# Suurandmed ja privaatsus

Juhendmaterjal organisatsioonidele

Juhend kehtestatakse isikuandmete kaitse seaduse § 33 lõike 1 punkti 5 alusel

Kinnitatud 19.01.2017

# Sisukord

Sissejuhatus ..... 3
1. Mis on suurandmed? ..... 5
   1.1. Maht ..... 5
   1.2. Andmevormingute paljusus ..... 5
   1.3. Töötlemise kiirus ..... 5
2. Algoritmid ..... 6
3. Andmete otstarbe muutumine ..... 6
4. Suurandmed ja isikuandmed ..... 6
5. Isikuandmete töötlemise alused ning nõuded ..... 7
   5.1. Nõusolek ..... 7
   5.2. Lepingud ja seaduslik alus ..... 8
   5.3. Andmete minimaalsus ..... 9
   5.4. Õigus saada teavet ..... 9
6. Infoturve ..... 9
7. Andmetöötluses osalejad ..... 10
8. Kokkuvõtteks ..... 10

# Sissejuhatus

Suurandmete (*ingl k. Big Data*) teema on päevakohane paljudes eluvaldkondades. Näiteks juhtimises, turunduses, teadusuuringutes, riigikaitses. Aga samuti avaliku sektori läbipaistvuse ja avaandmete (*ingl. k. OpenData*) ehk avaliku teabe kättesaadavusel. Nii avalik- kui erasektor kasutavad üha suuremal määral suurandmete analüüsi.

Paljudel juhtudel ei hõlma suurandmete analüüs üldse mingeid isikuandmeid. Näiteks kliima- ja ilmastikuandmete põhjal saab teha uusi avastusi ja parendada teenuseid ilma isikuandmeid kasutamata. Samas on mitmeid näiteid suurandmete analüüsist, kus töödeldakse ka isikuandmeid. Need saadakse sellistest allikatest nagu sotsiaalmeedia, kliendikaardid või tervisealased uuringud. Kui kasutatakse ehk töödeldakse isikuandmeid, siis peavad kõik osapooled tagama neile isikuandmete kaitse seadusest (edaspidi IKS) tulenevate kohustuste järgimise.

Üks põhiline andmekaitse nõue on tagada, et isikuandmete töötlemine on eesmärgikohane.[1] Eriti tähtis on see suurandmete kasutamisel üksikisikuid mõjutavate otsuste tegemisel. Näiteks puudutab see seda, kuidas isikuandmed saadakse.[2] Organisatsioonid[3] peavad andmete kogumise läbipaistvalt korraldama ja nende kasutamise selgitamine on tähtis osa andmekaitse põhimõtete järgimisel. Suurandmete analüüsimise keerukus ei ole ettekääne nõutava nõusoleku küsimata jätmiseks.

Suurandmete analüüs võib hõlmata ka isikuandmete kasutamist algselt märgitust erineval eesmärgil. Kui organisatsioon on kogunud isikuandmeid üheks otstarbeks ja siis otsustab analüüsida neid hoopis teiseks otstarbeks (või teha need kolmandatele osapooltele kättesaadavaks), peab ta oma teenuste kasutajaid sellest teavitama ning võtma neilt nõusolekud.[4] See on eriti tähtis, kui organisatsioon kavatseb kasutada andmeid eesmärgil, mis ei ole üksikisikule ilmne, sest puudub otsene seos temapoolse teenusekasutamisega.

Suurandmete üheks põhiliseks eripäraks on see, et andmetöötlejad kasutavad kõiki andmeid korraga. Mis on aga hoopis vastupidine andmete minimaalsuse põhimõttega.[5] See tõstatab küsimuse, kas suurandmete kasutamine tähendab ülemäärast isikuandmete kogumist ning kas kasutatavad isikuandmed on tegelikult asjakohased? Organisatsioonid saavad sellele vastata, teatades isikutele algusest peale selgelt, mida nad kavatsevad nende isikuandmete töötlemise teel teada saada või ära teha. Samuti peavad isikuandmete töötlejad veenduma, et nende eesmärgi mõistes on kõik andmed asjakohased ja isikuandmete kogumine ei ole ülemäärane.

