Joonas Kartano

TIEDON LUOKITTELU TIEDON SUOJAAMISEN JA LIIKETOIMINNAN NÄKÖKULMASTA

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2023

TIIVISTELMÄ

Kartano, Joonas
Tiedon luokittelu tiedon suojaamisen ja liiketoiminnan näkökulmasta
Jyväskylä: Jyväskylän yliopisto, 2023, 63 s.
Tietojärjestelmätiede, Pro Gradu
Ohjaaja: Siponen, Mikko

Tiedon luokittelu on tärkeä osa yritysten ja organisaatioiden datan hallintaa ja kokonaistietoturvaa. Nykypäivänä yritysten tarjoamat palvelut ovat sähköisiä, ja verkossa kulkee suuri määrä erilaista dataa, josta osa on palveluiden käyttäjien arkaluontoista ja luottamuksellista tietoa. Datamäärien kasvaessa yrityksiltä vaaditaan yhä useammin strategisen tason suunnittelua ja koko organisaation sitoutumista datan huolelliseen hallintaan. Tehokas ja täsmällinen tiedon luokittelu auttaa mm. kohdentamaan tietoturvaresursseja, vähentää manuaalista datan hallintatyötä, ja voi jopa nopeuttaa projektien toteuttamista poistamalla vääränlaisien pääsyoikeuksien aiheuttamaa viivettä. Vaikka tiedon luokittelu tarjoaa lukuisia etuja tiedon suojaamiselle ja liiketoiminnalle, on se syystä tai toisesta aliarvostettua yritysmaailmassa, eikä sen etuja osata hyödyntää, ja tästä syystä suuri osa kaikesta datasta täysin luokittelematonta dataa. Tätä tutkimusta varten haastateltiin viiden yrityksen tietoturvajohtajaa tiedon luokittelusta, ja haastattelujen avulla selvitetään, miten suomalaiset yritykset suhtautuvat tiedon luokitteluun, ja miten tärkeänä tiedon suojaamisen ja liiketoiminnan osana tiedon luokittelu nähdään, sekä millaisia etuja yritykset kokevat saavansa tiedon luokittelulla. Lisäksi haetaan mahdollisia selityksiä sille, millaista arvostusta tiedon luokittelu nauttii osana datan hallintaa ja tiedon suojaamista, sekä selityksiä mahdolliselle aliarvostukselle ja matalalle tiedon luokittelun hyödyntämiselle. Työn ensimmäisessä osassa käydään läpi aihealueeseen liittyvää edeltävää kirjallisuutta ja toinen osa käsittää tutkimuksen empiirisen osan. Tässä työssä tiedon luokittelulla tarkoitetaan yrityksen omistaman datan ja tiedon luokittelemista soveltamalla jotakin tunnettua, tai itse kehitettyä tiedon luokittelun mallia tai kaavaa. Tiedon luokittelun perimmäisenä tavoitteena on taata riittävä suojaus yrityksen omistamalle datalle ja tiedolle riippuen sen arvosta yritykselle.

Asiasanat: data, tiedon luokittelu, tiedon elinkaaren hallinta, datan pääsynhallinta, data datan häviämisen estäminen

ABSTRACT

Kartano, Joonas
Information classification from the perspective of data protection and business
Jyväskylä: University of Jyväskylä, 2023, 63 s.
Information systems science, Master's Thesis
Supervisor: Siponen, Mikko

Information classification is important part of data management and information security of companies and organizations. Today companies offer their services through the internet and there is vast amount of data travelling through internet every day, and part of all data includes also sensitive and confidential information about service users. Growing data volumes causes growing requirement for strategic level data management planning and commitment for organizations to follow the data management strategy in every part of organization. Effective and accurate information classification helps organizations and companies to target security resources more precisely, reduces manual data management and can even help companies to be faster with large projects removing delays caused by bad data access in project teams. Despite many advantages that information classification gives, it is underrated in business field, and companies are utilizing it in the best level, and it causes that big part the whole data in the world is still unclassified dark data. For this study five information security directors were interviewed about information classification, and findings are done about how companies think about information classification and how important part of information security and business information classification is seen and what kind of benefits companies feel to get from information classification. Answers are also sought for how high companies valuate information classification as part of data management and information security and possible explanation for possible underappreciation and low utilization. In the first part previous literature attached to the topic is gone through. The second part comprises the empirical part of the study. In this study information classification refers to actions that companies do to classify data and information that they own. Information classification can be done by using some known classification model or schema or by creating own model. The most important objective of information classification is to ensure that information receives an appropriate level of protection in accordance with its importance to the organization.

Keywords: data, information classification, information lifecycle management, data access control, data loss prevention

KUVIOT

TAULUKOT

SISÄLLYS

1 JOHDANTO

Datan rooli yritysten toiminnassa kasvaa jatkuvasti, ja jo nykyään datasta on tullut yritysten tärkein omaisuuserä (Bergström ym., 2021). Datan ja tiedon tärkeyttä yritystoiminnassa nostaa datan ja tiedon hallinnan tärkeään rooliin, sillä yritysten täytyy kyetä käsittelemään ja suojaamaan omistamaansa dataa jatkuvasti tehokkaammin ja täyttämään datan, erityisesti sovellusten ja palveluiden käyttäjien arkaluontoisen ja luottamuksellisen datan suojaamisen lailliset vaatimukset. Yksi keino datan ja tiedon hallinnan tehokkuuden ja vaatimustenmukaisuuden parantamiseen on tiedon tehokas luokittelu. Kaikki yrityksen omistama data ei ole keskenään saman arvoista, ja toisaalta on olemassa paljon sellaista, esimerkiksi asiakkaisiin liittyvää dataa, jonka säilyttämistä ja suojaamista säädellään vahvasti lakien ja standardien avulla. Luokittelematon data saattaa aiheuttaa yrityksissä monia ongelmia esimerkiksi tiedon saavutettavuuden sekä suojaamisen näkökulmasta. Luokittelematon data saattaa johtaa esimerkiksi tilanteeseen, jossa työntekijä ei pääse käsiksi sellaiseen tietoon, johon hänen kuuluisi päästä, tai vastaavasti hän voi päästä käsiksi sellaiseen tietoon, johon hänen ei kuuluisi päästä. (Bergström ym., 2021)

Vaikka data ja tieto on nykypäivänä yritysten tärkeintä omaisuutta, Veritas Technologiesin vuonna 2016 tekemän tutkimuksen mukaan vain 46 % kaikesta datasta luokiteltua ja tunnistettua dataa, ja jopa 54 % ns. pimeää, luokittelematonta dataa. Luokittelematon ”pimeä” data johtaa tilanteisiin, jossa data on joko yli- tai alisuojattua ja on myös huomattavasti haavoittuvaisempaa tietomurroille verrattuna dataan, jonka sisältö ja tärkeys on tunnistettu, ja joka on suojattu asianmukaisesti. (Veritas Technologiesin, 2016)

Tämän työn tavoitteena on tutkimuksen avulla pyrkiä selvittämään suomalaisten yritysten tiedon luokittelua tiedon suojaamisen ja liiketoiminnan näkökulmasta. Tutkimuksella pyritään muodostamaan kuva siitä, millä tasolla tiedon luokittelua harjoitetaan, mitkä ovat motiivit tiedon luokittelun taustalla, miten yritykset suhtautuvat ja miten korkealle ne arvostavat tiedon luokittelun osana tiedon suojaamista ja liiketoimintaa, ja millaisia etuja sillä tavoitellaan. Tutkimusta ohjaavat tutkimuskysymykset:

- Miten suomalaiset yritykset huolehtivat tiedon luokittelemisesta?
- Miten tärkeänä tiedon suojaamisen ja liiketoiminnan osana tiedon luokittelu nähdään?
- Mitkä ovat motiivit tiedon luokittelemiselle tietyllä tasolla?
- Millaisia malleja tai sääntöjä yritykset soveltavat tiedon luokittelemisessa?
- Millaisia työkaluja tiedon luokittelemisessa hyödynnetään?
- Millaisia tiedon luokittelun politiikkoja yritykset soveltavat toiminnassaan?

Tutkimuksen ensimmäisessä osassa käydään läpi tutkimuksen aiheeseen ja tärkeimpiin käsitteisiin liittyvää aiempaa kirjallisuutta. Jälkimmäinen osa käsittää tutkimuksen empiirisen osan.

Tutkimuksen lähteiden ja kirjallisuuden etsinnässä on käytetty seuraavia tietokantoja sekä hakukoneita: ACM Digital Library, IEEE Xplore Digital Library, ScienceDirect (Elsevier) sekä Google Scholar. Lähteiden etsinnässä käytettyjä hakutermejä olivat data, tiedon luokittelu, tiedon elinkaaren hallinta, datan pääsynhallinta, data datan häviämisen estäminen

2 DATA, INFORMAATIO JA TIETO

Tämä tutkielma käsittelee tiedon luokittelua ja sitä, miten tärkeänä osana tiedon suojaamista ja datan hallintaa yritykset pitävät tiedon luokittelua, ja millaisia etuja yritykset sillä tavoittelevat. Tiedon luokittelu on oleellinen osa tiedon suojaamisen kokonaisuutta, mutta sen lisäksi tiedon luokittelu tarjoaa myös etuja datan hallintaan, ja tehokas datan hallinta auttaa yrityksiä saavuttamaan kilpailuetua. Tässä luvussa käsitellään ensin dataa, informaatiota ja tietoa käsitteinä, jonka jälkeen käsitellään nykypäivän yhteiskunnan suuren datamäärän takana olevien ja yritysten kilpailun kannalta tärkeitä käsitteitä big data ja esineiden internet. Tämän jälkeen luvussa tarkastellaan vielä tarkemmin datan ja tiedon roolia yritystoiminnassa.

2.1 Data, informaatio ja tieto käsitteinä

Data-termi on nykypäivänä niin yleinen ja käytetty termi, että harva on luultavasti ikinä miettinyt, mitä sana data tarkalleen ottaen tarkoittaa. Ensimmäisenä datasta tulee varmasti mieleen taulukko tai kaavio, joka sisältää numeroita tai merkkejä. Data voi saada erilaisia merkityksiä riippuen asian yhteydestä, mutta nykypäivän laitteisiin, kuten puhelimiin ja tietokoneisiin sekä tietojenkäsittelyyn liittyen datan voidaan sanoa olevan sanoa olevan informaatiota, joka on käännetty sellaiseen muotoon, että sitä voidaan tehokkaasti siirtää ja käsitellä, eli binääriseen ja digitaaliseen muotoon (Vaughan, 2019). Tietojenkäsittelyn kontekstissa data -käsitteen on kehittänyt Claude E. Shannon, joka kehitti Informaatioteorian (Information Theory), ja julkaisi sen kaksiosaisessa tiedeartikkelissa Bell Labs-nimisessä tiedejournaalissa vuonna 1948. Shannon kehitti tavan ilmaista lähteen tuottaman informaation määrän yhtälöllä, joka muistuttaa termodynaamisen entropian määrittelevää yhtälöä fysiikassa. Yksinkertaistettuna Shannonin informationaalinen entropia tarkoittaa viestin koodaamiseen tarvittavien binääristen lukujen määrää (Collins, 2002). Shannonin kehittämä Informaatioteoria käsittelee kaksinapaisia kytkentäpiirejä, johon Shannon on soveltanut George

Boolen vuonna 1847 kehittämää Boolen Algebraa (Shannon, 1948, s. 59). Oleellisin osa, jota Shannon sovelsi Boolen algebrasta Informaatioteoriaan, on virtapiirien binääriset operaatiot, jotka voivat saada arvon 1 ”tosi”, tai 0 ”epätosi”, ja näin määräävät piirilevyn kytkimen asennon olevan joko auki, tai kiinni. Shannonin Informaatioteoria on tiettävästi ensimmäinen tieteellinen teos, jossa käytetään bitti, ”bit”-lyhennettä sanoille binäärinen luku, ”binary digit”. Bitti on pienin datan mittayksikkö (Collins, 2002).

Tietoa on säilötty ja se on kulkeutunut sukupolvelta toiselle suullisena perimätietona tai eri tavoin fyysiseen muotoon tallennettuna. Nykyaikaisen, digitaalisessa muodossa olevan tiedon käsittely sekä tallentaminen ovat sidoksissa mikropiirien keksimiseen ja kehittymiseen. 1940-luvulta 1960-luvulle sadat insinöörit ja tiedemiehet ympäri maailman työskentelivät mikropiirien, eli pienten, integroitujen ja elektronisten piirien kehitystyön parissa. Edeltävästä kehitystyöstä huolimatta monissa ammatillisissa historiallisissa julkaisuissa kirjoitetaan puolijohdemikropiirin olleen valtava kehitysloikka, joka tapahtui heinäkuussa 1958, kun Jack Kilbyn teki läpimurron puolijohdemikropiirien saralla. (Brock & Laws, 2012). Puolijohdemikropiirien keksimisen jälkeen niiden kehityksen parissa työskenteli huomattava määrä tutkijoita, ja tämän johdosta mikropiirit alkoivat kehittyä vuoden 1948 jälkeen hurjaa vauhtia. Gordon Moore esitti vuonna 1965 Mooren lakina tunnetun ennustuksen, jonka mukaan teknologian kehityksen ja komponenttien hinnan halpenemisen seurauksena transistorien määrä, joka yhteen mikropiirin on mahdollista sijoittaa, kaksinkertaistuu kahden vuoden välein. (Adee, 2008).

Mikropiirien keksimisen ansioista tietojen nopeutuva käsittely aiheutti kasvavan tarpeen tiedon säilyttämiselle jonkinlaiseen muistiin. Ensimmäisissä tietokoneissa tiedon säilömiseen käytettiin kaksitasoisia muistiratkaisuja, jotka koostuivat ferriittirengasmuistista, sekä rumpumuistista (Denning, 1996). Digitaalisten tietokoneiden nopeutuessa ilmeni kuitenkin tarve suuren kapasiteetin nopeasti toimiville muistiratkaisuille. Mancesterin yliopiston Atlas-tiimi kehitti vuonna 1959 kaksitasoisen ja käyttäjän manuaalisesti hallinnoiman muistijärjestelmän tilalle ensimmäisen yksitasoisen virtuaalimuistijärjestelmän, jossa tietokoneen käyttäjän ei tarvinnut tehdä manuaalista työtä siirtääkseen tietoja kahden eri muistityypin välillä, vaan käsitelty tieto siirtyi automaattisesti muistitasojen välillä. (Kilburn ym., 1962, s. 226).

Mikropiirien ja puolijohdelaitteiden keksimisen jälkeen elektroniikkateollisuus on kasvanut ja kehittynyt räjähdysmäisesti jokaisella vuosikymmenellä. Mikropiirien laskentateho kasvaa edelleen jatkuvasti, ja samalla piirien fyysinen koko pienenee, ja se mahdollistaa aina vain pienempien ja tehokkaampien laitteiden valmistamisen. Voidaan sanoa, että nykypäivänä kaikki tuotettu tieto on sähköisessä muodossa. Suurin osa nykypäivänä tuotetusta datasta koostuu kuvista, videoista, äänistä ja tekstistä. Tuoreen tiedon lisäksi myös vanhaa, fyysisessä muodossa olevaa tieteen tutkimusaineistoa siirretään jatkuvasti sähköiseen muotoon (Turkel ym. 2014.). Laitteiden kasvavan laskentatehon lisäksi vuonna 1991 kehitetty internet ja sen kehittyminen pisteeseen, jossa lähes mikä tahansa laite on mahdollista kytkeä internetiin langattomasti, on johtanut siihen, että

laiteet puhelimista kodinkoneisiin tuottavat jatkuvasti eri tasoista dataa. Ahmed & Ahmed (2020) mukaan vuonna 2020 oltiin tilanteessa, jossa maailmassa tuotettiin 2,5 triljoonaa tavua dataa päivittäin, ja vuosina 2018–2020 on tuotettu 90 % kaikesta maailman datasta. Lähteestä riippuen Gordon Mooren ennustus pitää edelleen paikkansa ja sen arvioidaan olevan voimassa edelleen seuraavan 10 vuoden ajan. Nykypäivän mikropiirit valmistetaan piikiekoista, sekä niiden päälle liimattavista ohuista kalvoista valottamalla kalvolle piirin haluttu kuvio (ASML, 2021). Nykytekniikalla mikropiirille on mahdollista mahduttaa jopa sata miljardia transistoria, ja maailman johtavan mikropiirien valmistajan TSMC:n ennusteiden mukaan vuonna 2030 mikropiirille voi mahtua jopa 300 miljardia transistoria (TSMC, 2021).

Informaatio termi on huomattavasti monitulkintaisempi ja vaikeampi selittää yksikäsitteisesti verrattuna dataan. Datan mielletään usein saavan arvoa ja muuttuvan informaatioksi ja sitä kautta tiedoksi sen jälkeen, kun raakadataan, esimerkiksi lukuarvoihin tai tekstiin liitetään tietty konteksti niin, että tiedetään, mihin data liittyy. Informaatio-termille ei ole olemassa yhtä universaalia määritelmää, mutta termiä on yritetty selittää ja sille on yritetty antaa tieteessä paljon mahdollisimman kuvaavia selityksiä. Michael Buckland (1991) esittelee kolme erilaista luokittelua informaatio-termille, joiden kautta termiä on mahdollista lähestyä. Nämä kolme luokittelua ovat:

- *Informaatio prosessina*: kun henkilöä informoidaan siten, että se, mitä henkilö tietää, muuttuu. Tällöin informaatio voidaan nähdä tekona tai prosessina, jossa informaatio tai ”uutinen” uudesta asiasta kerrotaan, ja joka näin ollen voi muuttaa näkemystä.
- *Informaatio tietona* voidaan nähdä termin informaatio prosessina-prosessina lopputulemana, kun tiettyyn tosiasiaan, kohteeseen tai tapahtumaan liittyvä tieto on ilmoitettu tai kerrottu.
- *Informaatio asiana*: informaatio-termi voidaan nähdä myös objektien, kuten datan tai dokumenttien ominaisuutena ja nähdä informaationa, joka kuvailee puolueettomasti tiettyä objektia.

Informaation ja tiedon välistä eroa on huomattavasti vaikeampi piirtää yksiselitteisesti verrattuna esimerkiksi datan ja informaation erotteluun. Bucklandin (1991) edellä mainittuun luokitteluun liittyy vahvasti kuitenkin informaation jakaminen aineelliseen ja aineettomaan informaatioon. *Informaatio prosessina* sekä *informaatio tietona* edustavat aineetonta informaatiota, kun *informaatio asiana* puolestaan edustaa aineellista informaatiota. Aineellisella informaatiolla tarkoitetaan Bucklandin mukaan sellaista informaatiota, joka on ilmaistu jollain tavalla fyysisessä muodossa, kuten signaalina, tekstinä tai kommunikoitu jollain tavalla. Informaatio-termi on itsessään kehittynyt kielen kehityksen mukana ja varsinkin tietojenkäsittelyn kehityksen myötä. Nyky-yhteiskunnassa *informaatio prosessina* sekä *informaatio tietona* ovat edelleen valideja termejä, mutta aineelliseksi luokiteltu *informaatio asiana* on aina enemmän läsnä ihmisten elämässä, sillä informaatiota on säilöttynä paljon älylaitteiden muistiin, palvelimille, tietokantoihin sekä

erilaisiin sovelluksiin, ja on näin ollen kenen tahansa hyödynnettävissä milloin tahansa.

Tieto terminä on informaation tavoin hankalasti selitettävissä yksiselitteisesti. On helppoa ajatella tiedon tarkoittavan yksinkertaisesti sitä, että tietää jotakin jostakin. Kun kuitenkin ihmistä pyydetään selittämään tieto-termi, voivat selitykset poiketakin jo melko paljon keskenään. Tietoa ja sen perimmäistä olemusta on pyritty ymmärtämään ja selittämään aina. Antiikin kreikan filosofi Platon on käsitellyt ja pyrkinyt selittämään tietoa dialogeissaan Meno ja Theaetetus, kuitenkaan yksiselitteisessä määrittelyssä. Perinteinen, Platonin Theaetetukseen perustuva tiedon määritelmä kuvaa tiedon olevan hyvin perusteltu tosi uskomus. (Lemos, 2007, s. 1) Tiedon määritteleminen on vahvasti filosofinen kysymys, ja Lemoksen (2007) mukaan osa filosofeista ajattelee, että tiedämme pitkälti ne asiat, joita uskomme tietävämme esimerkiksi ympäröivästä maailmasta kokemustemme perusteella. Toisaalta skeptisempi lähestyminen tietoon sanoo, että emme tiedä oikeastaan juuri mitään. Tämän työn tavoitteena ei kuitenkaan ole syventyä tarkemmin tiedon epistemologiaan, vaan lähestyä tietoa enemmän sen perinteisen määritelmän kautta ja tarkastella tiedon roolia modernissa maailmassa ja erityisesti liiketoiminnan osana.

