaan.[10]

Koon lisäksi henkilötietojen hintaan vaikuttaa toisekseen, kuinka tarkkoja ja ajan tasalla ne ovat. Pääsääntöisesti tuoreempi data on arvokkaampaa kuin vanhempi. Kolmanneksi hinnoissa on eroa henkilötietokategorioiden välillä. Henkilökohtaiset ja arkaluonteiset tiedot ovat muita henkilötietoja arvokkaampia, sillä niiden julkaisemisessa ollaan varovaisempia. Näin ollen tarjonnan ollessa pientä, hinnatkin ovat korkeampia. Neljänneksi hintaan vaikuttaa, kuinka hyvin yksilö voidaan tiedoista tunnistaa. Anonyymi data on arvokasta, mutta yrityksille on kannattavaa käyttää yksilöihin yhdistettävissä olevia tietoja personoituun markkinointiin ja mainontaan.[10] Seuraavaksi tässä tutkielmassa siirrytään erilaisiin metodeihin, joilla henkilötietoja voidaan hinnoitella.

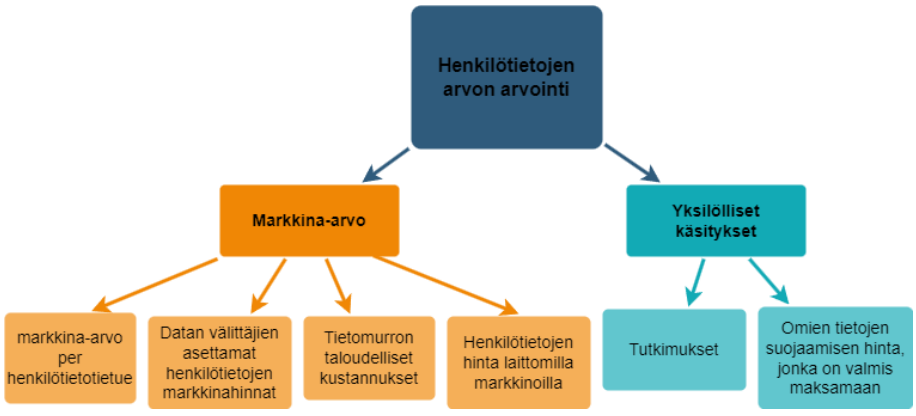
4.2 Hinnoittelumenetelmiä

Henkilötietodatalle voi määritellä arvoa eri tavoin. Henkilötietojen hinnoitteluun on ehdotettu ja kehitetty paljon erilaisia menetelmiä, joista tutkielma ottaa kantaa vain osaan. Tämä tutkielma perustuu Malgierin ja Custersin (2018) artikkelissaan esittämiin hinnoittelusysteemeihin, jotka kattavat OECD:n metodit, sekä ylhäältä alas ja alhaalta ylös -lähestymistavat.

Taloudellisen yhteistyön ja kehityksen järjestö eli **OECD** (Organisation for Economic Co-operation and Development) jakaa tutkimuksessaan (2011) henkilötietojen hinnan arviointimenetelmät kahteen luokkaan. Henkilötietojen hinnan voi perustella joko 1) markkina-arvon arvioilla tai 2) yksilöiden omilla arvioilla. Nämä tavat jakautuvat edelleen eri osa-alueisiin, joita tarkastellaan seuraavaksi tarkemmin.