Üks võimalik viis on andmete anonüümseks muutmine. Kui seda tehakse õigesti, ei loeta analüüsitavaid andmeid enam isikuandmeteks. See aitab suurandmeid analüüsida ning organisatsioonidel uuringuid teostada või tooteid ja teenuseid välja töötada. Samuti võimaldab see organisatsioonidel anda inimestele, kelle andmed koguti, kindlustunde, et organisatsioon ei kasuta selliseid andmeid, mille kaudu saaks neid suurandmete hulgast tuvastada.

Tänapäeva maailmas, kus kasutatakse mitmeid andmeallikaid koos, võib tõhus anonüümimine olla keeruline. Seega peaksid kõik organisatsioonid juba enne andmetöötluse algust hindama suurandmete töötlemise mõju isikuandmete kaitsele. Läbi tuleb viia andmekaitsealane mõjuhinnang.

Andmekaitse Inspektsiooni eesmärk on juhtida koostatud juhendiga tähelepanu eelkõige suurandmete kasutamisel tekkivatele andmekaitse- ja privaatsusriskidele. Seda nii organisatsioonidele, üksikisikutele kui ühiskonnale tervikuna. Nõustuda ei saa väitega, et andmekaitse põhimõtetel ei ole suurandmete kontekstis kohta. Oleme seda meelt, et andmekaitse nõuetele on pigem omane teatav paindlikkus. Neid ei tohiks näha edasimineku takistusena, vaid raamistikuna, mis aitab edendada privaatsusõigusi ja innustada avalikkuse teavitamiseks ja

[1] IKS § 6 p 2.
[2] IKS § 6 p 1 kohaselt võib isikuandmeid koguda vaid ausal ja seaduslikul teel.
[3] Avaliku sektori asutused, äriühingud, mittetulundusühingud, sihtasutused, avalik-õiguslik juriidilised isikud.
[4] IKS § 6 p 4.
[5] IKS § 6 p 3.

kaasamiseks uuenduslike lähenemiste väljatöötamist.

Juhendis selgitame suurandmete mõistet, kaalume, millised andmekaitse probleemid sellega kaasnevad ja teeme ettepanekuid, kuidas tagada andmekaitsenormide järgimine. Suurandmetest tulenevate kasude eest ei saa maksta lihtsalt privaatsusõigustest loobumisega.

Abimaterjalina oleme kasutanud Inglismaa andmekaitseasutuse vastavat arvamust.

# 1. Mis on suurandmed?

Keeruline on anda ühest definitsiooni, mis võimaldaks mingi konkreetse andmetöötluse kindlalt suurandmeteks või mitte suurandmeteks lugeda. Suurandmete kohta on öeldud, et see on pigem nähtus, mitte tehnoloogia. Kõige parem on ehk võtta seda kui lühikest sõna, mis kirjeldab teatud andmete omadusi ja nende töötlemise viisi. Suurandmeid võib näiteks sõnastada kui *suure mahuga, suure kiirusega ja suure varieeruvusega andmeid, mis nõuavad kulutõhusaid ja innovatiivseid andmetöötluse vorme, et saavutada parem ülevaade ja paremad otsused.* [6] Suurandmeid kirjeldatakse sageli kolme V-na: ***Volume*** (maht), ***Variety*** (andmevormingute paljusus) ja ***Velocity*** (andmete tekkimise ja töötlemise kiirus).

## 1.1.Maht

Suurandmetöötluses kasutatakse mahukaid andmete kogumeid. Näiteks meta-andmed internetiotsingutest, krediit- ja deebetkaardiostudest, sotsiaalmeedia postitustest, mobiiltelefonide asukohaandmetest või autode, targa kodu ja muude seadmete sensoritest.[7] Suurandmete tekkele ja levikule annavad hoogu erinevate elustiiliseadmete ja aktiivsusmonitoride kasutuselevõtt. Maailmas toodetavate andmete hulk kasvab pidevalt ja kiiresti. Tänu andmekandjate maksumuse vähenemise ja pilvepõhiste teenuste kättesaadavusele on võimalik talletada aina suuremaid andmemahtusid.