2.2 Big Data & IoT

Kuten tämän työn aiemmissa luvuissa on tuotu esille, on informaatioteorian ja mikropiirien syntymästä kuljettu pitkä matka, jonka varrella mikropiirit ovat pienentyneet, niiden laskentateho, sekä tuotetun datan määrä kasvaneet räjähdysmäistä vauhtia, kuten Gordon Moore on ennustanu jo 1960-luvun puolivälissä.

Big Data sekä esineiden internet (IoT) liittyvät oleellisesti toisiinsa, ja niiden kehitys kulkeekin suurelta osin käsi kädessä. Internetiin kytkettyjen laitteiden määrän kasvaessa tuotetun datan määrä kasvaa, ja kasvava datamäärä puolestaan tehostaa IoT :ta, sillä IoT-verkon toiminta, teho, sekä kehittyminen perustuvat kerätyn datan jatkuvasti tehokkaampaan analysointiin (Khare & Totaro, 2019).

Big data on terminä verrattain uusi, sillä O'Reilly Median Roger Mougalas käytti sitä ensimmäisen kerran julkaisussaan vuonna 2005 (Ahmed & Ahmed, 2020). Big datalle ei ole universaalia määritelmää, mutta esimerkiksi Ahmed & Ahmed (2020) artikkelin mukaan sillä tarkoitetaan suurta määrää strukturoitua ja strukturoimatonta dataa, joka muodostuu kaikkialla yhteiskunnassa laitteiden ja sovellusten sekä järjestelmien käyttäjien toimesta ääninä, videoina, kuvina, sensoritietona, sosiaalisen median sisältönä sekä interaktioina.

Koska datan määrä kasvaa jatkuvasti kiihtyvällä tahdilla, tarvitaan käsittelyyn ja hyödyntämiseen uudenlaisia teknologioita sekä tekniikoita, jotta sen sisältämä informaatio on mahdollista jalostaa analyysitarkoituksiin sopivaan muotoon ja ihmisten sekä yritysten hyödynnettäväksi (Kaur & Kushwaha, 2018). Sen lisäksi, että big data tarkoittaa suurta määrää eri lähteistä tulevaa dataa, useissa

tieteellisissä julkaisuissa luetellaan ominaisuuksia, jotka kuvaavat big datan luonnetta. Big datan ominaisuuksia kuvataan usein V-mallilla. V-mallista on olemassa erilaisia versioita, kuten 4V-, 5V-, sekä 6V-malli. 4V-malli on yksi useimmin big datan ominaisuuksien kuvaamisessa käytetty malli. 4 v-kirjainta tulevat Englannin kielen sanoista volume (määrä), variety (moninaisuus), velocity (vauhti), sekä veracity (todenmukaisuus). Khare & Totaro (2019)

- Volume tarkoittaa massiivista datamäärää, jota IoT-laitteet varastoivat. Yritysten IoT -laitteet ja sensorit voivat sisältää valtavan määrän dataa esimerkiksi työntekijöistä, varastosaldoista, laskuista, ostoista tai maksukorteista. Tällaisen datan lisäksi laitteet tallentavat erilaista metadataa, kuten sijaintitietoja tai aikaleimoja. Metadatalla rikastettu data on arvokasta yrityksille, ja siksi yritykset investoivat datan hyödyntämistä tehostaviin ohjelmistoihin sekä laitteistoihin.
- Variety tarkoittaa sitä, että haluttu data kertyy lukuisista erityyppisistä lähteistä eri muodoissa. Data voi kertyä esimerkiksi erilaisista sensoreista monessa eri muodossa, kuten esimerkiksi ääninä tai radiosignaaleina
- Velocity tarkoittaa, että dataa kertyy erilaisista IoT-laitteista erittäin korkealla nopeudella. Datan nopea kertyminen aiheuttaa haasteita, sillä kaikki kertyvä data pitää pystyä käsittelemään tarpeeksi nopeasti. Data kertyy usein erittäin nopeasti, mutta kertymisen nopeus ei ole kuitenkaan tasaista, vaan nopeus vaihtelee. Nopeuden vaihtelu edellyttääkin huolellista suunnittelua, jotta laskentateho ja muistin määrä riittää myös silloin, kun dataa kertyy normaalia enemmän, jotta datapiikki ei aiheuta katkoksia järjestelmissä.
- Veracity tarkoittaa kertyvän datan oikeellisuutta. Esimerkiksi erilaisissa sensoreissa on harvoin määriteltynä virhemarginaaleja, vaan sensorit syöttävät mittaamansa tiedon eteenpäin juuri sellaisena, kun se tulee sensoriin. Sensoreihin voi kuitenkin tulla erilaisia ongelmia, kuten yhteysongelmia, laitteisto-ongelmia ympäristön vaikutuksesta tai muusta syystä. Sensoreiden toimivuuden varmistaminen onkin erittäin tärkeää erityisesti päätöksenteon näkökulmasta, sillä monet tärkeät päätökset tehdään käytettävissä olevaan dataan perustuen.

Esineiden internet, eli IoT tarkoittaa tietoverkkoa, joka muodostuu erilaisista laitteista, jotka verkottuneet, ja yhteydessä toisiinsa internetin välityksellä. Toisiinsa yhteydessä olevat IoT-laitteet keräävät dataa, ja sitä kautta kertovat verkon muille laitteille, miten ne toimivat ja suoriutuvat niille osoitetuista tehtävistä. (Khan ym., 2021) IoT-termin keksijänä pidetään yleisesti Kevin Ashtonia, englantilaista teknologiapioneeria, joka käytti termiä ensimmäisen kerran Procter & Gamble nimiselle yritykselle pitämässään esityksessä vuonna 1999. BECS Technology on valmistanut IoT-laitteita yli kahden vuosikymmenen ajan vesikemian käyttöön. Ensimmäiset BECS Technologyn valmistamat IoT-laitteet tarvitsivat etäohjauksen mahdollistamiseksi erillisen modeemin sekä ohjelmiston turvallista internet-yhteyttä varten. Laitteet yhdistettiin modeemin ja ohjelmiston

avulla POTS-nimiseen puhelinverkkoon, ja laitteen käyttäjän tuli laitteen oikea puhelinnumero etäohjatakseen laitetta. (Chamberlain & Steinbrueck, 2020.)

IoT-ekosysteemi muodostuu dataa keräävistä IoT-laitteista, datan säilyttämiseen tarkoitetuista pilvipohjaisista tai fyysisitä palvelinratkaisuista, sekä datan käsittelyyn ja hyödyntämiseen tarkoitetuista analytiikkaohjelmistoista. Verkko-ominaisuuksilla varustettujen älykkäiden IoT-laitteiden, kuten prosessorien ja sensoreiden, tehtävänä on kerätä, käsitellä ja lähettää eteenpäin dataa, jota se on sille määritellyn tehtävän mukaisesti kerännyt. IoT-laitteiden keräämä data lähetetään laitteista eteenpäin on-premise tai pilvipalvelimille. Kolmas tärkeä osa IoT-ekosysteemiä ovat datan käsittelyyn, analysointiin ja hyödyntämiseen tarkoitetut ohjelmistot. Ohjelmistot ovat yhteydessä samoihin palvelimiin, joihin IoT-laitteet lähettävät dataa. (Khan ym., 2021, s. 2) Datan käsittely- ja analysointiohjelmistot ja sovellukset mahdollistavat suurien datamäärien käsittelyn, sekä erilaisten raporttien muodostamisen. Visuaalisessa muodossa olevat raportit puolestaan auttavat yrityksiä tekemään päätöksiä tietoon perustuen.

2.3 Datasta tiedoksi: DIKW-Malli

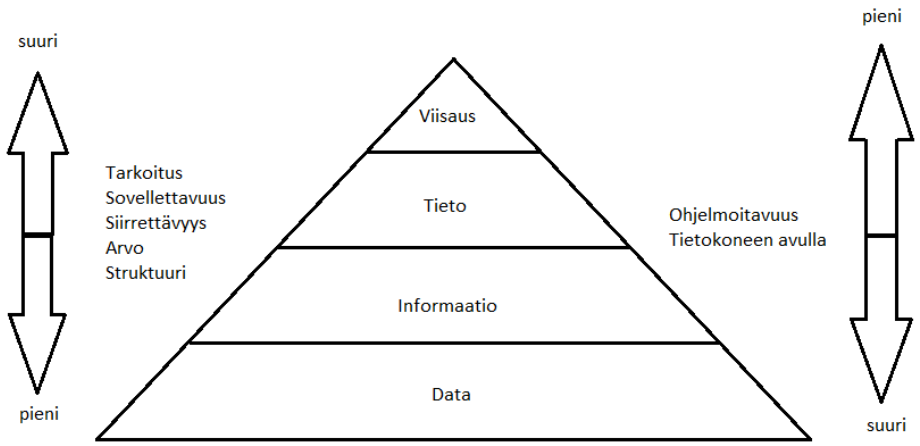
Ihmisten, samoin kun yritysten ja organisaatioiden jokapäiväinen toiminta perustuu päätöksien ja valintojen tekemiseen. Ennen päätösten tekemistä ihmiset ja organisaatiot hyödyntävät heiltä löytyvää tietoa päätöksenteon kohteena olevaan aiheeseen liittyen. Yritysten ja organisaatioiden valinnat vaikuttavat aina tulevaisuuteen, ja määrittelevät tulevia tapahtumaketjuja. Yritysten ja organisaatioiden kohdalla muuttujia on suuri määrä, ja tehtyjen valintojen seurauksia on vaikeaa ennustaa varmuudella. Siitä syystä yritykset pyrkivät mittaamaan omaa toimintaansa, ja lisäksi hankkimaan mahdollisimman paljon tietoa myös muusta toimintaympäristöstä, jotta strateginen ja operatiivinen johtaminen perustuisi mahdollisimman paljon tosiasioihin. Mitä enemmän tietoa löytyy ennen päätöksen tekemistä, sitä suuremmalla todennäköisyydellä tehty valinta osuu oikeaan. Tässä kappaleessa käsitellään tietämyksen hallintaa sekä tarkemmin tietämyksen hallinnan alalla käytettyä DIKW (data-information-knowledge-wisdom) -mallia.

DIKW-mallia hyödynnetään laajasti informaatiota ja tietoa sekä tietohallintaa käsittelevässä kirjallisuudessa. Mallin kuvaileva rooli on nostanut sen yhdeksi keskeisimmistä käsitteistä informaationhallinnassa, tietojenkäsittelyssä, sekä tiedonhallinnassa (Rowley, 2007). Russell Ackoffin vuonna 1989 kirjoittama *From data to wisdom* -artikkeli on uudemman ajan keskeinen ja hyvin lainattu DIWK-hierarkiaa käsittelevä teos. DIKW-malliin kuuluu sen sisältämien termien määrittely, mutta koska tämän työn aiemmassa luvussa käydään läpi, mitä data, informaatio ja tieto ovat, keskitytään tässä luvussa siihen, millainen on siirtymäprosessi datasta viisauteen (wisdom).

Kuvio 2 osoittaa, kuinka tiedon ja ymmärryksen saavuttaminen on prosessi. Se on prosessi, jossa datasta jalostetaan informaatiota liittämällä dataan, esimerkiksi numeerisiin tai aakkosellisiin arvoihin jokin aihe tai konteksti, jonka avulla

datasta tehdään merkityksellistä, arvokasta sekä haluttuun tarkoitukseen sopivaa. Prosessi, jossa informaatiosta jalostetaan tietoa, ei ole niin selkeä verrattuna datan ja informaation väliseen suhteeseen. Tietoa kuvaillaan usein toimivaksi informaatioksi, tai ymmärryksellä ja kyvykkyydellä varustetuksi informaatioksi. DIWK-hierarkian yli taso on viisaus, jonka katsotaan syntyvän tiedosta. Viisaudelle ei ole löydetty yhtä määritelmää, mutta esimerkiksi Baltes ja Kunzmann (2004) esittävät viisauden olevan ilmiö, joka sisältää kognitiivisia, tunnepohjaisia sekä vaikuttimellisia piirteitä, ja määrittelevät viisauden ”asiantuntemukseksi sekä käsitykseksi tärkeistä ja epävarmoista elämän kysymyksistä”. (Rowley, 2007, s. 14)

Tiedonhallinta on oleellinen osa yritysten ja organisaatioiden toimintaa, ja julkisen sekä yksityisen datamäärän kasvaessa täytyy datan hyödyntäminen olla tehokasta, jotta yrityksiä ja organisaatioita voidaan johtaa tietoon perustuen. Intezari ym. (2016) kirjoittavat artikkelissaan johdon päätöksenteon olevan dynaaminen tapahtuma, joka sisältää päätöksen tekijän (ihminen), päätöksenteon tilanteen (ongelma), sekä halutun ratkaisumallin kehittämisen (prosessi). Edellä lueteltujen tekijöiden johdosta DIWK-hierarkiaan pohjautuva tiedonhallinta auttaa yritysten tekemään mahdollisimman hyviä päätöksiä (Intezari ym., 2016, s. 3). Kuviossa 1 on kuvattuna DIKW-hierarkian malli, josta käy ilmi datan, infomaation, tiedon ja viisauden välinen hierarkinen suhde, sekä näiden ominaisuudet.



KUVIO 1 DIWK-hierarkia (Interazi ym., 2016)

2.4 Datan, informaation ja tiedon rooli yritystoiminnassa

Johtaminen kohtaa aina uusia haasteita, ja joskus uudet haasteet ovat vain vanhoja haasteita hieman uudessa kontekstissa. Silloin tällöin ilmestyy kuitenkin myös täysin uudenlaisia haasteita, joille ei ole olemassa ennakkotapausta. Aineettoman omaisuuden hallinnan voidaan katsoa olleen tällainen johtamisen haaste informaation aikakaudelle tultaessa. Tuloksen tekeminen useilla aloilla tiedonhallinnan avulla on kasvanut dramaattisesti informaation aikakaudella teknologisen kehityksen sekä aineettoman omaisuuden roolin muuttumisen ansioista. (Agatic ym. 2011)

Kuten jo aiemmissa luvuissa kerrotaan, on maailman datamäärä kasvanut hurjaa vauhtia, ja kasvu jatkuu edelleen kiihtyvällä tahdilla. Digitalisaatio, big data, sekä esineiden internet ovat vaikuttaneet vahvasti yritysten toimintaan lähes kaikilla aloilla. Riippumatta siitä, millä alalla yritys toimii, toiminnan perusedellytyksiä ovat taloudenpito, asiakkaalle lisäarvoa tuottavien tuotteiden tai palveluiden tuottamisen tai valmistamisen toiminnot, sekä asiakkaiden palveleminen. Yritystoiminnan prosessit ja toiminnot synnyttävät jatkuvasti numeerista dataa, josta on mahdollista jalostaa informaatiota, ja informaatiosta edelleen tietämystä.

Päätöksenteon näkökulmasta yritysten tärkeät rakenteet ja toiminnot ovat johdettavissa yksilön päätöksentekoprosessista ja rationaalisista yksilön valinnoista. Ideaalitilanteessa rationaalinen valinta edellyttäisi kaikkien mahdollisten ratkaisuvaihtoehtojen täydellistä tutkimista, luotettavaa tietoa vaihtoehtojen seurauksista, sekä johdonmukaisia preferenssejä lopputulosten arviointiin. Oikeassa elämässä tämä ei kuitenkaan ole mahdollista, sillä päätöksenteon tukena on hyvin harvoin kaikki mahdollinen valintaan vaikuttava tieto. (Choo, 1996) Vaikka oikeiden päätösten ja valintojen tekeminen on tärkein yrityksen menestykseen vaikuttava asia, ja organisaatiot ymmärtävät datan merkityksen toimintojen johtamisen kannalta, keskustellaan datan hallinnasta yrityksen omaisuutena liian harvoin, eikä sitä ole usein myöskään dokumentoitu riittävällä tasolla. Suurin osa organisaatioista ei myöskään laske omistamaansa dataa omaisuudeksi, niin kuin ne tekevät kaikelle muulle omaisuudelle. Yksi syy sille, miksi yritykset harvoin näkevät datan omaisuuseränä, voi olla se, että data on aineetonta, ja sitä käytetään eri tavalla verrattuna moneen muuhun yrityksen omaisuuteen. Lisäksi datalle on hankalaa laskea konkreettista arvoa verrattuna esimerkiksi koneisiin tai laitteisiin. (Fleckenstein & Fellows, 2018, s. 28–32)

Tässä luvussa käsitellään lisäksi dataa ja informaatiota yritysten omaisuuseränä sekä sitä, millaisia mahdollisuuksia ja kilpailuetuja yritykset voivat saavuttaa tehokkaan datan ja informaation hyödyntämisen avulla

2.4.1 Yrityksen informaatio-omaisuus

Yrityksen informaatio-omaisuus, jota kutsutaan kirjallisuudessa usein myös muun muassa organisaatiopääomaksi, tietopääomaksi tai osaamispääomaksi, on yrityksen omistamaa aineetonta omaisuutta. Vaikka informaatio-omaisuus ei ole

mitään konkreettista, kutsutaan sitä omaisuudeksi sen taloudellisen vaikutuksen vuoksi. Taloudellinen vaikutus syntyy arvonluonnin prosessina, jossa hyödynnetään yrityksen erilaista omaisuutta, sekä aineellista, että aineetonta, yhtä aikaa. (Martín-de-Castro ym., 2005, s.4)

Yrityksen omistama informaatio on hyvin monimuotoista, eikä kaikki informaatio ole keskenään saman arvoista. Informaation erilainen arvo on tunnistettu, ja useat eri tutkijat ovatkin kehittäneet erilaisia viitekehyksiä informaatio-omaisuuden luokittelemiseksi. Luokittelumallit ovat kehittyneet teknologisen kehityksen sekä tutkimustiedon lisääntymisen myötä. Ensimmäiset luokittelumallit ovat katsoneet aineettoman omaisuuden muodostuvat kolmesta komponentista: ihmispääoma, rakenteellinen pääoma, sekä asiakaspääoma. Luokittelumallit ovat kehittyneet ja tarkentuneet, ja Martín-de-Castro ym. (2005) kertovatkin artikkelissaan vuosina 2003 espanjan kielellä julkaistusta Intellectus-nimisestä mallista, joka on kehitetty tutkijoiden ja asiantuntijoiden toimesta edeltävään tutkimukseen sekä tekijöiden kokemukseen pohjautuen. Kyseisessä mallissa aineeton omaisuus rakentuu viidestä komponentista:

- *Ihmispääoma,* joka viittaa eksplisiittiseen tietoon, jota ihmiset omistavat, sekä myös heidän kykyynsä tuottaa tietoa, josta on hyötyä yrityksen tavoitteille. Sisältää myös arvot, asenteet, soveltuvuudet, sekä tietotaidon.
- *teknologinen pääoma,* joka viittaa teknisten järjestelmien sekä tiedon yhdistelmään, joka tähtää tuotteiden ja palveluiden hyödyntämiseen tähtäävien toimintojen ja prosessien kehittämiseen.
- *organisatorinen pääoma,* jolla tarkoitetaan eksplisiittistä ja implisiittistä, virallista ja epävirallista tietoa siitä, mikä on tehokas ja vaikuttava tapa strukturoida ja kehittää yrityksen organisatorista aktiivisuutta. Sisältää kulttuurin – implisiittinen ja epävirallinen tieto; rakenne – eksplisiittinen ja virallinen tieto; organisaation oppiminen – implisiittinen ja eksplisiittinen, virallinen ja epävirallinen tiedon uudistamisprosessi
- *liiketoimintapääoma,* joka viittaa niiden suhteiden arvoon, joita yrityksellä on muiden toimijoiden kanssa sen perusliiketoiminnan prosesseihin liittyen, kuten esimerkiksi asiakkaat, jakelijat, sekä kumppanit)
- *sosiaalinen pääoma,* joka tarkoittaa yrityksen muiden suhteiden arvoa muiden toimijoiden ja ympäristön kanssa

Tällainen aineettoman pääoman tarkka luokittelu mahdollistaa yritykselle paremman ymmärryksen aineettomasta pääomasta sekä organisatorisista tekijöistä, sekä mahdollisesti erilaisen aineettoman omaisuuden arvon määrittämisen. (Martín-de-Castro ym., 2005, s.4)

Yrityksen perusliiketoiminta, sen prosessit ja aktiviteetit tuottavat paljon informaatiota yrityksen käyttöön, mutta nykypäivän digitalisoituneessa toimintaympäristössä yritysten on oleellista huomioida se, että myös ulkoinen ympäristö tuottaa suuren määrän yrityksen toiminnan kannalta oleellista avointa

dataa. Ulkoisella ympäristöllä tarkoitetaan lähinnä sellaisia toimijoita, joilla on joko suoraan tai välillisesti vaikutusta yrityksen toimintaan. Yritykseen vaikuttavat ulkoiset tekijät voivat olla mitä tahansa, mutta yleisiä tällaisia tekijöitä voivat olla esimerkiksi nykyiset ja uudet asiakkaat, kilpailevat yritykset, täydentävät yritykset, informaatioympäristön toimijat (media, bloggarit, sekä muut mielipidevaikuttajat), sekä valtioilliset toimijat (verottaja, tulliviranomaiset jne.). (Dmitriev ym., 2021, s. 7)

2.4.2 Kilpailuetua datan avulla

Yrityksellä voidaan katsoa olevan kilpailuetua silloin, kun se toteuttaa sellaista arvonluonnin strategiaa, jota mikään muu yritys ei samanaikaisesti toteuta. Kestävää kilpailuetua yritys voi saavuttaa toteuttamalla sellaista lisäarvoa tuottavaa strategiaa, jota mikään nykyinen tai tuleva kilpaileva yritys toteuta, ja jonka etuja kilpailevat yritykset eivät myöskään pysty kopioimaan. (Barney, 1991)

Strategialla tarkoitetaan lyhyesti toimintasuunnitelmaa. Strategia on yritystoiminnan ja yrityksen menestyksen kannalta tärkeässä roolissa, ja nykypäivän globalisoituvassa maailmassa, jossa kaikilla aloilla on runsaasti toimijoita ja kova paikallinen sekä globaali kilpailu, strategian rooli korostuu entisestään. Shapiron (1989) mukaan John von Neumannin ja Oskar Morgensternin vuonna 1944 kehittämä (Ross, 2021) matemaattinen peliteoria on vaikuttanut teollisuuden alan organisaatiohin, ja niiden strategioiden analysointimenetelmiin 1980-luvulta lähtien. Lyhyesti määriteltynä peliteoria on strategisen vuorovaikutuksen oppi, jossa keskenään strategisessa vuorovaikutuksessa olevat, omaa etuaan tavoittelevat agentit pyrkivät toimintansa avulla tuottamaan sellaisia lopputuloksia, joista seuraa heille suurin mahdollinen hyöty. Peliteorian mukaan peliksi lasketaan sellaiset tilanteet, joissa vähintään yksi pelaaja voi pyrkiä maksimoimaan oman hyötynsä ennakoimalla vähintään yhden muun agentin vastineen omiin toimiinsa. Tärkeänä osana peliteorian mukaisiin pelitilanteisiin ja strategisiin valintoihin liittyy informaatio, joka pelaajalla on käytössään, kun hän valitsee strategioita. Peliteorian mukaan pelaajalla voi olla käyttössään täydellistä tai epätäydellistä informaatiota strategisia valintoja tehdessään. Tieto on täydellistä silloin, kun pelaajalla on valintaa tehdessään käytössään täydellinen tieto siitä, mitä pelissä on valinnan tekohetkeen mennessä tapahtunut. Epätäydellisen informaation pelissä puolestaan täytyy tehdä valintoja tietämättä kaikkea siihen mennessä tapahtunutta. (Ross, 2021) Shapiro (1989) puolestaan kirjoittaa liiketoimintastrategiateoriasta, jolla hän tarkoittaa joukkoa kilpailumalleja. Peliteorian lisäksi hän mainitsee liiketoimintastrategiateoriaan vaikuttaviksi malleiksi myös strategisen kilpailun teorian, sekä evoluutioteorian.