Markkina-arvolla voidaan ensinnäkin viitata markkina-arvoon tai -tuloihin yhtä henkilötietotietuetta kohden. Toisekseen se voi tarkoittaa hintaa, jolla henkilötietojen keräämiseen erikoistuneet yritykset (engl. data brokers) myyvät saamiaan tietoja markkinoilla. Kolmanneksi markkina-arvon määrittelyä voi lähestyä taloudellisista kustannuksista, joita tietomurrot aiheuttavat yrityksille ja yksilöille. Viimeiseksi markkina-arvot voivat viitata henkilötietojen arvottamiseen laittomilla markkinoilla. Markkina-arvojen määrittelyn eri metodeja yhdistää se, että ne kuvastavat henkilötietojen todellista taloudellista arvoa ja että ne on osittain helppo tunnistaa.[27]

Yksilöiden omia arvioita datan hinnoittelusta saadaan selville tutkimusten kautta. On mahdollista vertailla eri tutkimusten tuloksia keskenään ja päätellä yleistä johtopäätöstä henkilötietojen taloudellisesta arvosta. Lisäksi henkilötietojen hintaa voi arvioida sen mukaan, minkälaisia summia yksilöt raportoivat olevansa valmiita maksamaan suojatakseen omia tietojaan.[27]



Kuva 4.1: Kaavio OECD:n henkilötietojen arvoa arvioivista menetelmistä. Mukaillen lähdettä [27].

OECD:n henkilötietojen hinnoittelun kaksi arviointimenetelmäluokkaa, markkina-arvot ja yksilölliset käsitykset, on mahdollista tulkita myös pystysuuntaisella mallilla. Tällaiseen malliin voidaan viitata termeillä **ylhäältä alas** (engl. top-down) ja **alhaalta ylös** (engl. bottom-up). Ylhäältä alas -lähestymistapa viittaa digitaalisen datan kysyntään, jossa henkilötietojen arvo on se hinta, jonka yritykset tavallisesti niistä maksavat. Alhaalta ylös -lähestymistapa viittaa vastaavasti datan tarjontaan, jossa henkilötietojen arvo määräytyy yksilöiden yksityisyyteen kohdistuneista vahingoista.[10]

Mitä tulee henkilötietojen hinnan arvioimiseen, ylhäältä alas -tapa lähtee liikkeelle palvelua tuottavan yrityksen kokonaistulojen ja palvelun käyttäjien määrästä. Esimerkiksi Facebookin kokonaistulot olivat lähes 18 miljardia dollaria vuonna 2015, ja palvelua käyttäviä oli samana vuonna noin 1,6 miljardia. Suurin osa Facebookin kokonaistuloista oli peräisin mainonnasta, joten yritys nettosi noin 10 dollaria yksilöön kohdistuvasta mainonnasta yhtä vuotta kohden. Tämä viittaisi siihen, että yhden henkilön henkilötiedot olisivat arvoltaan noin 10 dollaria.[10]

Alhaalta ylös -lähestymistapa määrittää yksittäisten henkilötietojen arvon, kun niitä käytetään mainontaan. Mainonnan kohdistaminen eli personointi yksilöille on yrityksille kannattavampaa kuin mainostus suurelle yleisölle. Laskelmien mukaan esimerkiksi jokainen Facebookin personoitu mainos on arvoltaan noin 5–10 senttiä. Kun otetaan huomioon, kuinka usein keskimääräinen käyttäjä näkee mainoksia, saadaan personoidulle mainokselle kuukausihinnaksi 30–60 senttiä. Tietojen uudelleenmyynnin sisällytettäessä laskelmiin voidaan henkilötietojen hinnaksi laskea korkeintaan 10 dollaria. Tämä on linjassa ylhäältä alas -tavan vastaaviin tuloksiin.[10]

4.3 Hinnoittelun ongelmat

Edellisessä osiossa 4.2 tarkasteltiin erilaisia hinnoittelumekanismeja henkilötietojen arvon arvioimiseen. Seuraavaksi jatketaan OECD:n eri arviointitapojen, sekä alhaalta ylös ja ylhäältä alas -metodien käsittelyä keskittyen niiden ongelmallisiin puoliin. Lisäksi mainitaan lyhyesti aiheeseen liittyvistä eettisistä ongelmista.