Ka avalikust sektorist on saamas üks paljudest suurandmete allikatest. Näiteks on riigi- ja omavalitsusasutustel seadusest[8] tulenev kohustus teha nende valduses olev teave taaskasutamise eesmärgil avalikkusele masinloetaval kujul kättesaadavaks. Samas peab teabe avalikustamisel olema tagatud eraelu puutumatus, autoriõiguste kaitse, riigi julgeoleku kaitse ning äri- ja muu juurdepääsupiiranguga teabe kaitse.[9]

## 1.2.Andmevormingute paljusus

Sageli koguvad suurandmeid töötlevad organisatsioonid andmeid mitmest erinevast allikast. Korrastatud andmed võivad pärineda näiteks riigi- ja omavalitsusasutuste andmekogudest. [10] Avaandmetena teabe taaskasutusse andmisel järgivad asutused nõuetekohast vormingut.[11]

Kuid tegemist võib olla ka korrastamata või struktureerimata andmetega. Näiteks on võimalik kokku koguda kõik andmed, mis tulevad mõnest sotsiaalmeedia allikast, nagu näiteks Twitter, Facebook. Sageli kasutatakse seda nn „arvamuste analüüsiks“, st selleks, et analüüsida, mida inimesed mingi toote või organisatsiooni kohta arvavad. Jaekaupmees saab kombineerida need andmed oma müügiterminalidest ja kliendikaartidest saadud andmetega. Tulemuseks on rikkalikud ja üksikasjalikud andmed, mida turunduseks kasutada.

Kui ettevõte analüüsib omaenda kliendibaasi, siis isegi kui see andmebaas on väga suur, ei pruugi sellest tuleneda analüüsi ega andmekaitse osas midagi uut võrreldes tänaste kohustustega. Kui see ettevõte aga kombineerib oma teabe väljastpoolt saadud andmetega (olgu siis avalikkusele kättesaadavast allikast või mitte), siis teeb ta midagi kvalitatiivselt teistsugust, mida saab lugeda suurandmeteks.

## 1.3.Töötlemise kiirus

Mõnel juhul on tähtis analüüsida andmeid nii kiiresti kui võimalik, isegi reaalajas. Suurandmete analüüsi meetodeid saab kasutada andmete analüüsimiseks „liikumise pealt“, st sedamööda, kuidas neid luuakse või salvestatakse, või ka „paigal seisvate“ andmete, st andmebaasides olevate andmete

[6] http://www.gartner.com/it-glossary/big-data

[7] Sensor on tehniline seadis, mis reageerib mingile füüsilisele või keemilisele välistoimele ja muundab selle elektrisignaaliks, mida on hõlbus mõõta ja andmetöötluseks rakendada.

[8] Avaliku teabe seadus (AvTS) § $3^1$.

[9] AvTS § $3^1$ lg 3.

[10] AvTS § $43^1$.

[11] Avaandmete roheline raamat.

analüüsimiseks.

Näiteks võimaldavad riigi käsutuses olevad andmed riigil teha operatiivseid majandus-, haridus-, või sotsiaalpoliitilisi otsuseid. Sideettevõtted analüüsivad igapäevaselt klientide helistamise- ja andmeside harjumusi, et koostada optimaalsemaid teenuste pakette.

## 2. Algoritmid

Suurandmete analüüs tähendab sageli andmetele väga suure hulga algoritmide rakendamist, et leida seoseid, mitte testida mingit kindlat hüpoteesi. Kui asjakohased seosed on leitud, siis saab luua uue algoritmi, et seda konkreetsel juhul rakendada. See on masinõppimise vorm, sest süsteem „õpib" andmete analüüsimise käigus, millised on asjakohased kriteeriumid. Kuigi algoritmid ei ole iseenesest midagi uut, on nende sellisel viisil kasutamine omane just suurandmete analüüsile.

Suurandmete analüüsis seega kogutakse ja analüüsitakse pigem kõiki kättesaadavaid andmeid. Näiteks jaemüügis võib see tähendada kõigi kliendikaardiga tehtud ostude analüüsimist ja nende hulgas seoste leidmist. Selle asemel et paluda teatud hulgal ostjatel täita küsitlusankeete.

## 3. Andmete otstarbe muutumine

Suurandmete analüüsis juhtub sageli, et ühel kindlal eesmärgil kogutud andmeid kasutatakse ka teiseks otstarbeks ja mõningatel juhtudel antakse need kasutada mõnele teisele organisatsioonile. Andmed kogutakse näiteks sotsiaalmeedia portaalidest ning tehakse turu-uuringuteks või muuks otstarbeks kättesaadavaks. Sotsiaalmeedia andmeid saab kasutada näiteks isiku krediidireitingu hindamiseks, kui isik on sellest teadlik ning oma nõusoleku andnud.