Yritykset keräävät jatkuvasti enemmän ja enemmän dataa, ja investoivat runsaasti datan keräämiseen ja analysointiin niin osaamisen, kuin teknologian muodossa. Datasta on tullut niin tärkeä osa yritysten liiketoimintaa sekä keskinäistä kilpailua, että yhä useammat yritykset muodostavat erillisen datastrategian itselleen. Fleckenstein & Fellows (2018) ovat julkaisseet modernia datastrategiaa käsittelevän kirjan, jossa käsitellään kattavasti datastrategian etuja

sekä vaatimuksia. Heidän mukaansa kilpailuedun saavuttaminen datan avulla vaatii koko yritystä koskettavan datastrategian, joka vaatii yhteistyötä yrityksen liiketoiminta- ja it-osastoilta, jotta tavoitteet ja kyvykkyydet voidaan ymmärtää. Kirja painottaa sitä, kuinka tärkeää on, että datastrategia luodaan oikeiden ihmisten toimesta. Ideaalitilanteessa datastrategian tulisi olla luotu dataa hyvin ymmärtävän henkilön ohjaamana, tai strategian luomisessa tulisi olla mukana vähintään yksi dataosaaja. Dataosaamisen lisäksi toimivan ja tehokkaan datastragian luomiseen tulisi löytää dataosaajien lisäksi prosessi- ja teknologiaorientoituneita osaajia, jotta strategia pystytään huomioimaan kaikki oleellinen.

Huolellisesti suunniteltu ja tehokkaasti hyödynnetty datastrategia tarjoaa yritykselle useita konkreettisia hyötyjä. Datan avulla yritys voi parantaa päätöksentekoa, liiketoimintaprosesseja, riskienhallintaa, kustannustehokkuutta, tuottojen optimointia (Marr, 2017). Kasvanut suorituskyky- ja varmuus mahdollistavat edellisessä kappaleessa esiteltyjen aineettomien omaisuusluokkien kasvattamisen.

Yrityksen päätöksenteko on tärkein yksittäinen yrityksen menestykseen vaikuttava tekijä, ja datan tehokkaalla hyödyntämisellä pystytään varmistamaan, että tehdyt päätöksen pohjautuvat mahdollisimman pitkälle tosiasioihin. Marr (2017) kirjoittaa kirjassaan datan tärkeydestä ja siitä, että datan tulisi olla jokaisen yrityksen päätöksenteon ytimessa huolimatta yrityksen koosta, tai toimialasta. Jotta yritys voi saada mahdollisimman suuren hyödyn käytettävissä olevasta datasta, täytyy tietää, mitä dataa tarvitaan. Jotta yritys voi tunnistaa tärkeimmän tarvitsemansa datan, on sen asetettava itselleen liiketoiminnan avainkysymykset, jotka auttavat datan tarpeen tunnistamisessa. Avainkymysten tarkoitus on antaa ymmärrys siitä, mitä yrityksen tarvitsee tietää saavuttaakseen tavoitteet. Marr luettelee kirjassaan neljä avainaluetta, joiden toiminnan tavoitteet ja avainkysymykset olisi hyvä tunnistaa. Nämä neljä aluetta ovat 1) asiakkaat, markkinat ja kilpailu 2) talous 3) sisäiset operaatiot 4) ihmiset. Avainkysymysten muodostaminen on aina samat vaiheet sisältävä prosessi : ensin täytyy määritellä avainalueta koskevat tavoitteet, ja sen jälkeen määritellä tavoitteisiin liittyvät avainkysymykset, eli toisin sanoen se, mitä täytyy tietää saavuttaakseen asetetut tavoitteet. Kysymyksiä listatessa lyhyt ja hyvin mietitty listaus on parempi, kuin pitkä listaus. Hyvien avainkysymysten tulee olla selkeitä ja yksinkertaisia, jotta niistä on apua.

3 TIEDON LUOKITTELU TIEDON SUOJAAMINEN NÄKÖKULMASTA

Kuten aiemmissa luvuissa on käsitelty, datan, informaation ja tiedon rooli yhteiskunnassa ja yritystoiminnassa on kasvanut valtavasti, ja teknologisen kehityksen myötä rooli tulee kasvamaan myös jatkossa. Kaikki yrityksen omistama tieto ei ole kuitenkaan keskenään saman arvoista. Osa tiedosta voi olla sellaista, jota yritys hyödyntää omassa liiketoiminnassaan, mutta joka kuitenkaan ole millään tavalla salassa pidettävää, jonka vuotamista yrityksen ulkopuolisille tahoille täytyisi jollain tavalla suojella. Toisaalta data tai tieto voi olla myös erittäin oleellista, sellaista tietoa, joka tarjoaa yritykselle kilpailuetua, ja tällaisissa tilanteissa yrityksen täytyy suojella tietoa ja estää sitä vuotamasta yrityksen ulkopuolisten tahojen tietoon. Lisäksi on olemassa paljon informaatiota, esimerkiksi työntekijöihin tai asiakkaisiin liittyen, jonka salassa pitämistä ja suojaamista säädellään lainsäädännöllä. Lainsäädännön tarkoitus on suojella yksilöitä ja kuluttajia luottamuksellisen tiedon asianmukaisesta suojaamisesta. Lähihistoria tuntee kuitenkin tapauksia, joissa yritys on käsitellyt asiakkaisiinsa liittyvää luottamuksellista tietoa, ja joka on joutunut vääriin käsiin. Tällaisissa tapauksissa laki velvoittaa yritystä tiedottamaan tietovuodosta asianmukaisesti asianomistajille, ja lisäksi korjaamaan puutteet, joiden vuoksi tietovuoto on päässyt syntymään. Yksi tunnetuimmista tietovuototapauksista Suomessa on psykoterapiakeskusta Vastaamon tapaus vuodelta 2020, jossa yrityksen asiakastietokantaan tehtiin kaksi tietomurtoa, joiden seurauksena kymmenien tuhansien ihmisten henkilö- ja potilastietoja päätyi rikollisten käsiin. Kyseisen tietovuodon vuoksi tietosuojavaltuutettu katsoi, ettei henkilötietoja ollut asianmukaisesti suojattu luvattomalta ja lainvastaiselta käsittelyltä, ja määräsi Vastaamolle 608 000 euron hallinnollisen seuraamusmaksun. (Yle, 2020) Tiedon luokittelu ja suojaaminen ovat muuttuneet valtavasti digitalisaation myötä. Siinä missä ennen, tiedon ollessa fyysisessä muodossa mapitettuna arkistohuoneisiin, riitti, että kaikista salaisin tieto oli säilötty omaan huoneeseensa, ja avaimia huoneeseen annettu vain niille henkilöille, joilla oli asemansa puolesta oikeus tiedon käyttämiseen, nykypäivän digitaalisessa toimintaympäristössä periaatteet ovat täysin samat, mutta toimintaympäristö on muuttunut niin, että tietoa säilytetään palvelimilla ja tietokannoissa, huoneen oven ja

lukituksen ollessa sovellusohjelma, ja ohjelman vaatima käyttäjätunnus ja salasana, joiden avulla käyttäjän oikeudet automaattisesti tunnistetaan, ja käyttöoikeuksien riittäessä käyttäjä päästetään sisälle.

Tässä luvussa käsitellään tiedon luokittelua tiedon suojaamisen näkökulmasta. Luvussa käydään läpi mitä tiedon luokittelu sekä yrityksen tiedon luokittelua määrittelevä tietoturvan johtamisjärjestelmä tarkoittavat. Lisäksi käsitellään sitä, millaisia etuja tiedon luokittelu tarjoaa erilaisille tiedon suojaamisen osa-alueille, sekä millaiset lait ja standardit vaikuttavat tiedon luokitteluun yrityksissä.

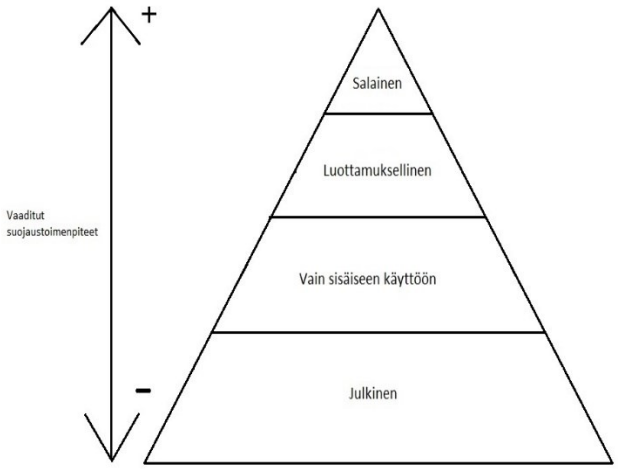
3.1 Tiedon luokittelu

Tiedon luokittelulla yritystoiminnan kontekstissa tarkoitetaan prosessia, jossa yritys arvioi ja luokittelee omistamansa datan. Yritykset voivat luokitella omistamaansa dataa millä perusteella tahansa, mutta yleisesti luokittelua tehdään perustuen datan liiketoiminnalliseen arvoon sekä luottamuksellisuuteen. Yksi tiedon luokittelun tärkeimmistä tavoitteista on varmistaa, että kaikelle yrityksen omistamalle datalle voidaan turvata riittävä suojaus riippuen sen tärkeydestä yritystoiminnalle (ISO/IEC 27002, 2017). Tiedon luokittelu lukeutuu yrityksen tietoturvakokonaisuudessa osaksi omaisuudenhallintaa, joka puolestaan lukeutuu osaksi tietoturvan johtamisjärjestelmää (ISMS, information security management system). (Bergström ym., 2021) Tietoturvan johtamisjärjestelmää sekä sen sisältämiä toimintaohjeita tiedon luokittelun suhteen käydään tarkemmin läpi tämän pääluvun seuraavassa kappaleessa.

Yritystoiminta synnyttää jatkuvasti uutta ja uudenlaista dataa, ja näin ollen tiedon luokitteleminen riittävän tietoturvan ja tietosuojan takaamiseksi onkin jatkuva prosessi. Koska yrityksen omistaman, erityisesti yksityishenkilöitä koskevan henkilökohtaisen datan säilyttämistä ja suojaamista säädellään laissa, on yrityksen kyettävä luokittelemaan dataa ja tietoa toiminnan alusta lähtien. Osana kokonaistietoturvaa, tulee tiedon luokittelulle myös luoda oma erillinen politiikkansa, jonka mukaan yritys toteuttaa tiedon luokittelua.

Tiedon luokittelun politiikkojen ja toimintatapojen rakentaminen on prosessi, joka koostuu eri vaiheista (Calder & Watkins, 2019). Calder & Watkins (2019) mukaan ensimmäisenä tulee tunnistaa kaikki informaatio-omaisuuserät, joiden tulee olla sisältyä osaksi tietoturvan johtamisjärjestelmää. Tunnistamisen yhteydessä tiedolle tulee myös nimetä henkilö tai tiimi, joka tiedon ”omistaa”. Tiedon tunnistamisen jälkeen tieto ryhmitellään riskiarviointia varten. Tiedon joukosta tulee pyrkiä löytämään ryhmiä, joiden suojausvaatimukset ovat keskenään samankaltaisia, jotta kaikelle tiedolle ei tarvitse erikseen määritellä suojaustoimenpiteitä. Tiedon ryhmittelyn jälkeen on tärkeätä tunnistaa tietojen keskinäisiä riippuvuuksia. Riippuvuudet täytyy tunnistaa siksi, että yksinään tietty tieto ei välttämättä yksin ole arvokasta, mutta yhdistettynä toiseen tietoon, ne voivat yhdessä muodostaa toiminnan kannalta erittäin oleellisen kokonaisuuden. Jonkin tärkeän prosessin tulosteiden (output) eheys voi olla riippuvainen siihen

tulevien syötteiden (input) eheydestä, joten jälkimmäisen tiedon laatu on riippuvainen ensimmäisen laadusta. Jokin korkeaa luottamuksellisuutta vaativa informaatio-omaisuus voi riippua toisesta, yksinään vähemmän tärkeästä tiedosta, ja siitä syystä myös vähemmän tärkeälle tiedolle täytyy taata korkeampi suojaustaso. Seuraavaksi kaikelle yrityksen tiedolle tulee määritellä omistaja, joko yksilö, tiimi tai liiketoimintayksikkö, joka on vastuussa tiedon suojaamisesta. Tiedolla voi usein olla useita käyttäjiä ja haltijoita, mutta tiedon omistaja on se taho, joka on viime kädessä vastuussa tiedon säilyttämisestä organisaation sisällä. Viimeisenä tiedon omistajan vastuulla on luokitella ”omistamansa” tieto arkaluontoisuuden perusteella. Tietoa luokiteltaessa omistajan täytyy ottaa huomioon tietoon liittyvät lailliset vaatimukset, arvo, kriittisyys sekä herkkyys luvattomalle levittämiselle tai muokkaamiselle. (Calder & Watkins, 2019) Yrityksen omistaman alkuperäisen tiedon lisäksi tiedosta otetut varmuuskopiot sekä duplikaatit vaativat myös suojausta. Calderin & Watkins (2019) mukaan varmuuskopiota sekä duplikaatteja voidaan käsitellä yhtenä tietoryhmänä. Kuvio 2 esittää yhden havainnollistavan esimerkin yritysten ja organisaatioiden omistaman luottamuksellisuudeltaan eri tasoisen datan jakauman, sekä sen, miten suojausvaatimukset muuttuvat, kun data muuttuu luottamuksellisemmaksi. Kuten kuviosta on nähtävissä, usein suurin osa omistetusta datasta on yleistä dataa, jonka luottamuksellisuus on pientä. Toisessa ääripäässä on puolestaan kaikista luottamuksellisin data, jollaista yleensä on pienin osa yrityksen kaikesta datasta. Kuvio myös osoittaa, että datan suojausvaatimukset kasvavat sitä mukaan, mitä luottamuksellisempaa data on. Tämä aiheuttaa myös sen, että kaikista luottamuksellisin data täytyy pystyä luokittelemaan tarkasti.



KUVIO 2 Tiedon luokat ja suojausvaatimukset

Tiedon luokittelun tarjoamista lukuisista hyödyistä huolimatta tiedon luokittelu on usein aliarvostettua. Veritas Technologiesin (2016) laatiman raportin

mukaan vuonna 2016 vain 48 % kaikesta datasta on luokiteltua ja 54 % luokittelematonta dataa. Bergström ym. (2021) mukaan luokittelematon data johtaa yrityksissä tilanteisiin, joissa työntekijä ei pääse käsiksi sellaiseen dataan, johon hänen kuuluisi päästä, tai vaihtoehtoisesti hänellä on pääsy sellaiseen dataan, johon pääsyä ei kuuluisi olla. Bergström ym. (2021) luettelevat aliarvostuksen perustuvan useaan eri syyhyn, joita ovat mm. kestävien sekä vaatimustenmukaisten tiedon luokittelumallien määrittämisen ja soveltamisen, sekä samanaikaisen riittävän joustavuuden tarjoamisen vaikeus, riittävän selkeän ohjeistuksen sekä viitekehysten puute, sekä tarjolla olevien ohjeiden ja standardien käytännön soveltamisen vaikeus. Tiedon luokittelun painoarvoa on pyritty nostamaan, ja esimerkiksi vuonna 2016 annettu EU:n yleinen tietosuoja-asetus GDPR onkin hieman nostanut tiedon luokittelun arvostusta. Standardeja ja lainsäädäntöä käsitellään tarkemmin vielä tämän luvun luvussa 3.3.

3.2 ISMS (information security management system)

ISMS (information security management system) tarkoittaa tietoturvan hallintaja johtamisjärjestelmää. ISMS:n tarkoituksena on toimia tietoturvan hallinnollisena työkaluna. ISMS on esitelty ensimmäisen kerran 1990-luvulla samaan aikaan ensimmäisen, puhtaasti tietoturvaa koskevan British Standard – BS7799 standardin kanssa. (Broderick, 2006) Nykypäivänä ISMS:n pohjana kuitenkin toimii ISO27001 standardi, jossa on lueteltu tarkat vaatimukset ISMS implementaatiolle (Calder, 2013).

ISMS on kokoelma toimintaperiaatteita sekä menettelytapoja, joiden tarkoituksena on taata yritykselle määrämuotoinen tapa hallita tietoturvaa, sekä osoittaa tietoturvan määrämuotoisuus ulospäin myös sidosryhmille. Yksi tärkeä osa ISMS:ssa on informaatio-omaisuuden hallinta, johon sisältyy organisaation informaatio-omaisuuden tunnistaminen, omistussuhteiden määrittely sekä informaatio-omaisuuden suojaamisen vastuut. (Bergström ym., 2021)

ISMS:n rooli yrityksen tietoturvan kannalta on toimia sisäisenä karttana ja ohjekirjana tietoturvan johtamiselle sovellettuna siihen ympäristöön ja niihin olosuhteisiin, joissa yritys toimii. Ensisijaisesti ISMS on tarkoitettu yrityksen johdon työkaluksi varmistamaan, että yrityksen dataa ja informaatiota hallitaan lakien ja vaatimusten mukaan sekä näyttämään toteen, että yritys täyttää tietoturvaan liittyvät asianmukaisuus- sekä luottamuksellisuusvaatimukset, joita heidän omistamalleen datalle ja informaatiolle asetetaan. (Broderick, 2006) Broderick (2006) toteaa myös, että vaikka ISMS on yritysjohdon työkalu, on yrityksen pystyttävä sitouttamaan työntekijät kaikilla tasoilla mukaan johdosta aina järjestelmien ylläpitäjiin ja käyttäjiin noudattamaan ISMS:n vaatimusten mukaisia toimintatapoja. Vaatimustenmukaisen toiminnan ylläpitäminen vaatii puolustautumisen ja prosessien harjoittelua datan ja informaation suojaamiseksi.