OECD:n markkina-arvoon perustuviin arviointimetodeihin on todettu edellä lukeutuvan 1) markkina-arvot yhtä henkilötietotietuetta kohden, 2) henkilötietodatan keräämiseen erikoistuneiden yritysten datasta pyytämä hinta, 3) tietomurtojen taloudelliset kustannukset ja 4) henkilötietodatan hinta laittomilla markkinoilla. Näiden metodien yhteiseksi ongelmaksi voidaan nimittää epätarkkuus, jolla henkilötietodatan arvoa arvioidaan. Seuraavaksi tarkastellaan yksityiskohtaisemmin kunkin OECD:n metodin haasteita epätarkkuudesta lähtien.

Henkilötietotietueisiin kohdistuvan markkina-arvon arviointi on epätarkkaa, sillä tosiasiassa yritysten tulot ovat peräisin henkilötietodatan lisäksi useilta muiltakin alueilta, kuten fyysisestä pääomasta ja asiantuntijuudesta. Tietuiden täsmällisen hinnan selvittämiseksi tulisi tietää, mikä osa yritysten liikevaihdosta on juuri henkilötietoihin liittyvää. Ongelmana on se, etteivät yritykset usein julkaise tällaisia tietoja ulkopuolisille. Metodissa on otettava huomioon, että yritysten taloudelliset tulokset ovat yleisesti riippuvaisia trendeistä, satunnaisilmiöistä ja spekulaatioista.[27]

Dataa keräävien yritysten henkilötiedoista pyytämä hinta ei ole täsmällinen, sillä todellisuudessa hintaan sisältyy myös datan haku- ja käsittelykustannukset. Lisäksi talousmarkkinoilla kaupattava henkilötietodata vaihtelee arvoltaan kontekstin mukaan. Kuten luvussa 4.1 käy ilmi, yksittäinen henkilötietoattribuutti, kuten puhelinnumero, ei ole yksinään yhtä arvokas kuin yhdistettynä muihin attribuutteihin. On myös todettava, että henkilötietodatan laatua kauppatavarana ei voi taata – data voi olla epäluotettavaa, epätarkkaa tai virheellistä, mikä voi vääristää sen arvoa.[27]

Tietomurroista aiheutuviin tappioihin perustuva henkilötietodatan arvon arviointi tarkoittaisi käytännössä sitä, että yhden henkilötietodatatietueen hinta muodostuisi sen varastamisen kustannuksista. Tässä on ongelmana se, ettei arvio oikeastaan viittaa datan arvoon, vaan vahinkojen markkina-arvoon. Metodin haasteena on sekin, ettei kustannusarvioihin sisälly yritysten mainehaitasta aiheutuvia kustannuksia.[27]

Kyberrikosten markkinoihin perustuva henkilötietojen arvotus on monin tavoin ongelmallista. Hintojen kerääminen on alun alkaen haastavaa laittomilta alustoilta. Laittomien kauppatavaroiden hinnoittelua pidetään salassa, jolloin tarkkoja arvioita on vaikeaa saada. Huomioon tulisi ottaa myös rikollisten saama hyöty datan jatkokäsittelystä, ja heidän ottamiensa riskien kokonaisvaikutukset datan arvoon.[27]

OECD:n yksilöiden omiin arvioihin perustuvista tutkimuksista voi saada selville, millä hinnalla yksilöt olisivat valmiita myymään omia tietojaan. Tutkimusten haasteena on niiden hypoteettisuus tai spekulatiivisuus, jota ei sellaisenaan voi soveltaa talousmarkkinoille. Lisäksi tutkimuksissa on havaittu, että tutkimuskysymysten asettelu ja niihin asetettu konteksti vaikuttaa merkittävästi henkilöiden antamiin vastauksiin. Esimerkiksi henkilötietojen myyntihintaan voi yksilöllisesti vaikuttaa se, mihin tarkoitukseen dataa tullaan käyttämään ja kuinka luottamuksellisesti sitä tullaan käsittelemään.[27]