Paljudel juhtudel loovad erinevad seadmed või infosüsteemid analüüsiks kasutatavad andmed automaatselt. Andmeid ei anna isikud teadlikult. Näiteks tänavale või poodidesse paigaldatud sensorid suudavad tuvastada mööduvate inimeste nutiseadmete unikaalse MAC aadressi (ingl. k. *Media Access Control Address*). Kuigi MAC aadress ei tuvasta iseenesest veel isikut, saab seda kasutada korduvate külastuste jälgimiseks.

## 4. Suurandmed ja isikuandmed

Isikuandmed on andmed, mis viitavad tuvastatud või tuvastatavale isikule. Isikuandmete mõiste on kirjas seaduses.[12] Tuvastatav tähendab, et isikut on võimalik nende andmete põhjal kindlaks teha. Kas eraldi või koos muude andmetega. Hindamaks, kas neid andmeid saab kombineerida teiste andmetega, mis võimaldaksid isikut tuvastada, tuleb kaaluda, milliseid meetmeid isiku tuvastamiseks mõistliku tõenäosusega kasutataks.

Andmekaitse tegeleb eelkõige isikuandmetega. Kuid on tähtis meeles pidada, et paljudel juhtudel ei hõlma suurandmete analüüs üldse isikuandmeid. Ilma isikuandmeteta suurandmeteks on näiteks ülemaailmsed kliima- ja ilmastikuandmed või GPS [13] -seadmetega varustatud busside asukohaandmete kasutamine busside saabumisaegade ennustamiseks. Samuti suurte raadioteleskoopide andmed või laevakonteinerite sensoritelt saadud andmed. Need kõik on sellised valdkonnad, kus suurandmete analüüs võimaldab teha uusi avastusi ning teenuseid ja äriprotsesse täiustada, ilma isikuandmeid kasutamata.

Samas on palju näiteid suurandmete analüüsist, kus on tegemist isikuandmete töötlemisega. Näiteks terviseuuringutes [14] osalevate patsientide terviseandmed, mobiiltelefonide kasutajate asukohaandmed, kliendikaartidega tehtud ostud. Suurandmete analüüsist saadud leide saab

[12] IKS § 4 lg 1.
[13] Ingl k. *Global Positioning System* ehk üleilmne asukoha määramise süsteem.
[14] Vt AKI juhendit isikuandmete töötlemisest teadusuuringus.

rakendada üksikisikule suunatud turunduses (nt „nende omadustega inimesed soovivad enamasti sel ajal neid tooteid või teenuseid“).
Suurandmete analüüsil on potentsiaal ka uusi isikuandmeid luua. Näiteks sotsiaalmeedia andmeid analüüsides saab luua profiili inimese elustiilist. See aga omakorda võib mõjutada tema krediidireitingut või annab aimu sellest, kas inimesel võib tulevikus tekkida terviseprobleem. Sarnaselt annavad autode sensorid hulga andmeid auto kohta, mida saab aga kasutada ka autot juhtiva inimese juhtimisstiili seaduspärade tuvastamiseks. Nii saab teha otsuseid näiteks tervise- ja liikluskindlustusmaksete kohta.

Kui isikuandmed on täielikult anonüümseks muudetud, siis ei ole need enam isikuandmed. St andmete alusel ega nende kombineerimisel teiste andmetega, võttes arvesse kõiki viise, mida võidaks mõistlikult inimeste tuvastamiseks kasutada, ei ole võimalik üksikisikuid tuvastada.
Näiteks kasutatakse linnaruumi planeerimisel koostöös mobiilsideettevõtetega anonüümseid mobiiltelefonide asukohaandmeid.

Anonüümitud andmeid kasutavad organisatsioonid peavad siiski suutma tõendada, et nad on seoses isikute tuvastamise riskiga läbi viinud usaldusväärse andmekaitsealase mõjuhinnangu ja võtnud kasutusele riskiga proportsionaalsed lahendused. See peab hõlmama valiku organisatoorsetest, füüsilistest ja infotehnilistest turvameetmetest.[15]

Anonüümimist ei tohiks näha pelgalt halduskoormuse vähendamise viisina, nihutamaks andmetöötluse andmekaitsenõuete mõjualast välja. See on pigem viis, kuidas vältida isikuandmete kogemata avaldamise või kaotsimineku riski ning on sellisena suurandmete analüüsis abiks ja aitab organisatsioonil teostada uuringuid ning arendada tooteid ja teenuseid. Samuti võimaldab see organisatsioonil kinnitada inimestele, kelle andmeid kogutakse, et organisatsioon ei kasuta oma suurandmete analüüsiks andmeid, mis võimaldaksid inimesi tuvastada.