3.2.1 ISM (information security management)

Tietoturvajohtamisella tarkoitetaan kokonaisvaltaista, liiketoiminnan etujen mukaista informaation suojaamista läpi koko organisaation. Tietoturvajohtamisesta on tullut strateginen osa yrityksen liiketoimintaa, jonka tavoitteena on tuottaa yritykselle liiketoiminnallista etua suojaamalla sekä helpottamalla informaation hallittua jakamista sekä hallitsemalla tietoturvariskejä jatkuvasti muuttuvassa toimintaympäristössä. (Ashenden, 2008) Yhteiskunnan digitalisaatio ja teknologinen kehitys aiheuttavat yrityksille jatkuvasti uusia haasteita ja vaatimuksia tietoturvan näkökulmasta. Tietoturvakysymyksiin liittyy vahvasti teknologinen näkökulma, mutta kehityksen jatkuessa tietoturvasta tulee kuitenkin yhä enemmän yrityksen sisäinen kokonaisvaltainen toiminto, joka koostuu teknologiasta, prosesseista ja ihmisistä (Ashenden, 2008).

Tietoturvajohtamisen tärkeys kasvaa jatkuvasti, ja Soomro ym. (2016) kirjoittavatkin, että johtamisesta on tullut tärkein yksittäinen tekijä yritysten tietoturvassa. Heidän mielestään tietoturvan johtamisen ei tulisi rajoittua vain teknisten- tai informaatioasiantuntijoiden vastuulle, vaan tietoturvan johtamista tulisi käsitellä yrityksen ylimmän johdon toimesta.

Tietoturvajohtamista käsittelevä kirjallisuus ja tutkimus arvostavat johtamisen jopa tärkeimmäksi yksittäiseksi osaksi yritysten tietoturvakokonaisuutta, mutta silti on hyvä muistaa, että teknologiset ratkaisut ovat myös oleellisessa osassa tietoturvan toiminnan kannalta. Tietoturvajohtamisen haaste onkin kyetä räätälöimään paras mahdollinen toimintatapojen ja prosessien sekä teknologisten ratkaisujen kokonaisuus, joka soveltuu juuri oman yrityksen tarpeeseen. Yrityksen tietoturvatarpeisiin vaikuttavat mm. yrityksen toimiala, koko sekä rakenne. Tehokas tietoturva vaatii toimiakseen sellaisen rakenteen, joka tukee raportointia, kommunikointia, selkeää määräysvaltaa sekä tehokkaita työnkulkuja. (Soomro ym., 2016)

Tietoturvanjohtamisen strateginen tärkeys korostuu lähdemateriaalissa, mutta johtaminen kätkee kuitenkin sisäänsä useita konkreettisia toimenpiteitä, jotka vaikuttavat siihen, miten tietoturvajohtaminen onnistuu. Tällaisia toimenpiteitä ovat

- Tietoturvabudjetin ja resurssien määrittely
- Tietoturva-arkkitehtuurin suunnitteleminen
- Teknologisten ja johdollisten toimenpiteiden yhteensovittaminen
- Yritysjohdon mukaan ottaminen tietoturvan liiketoimintaan liittyvien näkökohtien muotoiluun
- Informaatioriskien systemaattinen analysointi ja hoitaminen
- Tietoturvan hallintatoimenpiteiden implementointi sekä näiden toimenpiteiden mittaaminen, monitorointi ja tarkastelu
- Tietoturvakeskustelun ylläpitäminen sekä yhteydenpito sidosryhmien ja viranomaisten kanssa
- Tietoturvallisten toimenpiteiden ja toimintatapojen jalkauttaminen organisaatioon

- Tietoturvaympäristön jatkuva kehittäminen

(Ashenden, 2008.; Soomro ym., 2016; O'Hanley & Tiller, 2014)

3.2.2 DAC (Data access management)

Pääsynhallinta tarkoittaa sääntöpohjaista järjestelmää, joka päättää tunnistettujen käyttäjien datan käyttöoikeuksista (Kiran & Nalini, 2020). Pääsynhallinta on erittäin tärkeässä roolissa yrityksen kokonaistietoturvan kannalta, sillä väärin toteutettu pääsynhallinta avaa väärille käyttäjille mahdollisuuden käsitellä sellaista dataa, jonka käsittelyyn heillä ei pitäisi olla oikeutta, tai toisaalta estää sellaisen käyttäjän pääsyn dataan, johon hänellä kuuluisi olla pääsy.

Pääsynhallinnan toteuttamiselle on runsaasti erilaisia malleja ja periaatteita. Tiettävästi ensimmäisen pääsynhallintamallin, pääsymatriisin esitteli Butler W. Lampson vuonna 1969 (Barbosu ym. 2012). Nykypäivänä kuitenkin yleisin malli pääsynhallinnassa on rooliperusteinen pääsynhallinta. Rooliperusteisessa pääsynhallinnassa pääsyä erilaiseen ja eritasoiseen dataan on rajoitettu, ja käyttäjäoikeuksien jakamiseen hyödynnetään erilaisia rooleja. Datan käsittelyoikeuksia kontrolloidaan roolitasolla, eikä käyttäjätasolla, ja näin ollen tiettyyn roolin saava käyttäjä saa automaattisesti pääsyoikeuden roolin oikeuttamaan dataan. (Delphin Carolina Rani ym., 2022) Tietylle roolille voidaan antaa esimerkiksi lupa ainoastaan lukea tietyn turvaluokituksen dataa, mutta jättää kirjoitus- tai muokkausoikeus antamatta.

Tiedon luokittelu on erittäin tärkeää myös pääsynhallinnan näkökulmasta. Pääsynhallintaa voidaan tehdä ilman tiedon luokittelua, mutta tiedon luokittelu mahdollistaa huomattavasti paremman, kevyemmin hallittavan sekä hienojakoisemman pääsynhallinnan. Organisaatio voi esimerkiksi luokitella sen omistamaa tiettyä dataa luottamukselliseksi, ja antaa vain tietylle roolille pääsyn kaikista luottamuksellisimpaan dataan. Tämän jälkeen, kun jokin dokumentti luokitellaan luottamukselliseksi, rajaa se automaattisesti pääsyn tietylle käyttäjäryhmälle. Ilman tiedon erilaisia luokkia, pääsynhallinta vaatii huomattavasti enemmän käsityötä, ja käyttäjäryhmien pääsyoikeuksia voidaan joutua määrittelemään manuaalisesti jopa yksittäisille dokumenteille. Tästä syystä tiedon luokittelu on äärimmäisen tärkeää tehokkaan ja toimivan pääsynhallinnan näkökulmasta

3.2.3 ILM (information lifecycle management)

Yritystoiminnassa tiedon elinkaaren hallinnalla tarkoitetaan yrityksen tavoitteista johdettua strategiaa datan ja tiedon hallinnalle datan elinkaaren eri vaiheissa. Datan monimuotoisuuden ja tiedon käsittelyyn käytettävien ohjelmistojen määrän kasvaessa sekä tiedon varastointiratkaisujen kehittyessä on tärkeää ymmärtää, millaisesta datasta yrityksen informaatio-omaisuus koostuu. Lisäksi keskenään eriarvoiselle datalle tärkeää räätälöidä suojaus- ja käsittelypolitiikkoja sen tärkeyden sekä elinkaaren vaiheen mukaan. (Barta ym., 2004, s. 1-3)

Datan elinkaarelle on annettu erilaisissa lähteissä runsaasti hieman toisistaan poikkeavia vaiheita, vaiheiden nimiä sekä sisältöjä. Yleisesti datalle on kuitenkin tunnistettavissa viisi erilaista elinkaaren vaihetta (Sheldon, 2022).

- Datan luominen: Data luodaan käyttäjien, laitteiden, sovellusten sekä koneiden toimesta
- Datan säilyttäminen: Luotu data täytyy säilöä ympäristöön, esimerkiksi tietokantaan, jossa se täytyy säilyttää varmistaen sen yhtenäisyys, turvallisuus ja suojaus.
- Datan käyttäminen ja jakaminen: Tässä vaiheessa käyttäjät lukevat, muokkaavat, yhdistävät, jakavat ja uudelleenkirjoittavat dataa. Dataa voidaan käyttää esimerkiksi analytiikkaan tai visualisointiin
- Datan arkistointi: Data arkistoidaan, kun sille ei ole tarvetta enää yrityksen aktiivisessa, päivittäisessä toiminnassa. Arkistointi tapahtuu siirtämällä data pitkäaikaiseen säilöön, turvattuun erilliseen levylle esimerkiksi pilvipalvelussa. Arkistoidun datan säilöminen on edullisempaa verrattuna käytössä olevan datan säilyttämiseen, mutta arkistoidun datan lukeminen kestää kauemmin käytössä olevaan dataan verrattuna.
- Datan tuhoaminen: Datalle voidaan määritellä aika, miten kauan sitä säilytetään, ja kun aika tulee täyteen, data tuhotaan. Arkistointipolitiikoilla voidaan määrittää, kuinka kauan dataa säilytetään ennen hävittämistä

Dataan ja sen suojaamiseen kohdistuu erilaisia suojaamisen piirteitä erilaisissa elinkaaren vaiheissa. Taulukko 1 näyttää, millaisia tietoturvan erityispiirteitä liittyy datan luomiseen, säilömiseen ja tuhoamiseen. Lee & Yong (2021, s. 13-17) ovat listanneet erityisesti IoT-laitteiden tuottamasta arkaluontoisen henkilödatan erilaisten elinkaaren vaiheiden tietoturvaan liittyviä erityispiirteitä, piirteet ovat johdettavissa myös muuhun, kun IoT-laitteiden luomaan, luottamukselliseen dataan.

TAULUKKO 1 Datan elinkaaren vaiheisiin liittyvät tietoturvapiirteet

Elinkaaren vaihe	Tietoturvapiirteet
Luominen	Sanomien autentikaatio Kiistämättömyys Integriteetti Luottamuksellisuus
Säilyttäminen	Pääsyoikeuksien hallinta Datan omistajuus Luottamuksellisuus Saavutettavuus
Tuhoaminen	Datan omistajuus Pääsyn hallinta

Jatkuvasti kasvava ja monipuolistuva data aiheuttaa sen, että tietyn datan arvo saattaa vaihdella ajan saatossa, eikä sen arvon kehitystä välttämättä pystytä ennakoimaan (Barta ym., 2004, s. 1-3). Barta ym. (2004) kirjoittavat, että arvon kehityksen ennakoinnin vaikeuden vuoksi datan hallinta täytyy olla proaktiivista, ja että hyvä elinkaaren hallintastrategia täytyy olla liiketoimintalähtöinen, politiikkoihin pohjautuva, keskitetysti hallittu, heterogeeninen sekä linjassa datan arvon kanssa. Myös Garcia ym. (2016, s. 101-103) kirjoittavat nykypäivän datakeskeisen ympäristön aiheuttavan kasvavan tarpeen datan elinkaaren hallinnalle. Elinkaaren hallintamalli tarjoaa lukuisia etuja organisaatiolle, Garcia ym. (2016) listaavat konkreettisimmiksi hyödyiksi monimutkaisen datan hallinnan ja suunnittelun helpottuminen koko organisaatiossa, laadukkaat datatuotteet loppukäyttäjille, korkealaatuinen data ilman hukkaa tai hälyä, tärkeiden aktiviteettien tarpeen tunnistaminen läpi koko datan elinkaaren sekä ohjelmoijien kestävien ja tehokkaiden ohjelmistojen luomisen helpottuminen.

Elinkaaren hallinta alkaa datan ja tiedon luokittelulla perustuen datan liiketoiminnalliseen arvoon sekä luottamuksellisuuteen (Barta ym., 2004). Barta ym. (2004) puhuvat tiedon luokittelun yhteydessä informaatioryhmistä, joissa kuvataan tiedostot, hakemistot, levyasemat, tietokannat sekä taulut, joissa tietyn ryhmä dataa sijaitsee. Tämän jälkeen informaatioryhmille määritellään mm. miten nopeasti informaatioryhmän data täytyy saada palautettua, jos se häviää, miten data täytyy suojata fyysisellä ja loogisella tasolla, miten nopeasti data täytyy olla saatavilla arkistosta, täytyykö ryhmän dataa arkistoida erikseen viranomaistarpeita varten, sekä se, miten kauan yritystoiminta voi jatkua kannattavana ilman tietyn ryhmän dataa. Tämän vuoksi tiedon luokittelulle on erittäin oleellinen rooli sujuvassa datan elinkaaren hallinnassa. Elinkaaren hallinnan tavoin tiedon luokittelulle täytyy olla strategiassa määritellyt, säännöllisesti tehtävät ylläpitotoimenpiteet.

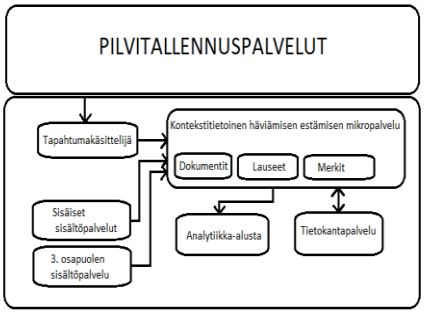
3.2.4 DLP (Data loss prevention)

Datan häviämisen estäminen tarkoittaa toimintamalleja, strategioita ja teknologioita, joiden yhteistoiminnan avulla pyritään estämään arkaluontoisen ja luottamuksellisen datan häviäminen, vuotaminen sekä varastaminen (Cao ym. 2020, s. 401). Datan häviämisen estäminen on erityisen tärkeää sellaisissa yrityksissä ja organisaatioissa, jotka käsittelevät yksityishenkilöihin yksilöitävissä olevaa arkaluontoista dataa, kun taloudellista tai terveydellistä dataa.

Datan häviämisen estäminen tapahtuu usein hyödyntämällä teknologioita ja tekoälyä, jotka osaavat tunnistaa yrityksen datasta automaattisesti arkaluontoisen datan, kuten henkilötunnukset, tai maksukorttien tiedot. Teknologiat mahdollistavat esimerkiksi luottamuksellisuutta kuvaavien avainsanojen automaattisen etsimisen, säännöllisten ilmausten (regular experssions) hyödyntämisen, poikkeamien tunnistamisen, koneoppimismallien kouluttamisen tunnistamaan käyttäjien epäilyttävän toiminnan. (Cao ym. 2020)

Ong ym. (2017) esittelevät pilvialustoille soveltuvan kontekstitietoisen datan häviämisen estämisen mallin (Context Aware Data Loss Prevention), joka on

rakennettu mikropalveluiden avulla, jotka pystyvät reaaliaikaisesti palauttamaan datan tunnistustulokset joko dokumentti-, lause-, tai merkkitasolla käyttäjän tarpeen mukaan. Datan häviämisen estämisen on tarjolla runsaasti erilaisia työkaluja. Ong ym. (2019) sekä Cao ym. (2020) esittelevät erityisesti pilvialustoille sopivia DLP-työkaluja, ja näiden molempien arkkitehtuurit sisältävät samanlaisia komponentteja. Pilviympäristöön tarkoitetut datan häviämisen estämisen työkalut koostuvat arkkitehtuuritasolla mm. tapahtumien ja aktiviteettien käsittelijöistä, analytiikka-alustasta, joka tunnistaa datan sisältöä, sekä tunnistettujen tapahtumien ja sisältöjen luokittelijoista. Lisäksi kaikkiin työkaluihin sisältyy omia erityispiirteitään sekä keskinäisriippuvuuksia.



KUVIO 3 Konstekstitietoinen datan häviämisen estämisen arkkitehtuuri (Ong ym., 2017)

Datan häviämisen estämisen yksi tärkeä osa on tiedon ja datan luokittelujen luominen ja ylläpitäminen, sillä ilman luokkia järjestelmät eivät voi kohdentaa suojaus ja käsittelytoimenpiteitä, vaikka ne tunnistaisivatkin kaiken käsiteltävän datan joukosta arkaluontoista tai luottamuksellista dataa. Pilvialustoille suunnitellut datan häviämisen estämisen luokittelijat hyödyntävät myös koneoppimista luokittelujen automaattiseen tekemiseen (Cao ym., 2020)

3.3 Lainsäädäntö ja standardit

Lait ja säädökset luovat perustan yhteiskunnan toiminnalle. Lakien avulla määritellään yleisesti se, mitä saa tehdä ja mitä. Myös yritystoimintaa säädellään lakien avulla, mutta sen lisäksi yritysten toimintaa ohjaamaan on olemassa myös erilaisia standardeja, joissa määritellään tietyn toiminnan yhteisesti sovittuja vaatimuksia tai suosituksia. Suomen Standardisoimisliitto SFS RY kuvailee standardeja kotisivuillaan seuraavasti:

- *"Standardi on kirjallinen julkaisu, jossa määritetään esimerkiksi tuotteiden ja palvelujen ominaisuuksia ja vaatimuksia tai järjestelmien toimintaa."*
- *"Standardisointi on yhteisten toimintatapojen – hyvien käytäntöjen, ratkaisujen ja vaatimusten – laatimista. Standardisointiin saa osallistua kuka tahansa alan asiantuntija, ja standardisoinnin tuloksena syntyy edellä mainittuja asiakirjoja."*

Tämän työn aiheena on tiedon luokittelu tiedon suojaamisen ja liiketoiminnan näkökulmasta. Myös yritysten harjoittamaa tiedon luokittelua ohjaamaan on kehitetty erilaisia lakeja ja standardeja, joiden tavoitteena luoda puitteet ja säännöt sille, miten yritysten tulee luokitella ja suojata omistamaansa dataa.

Ainoastaan tiedon luokittelua ja eri tasoisen tiedon suojaamiseen liittyvää standardia ei ole olemassa, vaan tiedon luokitteluun ja suojaamiseen liittyvät standardimuotoiset ohjeet ja viitekehykset koskevat laajemmin tietoturvaa sekä tietoturvan johtamisjärjestelmää ISMS:ää. Ensimmäinen ISMS:ää koskeva standardi, BS-7799, on julkaistu vuonna 1995. Ennen BS-7799:ää tietoturvan johtamisjärjestelmän kontekstiin sovellettiin BS-5750 sekä ISO-9000 standardeja. Ne olivat kuitenkin vain osittain sovellettavissa tietoturvan johtamisjärjestelmään, joten BS-7799 oli ensimmäinen täysin ISMS:ää määrittelevä standardi. (Broderick, 2006)

Kansainvälisesti sovelletuin ja tunnetuin tietoturvan standardiperhe on ISO27000-standardiperhe. Standardiperheeseen kuuluvat ISO27001, ISO27002, ISO27003, ISO27004, ISO27005 sekä ISO27006 standardit. (Broderick, 2006)

- ISO27001:2013 standardi on uusin ISMS:ää määrittelevä standardi. Se on toimittajavapaa sekä teknologiariippumaton standardi, ja sen tarkoituksena on olla sovellettavissa millaiselle organisaatiolle tahansa kaikilla sektoreilla ja missä päin maailmaa tahansa.

- ISO27002:2013 standardi on tietoturvatekniikoihin keskittyvä standardi, joka sisältää tietoturvan johtamisen menettelyohjeita

- ISO 27003:2010 standardi käsittelee tietoturvan johtamisjärjestelmän implementointia.