Mitä tulee summiin, joita yksilöt olisivat valmiita maksamaan tietojensa suojaamiseksi, ovat nekin spekulatiivisia, eivätkä suoraan sovellu markkinoille. Samoin kuin tietomurroista aiheutuvien kustannusten kohdalla, tämäkin metodi arvioi oikeastaan vahingoista aiheutuvia kustannuksia, ei datan arvoa itsessään. On lisäksi kyseenalaistettava, kuinka hyvin tietoja suojaavat vakuutukset tai sovellukset toimivat tehtävässään.[27]

Ylhäältä alas ja **alhaalta ylös** -lähestymistavat eivät ota huomioon hintaa, jolla yksilöt paljastaisivat omia tietojaan. Näitä metodeja voi kuitenkin käyttää tilan-

teissa, jossa henkilötietodatan arvo määrätään ns. käänteisen vastuun (engl. reverse liability) avulla. Silloin lasketaan korvaus, jonka mahdollinen loukkaaja maksaa etukäteen yksilölle saadakseen luvan suorittaa todennäköisesti haitallista toimintaa, kuten henkilötietojen käsittelyä.[10]

Edellä käsitellyistä hinnoittelumetodeista voi todeta, että kukin on vain arvio henkilötietodatan arvosta. Tosiasiassa henkilötietojen todellista hintaa on lähes mahdotonta yksimielisesti määrittää. Ongelmia ilmenee sen mukaan, kuka todellisuudessa toteuttaa hinnoittelun ja miten markkinoita tai henkilötietojen käsittelyä valvotaan. Haasteita hinnoitteluun lisää se, että osa henkilötietodatasta voi olla henkilön itsensä julkaisemaa jollain sosiaalisen median alustalla. Tällainen data voi olla hallitsemattomasti leviävää, jolloin sille on hankalaa asettaa pysyvää hintaa.[10]

Eettisiin ongelmiin liittyy se, onko yksilön henkilötietoja oikeastaan mahdollista täysin erottaa yksilön entiteetistä [8]. On ongelmallista rinnastaa yksityisyyttä ja muita ihmisarvoja kauppatavaroihin [10], joita muut ihmiset voivat ostaa, myydä ja omistaa [5]. Voidaan myös väittää, että yhteiskunnassa eriarvoisuus lisääntyy, kun toisten henkilöiden data on arvokkaampaa kuin toisten. Tällaisessa ns. hintasyrjinnässä köyhempien ihmisten henkilötiedot eivät ole yhtä arvokkaita kuin varakkaiden ihmisten.[10]

5 Pohdinta

Tutkielman luvussa 3 todettiin, että kuluttajien henkilötietojen riskien tuntemuksella on vaikutusta sosiaalisen median käyttöön ja täten myös digitaalisten jalanjälkien määrään. Intuitiivisesti väitteen voisi ajatella olevan päinvastoin: kuluttajat saattavat olla tietoisia riskeistä, mutta eivät vain pidä niitä varteenotettavina. Tätä vastaväitettä voi tukea kunkin omat kokemukset sosiaalisesta mediasta, jonne julkaistaan nimiä, terveystietoja, kuvia itsestä, perheestä, asuinalueista ja lomamatkoista sekä tietoja harrastustoiminnasta. Tällaisten tietojen julkaisemista todennäköisesti harkittaisiin enemmän, jos yksityisyyttä koskevat riskit otettaisiin todesta.

Luvussa 4 käsiteltyjen henkilötietojen hinnoittelutapojen toivoisi lisäävän kuluttajien motivaatiota omien, arvokkaiden tietojensa suojaamiseen. Henkilötietojen arvoa tulisi tuoda esille enemmän ja kampanjoida yksilöiden oikeuksien puolesta. Vaikka tällaisia isoja kampanjoita tai uutisia ei olla vielä nähty, on tietosuojalakien ja -direktiivien saralla onneksi nähty edistystä, kuten luvussa 3 esitellyn GDPR:n osalta. Näissä laeissa myönnetään, että henkilötiedoilla on todellista arvoa ja siksi niitä tulisi suojella.