Erinevate anonüümimistehnikate kohta saab lugeda Euroopa andmekaitseasutuste ühisarvamusest.[16]

# 5. Isikuandmete töötlemise alused ning nõuded

Isikuandmete kaitse põhimõtete kohaselt on isikuandmeid lubatud töödelda kas seaduse-, nõusoleku alusel või inimesega sõlmitud lepingu täitmise tagamiseks.[17]

Seaduse alusel kogub riik näiteks rahvastikuandmeid või korraldab riiklikku statistikat.

Lepingu alusel andmetöötlus toimub näiteks laenulepingu, kõnesideteenuse või e-poes kauba müügilepingu puhul.

Nõusolek on vajalik näiteks inimesele otseturundussõnumite saatmisel. Samuti on vajalik võtta inimeselt täiendav nõusolek siis, kui tema kohta kogutud andmeid soovitakse kasutada esialgsest eesmärgist erinevalt.

Mõnikord viitavad organisatsioonid üldiselt, et töötlevad isikuandmeid isikuandmete kaitse seaduse alusel. Selline viide ei ole korrektne. Märkima peaks lisaks, kas õiguslikuks aluseks on nõusolek, lepinguline suhe või mõnest eriseadusest tulenev alus.

Suurandmete analüüsil on tõenäoliselt kõige olulisemad nõusolek ning see, kas andmetöötlus on vajalik lepingu täitmiseks. Seda siis juhul, kui analüüsitakse isikuandmeid. Nagu varasemalt selgitasime, isikustamata ehk anonüümitud andmete töötlus ei käi andmekaitseregulatsiooni alla.

## 5.1. Nõusolek

Kui organisatsioon toetub inimese nõusolekule tema isikuandmete töötlemiseks, siis peab see nõusolek olema antud vabal tahtel, konkreetne ja teadmistel põhinev. See tähendab ,et inimene peab

[15] IKS § 25.
[16] Arvamus nr 05/2014 anonüümimistehnikate kohta.
[17] IKS § 10, § 14 lg 1.

suutma aru saada, mida organisatsioon tema andmetega teeb ja tema nõusolek selliseks tegevuseks peab olema selgelt väljendatud.[18] Kui organisatsioon on kogunud isikuandmeid üheks otstarbeks ja siis otsustab analüüsida neid hoopis teiseks otstarbeks (või teha need teistele selleks kättesaadavaks), peab ta oma teenuste kasutajaid sellest teavitama ning saama sellekohase nõusoleku. See on eriti tähtis, kui organisatsioon kavatseb kasutada andmeid eesmärgil, mis ei ole üksikisikule ilmne, sest puudub otsene seos temapoolse teenusekasutamisega. Näiteks kui sotsiaalmeedia ettevõte hakkab oma kasutajate kohta kogunenud isikuandmeid muudeks eesmärkideks teistele ettevõtetele müüma.

Võiks arvata, et kui inimene on avaldanud sotsiaalmeedias enda kohta teavet, siis on lubatud piiramatult seda teavet kasutada, sh selle isiku kohta suurandmete analüüsi tegemiseks. Seda justkui õigustaks IKS § 11 lõige 1. Samas ei ole IKS § 11 lõige 1 eraldiseisev isikuandmete töötlemise alus[19] ning ka isiku enda sotsiaalmeedias avaldatud andmete kasutamiseks on vajalik õiguslikku alust – kas isiku nõusolekut või mõnda muud seadusest tulenevat alust. Kui isiku avaldatud teavet kasutatakse isikustamata kujul (nt mõne toote kohta sotsiaalmeedias esitatud arvamust kasutatakse sama toote parendamiseks), siis pole selline tegevus isikuandmete töötlemine.

Suurandmete töötluse eeldusena nõusoleku küsimisel peab andmetöötleja olema eelnevalt kindel, et see on sobilik tingimus. Ehk, kui inimesel ei ole reaalset valikut ja ta ei saa oma nõusolekut soovi korral tagasi võtta, siis ei vasta selline nõusolek isikuandmete kaitse seaduse nõuetele.