- ISO27004:2009 standardi on kehitetty auttamaan organisaatioita vastaamaan tehokkaammin ISO27001 standardissa asetettuihin vaatimuksiin koskien tietoturvan valvontaa

- ISO27005:2011 standardi on tietoturvan johtamista käsittelevä standardi

- ISO27006:2011 standardi asettaa vaatimuksia toimijoille, jotka tarjoavat tietoturvan johtamisjärjestelmän auditointeja ja sertifikointeja

Standardien lisäksi yritysten tiedon käsittelyä, luokittelua ja säilyttämistä valvotaan ja ohjataan lainsäädännön avulla. Lainsäätäjien rooli on haastava, sillä teknologisen kehityksen perässä pysyminen ja lainsäädännön ajan tasalla pitäminen on vaikeaa. Uusin ja tällä hetkellä kattavin yksilönsuojaa ja tiedon suojaamista käsittelevä laki on EU:n vuonna 2018 asettama yleinen tietosuoja-asetus, The General Data Protection Regulation, GDPR. (Barber & Zaeem, 2020) Yleinen tietosuoja-asetus asettaa tarkkoja vaatimuksia yrityksille ja organisaatioille henkilötietojen keräämistä, säilytystä ja hallinnointia koskien. Asetuksen vaatimuksia sovelletaan sekä eurooppalaisiin organisaatioihin, jotka käsittelevät ihmisten henkilötietoja EU:ssa, että EU:n ulkopuolisiin organisaatioihin, joiden suorittama tietojen käsittely kohdistuu EU:n alueella asuviin ihmisiin. (Your Europe, 2021) Tietosuoja-asetuksessa määritellään tarkat ohjeet sille, milloin ja miten yritys saa käsitellä ja säilöä henkilötietoja. Lisäksi asetus määrittelee myös sanktiot ja toimenpiteet, joita tietosuoja-asetuksen rikkominen aiheuttaa. Tärkeimpiä kohtia tietosuoja-asetuksessa ovat mm. henkilötietojen sisältö, erityiset tietoryhmät, henkilötietojen käsittelyn valvonta yrityksessä, tietojen siirto EU:n ulkopuolelle, tietojenkäsittelyn edellytykset, tietojen käsittelyn avoin tiedottaminen, lapsia koskevat erityissäännöt, tietoihin pääsemisen ja tietojen siirtämisen oikeus, tietojen korjaamisen oikeus, tietojen poistamisen oikeus, profilointi, tietoturvaloukkaus, loukkauksen vaikutustenarviointi ja tietojen käsittelyn dokumentointi ja sääntöjen rikkominen ja sen seuraamukset (Your Europe, 2021)

4 TEORIAN YHTEENVETO

Maailmassa tuotetun datamäärän kiihtyvä kasvu avaa yrityksille runsaasti uusia liiketoiminnan mahdollisuuksia. Samaan aikaan tilanne on kuitenkin se, että kaiken, mukaan lukien tärkeiden yritystason salaisuuksien sekä tavallisten kuluttajien arkaluontoisen datan ollessa sähköisessä muodossa, aiheuttaa se yrityksille paljon vaatimuksia tiedon käsittelyn, säilyttämisen ja suojaamisen suhteen. Yksi hyödyllinen, mutta aliarvostettu menetelmä tietosuojan parantamiseen on tiedon luokittelu. Tiedon luokittelun avulla yritysten on mahdollista kohdentaa rajallisia tietoturvaresurssejaan tehokkaammin, kun kaikesta omistuksessa olevasta tiedosta kyetään erottomaan korkean ja matalan suojausvaatimuksen tietoryhmät (Calder & Watkins, 2019)

TAULUKKO 2 Avainkäsitteet

Käsite	**Määritelmä**	**Lähde**
Data	Joukko arvoja, jotka muunnettu siirrettävään ja käsiteltävään, eli binääriseen ja digitaaliseen muotoon	Brock, D. & Laws, D. (2012)
Informaatio	Tiettyyn kontekstiin liittämällä merkityksen saanut data	Buckland, 1991
Tiedon luokittelu	Tiedon jakamista ryhmiin luokittelemalla sitä jonkin ominaisuuden (esim. luottamuksellisuus) perusteella	Bergström ym., 2017

	soveltamalla jotakin ennalta valittua mallia tai kaavaa	
Tietosuojan hallintajärjestelmä	Kokoelma toimintaperiaatteita sekä menettelytapoja, joiden tarkoituksena on taata yritykselle määrämuotoinen tapa hallita tietoturvaa	Broderick, 2005
Tiedon elinkaaren hallinta	Kokoelma toimintatapoja, joiden avulla hallitaan yrityksen järjestelmien ja prosessien tuottamaa dataa koko elinkaaren ajan datan luomisesta ja siihen asti, kunnes datasta tulee tarpeetonta	Chen, Y. (2005)
Datan pääsyn hallinta	Sääntöpohjainen järjestelmää, joka päättää tunnistettujen käyttäjien datan käyttöoikeuksista	(Kiran & Nalini, 2020).
Datan häviämisen estäminen	toimintamalleja, strategioita ja teknologioita, joiden yhteistoiminnan avulla pyritään estämään arkaluontoisen ja luottamuksellisen datan häviäminen, vuotaminen sekä varastaminen	(Han ym. 2020)

5 TUTKIMUKSEN TOTEUTUS

Tässä luvussa esitellään tutkimuksen empiiristä osuutta. Luvussa kerrotaan tutkimuksen lähtökohdista, tavoitteista, tiedonkeruu- sekä analyysimenetelmistä, sekä lisäksi perustellaan tehtyjä valintoja tutkimuksen eri vaiheita koskien. Luvun tavoitteena on kertoa empiirisen osan toteuttamisesta, mutta pyrkiä myös tarkastelemaan objektiivisesti tehtyjä valintoja ja tunnistaa vahvuudet ja mahdolliset heikkoudet, joita tehtyihin valintoihin liittyy.

5.1 Tutkimusmenetelmä ja tavoitteet

Tämän tutkimuksen tavoitteena on saavuttaa syvällistä ymmärrystä tiedon luokittelusta tiedon suojaamisen näkökulmasta ja tiedon luokittelun merkityksestä yritysten liiketoiminnalle. Tutkimuksen ja valittujen menetelmien avulla pyrittiin saavuttamaan ymmärrys siitä, mitä yritysten tiedon luokittelusta vastuussa olevat henkilöt ajattelevat tiedon luokittelusta: millaisella tasolla tiedon luokittelua tehdään ja miksi, miten tiedon luokittelua hallitaan ja ylläpidetään, miten tärkeänä tiedon suojaamisen osana tiedon luokittelu nähdään, millaisia liiketoiminnallisia hyötyjä tiedon luokittelun koetaan tarjoavan sekä mahdollisia muita motiiveja, jotka vaikuttavat siihen, miten tiedon luokittelua tehdään. Tavoitteena oli käydä vuoropuhelua kerätyn tutkimusaineiston ja siitä tehtyjen löydösten, sekä kirjallisuuskatsauksesta johdettujen teemojen välillä, ja pyrkiä löytämään niiden välillä toisiaan vahvistavia tekijöitä, mutta myös mahdollisia eroja teorian ja empiiristen löydösten välillä.

Tutkimuksen tavoitteiden perusteella tutkimusmenetelmäksi valittiin kvalitatiivinen, eli laadullinen tutkimusmenetelmä. Laadulliselle tutkimukselle tyypillistä on, että siinä ei pyritä tilastollisiin yleistyksiin, vaan sen sijaan pyritään kuvaamaan ilmiötä tai tapahtumaa, ymmärtämään tiettyä toimintaa tai antamaan teoreettisesti mielekäs tulkinta jollekin ilmiölle (Tuomi & Sarajärvi, 2018, s. 73). Määrällisten eli kvantitatiivisten menetelmien avulla puolestaan selvitetään lukumääriin ja prosenttiosuuksiin liittyviä kysymyksiä. Määrällisen

tutkimuksen avulla voidaan testata tutkimusaiheeseen liittyviä hypoteeseja ja aineistojen suurten kokojen vuoksi tehdä yleistyksiä tutkittuja havaintoyksiköitä laajempaan joukkoon tilastollisen päättelyn keinoin. (Heikkilä, 2014) Edellä mainittujen tutkimusmenetelmien välisten erojen vuoksi kvalitatiivisen tutkimusotteen valinta oli luonnollinen, sillä työn tavoitteena on saavuttaa syvällinen ymmärrys tiedon luokittelusta ilmiönä, sekä pyrkiä löytämään erilaisia motiiveja, sekä selittäviä tekijöitä sille, miksi yritykset tekevät toteuttavat tiedon luokittelua niin kuin toteuttavat.

5.2 Haastateltavat ja haastattelut

Laadullisen tutkimuksen aineistonkeruussa on tärkeää haastatella sellaisia henkilöitä, jotka tietävät tutkittavasta ilmiöistä mahdollisimman paljon tai heillä on kokemusta asiasta (Tuomi & Sarajärvi, 2018). Tämän työn aiheen osalta parhaiten sopivia ja ilmiöistä tietoa omaavia henkilöitä ovat yritysten tietotuvan kanssa työskentelevät tai siitä vastaavat henkilöt. Haastattelujen määrän osalta tavoitteeksi asetettiin 5-8 haastattelua, sillä Hirsijärvi & Hurme (2008) kirjoittavat kvalitatiivisen tutkimuksen haastateltavien määrän olevan tavallisesti 15, mutta Eskola & Suoranta (1998) puolestaan kirjoittaa vielä tarkemmin laadullisista opinnäytetöistä, ja toteaa 6-8 haastattelun olevan jo varsin hyvä määrä. Sopivien haastateltavien valinnassa ja tavoittamisessa hyödynnettiin Sulava Oy:n verkostoa, ja yhteistyön tuloksena haastatteluja saatiin kerättyä yhteensä 5 kappaletta.

Laadullisen tutkimuksen yleisimmät aineistonkeruumenetelmät ovat haastattelu, kysely, havainnointi ja erilaisista dokumenteista koottu tieto. Mainittuja menetelmiä voidaan käyttää vaihtoehtoisesti, rinnan tai eri tavoin yhdisteltynä riippuen tutkittavasta ongelmasta sekä tutkimusresursseista. (Sarajärvi & Tuomi, 2018) Yleisimpänä tapana kerätä laadullista aineistoa pidetään haastattelua, sillä se on yksinkertainen ja järkevä tapa selvittää, mitä ihminen ajattelee ja mitä hänellä on mielessään tietystä aiheesta. Haastattelut voidaan jakaa edelleen erilaisiin alatyyppeihin esimerkiksi niiden kysymysten muotoilun kiinteyden sekä haastattelijan haastattelutilanteen jäsentämisen mukaan. Määrämuotoisimmasta vapaamuotoisimpaan lueteltuna haastattelut voidaan jakaa strukturoituihin, puolistrukturoituihin haastatteluihin sekä teemahaastatteluihin ja avoimiin haastatteluihin. (Eskola & Suoranta, 1998)

Tämän tutkimuksen aineistonkeruumenetelmän valinta pohjautui työlle asetettuihin tavoitteisiin, sekä käytettävissä oleviin resursseihin. Aineistonkeruumenetelmäksi valittiin puolistrukturoitu teemahaastattelu, joka tarkoittaa, että teemahaastattelulle ominaisesti haastattelu etenee tiettyjen keskeisten, etukäteen valittujen teemojen mukaisesti (Hirsijärvi & Hurme, 2008; Eskola & Suoranta, 1998), ja että kysymyksiin ei annettu valmiita vastausvaihtoehtoja, vaan haastateltavat saivat vastata kysymyksiin omin sanoin. Haastattelun teemat ja kysymykset johdettiin kirjallisuuskatsauksessa käsitellyistä teemoista. Eskola & Suoranta (1998) korostavat haastattelukäytäntöjen ja -välineiden testaamisen tärkeyttä ennen varsinaisten haastattelujen keräämisen aloittamista, jotta voidaan

varmistaa haastattelujen onnistuminen. Tämän työn osalta haastattelukysymykset sekä välineet testattiin suorittamalla harjoitushaastattelu yhdelle henkilölle. Harjoitushaastattelusta kerätyn palautteen perusteella haastattelukysymysten asettelua ja järjestystä korjattiin, ja näin pyrittiin muodostamaan haastattelusta mahdollisimman looginen kokonaisuus.

Kaikki haastattelut toteutettiin etäyhteyksiä hyödyntäen Microsoft Teams-virtuaalikokousten avulla. Haastattelujen pitotavassa kuunneltiin vahvasti haastateltavien mielipidettä, ja kaikki haastateltavat halusivat lähtökohtaisesti suorittaa haastattelut etäyhteyksien avulla, sillä se on paikkariippumattomuutensa vuoksi haastattelun molempien osapuolten osalta perinteiseen haastatteluun verrattuna huomattavasti joustavampi tapa. Litterointia varten haastattelut nauhoitettiin, ja nauhoittamisen onnistumisen takaamiseksi nauhoitus tehtiin kahdella eri tavalla, Microsoft Teamsista löytyvän nauhoitusominaisuuden avulla, sekä lisäksi älypuhelimen ääninauhurin avulla.

5.3 Aineiston purku

Ennen analyysin kirjoittamista aineisto järjesteltiin teemoittelun ja tyypittelyn avulla. Teemoittelussa on kyse laadullisen aineiston pilkkomisesta ja ryhmittelystä ylä- ja alakategorioihin aihepiirien mukaan. (Tuomi & Sarajärvi, 2018) Teemoittelun perusteella valitut kategoriat johdettiin kirjallisuuskatsauksen teoriasta, ja ne kategorioille annettiin sisältöä kuvaavat nimet. Yläkategorioiksi määräytyivät (1) Datan ja tiedon rooli liiketoiminnassa, (2) Tiedon luokittelumallit, (3) Motiivit tiedon luokittelulle, (4) Tiedon luokittelun hyödyt tiedon suojaamiselle, (5) Tiedon luokittelun hyödyt liiketoiminnalle, (6) Tiedon luokittelun hallinta, (7) Lainsäädännön ja standardien vaikutus tiedon luokitteluun. Lisäksi yläkategoriat jaettiin tarvittaessa tarkentaviin alakategorioihin, esimerkiksi Tiedon luokittelun hallinta-kategorian alakategoriat olivat: (a) ylläpito, (b) roolitus ja vastuut, (c) työkalut.

Tyypittelyssä teemoitettu aineisto Tuomen ja Sarajärven (2018) kuvauksen mukaisesti ryhmiteltiin tietyiksi tyypeiksi, teemojen sisältä etsittiin näkemyksille yhteisiä ominaisuuksia ja tiivistettiin joukosta tiettyä teemaa koskevia yhtenäisiä näkemyksiä tyyppiesimerkkejä, eli yleistyksiä.

5.4 Aineiston analysointi

Asiantuntijahaastattelujen analyysin päätavoitteena on houkutella esiin faktoja (Hyvärinen ym., 2010). Aineiston analysointivaiheessa kerätystä aineistosta pyritään löytämään vastauksia tutkimusongelmaan sekä tutkimukselle asetettuihin tutkimuskysymyksiin (Tuomi & Sarajärvi, 2018). Laadullisessa tutkimuksessa aineistoa edustavat haastateltavien vastaukset haastattelussa esitettyihin kysymyksiin. Tutkimuksen tulokset sekä laatu riippuvat oleellisesti siitä, miten hyvin haastattelujen avulla kerätty aineisto onnistutaan purkamaan ja analysoimaan. Laadukkaan analyysin tuottamiseksi tutkijan täytyy suorittaa myös analyysiä edeltävät tutkimuksen vaiheet laadukkaasti.

Tässä tutkimuksessa analysoitava aineisto kerättiin edeltävässä luvussa esiteltyjen menetelmien avulla. Aineiston keruun jälkeen analysoinnin mahdollistamiseksi aineisto litteroitiin. Litterointi tarkoittaa video- tai äänimuodossa olevan aineiston tekstimuotoon muuttamista (Hyvärinen, ym., 2010). Ruusuvuoren ym. (2010) mukaan kiinnostuksen kohdistuessa haastattelujen asiasisältöön, riittävää haastattelijan sekä haastateltavien puheenvuorojen litterointi ilman tarkempaa, esimerkiksi taukojen pituuksien tai äänensävyjen litterointia. Tässä tutkimuksessa kiinnostus aineiston osalta kohdistui asiasisältöön, joten litteroinnin tasoksi valittiin haastattelijan ja haastateltavien puheenvuorojen litterointi.

Tekstimuotoon litteroidun aineiston analysoinnissa noudatettiin teoriaohjaavan analysoinnin periaatteita. Teoriaohjaava analyysi valittiin analysointimenetelmäksi, koska tiedon luokittelu itsessään ei ole uusi ilmiö, mutta datan määrän kiihtyvän kasvun ja teknologisen kehityksen seurauksena tiedon luokittelun rooli ja toteutustavat muuttuvat jatkuvasti, ja siitä syystä edeltävän teorian ja aineiston analyysi- ja ajattelutapaa ohjaavat vaihtelevasti valmiit mallit sekä tutkimusaineisto. Teoriaohjaavassa analyysissa tehdään teoreettisia kytkentöjä siten, että teoria toimii apuna, mutta analyysi ei pohjaudu suoraan teoriaan. Teoriaohjaavasta analyysista on tunnistettavissa aikaisemman tiedon vaikutus, mutta aikaisemman tiedon merkitys ei ole teoriaa testaavaa, vaan enemmänkin uusia ajatusuria aukova. (Tuomi & Sarajärvi, 2018, s. 81) Teoriaohjaavan analysoinnin valinnalla pyrittiin toisaalta löytämään vahvistusta tiedon luokittelun teorioiden toteutumiselle yritystoiminnassa, mutta toisaalta myös antamaan tilaa aineistosta nouseville tulkinnoille tutkittavasta ilmiöstä. Lisäksi ennen analysoinnin aloittamista tehtiin valinta, että aineistosta haetaan erilaisuutta vastausten välillä ja sitä kautta löytämään mahdollisia sävyeroja edeltävän teorian toteutumisessa vastaajien välillä.

6 EMPIIRISEN TUTKIMUKSEN TULOKSET

Tässä luvussa käsitellään empiirisen tutkimuksen tuloksia. Luku jakautuu analyysia varten suoritetun teemoittelun mukaisiin teemoihin. Luku 6.1 kuvailee datan ja tiedon merkitystä ja roolia organisaatioissa. Luvussa 6.2 tarkastellaan aineistosta esiin nousseita erilaisia tiedon luokittelun malleja. Luku 6.3 käy läpi esille nousevia motiiveja, joiden vuoksi organisaatiot luokittelevat tietoa juuri sillä tasolla, kun luokittelevat. Luvuissa 6.4 ja 6.5 tarkastellaan tiedon luokitteluun liittyviä koettuja hyötyjä tiedon suojaamisen, sekä liiketoiminnan näkökulmasta. Luku 6.5 käsittelee tiedon organisaatioiden tiedon luokittelun hallintaa ja ylläpitoa, sekä siihen liittyviä erilaisia tekijöitä, kuten roolitusta ja vastuita, sekä tiedon luokittelussa käytettäviä työkaluja ja teknologioita. Viimeinen luku 6.7 käsittelee lainsäädännön ja standardien vaikutusta organisaatioiden harjoittamaan tiedon luokitteluun ja sen tasoon.

6.1 Datan ja tiedon rooli liiketoiminnassa

Tutkimusaineiston perusteella datalla ja tiedolla on poikkeuksetta tärkeä rooli liiketoiminnassa. Jokaisessa vastauksessa roolia kuvailtiin vähintään tärkeäksi, mutta kolmessa vastauksessa viidestä datan ja tiedon kuvailtiin olevan jopa aivan keskiössä liiketoiminnan kannalta. Aineiston perusteella tärkeyden lisäksi datan ja tiedon roolin kehitystä kuvailtiin jatkuvasti kasvavaksi.

Teollisuusteknologian alalla toimivan H1 kuvailee yritystoiminnan yleensäkin perustuvan yrityksen tieto-omaisuuteen, joka syntyy kehitetystä tuotteesta tai palvelusta, johon puolestaan liittyy tarkasti varjeltavaa dataa ja tietoa. Kehitettyyn palveluun liittyvää tieto-omaisuutta pyritään suojaamaan patentein. Konsultointia harjoittavaa, puhtaasti digitaalista liiketoimintaa harjoittava H2 puolestaan kuvailee heidän kilpailukykynsä syntyvän siitä, miten hyvä kyky heillä on kerätä, hyödyntää, jalostaa ja jakaa dataa ja tietoa. Päivittäistavara- ja kuluttajaliiketoimintaa harjoittavat H3, H4 ja H5 korostavat mittaamisesta syntyvät toiminnan datan tärkeyttä liiketoiminnan jatkuvan kehittämisen ja tiedolla

johtamisen näkökulmasta. H3 kertoo, että data toimii pohjana kuluttajille suunnattujen alennusten ja kampanjoiden räätälöinnissä, sekä edellä mainittujen onnistumisten mittaamisessa.

> No tota siis datalla ylipäätään niin se on aivan keskiössä mitä me tehdään ja onnistutaanko me meidän liiketoiminnan prosesseissa, niin se data on aivan keskiössä siellä, ja esimerkiksi kun ajatellaan jotain meidän päivittäistavarakaupan valikoima-asioita tai meidän kanta-asiakkaan saamia tarjouksia, niin kaikki ne perustuu dataan, siihen mitä me katsotaan mitä meidän kuluttajat on ostaneet meiltä, mitä yksittäinen kuluttaja, mitä tuotteita se yleensä ostaa ja mitä hän hyvin todennäköisesti haluaa nähdä tarjouksessa. (H3)

> Vuosi vuodelta ja kuukausi kuukaudelta isompi rooli, että siis selkeästi on nähtävissä se, että halutaan tehdä enemmän ja enemmän tiedolla johtamista ja perustaa niitä päätöksiä, perustaa sitä strategiaa enemmän kerättyyn dataan mitä meillä on, että analytiikka on selkeästi niin kuin koko ajan enemmän ja enemmän keskiössä (H5)

6.2 Tiedon luokittelumallit ja luokittelun taso

Tiedon luokittelun harjoittamisen tasossa ja luokittelumalleissa löytyi runsaasti eroavaisuuksia vastausten välillä. Tiedon luokittelun harjoittamisen tasoon vaikuttivat monet asiat yritystoiminnan taustalla, kuten liiketoiminnan tuottaman tiedon kriittisyys, tiedon suojaamisen kulttuuri sekä teknologian hyödyntäminen. Tiedon luokittelumalleissa sen sijaan löytyi samankaltaisuuksia tiedon luokittelumallien sisältämissä luokissa ja luokkien nimissä.

Tämän tutkimuksen luvussa 3.1 kerrotaan, että yritykset voivat luokitella omistamaansa tietoa millä perusteella tahansa, mutta yleisimmin luokittelua tehdään perustuen datan liiketoiminnalliseen arvoon ja luottamuksellisuuteen. Tämä ilmeni hyvin haastattelujen vastauksista, sillä jokaisessa vastauksessa luokitteluperusteeksi kerrottiin tiedon joko tiedon liiketoiminnallinen arvo, luottamuksellisuus, tai molemmat. Vaikka luokitteluperusteet olivat kaikilla samoja, oli luokkien välillä hieman eroavaisuuksia.