Tietosuojalakien tuoma panostus yksityisyyden suojaamiseen ja epäsuorasti henkilötietojen arvosta tiedottamiseen ei kuitenkaan riitä. Tämän pohdintaosion alussa mainitun vastaväitteen perusteella voi tulla siihen johtopäätökseen, ettei tavallisilla kansalaisilla ole riittävästi lakitietoisuutta ymmärtääkseen henkilötietoihin liittyvän kaupallistamisen vakavuutta. Johtopäätös voi olla myös se, että vaikka lakeja ymmärtäisi, voi niiden tuottamaan turvallisuuden tunteeseen helposti tuudittautua ja sivuuttaa omat velvollisuutensa yksityisyytensä suojelemiseen.

Tietosuojalakien riittämättömyyden vuoksi tässäkin tutkielmassa kannatetaan ajatusta kuluttajien hintatietoisuuden lisäämisestä henkilötiedoistaan. Rahan voi intuitiivisesti ajatella olevan yksi isoimmista motivaattoreista tavallisten kansalaisten jokapäiväisessä elämässä. Jos saataisiin yleiseen tietoon, kuinka paljon yritykset saavat tulosta keräämillään henkilötiedoilla, voisivat yksittäiset henkilöt huomata henkilötietojensa arvon tiedotuksen epäsymmetrian. Keinoja symmetrisempään tiedottamiseen tulisi tutkia jatkotutkimuksissa. Joitakin keinoja on nyt jo tuotu esille, esimerkiksi Malgierin ja Custersin (2018) artikkelissa ehdotetaan, että tietosuojalakiin lisättäisiin kohta, jonka mukaan ihmisillä olisi oikeus vastaanottaa dataa käsitteleviltä toimijoilta tietoa omien henkilötietojensa rahallisesta arvosta. Tämä voisi olla yksi varteenotettava ratkaisu hintatietoisuuden lisäämiseen ja täten yksityisyyden suojan lisäämiseen.

Edellä mainitusta ehdotuksesta voi kuitenkin tulla mieleen ainakin yksi selkeä ongelma. Jos henkilötietojen hinta kerrotaan kuluttajalle, se tehdään luultavasti evästeissä. Yksinkertaisesti, kun kuluttaja käy verkkosivuilla, saa hän ilmoituksen, jossa pyydetään kieltämään tai hyväksymään evästeet ja jossa lisätietoja evästeistä tarjotaan esimerkiksi erillisen linkin kautta. Hinnoittelutiedot voivat sisältyä linkin takana sijaitseviin lisätietoihin. Ongelma muodostuu siitä, kuinka moni kuluttaja tosi asiassa on valmis käyttämään aikaa ja vaivaa lisätietojen lukemiseen. Intuitiivisesti, useimmat luultavasti tyytyvät suoraan hyväksymään evästeet niitä sen kummemmin tutkimatta.

Ongelmaksi ehdotukselle ja koko hintatietoisuuden lisäämiselle muodostuu myös se, miten henkilötiedot loppujen lopuksi pystytään hinnoittelemaan. Kuten luvussa 4 todettiin, henkilötietojen todellinen hinnoittelu ei ole mitenkään yksinkertaista ja siihen sisältyy monenlaisia ongelmia. Tässä tutkielmassa nyt ehdotetaankin, että alustavilla ja epätarkoilla hinta-arvioillakin voi olla merkitystä kuluttajien hintatietoisuuden lisäämiseen ja epäsymmetrian kaventamiseen.

6 Yhteenveto

Nykymaailman talous nojaa yhä enenevissä määrin dataan – niin paljon, että aikakauttamme voi kutsua big datan vallankumouksen ajaksi. Tätä valtavaa määrää dataa, big dataa, syntyy jatkuvasti ja sitä pystytään keräämään myös aiempaa tehokkaammin. Dataa, mukaan lukien henkilötietodataa, hyödynnetään talousmaailmassa yritysten kilpailuaseena ja apuna päätöstentekoon.