Kui organisatsioon ostab analüüsimiseks suure isikuandmete kogumi, siis saab temast nende andmete vastutav töötleja.[20] Organisatsioon peab olema kindel, et ta vastab nende andmete edasiseks kasutuseks isikuandmete kaitse seaduse nõuetele. Kui ta toetub selle tingimusena andmete algse andja hangitud nõusolekule, siis peab ta veenduma, et see nõusolek hõlmab tema planeeritavat täiendavat andmetöötlust.

Suurandmete analüüsi võimalik keerukus ei tohi saada ettekäändeks nõutava nõusoleku hankimata jätmisele. Organisatsioonid peavad leidma sobiva viisi, kus ja kuidas analüüsi eesmärki ja kasusid inimestele selgitada ja nõusoleku võtmist korraldada.

## 5.2.Lepingud ja seaduslik alus

Nõusolek on üks isikuandmete töötlemise tingimusi, kuid mitte ainus. Organisatsioonid ei pea kõigil juhtudel nõusolekut hankima ja teatud juhtudel võivad olla asjakohased muud tingimused.

Isikuandmeid võib töödelda ka inimesega sõlmitud lepingu täitmise tagamiseks, välja arvatud delikaatsed isikuandmed.[21] Näiteks teeb inimene taksoettevõttega püsilepingu hommikul kindlal kellaajal tööle sõitmiseks. Selleks on taksoettevõttel vaja teada inimese nime, sõidu alguse aadressi, telefoni numbrit (taksojuht helistab hilinemise korral). Arve saatmiseks ka e-posti aadressi. Lepingus on näiteks ka kirjas, et taksoettevõte kasutab neid andmeid sisemise töökorralduse parandamiseks ja autopargi optimeerimiseks. Selline leping on oma olemuselt sobilik.

Kui taksoettevõte otsustab sõitjate andmeid aga jagada näiteks omavalitsusega, kes omakorda liidab need nn ühisesse potti ühistranspordi kasutajatega ja analüüsib neid näiteks transpordivõrgu planeerimiseks, on juba tegemist suurandmetega. Ehk, taksoettevõtte algselt kogutud andmete kasutamise eesmärk muutub ning läheb kaugemale kui taksoteenuse osutamine inimesega sõlmitud lepingu täitmiseks. Seetõttu võib olla taksoettevõttel keeruline oma klientidele põhjendada, et suurandmete analüüs on lepingu täitmiseks rangelt vajalik. Kui aga taksoettevõte edastab oma klientide teenuse kasutuse andmed omavalitsusele isikustamata kujul (anonüümimine), probleemi ei ole.

Andmetöötlus võib olla vajalik ka andmeid koguva organisatsiooni seadusest tulenevate ülesannete

[18] IKS § 12.
[19] Nimetatud seisukohale on Riigikohus jõudnud 12.06.2012 otsuses nr 3-3-1-3-12 (vt otsuse punkte 21-29), 18.02.2015 otsuses nr 3-2-1-159-14 (vt otsuse punkti 14) ning 23.03.2016 otsuses nr 3-3-1-85-15 (vt otsuse punkti 18).
[20] IKS § 7.
[21] Delikaatsete isikuandmete mõiste on kirjas IKS § 4 lg 2.

või õiguste täitmiseks. Näiteks Statistikaamet analüüsib nii avaliku- kui erasektori tegevust riikliku statistika korraldamiseks. Õiguse selleks annab riikliku statistika seadus. Üldkasutatava elektroonilise sideteenuse osutajad võivad elektroonilise side seaduse alusel töödelda oma klientide isikuandmeid näiteks äritegevuse käigus tehtud tehingute dokumenteerimiseks ja muuks äriliseks infovahetuseks.[22]

## 5.3.Andmete minimaalsus

Isikuandmete töötlemise üks olulisi põhimõtteid on andmete minimaalsus. Ehk, isikuandmeid võib koguda vaid ulatuses, mis on vajalik määratletud eesmärkide saavutamiseks.[23] St organisatsioonid peavad minimeerima kogutavate ja töödeldavate andmete hulka ning nende säilitamise kestust.