H3 kertoi heidän organisaationsa soveltavan luokittelussaan Microsoftin tarjoamia oletusluokkia. Microsoft luokittelun yläluokat ovat personal, public, general, confidential, sekä highly confidential. Edellä mainittuihin luokkiin sisältyvät myös lisäluokittelut, joilla tietyn luokan tietoa voidaan organisaation sisällä joko kaikille työntekijöille, tai rajatulle joukolle ihmisiä. Myös H1, H2, H4 ja H5 kertoivat heidän luokitteluistaan löytyvän korkeimpina luokkina luottamuksellinen (confidential) sekä salainen (secret).

> Meillä on hyvin tyypillinen neliportainen luokittelu, eli meillä se menee sillä tavalla että meillä on tämmöinen niin kuin, käytän englantilaisia termejä, kun me ollaan englanninkielinen yritys siinä mielessä että meillä ei suomenkielisiä politikointitekstejä ole, niin on tällanen kun unrestricted sitten on restricted, sitten on condicential ja sitten on secret ja kaikkihan lähtee siitä että me ollaan ohjeistettu ja laitettu esimerkkejä että mitä näillä eri tasoilla se tieto voisi olla, että se tiedon luoja osaisi luokitella sen tiedon oikein jotta sitten kun sitä tietoa hyödynnetään niin ihmiset jotka sitä käyttävät, ymmärtävät omassa työssään, mitkä sen esimerkiksi rajoitteet ovat (H1)

> Meillä on käytännössä Tällä hetkellä 3 tasoinen tiedon luokittelu, me puhutaan ehkä enemmän kyllä dokumentaation luokittelusta, ja siellä public-tasoisesta dokumentaatiosta ja sitten puhutaan company confidential tasoisesta, joka käytännössä on se standardi mihinkä meillä menee kaikki ja sitten on confidential&restricted tasoinen, joka sitten on se tavallaan se kaikista arkaluontoisiin ja nyt ollaan otettu siihen keskusteluihin mukaan, ei olla vielä virallisesti tehty neljättä tiedon luokkaa, mutta sitä mietitään että otetaanko ja se siinä puhutaan niinku trade secreteista, että meillä on meillä on nyt alkanut olemaan enemmän ja enemmän intoa (H5)

Luokittelujen lisäksi H4 ja H5 kertovat, että osa yrityksen prosesseista on määritelty kriittiseksi liiketoiminnan kannalta, ja näin ollen kyseisten prosessien tuottama data saa oletuksena normaalia korkeamman turvaluokittelun.

> Tietenkin koska ollaan pörssiyhtiö ja niin kuin muutenkin finanssiprosesseissa sen tyyppisissä on tietenkin ja ehkä jossain hr-prosesseissa, joissa on perinteisesti käsitelty luottamuksellista dataa ja on sisäpiiritietoja ja niin edelleen niin siellä on toki ollut sitten omat käytännöt pidempään (H4)

6.3 Motiivit tiedon luokittelulle

Tiedon luokittelun harjoittamisen motiiveista kysyttäessä, suurin yksittäinen esittele noussut tekijä oli compliance, eli tiedon säilyttämisen ja suojaamisen vaatimustenmukaisuus. Vaatimustenmukaisuudesta puhuttaessa kaikki haastateltavat toivat ilmi taloustiedot sekä henkilötiedot. Taloustietojen luokittelun tarkkuudelle kaikissa paitsi H5:n vastauksessa perusteeksi kerrottiin pörssilainsäädäntö, sekä siihen liittyvät salassapitosäännöt yhtiön tulokseen sekä sitä kautta osakkeen hintaan vaikuttaviin tietoihin liittyen. Pörssiyhtiön on salattava esimerkiksi osavuositulokseen liittyviä tietoja siihen saakka, kunnes tuloksesta annetaan virallinen pörssitiedote. Ennen virallisen tiedotteen julkistamista julkisuuteen vuotaneista tiedoista seuraa aina rangaistus, ja siitä syystä yhtiöt luokittelevat tällaiset tiedot tiukasti, ja kohdistavat tietoihin myös erityisen tarkkoja suojaustoimenpiteitä.

Toinen tietoluokka, jonka kanssa kaikki haastateltavat kertoivat olevansa erityisen tarkkana, on henkilötiedot. Yleinen tietosuoja-asetus (GDPR) asettaa tarkkoja vaatimuksia liiketoiminnan yhteydessä tapahtuvasta henkilötietojen käsittelystä ja sen laajuudesta, kuten tämän tuon luvussa 3.3 kerrotaan. GDPR:n vaatimusten rikkomisesta seuraa ankaria taloudellisia sanktioita, ja siitä syystä se motivoi kaikkia vastaajia luokittelemaan henkilötietoja tiukasti, ja noudattamaan henkilötietojen käsittelyyn liittyviä vaatimuksia.

Ainoastaan H3 kertoi haastattelussa tiedon luokittelun motiiviksi liiketoiminnan tehostamisen ja sen, että tehokkaalla luokittelulla haetaan maksimaalista datan jatkojalostamista ja hyödyntämistä.

> Kyllähän me se myöskin tämä niin kun, että me olemme yrityksenä nähneet sen datan arvo ja se, että jos ei me sitä luokitella niin ei me voi sitä hyödyntääkään. Eli se lähtee sieltä, että toisaalta se, että me halutaan olla vastuullinen toimija, me halutaan olla compliant toimija, mutta toisaalta me halutaan myöskin maksimaalisesti pystyä hyödyntämään se data, mikä meillä on mahdollista kerätä ja hyödyntää sen jälkeen ja se tarvitsee sen, että me tunnemme sitä dataa tai tiedetään mitä sieltä kerätään, että miten sitä voidaan luokitella, jotta me voidaan sitä jatkojalostaa ja hyödyntää sitten sen jälkeen. (H3)

6.4 Tiedon luokittelun hyödyt tiedon suojaamiselle

Kirjallisuuskatsaus osoitti, että tehokas tiedon luokittelu tarjoaa etuja tiedon suojaamiselle helpottamalla useita suojaamiseen liittyviä näkökulmia, kuten pääsynhallintaa, elinkaarenhallintaa, tiedon häviämisen estämistä, sekä tietoturvaa. Luvuissa 6.4.1–6.4.4 käsitellään sitä, millaisia hyötyjä vastaajat kokivat tiedon luokittelun tarjoavan näiden tiedon suojaamisen osa-alueiden näkökulmasta. Haastattelussa kysyttiin tiedon luokittelun hyödyistä yleisesti tiedon suojaamiselle, jotta saataisiin mahdollisimman todenmukainen kuvaus siitä, miltä osin haastateltavat kokivat tiedon luokittelun auttavan tiedon suojaamisessa.

6.4.1 Pääsynhallinta (access control)

Viidestä haastateltavasta kolme listasi tiedon luokittelun auttavan tiedon suojaamista tehostamalla ja helpottamalla pääsynhallintaa. H1 kertoo pääsynhallinnan olevan yksi tärkeimmistä tiedon suojaamisen näkökulmista, joissa tiedon luokittelu auttaa. Varsinkin yrityksen tärkeimmän ja salaisimman tieto-omaisuuden tunnistaminen ja luokittelu täytyy toimia virheettömästi, jotta tiedon käsittelyn kontrolleja, prosesseja, rooleja sekä omistajuutta on mahdollista räätälöidä tietoluokalle sopivaksi. Ilman toimivaa luokittelua salainen tieto saattaisi päätyä sellaisten käyttäjien käyttöön, joiden ei kuuluisi päästä näkemään tietoa tai toisin päin käyttäjä, jonka kuuluisi nähdä tietty salainen asiakirja, ei välttämättä pääse

sitä näkemään. Heikosti toteutettu tiedon luokittelu vaikeuttaa pääsynhallintaa ja aiheuttaa turhaa manuaalista työtä.

H1 ja H4 toivat esille pääsynhallinnan yhteydessä myös tiedon päätymisen väärin perustein organisaation ulkopuolelle esimerkiksi organisaation jaetuista pilvipalveluista (esim. Microsoft OneDrive, Microsoft Sharepoint). Mikäli tietoa on huonosti tai väärin luokiteltu, yrityksen sisäinen käyttäjä saattaa vahingossa jakaa yhteisessä tallennustilassa sijaitsevaa, vain organisaation työntekijöille tarkoitettua tietoa organisaation ulkopuolelle ja näin aiheuttaa vahinkoa organisaatiolle.

> Pyrimme myös teknologian avulla suunnittelemaan kontrolleja eri tietoluokille, hyvänä esimerkkinä on se, että minkälaisia tietoluokkia saa jakaa esimerkiksi OneDrivesta tai Sharepointista ulkopuolelelle organisaatiota. Tämä on hyvä esimerkki, eli pyritään teknologian avulla ilman että me tavallaan kosketaan asiasisältöön, luokkatiedon perusteella tekemään poliitikoita että, jos on merkitty secretiksi niin että mehän ei lueta silloin sitä itse sisältöä lainkaan, emme me tiedä, mitä siellä on sisällä vaan tiedon omistaja tai luoja on merkinnyt esimerkiksi wordin metatietoihin että tämä on nyt secret dataa, niin silloin me voidaan luoda tämmöinen tekninen kontrolli siihen että sitä ei saa jakaa esimerkiksi Sharepointissa tai OneDrivella ulkopuolisille, vaan sitten ihan normaalisti tulee pop-up ja kysyy, että ootko sä nyt ihan varma. (H1)

H2 mainitsee pääsynhallinnan hyötyvän tiedon tehokkaasta luokittelusta, mutta samalla mainitsee haasteeksi sen, että tiedon moninaisuus ja eri järjestelmät vaikeuttavat huomattavasti täydellisen tiedon luokittelun ja kontrollien saavuttamista. Esimerkkinä hän mainitsee sähköpostiviestit, jotka ovat sisältönsä osalta tiedon luokittelun saavuttamattomissa.

> Sitten meillä on tietojärjestelmiä jotka ei sitten oikein tunnu tukevan tämmöistä niin kun tiedonkäsittelyn käsittelypolun tai tietoon kohdistettua suojan erittelyä sen leiman perusteella että siellä on semmoisia tietoturvavastaavan näkökulmasta kiusallisia tilanteita, että siellä on eri tavalla leimattua dataa, mutta sitten ne leijuu siellä yhdessä ja samassa paikassa, kuten intranet, esimerkiksi niin siellä ei ole pakotetusti käyttövaltuusrajauksia laitettu leimauksen perusteella tai sähköposti nyt on tämmöinen ikuinen ihan paska esimerkki että, sähköposti ja turvaluokitus ei vaan nyt tunnu sopivan yhteen, ellei sitten ole eriytetyt järjestelmät. (H2)

6.4.2 Datan elinkaaren hallinta (data lifecycle management)

Haastattelujen perusteella tiedon luokittelun koettiin auttavan myös tiedon elinkaaren hallinnassa. Tiedon luokittelun kerrottiin auttavan erityisesti niissä tapauksissa, kun teknologiaa hyödyntämällä on pyritty automatisoimaan tietoluokkien muutoksia ajan kuluessa. Neljä haastateltavaa viidestä ottivat talous- ja tilinpäätöstiedot esimerkiksi tietoluokasta, jonka leimaus muuttuu hyvin radikaalisti. Pörssiyhtiön osavuosikatsaus tiedot ovat korkeimman salausluokan tietoa ennen tiedon julkistamista, mutta heti tiedotteen julkaisun jälkeen tieto muuttuu julkiseksi.

Tiedon luokittelu yhdistettynä teknologian mahdollistamaan automaattisiin tiedon luokkien muutoksiin auttaa yrityksiä ylläpitämään tiedon elinkaaren hallintaa. H1 kertoi, että he ovat pyrkineet edesauttamaan elinkaaren hallintaa automatisoimalla tiedon luokittelujen muutoksia ajan kuluessa, ja lisäksi määrittämään säilytysaikoja tietoluokille.

> Ollaan pyritty sillä tavalla luokkien avulla keskittymään tietenkin meille tärkeisiin tietoluokkiin, eli noihin secrettiin ja confidential-tasoon, eli kaikkie sulla ei vaan yksinkertaisesti siirtä aika, kun sun pitää automatisoida paljon asioita, mutta kaikkien tietojen elinkaareen ei ehkä kannatakaan puuttua muuta kun yleisellä tasolla, että sulla on erilaisilla teknologia-alustoilla sitten tällaisia automaatisia retention-policyja, kuten vaikka Sharepointissa, OneDrivessä, sähköposteissa, että siinä automatiikan avulla hävitetään koko ajan sitä vanhaa tietoa siellä et sitä ei niin kun jää sinne vahingossakaan, että sitten taas sitten näihin korkeamman luokan asioihin niin sinne sitten määritellään omia asioita koska siinä ne on vähän erilaisia, sinne ei ehkä automatiikkaa ole ei ole mahdollista edes tehdä siihen elinkaareen vaan sitten vaan sitten manuaalisesti hallitaan sitä elinkaarta. (H1)

H3 puolestaan kertoi heillä loppukäyttäjien tekemiin tiedon luokitteluihin ei sovelleta säilytysaikoja, mutta järjestelmäkohtaisesti säilytysaikoja määritellään riippuen järjestelmien tuottamien tietomassojen luokista, esimerkiksi talous- ja hr-järjestelmät.

> Loppukäyttäjien, eli tietotyöläisten tekemissä luokissa ei mitenkään, että siellä retentioaikoja ei ole laitettu sinne tiedolle tai näihin luokkiin millään tavalla, muissa tämmöisissä järjestelmäkohtaisissa tiedon luokittelussa, niin siellä on sitten retentioajat riippuen tietomassasta tai tiedon luokista, että mitä siellä säilytetään. Meillä on sitten niin kun tilinpäätöstietoja ja muita, niin niitä säilytetään tietty se verohallituksen ja näiden finanssipolitiikkojen mukainen aika on se sitten 7+1 tai 10 vuotta, tai jos se on jotain niin kun hr-dataa niin se on sitten taas niin kun tietyn ajan hr-järjestelmissä tai arkistossa säilytettynä, mutta lähtökohtaisesti näissä niin kun loppukäyttäjien luokituksissa niin niissä ei ole otettu retentioaikaa millään tavalla. (H3)

6.4.3 Datan häviämisen estäminen (data loss prevention)

Haastatteluissa annettujen vastausten perusteella tiedon luokittelun ei katsottu tarjoavan suurta hyötyä tiedon häviämisen estämiselle. H2 kertoi, että hyödyntävät pieneltä osin tiedon häviämisen estämisen tekniikoita ja työkaluja Microsoft 365-ympäristössään, jossa he ovat määritelleet tiettyjä automaattisia suojausmekanismeja sen mukaan, miten korkean turvaluokitusleimauksen tiedon luoja, tai käsittelijä dokumentille antaa. Tämän lisäksi H1 kertoi, että he pyrkivät pitämään huolen siitä, että korkealla turvaluokituksella varustettuja dokumentteja ei jaeta koko yritykselle, tai yrityksen ulkopuolelle, ja puuttuvat välittömästi asiaan, mikäli heidän DLP (data loss prevention) järjestelmä hälyttää korkean turvaluokituksen dokumentin liian laajasta jakamisesta.

H4 puolestaan kertoi, että heillä tiedon luokittelun hyödyt tiedon häviämisen estämisessä näkyvät asiakkaiden maksutietojen, erityisesti maksukorttitietojen korkeammassa suojaamisen tasossa. Mikäli tietty tieto heidän järjestelmissään kyetään tunnistamaan korttitiedoksi, se luokitellaan välittömästi korkean turvaluokan tiedoksi, ja siihen kohdistetaan korkeampia suojaustoimenpiteitä, kuin muuhun, vähemmän salaiseen tietoon. Näin pyritään pitämään huoli siitä, että asiakkaiden maksutiedot eivät vuoda vääriin käsiin yrityksen järjestelmistä.

> Me pyrimme automatisoidusti tai puoliautomatisoidusti tunnistaa esimerkiksi, että hei, tämä voisin nyt näyttää joltain tietyltä tietoaineistolta ja sitä kautta sitä luokiteltaisiin niin ei sen tyyppistä juurikaan tehdä. Meillä on muutama tommoinen tota data loss prevention (DLP) tyyppinen sääntö korttitietoihin liittyen tota lähinnä kuin Yhdysyhdysvalloissa ja että on pyritty välttämään sitä, ettei sitten jos asiakaspalvelukanavien tai muiden kautta tulisi korttitietoja meille. (H4)

Yleisellä tasolla vastauksista oli havaittavissa, että tiedon luokittelu ei ole kovinkaan suuressa roolissa tiedon häviämisen estämisessä, kun vain kaksi haastateltavaa viidestä mainitsi erikseen tiedon luokittelusta olevan hyötyä tiedon häviämisen estämisessä.

6.5 Tiedon luokittelun hyödyt liiketoiminnalle

Haastattelujen perusteella tiedon luokittelun ei nähdä toimintana, joka ensisijaisesti tarjoaa suoria liiketoiminnallisia hyötyjä, vaikka tiedon suojaamisen hyödyt ovat myös vähintään välillisesti liiketoimintahyötyjä. Tämän työn johdannossa puhutaan lyhyesti siitä, millaisia haittavaikutuksia luokittelemattomalla, eli ”pimeällä” datalla. Yksi suuri haittavaikutus on pimeän data yli- tai alimitoitetut suojaustoimenpiteet ja näin ollen väärin kohdennetut suojausresurssit. Lisäksi puutteellinen luokittelu aiheuttaa tiedon omistajille manuaalista työtä pääsyoikeuksien manuaalisesta säätämisestä, sekä työntekijöille mahdollisia viivästyksiä työssään, mikäli väärien luokitusten vuoksi ei päästä tarvittavaan tietoon käsiksi. Kääntäen edellä luetellut asiat ovat käännettävissä liiketoiminnan kilpailueduiksi, koska esteet poistamalla työntekijöiden tyytyväisyys sekä tehokkuus kasvavat.

Haastatteluista ainoastaan H3 kertoi organisaation hakevan liiketoiminnallista hyötyä ja kilpailuetua tiedon tehokkaan luokittelun avulla. H3 kertoi, että he haluavat olla mahdollisimman tehokkaita datan hyödyntäjiä liiketoiminnan kehittämisessä, ja kertoi, ettei datan maksimaalinen hyödyntäminen olisi mahdollista ilman, että tunnetaan täysin, mistä informaatio-omaisuus muodostuu. H3:n mukaan datan kaikenlainen hyödyntäminen mahdollistetaan tehokkaalla luokittelulla.

Muut haastateltavat lähestyivät tiedon luokittelun liiketoimintahyötyjä enemmän tiedon suojaamisen kautta tulevista hyödyistä, sekä juridisesta turvasta, mitä tiettyjen tietojen tarkka luokittelu, sekä riittävä suojaaminen tuovat.

> Kyllä sanoisin tietoturvajohtajan hattu päässä, että se synnyttää semmoisen varmuuden että että me pystytään tunnistamaan milloin on toimittu oikein ja tunnistetaan myös milloin on toimittu väärin ettei jää niin paljon väittelyntilaa, ja että voimme paremmin osoittaa ulkopuoliselle arvioijalle, on se sitten auditori tai asiakas, että millä pelisäännöillä meillä tietoa käsitellään, jaetaan ja suojataan, että se varmaan se keskeinen asia, olisiko se semmoinen assurance ja confidence tähän tekemiseen se kaikkein selkein ja sitten tosiaan siltä osin kun on viranomaissysteemit niin tuota, ei joudu vankilaan se on semmoinen, että pysyy kaltereiden tällä puolella kun ei että tee tyhmyyksiä (H2)

6.6 Tiedon luokittelun hallinta

Tässä luvussa käsitellään tiedon luokittelun hallintaa ja sitä, miten haastateltavat organisaatiot hallitsevat ja ylläpitävät tiedonluokittelua, millaisia roolituksia ja vastuita tiedon luokitteluun liittyy, sekä millainen rooli erilaisilla työkaluilla ja teknologioilla on tiedon luokittelussa haastateltavissa organisaatioissa. Ensimmäisessä luvussa 6.6.1 käydään läpi tiedon luokittelun ylläpitoa, luku 6.6.2 käsittelee puolestaan tiedon luokittelun roolituksia ja vastuita, ja viimeinen luku 6.6.3 käsittelee teknologioiden ja työkalujen merkitystä tiedon luokittelussa

6.6.1 Ylläpito

Tiedon luokittelun toimivuuden ja ajantasaisuuden varmistaminen vaatii ylläpitoa sekä säännöllistä tarkastelua, jotta luokittelusäännöt, sekä luokat ovat ajan tasalla, ja sopivia sille datalla, jota organisaatio kulloinkin tuottaa. Haastatteluissa ylläpidon vahvana osana nähtiin tiedon luokittelun kulttuurin synnyttämistä sekä ylläpitämistä. Tiedon luokittelun säännöistä, sekä erilaisten tietoluokkien teknisestä suojaamisesta vastaavat tietoturvatiimit, mutta suurin vastuu tiedon luokittelusta on loppukäyttäjillä, eli heillä, jotka tietoa luovat ja käsittelevät. H1 kertoi heidän ajattelevan, että tiedon luokittelun tietoisuuden ja kulttuurin rakentaminen sekä ylläpitäminen on yksi tärkeimmistä tekijöistä tiedon luokittelun ylläpidossa. Kulttuurin rakentamisen ja kehittämisen lisäksi H1 näki tärkeänä myös teknologisen puolen ylläpitämisestä, eli erilaisten automaattisten suojaustoimenpiteiden luomisesta, sekä suojaussääntöjen päivittämisestä säännöllisesti.