Datan kaupallistaminen on johtanut siihen, että henkilötietojen arvosta tietoiset käyvät niillä kauppaa, kun taas henkilötietojen lähteet, kuluttajat, on suljettu markkinoiden ulkopuolelle. Kuluttajilla ei ole samassa suhteessa informaatiota henkilötietojen hinnoittelusta ja arvosta kuin niillä, jotka osallistuvat datamarkkinoille myymään ja ostamaan dataa. Kuluttajat ovat jo pitkään, todennäköisesti tietämättään, maksaneet ilmaisiksi ajatelluista verkkopalveluista henkilötiedoillaan. Moni sosiaalisen median tai muiden digitaalisten palveluiden käyttäjä ei ole lainkaan tietoinen, millaisia digitaalisia jalanjälkiä jättää jälkeensä käyttämillään alustoilla.

Yksilöiden yksityisyyden suojan vahvistamiseksi on niin Euroopassa kuin muuallakin maailmassa yksittäiset valtiot luoneet tietosuojalakeja. Tässä tutkielmassa on erityisesti nostettu esille Euroopan unionin tietosuojalaki eli GDPR, joka määrittelee henkilötiedot kattavasti lain silmissä ja asettaa niiden käsittelyyn rajoituksia. Tämä laki määrää, että yksilöllä on oikeus saada tietää, mihin tarkoitukseen tai uudelleenkäyttöön henkilötietoja kerätään, oikaista virheelliset tiedot ja poistattaa tiedot eli tulla unohdetuksi. Sen lisäksi datan minimointi ja pseudonymisointi ovat käsittelyitä, joita yksilöt ovat lain mukaan oikeutettuja vaatimaan.

Tietosuojalait eivät yksinään riitä suojaamaan yksilöiden henkilötietoja, sillä jokainen on myös itse vastuussa omien tietojensa salassapidosta. Henkilötietojen arvotus on yksi keino kannustaa yksilöitä tietojensa suojaamiseen. Henkilötietodatan arvoa ei kuitenkaan ole helppoa määritellä. Kuten luvussa 4 käy ilmi, hinnoittelutapoja on monia, joista suurin osa ulottuu tämän tutkielman ulkopuolelle. Kaikilla metodeilla on omat heikkoutensa ja vahvuutensa, mukaan lukien tutkielmassa esitellyillä OECD:n menetelmillä sekä ylhäältä alas ja alhaalta ylös -lähestymistavoilla. Lisäksi itse henkilötietodata vaikuttaa siitä arvioitavaan hintaan niin tyyppinsä, kokonsa, ajantasaisuutensa kuin tunnistettavuutensa osalta.

Jos kuluttajat tietäisivät henkilötietojensa hinnan ja millaista kauppaa tiedoilla käydään, he saattaisivat pitää omista tiedoistaan parempaa huolta. Tätä varten tulisi tehdä jatkotutkimuksia siitä, miten lisätä kuluttajien tietoisuutta henkilötietodatansa arvosta. Yhdeksi keinoksi tässä tutkielmassa mainitaan ehdotus, jossa tietosuojalakiin lisättäisiin yksilöille oikeus saada tietää omien tietojensa rahallinen arvo. Ehdotus ei ole täysin ongelmaton, mutta varteenotettava yritys tasapainottaa hinnoittelutietoisuutta kuluttajien ja datalla kauppaa käyvien välillä.