Suurandmete analüüsi korral seevastu kogutakse enamasti võimalikult palju andmeid. Esiteks suurenevad andmete säilitamise võimalused pidevalt ja nende säilitamise kulud aina vähenevad. Teiseks võib suurandmete analüüsi võime töödelda väga suuri andmehulki julgustada andmete vastutavat töötlejat hoidma alles väga pikka andmete ajalugu, mis ulatub kaugemale tavalisteks ärieesmärkideks nõutavast perioodist. Kuigi raamatupidamise algdokumente tuleb vastavalt seadusele[24] säilitada 7 aastat, pole eraisikule esitataval arvel vaja näidata tema nime või muid kontaktandmeid.[25] Seepärast pole sellise arve puhul ka alust inimese isikuandmeid 7 aastat säilitada.

Kui organisatsioonid soovivad andmeid suurandmete analüüsi jaoks pikemat aega säilitada, peaksid nad suutma nende andmete potentsiaalseid kasutusviise ja kasusid teatud määral väljendada ja ette näha, isegi kui konkreetsed üksikasjad on teadmata. Säilitamise perioodid on üldistes dokumendihalduse reeglites ja raamatupidamiseeskirjades ära määratud. Samuti on mõnedes ettevõtlusvaldkondades kehtestatud regulatiivsed nõuded sellele, kui kaua peab andmeid säilitama. Teatud uuringute osas, näiteks kliinilistes katsetes, on aga pikaajalised uuringud tähtsad.

Tähtis on meeles pidada, et andmekaitse põhimõtted kohalduvad ainult isikuandmetele. Seega on anonüümimistehnikad paindlikuks meetmeks andmete innovaatilisel analüüsil või pikemal säilitamisel.

## 5.4.Õigus saada teavet

Inimesel on õigus saada teavet isikuandmetest, mida tema kohta kogutakse, mis eesmärgil seda tehakse ning kellele neid võidakse avaldada. Õigus on saada koopia andmetest, mida nende isikuandmeteks loetakse, samuti teavet andmeallikate kohta.[26] Need reeglid kehtivad ka suurandmete maailmas.

Võib arvata, et suurandmete mahukus ja mitmekülgsus ning nende analüüsimise keerukus teeb organisatsioonidel selle nõude täitmise raskeks. Samas ei saa see olla ettekäändeks seaduslike nõuete eiramiseks. Nõude täitmisel võib probleemiks osutada see, et teavet hoitakse erinevates kohtades. Samas on andmete erinevates andmebaasides hoidmine just lisatagatis inimeste privaatsusõiguse tagamisel. Ka suurandmete analüüsis ei soovita me koondada andmeid ühte nn superandmebaasi, vaid võimalusel töödelda neid hajusalt. Meetodeid selleks leidub.

# 6. Infoturve

Isikuandmete töötlemisel peab olema tagatud turvameetmete rakendamine.[27] Seda ka inimest tuvastada võimaldaval suurandmete analüüsil. Suurandmete puhul kasutatakse mahukaid

[22] Elektroonilise side seadus § 102 lg 4.
[23] IKS § 6 p 3.
[24] Raamatupidamise seadus (RPS) § 12 lg 1.
[25] RPS § 7 lg 2, lg3.
[26] IKS § 19.
[27] IKS § 25.

andmemassiive, mida võidakse hoida „pilves“.[28] Andmetele võidakse rakendada mitmes serveris toimuvat hajutatud andmetöötlust. Sel juhul peavad organisatsioonid saama pilveteenuse pakkujalt piisava garantii selle teenuse puhul kasutatavatest turvameetmetest. Selles osas ei erine pilvepõhine suurandmete analüüs pilveteenuse kasutamisest muuks andmete hoidmiseks ja töötlemiseks.

Koos suurandmete laialdasema levikuga võivad suureneda ka infoturbealased ohud. Näiteks kirjeldab Euroopa Liidu Võrgu- ja Infoturbe Amet (ENISA) oma aruandes[29] ohte, mis tulenevad suurandmete potentsiaalsest väärkasutamisest. Dokumendi sõnul on inimeste andmete kontrollimatu kogumine, kasutamine ja levitamine ideaalne mängumaa pahatahtlikele tegevustele. Tekib küsimus, millise piirini võib suurandmete kasvu nimetada veel kontrolli all olevaks.

Samas selgitab ENISA oma hilisemas analüüsis[30] põhjalikumalt, et suurandmeid saab kasutada ka infoturbe täiustamiseks. Suurandmete analüüsi võime tegeleda väga kiiresti väga suurte andmemahtudega tähendab, et seda saab kasutada võrguliikluse, tehingute ja logifailide analüüsimiseks, mis on tavaliste tehnoloogiate jaoks liiga suured. Nii on võimalik leida seaduspärasid, mustreid ja anomaaliaid ning seega kiiresti turvaohtusid tuvastada.