Vaikka loppukäyttäjillä on suuri rooli tiedon luokittelun toteutumisessa ja ylläpitämisessä, vaatii ylläpitäminen myös ylätasolla jatkuvaa tarkastelua, analysointia ja päivittämistä. Luokittelumalli vaatii ylläpitämistä sen ajantasaisuuden varmistamiseksi, mutta toinen tärkeä seikka, joka asettaa ylätason ylläpidollisia tarpeita, on teknologiaan tulevat päivitykset. H3 mainitseekin, että he käyvät säännöllisin väliajoin läpi itse luokittelumallia, mutta lisäksi tutkivat tarkasti heidän hyödyntämänsä teknologiatarjoajan Microsoftin puolelta, mikäli heille on tullut merkittäviä muutoksia malleihin tai suojausmekanismeihin, ja päivittävät omat mallinsa ja suojauksensa vastaamaan Microsoftin uusimpia sisältöjä mahdollisimman nopeasti.

> No tiedon luokittelun ajantasaisuudesta niin jos ajatellaan sitten tiedon luokittelun mallin ajantasaisuutta ensimmäisenä, niin sitä käydään tasaisin väliajoin läpi, että nyt juuri viime vuoden syksyllä käytiin viimeksi läpi ja se poiki tietyn määrän muutostöitä, jotka nyt taidettiin saada viime kuun lopulla vietyä viimeisetkin tuotantoon, joten sitä käydään tietyn väliajoin sitten katsotaan, että se tiedon luokittelumalli on edelleen validi, onko Microsoftin puolelta tullut muutoksia tai onko siellä road mapissa jotain joka pakottaa meidät muuttamaan sitä mallia tai suojauksia tai tekemään jotain muuta sinne. (H3)

Tiedon luokittelun hallinnan ja ylläpitämisen säännöllisten toimenpiteiden suorittamiselle on usein nimetyt henkilöt ja/tai tiimit, jotka vastaavat ylläpidollisista toimenpiteistä H3 kertoi, että heidän organisaatiossaan tiedon luokittelusta huolehditaan kolmikannassa kolmen eri tiimin voimin. Ylläpitoon osallistuvat loppukäyttäjien user services-tiimi, kyberturvatiimi, sekä riskienhallinnantiimi, ja nämä tiimit yhdessä katsovat jokaisen tiimin näkökulmasta tiedon luokittelun ajantasaisuutta ja muutostarpeita. Lisäksi he käyttävät myös ulkopuolisia toimijoita tarpeen mukaan muutosten ja ylläpitotöiden toteuttamisessa sekä tuotantoon viennissä.

> Vähän niinku avasinkin sitä, että me tehdään sitä kolmikannassa sisäisesti, että meillä on niin kun and user services, mun tiimi kybertiimi ja riskienhallinta, joka katsoo sitä niinku kokonaisuutta, että onko meillä oikeat luokat, tarvitseeko niin tehdä muutoksia, tarvitseeko käyttäjiä kouluttaa tai dokumentointia muuttaa ja tota sitten joko pystytään tekemään se itse tai palvelukumppanin kanssa tai sitten ulkoisen konsulttitalon kanssa niin kun Sulava niin tota tehdään niitä muutoksia, testataan, pilotoidaan ja viedään niitä eteenpäin että se on niin kun kolmikannassa sisäisesti ja sitten tarvittaessa otetaan ulkopuolisia toimittajia tekemään näitä niin kun itse itse muutostöitä. (H3)

H4 puolestaan kertoi, että heidän organisaationsa on myös rakentanut ylätason ylläpidosta vastaavan ryhmän niin, että edustettuna on useampi tiimi. Tämän lisäksi H4 kertoi heillä jatkuvan palautteen keräämisen ja jatkuvan kehittämisen kulttuurin olevan vahvassa roolissa tiedon luokittelun ylläpitämisessä ja kehittämisessä. He kysyvät tiedon loppukäyttäjiltä ja käsittelijöitä jatkuvasti palautetta luokkien toimivuudesta, sekä tiedon leimaamisen käyttöliittymän käyttämisestä, ja pyrkivät jatkuvasti ottamaan palautteita huomioon, ja kehittämään luokittelumallin toimivuutta, sekä leimaamisen käyttäjäkokemusta.

6.6.2 Roolitus ja vastuut

Tiedon luokittelun ja suojaamisen liittyvistä roolituksista ja vastuista ajateltiin hyvin samalla tavalla kaikissa haastatelluissa organisaatioissa, eikä yhdenkään organisaation lähestymistapa sekä roolien ja vastuiden jakautumien eronnut merkittävälle tavalla muista vastauksista.

Tiedon luokittelun roolituksen osalta vastauksista esille nousi selkeästi kolme erilaista roolia, ja jokaiseen rooliin liittyy myös erilaisia vastuita. Ensimmäinen rooli tiedon luokittelu teknisen ja strategisen toteuttamisen rooli, johon yleensä kuuluvat tietojärjestelmien ylläpito- ja kehitystiimit, sekä tietoturvatiimit, sekä näiden toiminnasta korkeammalla tasolla vastaavat johtohenkilöt. Tiedon luokitteluun on usein osa tietoturvan johtamisjärjestelmää ISMS, josta kerrotaan myös tämän työn luvussa 3.2, ja osana tietoturvan johtamisjärjestelmää tiedon luokittelusta on tehty politiikka, tai direktiivitason dokumentti, josta haastattelujen perusteella vastaa usein tietoturvajohtaja. Lisäksi tietoturvan alle kuuluvat tietoturva- ja järjestelmätiimit yleensä vastaavat tiedon luokittelun ja

suojaamisen teknisestä toteuttamista. Tietoturvajohtajana organisaatiossaan toimiva H1 kertoo, että he vastaavat tiedon luokittelun politiikoista, sekä luokittelun teknisestä toteutuksesta, mutta kertoo vastuun luokittelun kokonaisuudesta kuuluvan laajemmalle joukolle, ja myös, että vastuukysymykset luokittelun vastuista monimutkaisten järjestelmäkokonaisuuksien vuoksi olla joskus myös epäselviä.

> Minun tiimi vastaa direktiivistä tai politiikoista, niin me vastataan sitten itse että se myös luokittelu siellä teknisellä tasolla tapahtuu, niin se ei vaan mene niin, ja varmaan isoimpia haasteita koko tiedon luokitteluun ja tähän vastuukysymyksiin liittyy siihen, että kenelle se vastuu siitä, kun sitä dataa yhdistetään toiseen tietoon tai useampaan tietolähteeseen, eli puhutaan näistä data management platformeista ja tai data-analytiikkapalveluista, eli viedään paljon dataa jonnekin yhteiseen tekniseen alustaan, niin kuka sen jälkeen on vastuussa siitä luokittelusta, kun se data muuttuu, siihen lisätään tai poistetaan vaikka osia tai jotain muuta, niin sittenhän siitä se vastuukysymys onkin vähän epäselvä. (H1)

Myös muut haastateltavat kertoivat luokittelun strategisen ja teknisen toteuttamisen roolin kuuluvat tietoturvajohdolle, sekä tietoturvasta ja järjestelmistä vastaaville tiimeille. H3 mainitsi lisäksi heidän organisaatiossaan ylätason luokittelussa olevan mukana myös loppukäyttäjäpalveluiden sekä riskienhallinnan tiimit, ja H4 puolestaan kertoi, että heidän organisaatiossaan ylätason luokittelun toteuttamisessa mukava olevan legal&compliance -osasto.

Toinen tiedon luokittelun rooli, joka haastatteluissa nousi esille, on tietojen käsittelyn sekä liiketoimintaprosessien omistajat. Tähän rooliin liittyy liiketoimintaprosessin tuottaman datan luottamuksellisuuden arviointi, sekä tarkempien, prosessikohtaisten tiedon luokittelusääntöjen luominen. Liiketoimintaprosessien omistaja on myös vastuussa prosessinsa tuottaman datan oikeellisuudesta, sekä tietojen oikeanlaisesta luokittelusta. Haastatelluista H2, H4 ja H5 mainitsivat suoraan prosessien omistajat osaksi tiedon luokittelun roolitusta, ja myös H3 ja H1 vastausten muotoilusta oli pääteltävissä, että prosessien omistajilla on oma roolinsa kokonaisuudessa.

Kolmas tiedon luokittelun rooli on tiedon luokittelun toteuttajat, johon lukeutuvat kaikki tiedon loppukäyttäjät, jotka käsittelevät, muokkaavat ja jakavat tietoa. Kaikki haasteltavat painottivat vastauksissaan loppukäyttäjien roolia tiedon luokittelun toimivuudessa. Kaikissa haastatteluissa jossakin kohdassa korostettiin sitä, että strateginen ja tekninen taso pitää olla kunnossa, jotta tiedon luokittelua on mahdollista tehdä, mutta siitä huolimatta loppukäyttäjä on se, joka joko leimaa tai jättää leimaamatta käsittelemänsä dokumentin.

> Tietohallinto tuottaa tietojärjestelmät ja niissä tietojärjestelmissä on määritelty, että mikä on korkein turvaluokka, mitä niissä saa käsitellä ja sitten loppukäyttäjän vastuulle jää, että noudattaa saamiaan ohjeita. (H2)

H1 puhuu tiedon luokittelun osalta yhteisvastuussa, jossa kaikilla kolmella roolilla on yhtä suuri vastuu tiedon luokittelun onnistumisesta. H2 puolestaan korosti loppukäyttäjien vastuuta, sillä he ovat ainoa ryhmä, jotka luokittelua käytännön tasolla toteuttavat. Organisaatiot voivat teknologian avulla pyrkiä tekemään luokittelusta mahdollisimman vaivatonta, mutta siitä huolimatta haastattelujen perusteella aukotonta tiedon luokittelun kokonaisuutta on erittäin vaikea saada rakennettua.

6.6.3 Teknologia ja työkalut

Tiedon luokittelun ja työkalujen ja teknologioiden roolia selvitettiin tässä tutkimuksessa niiden merkityksen kannalta ja osana tiedon luokittelun kokonaisuutta, eikä niinkään sitä, miten teknologioita ja työkaluja teknisesti käytetään. Tutkimuksessa haluttiin tietää, miten tärkeänä teknologioiden rooli nähtiin tiedon luokittelussa.

Haastattelujen perusteella teknologian ja erilaisten teknologisten työkalujen roolilla ja merkityksellä oli kaksi ääripäätä. Ne nähtiin joko täysin keskeisenä tiedon luokittelun ja suojaamisen kannalta, tai sitten merkitys oli hyvin vähäinen. Tämä selittyy suurimmaksi osaksi haastateltavan organisaation teknologioiden hyödyntämisen kypsyysasteella. Haastateltavista H4 ja H5 kertoivat, että teknologioiden roolin olevan hyvin pieni. He kyllä tiedostivat teknologioiden tarjoamat mahdollisuudet tiedon luokittelun prosessien automatisoinnissa, mutta juuri tällä alueella teknologioiden implementointi ei ole ollut prioriteettilistan kärjessä, vaikka sekä H4 ja H5 kertoivat, että kyseisiä teknologioita tullaan todennäköisesti tulevaisuudessa hyödyntämään.

> Teknologioiden rooli on hyvin pieni tällä hetkellä, on tunnistettuja tarpeita, että meidän pitäisi enemmän ja enemmän pystyä käyttämään teknologiaa myös tiedon luokittelun ja tiedon suojaamiseen, mutta tällä hetkellä sitä meillä ollaan tosi perusjutuissa. (H5)

Myös H4 kertoi organisaation tunnistavan teknologioiden tarjoamat mahdollisuudet tiedon luokittelulle, mutta kertoi, että he ovat teknologisella matkalla vielä alkuvaiheessa, ja tiedon luokittelutyökalujen käyttöönotto ei ole ollut heillä prioriteettilistalla niin korkealla, että niitä olisi vielä otettu käyttöön.

Suuri rooli teknologioilla tiedon luokittelussa oli puolestaan H1, H2 ja H3 organisaatioissa, jotka puolestaan kertoivat teknologian olevan täysin keskeisessä roolissa tiedon luokittelun ja siihen liittyvien suojaustoimenpiteiden toteuttamisessa. H1 kertoi, että jatkuvasti tuotetun datan määrä on niin suuri, että he ovat pyrkineet teknologian avulla automatisoimaan mahdollisimman pitkälle dokumenttien luokittelua, ja luokittelun mukanaan tuomia suojauskontrolleja. He tunnistavat tosiasian, että kaikkea ei voida koskaan täysin ulkoistaa ja automatisoida teknologian avulla, mutta he pyrkivät toteuttamaan kaiken teknologioiden avulla, kun mahdollista. H2 puolestaan kertoi myös teknologian roolin olevan keskeinen, ja sanoi jopa olevan työsuojeluongelman, että heillä on

velvoittavia määräyksiä tiedon luokittelun suhteen, mutta työkalut eivät ainakaan tällä hetkellä täysin tue tiedon käsittelijöiden tiedon luokittelua.

> Teknologian rooli on keskeinen, et kaikki muu on toiveajattelua ja jopa näin, että minä koen, että se on yhden sortin työsuojeluongelma, jos meillä on velvoittavia määräyksiä tiedonkäsittelystä ja sitten työkalut ei oikeasti tue sitä, että joo, kyllä nostaisin sen kyllä niin kun hyvin keskeiseenkin asemaan. (H2)

Myös H3 näki teknologian roolin keskeisenä osana heidän tiedon luokittelun prosessejaan, ja hän kertoi heidän tunnistavansa, että loppukäyttäjien tiedon luokitteleminen on täysin riippuvaista teknologiasta ja siitä, että he pyrkivät tekemään dokumentin luokkien määrittelyn loppukäyttäjille niin helpoksi, kuin mahdollista. Mikäli dokumenttien luokittelu olisi yhtään vaikeampaa, jäisi se huomattavasti useammin tekemättä.

6.7 Lainsäädännön ja standardien vaikutus tiedon luokitteluun

Haastattelujen perusteella lainsäädännöllä on vaikutusta liiketoiminnan toimialasta riippumatta tiedon luokitteluun, sillä tietyt lait ja säädökset asettavat vaatimuksia tietyn tyyppisen datan ja tiedon säilyttämiselle. Lisäksi haastatteluista kävi myös ilmi, että lakien lisäksi myös erilaiset standardit saattavat vaikuttaa ja määritellä, sekä asettaa tietyn tyyppisiä laatutavoitteita yritysten tiedon luokittelulle ja suojaamiselle. Standardien vaikutus tiedon luokittelulle ei ole haastattelujen perusteella kuitenkaan niin vahvaa verrattuna lainsäädäntöön.

Lainsäädännöstä puhuttaessa selkeästi suurin vaikutus yritysten tiedon luokittelulle nykypäivänä on yleisellä tietosuoja-asetuksella (General Data Protection Regulation, GDPR). Kuten tämän työn luvussa 3.3 käydään läpi, vuonna 2018 EU:n asettama yleinen tietosuoja-asetus asettaa tarkat rajat yritysten henkilötietojen käsittelylle ja säilyttämiselle. Käsittelyn rajoitusten lisäksi asetuksessa on määritelty sanktiot, joita yrityksille koituu henkilötietojen käsitellyrikkomuksista. Kaikki haastateltavat mainitsivat GBPR:n lakina, joka vaikuttaa tiedon luokitteluun niin, että kaikki yksityishenkilöiden yksilöiviä tietoja sisältävä data on luokiteltava huolellisesti, ja data käsittelylle on asetettava vahvat käsittely-, suojaus, sekä säilytyspolitiikat.

Haastatelluista organisaatioista 4/5 on julkisesti noteerattuja pörssiyhtiöitä, ja näiden yritysten kohdalla pörssilainsäädäntö vaikuttaa vahvasti tiedon luokitteluun ja suojaamiseen. H2 kertoi, että pörssilainsäädännön mukaisesti julkisesti noteeratun pörssiyhtiön tuloksentekokykyyn tai markkina-arvoon vaikuttavia taloustietoja ei saa vuotaa ennen virallista tiedon julkistamista. Myös taloustietojen vuotamisesta, tai tietoisesta väärinkäyttämisestä määrätään taloudellisia sanktioita yrityksille, joten siitä syystä kaikki taloustiedot täytyy luokitella vahvasti ennen julkistamista, ja tiedoille täytyy myös määritellä tarkat käsittely-, jakamis-, sekä suojaussäännöt.

Muita tiedon luokitteluun vaikuttavia lakeja ja asetuksia, joiden kerrottiin vaikuttavan tiedon luokitteluun ja suojaamiseen olivat yksityisyyden suoja, yksityisyydensuoja työelämässä, sekä kirjesalaisuus. H1 mainitsi yksityisyydensuojan ja kirjesalaisuuden vaikuttavan heidän tiedon luokitteluunsa niin, että he eivät ole voineet ulottaa automaattisia työkaluja ja toimenpiteitä työntekijöiden sähköpostiviestintään, ja näin ollen se tekee sähköpostiviestinnän valvomisen huomattavan vaikeaksi sähköpostin ollessa yksi yleisimmistä kanavista, jossa turvaluokiteltuja tietoja voidaan käsitellä tai jakaa väärin.

Standardien vaikutus tiedon luokitteluun ja suojaamiseen on lainsäädäntöön verrattuna enemmän toimenpiteiden, ylläpidon ja kehittämisen laatua ohjaava, kun velvoittava. H1 ja H2 kertoivat, että heidän organisaatioissaan on seurattu enemmän tai vähemmän ISO27000-standardiperheeseen kuuluvaa ISO27001 tietoturvastandardia, josta kerrottaan myös tämän työn luvussa 3.3. ISO27001 on tietoturvan johtamisjärjestelmää käsittelevä standardi, ja yhtenä sen osana käsitellään myös tiedon luokittelua, ja sen ylläpitoa. Tämän lisäksi H1 mainitsi heidän organisaationsa noudattavansa toiminnassaan myös ISO9001-laatustandardin mukaisia toimintamalleja mm. tiedon luokittelun osalta, joiden avulla voidaan osoittaa yhteistyökumppaneille toiminnan laadukkuuden olevan korkealla tasolla.

7 POHDINTA JA JOHTOPÄÄTÖKSET

Tämän tutkielman tarkoituksena oli tutkia tiedon luokittelua tiedon suojaamisen näkökulmasta, sekä laadullisin tutkimusmenetelmin toteutetun tutkimuksen avulla pyrkiä selvittämään syvällisesti sitä, miten yritykset suhtautuvat tiedon luokitteluun, ja millaisia hyötyjä tiedon luokittelun koetaan tarjoavan tiedon suojaamiselle, sekä liiketoiminnalle yleisesti. Tämän työn kirjallisuuskatsauksen luvussa 3.1 kerrotaan lyhyesti, kuinka tiedon luokittelu on varsin usein aliarvostettua yritysten näkökulmasta, vaikka, kuten tämänkin työn luvuissa 2.5.1 ja 2.5.2 kerrotaan, se tarjoaa lukuisia yrityksille lukuisia hyötyjä. Tästä syystä tämän työn tärkeimpänä tavoitteena oli pyrkiä lisäämään ymmärrystä siitä, millaisena liiketoiminnan ja tiedon suojaamisen osana tiedon luokittelu nähdään. Laadullisella tutkimusotteella pyrittiin saavuttamaan syvällinen ymmärrys yritysten suhtautumisesta tiedon luokitteluun, vertaamaan tutkimuksen löydöksiä edeltävään tietoon aiheesta, ja joko vahvistamaan edeltävää kirjoitettua tietoa, tai löytämään aineistosta mahdollisesti esille nousevia uusia näkökulmia tiedon luokitteluun.

7.1 Pohdinta

Datalla ja tiedolla on jatkuvasti kasvava rooli yritysten liiketoiminnassa, ja nykypäivänä yrityksiltä vaaditaan yhä useammin aineettoman omaisuuden, joksi myös data lasketaan, hallintaan laadittu strategia, jonka avulla yritysten on mahdollista parantaa päätöksentekoa, liiketoimintaprosesseja, riskienhallintaa, kustannustehokkuutta ja tuottoja (Marr, 2017). Vaikka tietoon perustuvan päätöksenteon tärkeys kasvaa jatkuvasti, yritykset eivät osaa laskea tarpeeksi arvoa omistamalleen datalle, sekä aineettomalle omaisuudelle yleisemmin. (Fleckenstein & Fellows, 2018). Tämän työn empiirisessä osassa lähdettiin lähestymään tiedon luokittelua siitä, miten haastateltavat näkivät datan roolin heidän liiketoiminnassaan, ja annettujen vastausten perusteella datan suuri rooli ja tärkeys liiketoiminnan kannalta osattiin nähdä hyvin selkeästi. Nykypäivän pilviteknologiat ja kehittyvät datan käsittelyn työkalut kehittyvät valtavalla vauhdilla, ja

datasta pystytään jalostamaan helpommin konkreettista lisäarvoa. Tästä syystä tuntuu, että datan tärkeys ja arvo kasvaa jatkuvasti, ja alkaa olemaan jo nykypäivänä melko korkealla tasolla. Datan hyödyntämisen kehittymisessä yksi vuosikin on jo todella pitkä aika, ja yritysten edetessä teknologisella matkallaan, data tulee varmasti vain kasvattamaan arvoa tulevaisuudessa.