Tässä tutkielmassa on esitetty, että big datana kerätyillä henkilötiedoilla on rahallista arvoa ja että kerättyä dataa voidaan käyttää valuuttana digitaalisessa maailmassa. Tutkielmassa käytetyn aineiston perusteella henkilötietodatan todellista arvoa on tämänhetkisillä menetelmillä lähes mahdotonta arvioida tarkasti ja monipuolisesti eri tekijät huomioon ottaen. Aiheen laajempi tarkastelu voi olla paikallaan muiden menetelmien ja erityisesti uudempien metodien kattavuuden arvioimiseksi. Tutkielman tietojen pohjalta on kuitenkin todettava, että epätarkoillakin hinta-arvioilla voi olla mahdollista suuntaa-antavasti tiedottaa kuluttajia henkilötietojen arvosta ja kannustaa niiden huolellisempaan suojaamiseen. Tietosuojalakien kaltaisten passiivisten puolustuskeinojen lisäksi tulisi kannustaa yksilöitä panostamaan aktiivisesti omien henkilötietojen suojelemiseen.

Lähdeluettelo

[1] W. Pietsch, *Big data* (Cambridge elements. Elements in the philosophy of science). Cambridge: Cambridge University Press, 2021. DOI: 10.1017/9781108588676.

[2] N. Eagle ja K. Greene, *Reality mining: using big data to engineer a better world*. Cambridge, Massachusetts: MIT Press, 2014.

[3] P. C. K. Hung, *Big data applications and use cases* (International series on computer entertainment and media technology). Switzerland: Springer, 2016. DOI: 10.1007/978-3-319-30146-4.

[4] Statista ja IDC, *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025 (in zettabytes)*, Haettu 12.3.2024, kesäkuu 2021. url: https://www.statista.com/statistics/871513/worldwide-data-created/.

[5] M. Tufiş ja L. Boratto, "Toward a Complete Data Valuation Process. Challenges of Personal Data", *J. Data and Information Quality*, vol. 13, nro 4, elokuu 2021. DOI: 10.1145/3447269.

[6] C. F. Libaque-Sáenz, S. F. Wong, Y. Chang ja E. R. Bravo, "The effect of Fair information practices and data collection methods on privacy-related behaviors: A study of Mobile apps", *Information Management*, vol. 58, nro 1, s. 103–284, 2021. DOI: 10.1016/j.im.2020.103284.

[7] S. S. Muhammad, B. L. Dey ja V. Weerakkody, "Analysis of Factors that Influence Customers' Willingness to Leave Big Data Digital Footprints on Social Media: A Systematic Review of Literature", *Information Systems Frontiers*, vol. 20, nro 3, s. 559–576, 2018. DOI: 10.1007/s10796-017-9802-y.

[8] S. Spiekermann, A. Acquisti, R. Böhme ja K.-l. Hui, "The challenges of personal data markets and privacy", *Electronic Markets*, vol. 25, nro 2, s. 161–167, 2015. DOI: 10.1007/s12525-015-0191-0.

[9] Z. Zhang, W. Song ja Y. Shen, "A Reasonable Data Pricing Mechanism for Personal Data Transactions with Privacy Concern", teoksessa *Web and Big Data*, L. H. U, M. Spaniol, Y. Sakurai ja J. Chen, toim., sarja Lecture Notes in Computer Science, Cham: Springer International Publishing, 2021, s. 64–71. DOI: 10.1007/978-3-030-85899-5_5.

[10] G. Malgieri ja B. Custers, "Pricing privacy – the right to know the value of your personal data", *Computer Law Security Review*, vol. 34, nro 2, s. 289–303, huhtikuu 2018. DOI: 10.1016/j.clsr.2017.08.006.

[11] A. A. Becerril, "The value of our personal data in the Big Data and the Internet of all Things Era", *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 7, nro 2, s. 71–80, 2018. DOI: 10.14201/ADCAIJ2018727180.

[12] B. Franks ja T. H. Davenport, *Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*, B. Franks, toim. Newark, United States: John Wiley Sons, Incorporated, 2012. DOI: 10.1002/9781119204275.

[13] Merriam-Webster, *Big data*, Haettu 2.2.2024. url: https://www.merriam-webster.com/dictionary/big%20data.