Infoturve sõltub korralikust riskihindamisest. Kui suurandmeid analüüsivad organisatsioonid rakendavad uute andmete hankimisel või olemasolevate kasutamisel oma tavalisi riskijuhtimise poliitikaid ja protseduure, siis ei tohiks seda „kontrollimatuks“ lugeda. Kombineerides väljastpoolt hangitud andmeid organisatsioonisiseste andmetega kehtivad ka väljastpoolt hangitud andmetele organisatsioonisisesed riskijuhtimise poliitikad ja turvaprotseduurid. Samas tuleb tehnoloogia kiire arengu tõttu organisatsiooni infoturvet ka perioodiliselt täiendada ning uuendada.

# 7. Andmetöötluses osalejad

Suurandmete analüüs võib olla uudne andmetöötluse vorm, mille organisatsioon lepingu alusel väljastpoolt sisse ostab, kuna tal ei ole piisavaid oskusi, et seda organisatsioonisiseselt teha.

Andmekaitsereeglid näevad ette konkreetsed nõuded, kui isikuandmeid soovitakse edastada töötlemiseks organisatsioonivälistele isikutele. Andmete valdaja ehk vastutav töötleja vastutab turvalise andmetöötluse eest. St lisaks sellele, et ta peab ise suutma rakendada sobilikke tehnilisi ja organisatsioonilisi turvameetmed, peab ta ka välise andmetöötleja (volitatud töötleja) valimisel olema kindel turvagarantiide osas. Vastutaval töötlejal peab olema võimalus ka reaalselt neid turvameetmeid kontrollida. Vastutav töötleja peab olema kindel, et volitatud töötleja tegutseb ainult andmete vastutava töötleja juhistest lähtuvalt. Kõige selle aluseks on hästi läbi mõeldud koostööleping.

**Kui organisatsioon ostab isikuandmete analüüsi organisatsioonivälisel isikult sisse, siis on organisatsioonil kohustus tagada, et see isik kohaldab sobivaid turvameetmeid.**

# 8. Kokkuvõtteks

Isikuandmete kaitse põhimõtted sobivad ka suurandmete jaoks. Arvamus, et kehtivad põhimõtted on ebapiisavad, alahindab neile omast paindlikkust. Põhimõtete kohaldamine hõlmab hindamist, millist mõju andmetöötlus üksikisikutele avaldab ja kas see on igal konkreetsel juhul eesmärkide suhtes proportsionaalne. See hindamine hõlmab ka asjaolu, kas isikuandmete uus kasutus on nende algse otstarbega kokkusobimatu ja millal on vaja küsida inimeste nõusolekut.

Suurandmed on valdkond, mille abil tehakse avastusi ja areneb innovatsioon. Samas on ülimalt tähtis välistada võimalus, et keegi organisatsioonis võiks kasutada isikuandmeid volitamata ja privaatsust rikkuval moel. See võib kaas tuua nii õiguskaitseasutuste järelevalvemeetmed kui mainekahju. Korrektselt ja arusaadavalt vormistatud töökorralduse reeglid ja sisekorra- ning

[28] Pilvandmetöötluse kohta vt Pilvandmetöötlus – andmetöötluse loomulik areng.
[29] ENISA Threat Landscape 2013 – Overview of current and emerging cyber-threats. December 2013.
[30] Big Data Threat Landscape and Good Practise Guide, ENISA, January 2016.

turvaeeskirjad aitavad isikuandmete alastest rikkumistest hoiduda.

Suurandmete analüüsist võib saada palju ühiskondlikku kasu. Lisanduvad paremad tooted ja teenused tarbijatele ning ärilised eelised ettevõtetele. Samal ajal ei tohiks neid eeliseid saada ebaausa privaatsuserikkumise teel. Andmekaitse põhimõtted peaksid olema pigem stiimuliks töötada välja innovatiivseid viise avalikkuse teavitamiseks ja kaasamiseks. Suurandmete töötlemise eesmärgi ja kaasnevate mõjude läbipaistvuse tagamine ei ole pelgalt juriidiline nõue. See aitab inimestele kui digitaalsetele kodanikele luua suurandmete maailmas kindlustunde.