[14] D. Laney et al., ”3D data management: Controlling data volume, velocity and variety”, *META group research note*, vol. 6, nro 70, s. 1, 2001, Haettu 2.2.2024 osoitteesta https://community.aiim.org/blogs/doug-laney/2012/08/25/deja-vvvu-gartners-original-volume-velocity-variety-definition-of-big-data.

[15] R. Guay ja K. Birch, ”A comparative analysis of data governance: Socio-technical imaginaries of digital personal data in the USA and EU (2008–2016)”, *Big Data Society*, vol. 9, nro 2, 2022. DOI: 10.1177/20539517221112925.

[16] P. M. Leonardi, ”COVID-19 and the New Technologies of Organizing: Digital Exhaust, Digital Footprints, and Artificial Intelligence in the Wake of Remote Work”, *Journal of Management Studies*, vol. 58, nro 1, s. 249–253, 2021. DOI: 10.1111/joms.12648.

[17] S. S. e. Zainab ja T. Kechadi, ”Sensitive and Private Data Analysis: A Systematic Review”, teoksessa *Proceedings of the 3rd International Conference on Future Networks and Distributed Systems*, sarja ICFNDS ’19, New York, NY, USA: Association for Computing Machinery, heinäkuu 2019, s. 1–11. DOI: 10.1145/3341325.3342002.

[18] Tietosuojavaltuutetun-toimisto, *Mikä on henkilötieto?*, Haettu 21.2.2024. url: https://tietosuoja.fi/mika-on-henkilotieto.

[19] SurfShark, *Share of data points collected and used to track iOS apps users worldwide as of May 2023, by category*, Haettu 22.2.2024, joulukuu 2023. url: https://www.statista.com/statistics/1440804/collection-and-tracking-ios-apps-worldwide/.

[20] IAPP, *How would you rate your current level of GDPR compliance?*, Haettu 22.2.2024, lokakuu 2019. url: https://www.statista.com/statistics/1172852/gdpr-compliance-among-eu-and-us-firms/.

[21] Y. Jiang ja T. Syn, ”Online Privacy Policy Disclosure: An Empirical Investigation”, *Journal of Computer Information Systems*, vol. 63, nro 3, s. 663–680, 2023. DOI: 10.1080/08874417.2022.2095542.

[22] Euroopan unionin neuvosto, *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) (ETA_2016)*. huhtikuu 2016, vol. 119, Legislative Body: EP, CONSIL. url: http://data.europa.eu/eli/reg/2016/679/oj/fin.

[23] M. Armstrong, *EU Data Protection Fines Hit Record High in 2023*, Haettu 22.2.2024, tammikuu 2024. url: https://www.statista.com/chart/30053/gdpr-data-protection-fines-timeline.

[24] Euroopan unionin neuvosto, *Euroopan parlamentin ja neuvoston direktiivi (EU) 2019/770, annettu 20 päivänä toukokuuta 2019, tietyistä digitaalisen sisällön ja digitaalisten palvelujen toimittamista koskeviin sopimuksiin liittyvistä seikoista (ETA_2019)*. toukokuu 2019, vol. 136. url: http://data.europa.eu/eli/dir/2019/770/oj/fin.

[25] E. Steel, C. Locke, C. Emily ja B. Freese, *How much is your personal data worth?*, Haettu 22.2.2024, kesäkuu 2013. url: https://ig.ft.com/how-much-is-your-personal-data-worth/.

[26] J. P. Carrascal, C. Riederer, V. Erramilli, M. Cherubini ja R. de Oliveira, ”Your browsing behavior for a big mac: economics of personal information online”, teoksessa *Proceedings of the 22nd international conference on World Wide Web*, sarja WWW ’13, New York, NY, USA: Association for Computing Machinery, toukokuu 2013, s. 189–200. DOI: 10.1145/2488388.2488406.

[27] OECD, *Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value*, huhtikuu 2013. DOI: 10.1787/5k486qtxldmq-en.