Datan ja tiedon tehokkaan hyödyntämisen ensimmäisenä vaatimuksena on se, että tiedetään, millaista dataa omistetaan. Parhaan lisäarvon saavuttaminen vaatii systemaattista toimintaa. Kaikki yrityksen omistama data ei ole yhtä tärkeään, joten tiedon luokitteleminen ominaisuuksien, käyttötarkoituksen, sekä muiden attribuuttien mukaan luo perustan tehokkaalle hyödyntämiselle. Tietoa luokitellaan yrityksissä yleisimmin liiketoiminnallisen arvon, sekä luottamuksellisuuden perusteella, ja luokitukset auttavat kohdentamaan mm. tiedon suojaamiseen käytettäviä resursseja. Tiedon luokittelu lukeutuu usein osaksi yritysten tietoturvan johtamisjärjestelmää (Information Security Management System, ISMS), ja tiedon luokat nimetään usein juuri luottamuksellisuuden mukaan. Tämän työn tulokset vahvistavat edeltävää tietoa siitä, miten tiedon luokittelua lähestytään yrityksissä. Tuloksista kävi vahvasti ilmi, että yritysten tiedon luokittelu tapahtuu vahvasti luottamuksellisuuden näkökulmasta. Tuloksista on havaittavissa se, että varsinkin luottamuksellisimpia dokumentteja pyritään luokittelemaan tehokkaasti, ja luokat on jaettu usein luottamukselliseen, ja salaiseen tietoon. Toisaalta tutkimuksen tulokset vahvistivat myös tässä työssä läpikäytyä, Veritas Technologiesin julkaisemaa tutkimustietoa siitä, että maailman kaikesta datasta yli puolet on luokittelematonta ”pimeää” dataa. Haastatteluvastauksista kävi ilmi, että historian saatossa kertynyt valtava datamäärä, ja lukemattomat järjestelmät aiheuttavat sen, että kaikkea dataa ei ole ollut yksinkertaisesti mahdollista luokitella halutulla tavalla.

Tiedon tehokas luokittelu tarjoaa runsaasti mahdollisuuksia datan tehokkaammalle hyödyntämiselle ja liiketoiminnan tehostamiselle. Lisäksi yritystoiminta tuottaa myös mm. asiakkaista sellaista henkilökohtaista tietoa, jonka säilyttämistä ja suojaamista säädellään lakien avulla. Vaikka tiedon luokittelun hyödyt ovat kiistattomia, voi tiedon luokittelun taustalla olla erilaisia motiiveja. Motiiveja voivat olla esimerkiksi toiminnan ja resurssien käytön tehostaminen, tai se, että pyritään ainoastaan täyttämään lailliset vaatimukset ja näin välttymään sanktioilta. Tämä tutkimus osoitti, että tärkein motiivit tiedon luokittelulle on vaatimustenmukaisuus (compliance), joka tarkoittaa sitä, että liiketoiminnan datasta halutaan tunnistaa se data, jonka käsittely on luottamuksellista, ja jonka päätyminen ulos organisaatioista aiheuttaisi laillisia seuraamuksia. Vuonna 2018 voimaan tullut yleinen tietosuoja-asetus on vaikuttanut vahvasti siihen, että jokaisessa organisaatiossa tiedon luokittelua on täytynyt ruveta tekemään vähintään yksityishenkilöiden luottamuksellisen datan osalta. Liiketoiminnallisen hyödyn tavoittelu tiedon luokittelun toimenpiteillä sen sijaan on pienessä roolissa, eikä sitä nähty kuin yhdessä organisaatiossa funktioina, jonka avulla pyritään saavuttamaan kilpailuetua.

Tiedon suojaaminen ja tietoturva on liiketoiminnan osa, jonka tärkeys kasvaa jatkuvasti samalla, kun teknologinen kehitys etenee ja datan määrä kasvaa.

Pääsynhallinta, elinkaaren hallinta sekä häviämisen estäminen ovat kyberturvan osa-alueita, joita voidaan tehostaa tiedon luokittelun avulla. Tutkimushaastattelujen perusteella yritykset näkevät tiedon luokittelun hyödyt tiedon suojaamisessa. Tiedon suojaamisen oleellinen osa on pääsynhallinta, ja tutkimuksen perusteella yrityksen pitivät tiedon luokittelua tärkeänä pääsynhallinnan keinona varsinkin luottamuksellisimmissa tietoluokissa, ja tärkeää oli tiedon tarkan luokittelun avulla varmistaa, ettei luottamukselliseen tietoon ole pääsyä organisaation ulkopuolisilla käyttäjillä. Pääsynhallinnan tavoin datan häviämisen estämiseen on mahdollista saavuttaa tehokkuutta tiedon luokittelulla, mutta tutkimuksen perusteella hyödyntäminen on melko pienimuotoista. Tiedon elinkaaren hallinta nähtiin tärkeänä osana yrityksen datakokonaisuuden hallintaa, ja tutkimuksen perusteella tiedon luokittelu tarjoaa varsinkin talousdatan elinkaaren hallintaan suurta apua, sillä varsinkin julkisten pörssiyhtiöiden taloustietojen säilyttämiseen liittyy paljon laillisia vaatimuksia. Kaikkiaan tutkimuksen perusteella voidaan sanoa, että yritykset kyllä näkevät tiedon luokittelun potentiaalin tiedon suojaamisen apuvälineenä, mutta hyödyntämisessä ollaan suurelta osin vielä matkan alussa, ja tiedon luokittelun hyötyjä suojaamisen eri osa-alueisiin ollaan vasta selvittämässä ja ottamassa käyttöön.

Tutkimuksessa selvitettiin myös, miten vahvasti liiketoiminnalliset edut tulevat esille suhtautumisessa tiedon luokitteluun, sillä tässäkin työssä käytyjen näkökulmien kautta on selvää, että tehokkaalla tiedon luokittelemisella on mahdollista saavuttaa liiketoiminnallisia etuja. Tehokas tiedon luokittelu auttaa pitämään tiedon suojassa, ja vähentää huomattavasti datan ja dokumenttien käsittelyyn liittyvien virheiden manuaalisen käsittelyn määrää, ja mm. poistaa viivettä tärkeissä projekteissa, kun oikeat ihmiset pääsevät käsiksi tietoon, johon heidän kuuluu päästä. Tutkimuksen tulosten perusteella tiedon luokittelua ei kuitenkaan lähestytä kovinkaan usein liiketoiminnallisten etujen näkökulmasta. Aineistosta tuli kuitenkin ilmi myös liiketoiminnallisten etujen näkökulmaa, joten suhtautuminen on enemmän riippuvainen organisaatioista ja henkilöistä, jotka vastaavat tiedon luokittelun hyödyntämiseen liittyvistä kysymyksistä.

Tiedon luokittelu ei ole staattinen toimi, vaan vaatii ylläpitoa ja hallintaa, koska datan määrä ja monimuotoisuus, sekä erilaisten järjestelmien määrä kasvaa jatkuvasti. Kuten Bergström ym. (2021) kirjoittavat, on tiedon luokittelu osa tietoturvan johtamisjärjestelmää, joka toimii karttana yrityksen johdolle tietoturvan johtamisessa. Tutkimuksen perusteella johtamisen osana organisaatioissa toimivan ja tehokkaan tiedon luokittelun kannalta erityisen tärkeää on luokittelun kulttuurin rakentaminen, sillä data ja dokumenttien loppukäyttäjät ovat ne, jotka loppujen lopuksi asettavat oikeat luokitukset käsittelemilleen dokumenteille, tai jättävät asettamatta. Tehokkaan ylläpidon ja hallinnan kannalta oleellista on tehdä luokittelun suorittaminen teknologioiden avulla mahdollisimman helpoksi, mutta siitä huolimatta täydellisyyttä luokittelussa on vaikea saavuttaa. Hallinta vaatii myös luokittelupolitiikkojen sekä luokittelujen säännöllistä tarkastelua, ja tutkimuksen perusteella niissä organisaatioissa, joissa oltiin luokittelussa pisimmällä, myös ylläpitoa ja tarkastelua tehtiin säännöllisesti, ja

tarkasteluprosessiin osallistui oikeat henkilöt, kuten eri liiketoiminta-alueiden johto, tiedon omistajat, sekä joskus myös loppukäyttäjät.

Kuten kaikkea liiketoimintaa, myös tiedon, varsinkin arkaluontoisen, yksityishenkilöihin liittyvät käsittelyä säädellään lakien avulla. Lakien lisäksi liiketoiminnan laadunhallintaan ja laadukkuuden osoittamista varten on olemassa erilaisia standardeja. Tiedon luokitteluun vaikuttavan lainsäädännön osalta tässä tutkimuksessa erityisesti yleinen tietosuoja-asetus (GDPR) nousi erittäin vahvasti esille, sillä se on asettanut eurooppalaisille yrityksille tarkkoja vaatimuksia henkilötietojen keräämiselle, säilyttämiselle sekä hallinnoimiselle (Your Europe, 2021). Lisäksi tutkimus osoitti, että mm. yksityisyyden suoja, yksityisyydensuoja työelämässä, sekä kirjesalaisuus ovat lakeja, jotka ohjaavat ja asettavat rajoitteita tiedon luokittelulle, sillä esimerkiksi yksityisyydensuoja ja kirjesalaisuus estävät teknologioiden käytön työntekijöiden sähköpostiviestien sisällön automaattiseen lukemiseen ja luokittelemiseen. Tiedon luokitteluun liittyvät tunnetuimmat, ISO27000-perheen standardit (Broderick, 2006) eivät tutkimuksen perusteella määrittele tai ohjaa tiedon luokittelun toimenpiteitä kovinkaan yleisesti, vaan tiedon luokittelun lähestyminen tapahtuu useammin omista tarpeista ja käytettävissä olevista teknologioista.

7.2 Johtopäätökset ja jatkotutkimusaiheet

Tätä tutkielmaa ei ole tehty tilaajalle toimeksiantona, mutta idea työn aiheeseen tuli tutkielman tekijän työnantajalta Sulava Oy:ltä. Sulava on konsulttiyhtiö, joka tarjoaa konsultointia pilviteknologioiden hyödyntämiseen, ja yksi konsultoinnin osa-alue on tietoturva, ja tiedon suojaaminen, jonka yksi osa tiedon luokittelu on. Työn kirjallisuuskatsauksessa käsitellään tiedon luokittelua tiedon suojaamisen näkökulmasta, mutta myös sitä, miten tiedon luokittelu sekä datan tehokas hallinta voivat auttaa myös monilla muilla tavoilla yrityksen liiketoimintaa, ja saavuttamaan kilpailuetua.

Tiedon luokittelusta kirjoitetun tiedon perusteella kirjallisuuskatsauksen pääviestit olivat se, että datan määrä, sekä datan käsittelyyn hyödynnettävien ohjelmistojen määrä kasvaa jatkuvasti kiihtyvällä tahdilla, ja se asettaa yrityksen datan hallinnalle ja suojaamiselle jatkuvasti uusia vaatimuksia. Hallinnan kasvavaan haasteeseen yksi vastaus on tiedon luokittelu, jonka avulla voi saavuttaa tehoa tiedon suojaamiseen, sekä liiketoimintaan yleisesti.

Lukuisista eduista huolimatta edeltävän tiedon valossa tiedon luokittelu on kuitenkin aliarvostettua, ja iso osa nykypäivän datasta on luokittelematonta, ”pimeää” dataa. Empiirisessä osassa tavoitteena oli selvittää yritysten suhtautumista tiedon luokitteluun osana tiedon suojaamista ja liiketoimintaa, sekä pyrkiä löytämään syitä, miksi yritykset luokittelevat tietoa sillä tasolla, kun ne tekevät, sekä millaisia etuja tiedon luokittelun nähtiin tarjoavan.

Tutkimuksen empiirinen osa vahvisti edeltävää tietoa aiheesta, ja tulosten perusteella näytti siltä, että yritykset tiedostivat tiedon luokittelun edut tiedon suojaamiselle hyvin, mutta muita etuja ei ole osattu ajatella niin syvällisesti.

Lisäksi datan ja järjestelmien määrä, tiedon luokittelun kulttuurin rakentamisen vaikeus vaikuttivat siihen, ettei sitä pystyttäisi hyödyntämään täysimittaisesti yritystoiminnassa. Toisaalta tutkimus osoitti myös, että esimerkiksi pilviteknologioiden hyödyntämisen kypsyysaste vaikuttaa myös tiedon luokittelun hyödyntämiseen. Pilvisiirtymän alkuvaiheessa olevat yritykset priorisoivat selkeästi muita toimintoja tiedon luokittelun edelle. Tiedon suojaamisessa on paljon erilaisia näkökulmia, mutta tiedon luokittelu on harvoin korkean prioriteetin toimi siitä syystä, että se vaatii täydellisesti toimiakseen loppukäyttäjiltä suurta huolellisuutta, ja kulttuurin rakentaminen on hyvin pitkä prosessi. Tämä todennäköisesti selittää sen, että vielä vuonna 2016 yli puolet kaikesta datasta oli luokittelematonta dataa (Veritas Technologies, 2016).

Tutkimuksessa esille nousseena mahdollisena jatkotutkimuksen aiheena voisivat olla tiedon luokittelun ylläpito ja hallinta, sekä se, millaiset toimintamallit tukisivat parhaiten tiedon luokittelun ajantasaisuuden hallintaa. Tämä olisi tarpeellinen aihe tutkia siitä syystä, että yhtenä haasteena tutkimuksessa esille nousi se, että dataa käsitellään jatkuvasti useammissa paikoissa, alkuperäistä dataa jalostetaan ja rikastetaan jatkuvasti, sekä siirretään uusille alustoille ja uusiin datavarastoihin. Tämä tuo huomattavan haasteen tiedon luokittelun ajantasaisuuden ylläpitämiselle, koska pelkästään uuden monimuotoistuvan datan omistajuuden ylläpitäminen on haastavaa.

LÄHTEET

Adee, S. (2008). 37 Years of Moore's Law. IEEE Spectrum. https://doi.org/10.1109/MSPEC.2008.4505312

Agatic, A., Aksentijevic, S., Tijan, E. (2011). Information security as utilization tool of enterprise information capital. Conference: MIPRO, 2011 Proceedings of the 34th International Convention. https://www.researchgate.net/publication/221412859_Information_security_as_utilization_tool_of_enterprise_information_capital

Ahmed, J., & Ahmed, M. (2020). A Study of Big Data and Classification of NoSQL Databases," 2020 IEEE International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET), 2020, pp. 1-8, doi: 10.1109/TEMSMET51618.2020.9557566.

Ashenden, D. (2008). Information Security management: A human challenge? Information Security Technical Report Volume 13, Issue 4, November 2008, Pages 195-201. https://doi.org/10.1016/j.istr.2008.10.006

ASML. (2021). How microchips are made. https://www.asml.com/en/technology/all-about-microchips/how-microchips-are-made

Baltes, P., Kunzmann, U. (2004). The Two Faces of Wisdom: Wisdom as a General Theory of Knowledge and Judgment about Excellence in Mind and Virtue vs. Wisdom as Everyday Realization in People and Products. Human Development 2004;47:290–299 DOI: 10.1159/000079156

Barber, S., Zaeem, R. (2020). The Effect of the GDPR on Privacy Policies: Recent Progress and Future Promise. ACM Transactions on Management Information SystemsVolume 12Issue 1March 2021 Article No.: 2pp 1–20. https://doi.org/10.1145/3389685

Barbosu, M., Doll, T., Hung Le, X., Wang, D., (2012). An enhancement of the Role-Based Access Control model to facilitate information access management in context of team collaboration and workflow. Journal of Biomedical Informatics. https://doi.org/10.1016/j.jbi.2012.06.001

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. Journal of Management. https://journals.sagepub.com/doi/10.1177/014920639101700108

Barta, D., Lenaghan, M., Press, G., Reiner, D. (2004). Information Lifecycle Management: The EMC Perspective. International Conference on Data Engineering (ICDE'04). https://doi.org/10.1109/ICDE.2004.1320052

Bergström, E., Karlsson, F., Åhlfeldt, R. (2021). Developing an information classification method. Information & Computer Security Vol. 29 No. 2, 2021 pp. 209-239 © Emerald Publishing Limited 2056-4961. DOI 10.1108/ICS-07-2020-0110

Brock, D. & Laws, D. (2012). The Early History of Microcircuitry: An Overview. IEEE Annals of the History of Computing. https://doi.org/10.1109/MAHC.2011.85.

Broderick, J. (2006). ISMS, security standards and security regulations. Information Security Technical Report Volume 11, Issue 1, 2006, Pages 26-31. https://doi.org/10.1016/j.istr.2005.12.001

Buckland, M. (1991). Information as Thing. Journal of the American Society for Information Science (1986-1998); Jun 1991; 42, 5. https://skat.ihmc.us/rid=1KR7VC4CQ-SLX5RG-5T39/BUCKLAND(1991)-informationasthing.pdf.

Calder, A. (2016). Nine Steps to Success: An ISO27001:2013 Implementation Overview. Third edition. Ely, Cambridgeshire, United Kingdom: ITGP.

Calder, A., Watkins, S. (2019). Information Security Risk Management for ISO 27001/ISO 27002, third edition. IT Governance Publishing

Cao, J., Cao, Z., Duan, S., Fang, B., Han, P., Liu, C., Pan, H. (2020). CloudDLP: Transparent and Scalable Data Sanitization for Browser-Based Cloud Storage. https://doi.org/10.1109/ACCESS.2020.2985870

Chamberlain, R., Steinbrueck, T. (2020). Demo Abstract: More Than Two Decades of IoT. IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation. DOI 10.1109/IoTDI49375.2020.00041

Chen, Y. (2005). Information valuation for Information Lifecycle Management. Second International Conference on Autonomic Computing (ICAC'05). https://doi.org/10.1109/ICAC.2005.35

Choo, C. (1996). The Knowing Organization: How Organizations Use Information to Construct Meaning, Create Knowledge and Make Decisions. International Journal of Information Management, Vol. 16, No. 5, pp. 329-340, 1996. http://choo.fis.utoronto.ca/FIS/respub/IJIM1996.pdf.

Collins, G., (14.10.2002). Claude E. Shannon: Founder of Information Theory. Scietific American. https://www.scientificamerican.com/article/claude-e-shannon-founder/

Delphin Carolina Rani, A., Jamuna, K., Keerthi, V., Punithavathi, R., Poornima, K. P., Raajeshwari, K. (2022). Preserving Data by Role based Access Control and Query Auditing System. International Conference on Sustainable Computing and Data Communication Systems https://doi.org/10.1109/ICSCDS53736.2022.9760836

Denning, P. J. (1996). Before memory was virtual. http://denninginstitute.com/pjd/PUBS/bvm.pdf

Dmitriev, N., Dubolazova, Y., Konnikov, E., Radionov, D., Zaytsev, A. (2021). Modeling Changes in the Enterprise Information Capital in the Digital Economy. Journal of Open Innovation: Technology, Market, and Complexity. https://doi.org/10.3390/joitmc7030166

Eskola, J., Suoranta, J. (1998). Johdatus laadulliseen tutkimukseen. Vastapaino.

Fleckenstein, M., Fellows. (2018). Modern Data Strategy. Springer. https://doi.org/10.1007/978-3-319-68993-7

Intezari, A., Pauleen, D., Taskin, N. (2016). The DIKW Hierarchy and Management Decision-Making. 2016 49th Hawaii International Conference on System Sciences (HICSS), 2016, pp. 4193-4201, doi: 10.1109/HICSS.2016.520.

Hirsijärvi, S., Hurme, H. (2008). Tutkimushaastattelu: Teemahaastattelun teoria ja käytäntö. Gaudeamus.

Heikkilä, T. (2014). Tilastollinen Tutkimus. Edita Publishing Oy.

Suomen Standardisoimisliitto SFS ry (2014). SFS-ISO/IEC 27001:2013/Cor 1:2014.

Kaur, H. & Kushawa, A. (2018). A review on integration of big data and IoT. 2018 4th International Conference on Computing Sciences. https://doi.org/10.1109/ICCS.2018.00040

Khan, A., Laghari, A., Laghari R. (2021). A Review and State of Art of Internet of Things (IoT). Archives of Computational Methods in Engineering. DOI: 10.1007/s11831-021-09622-6

Khare, S. & Totaro, M. (2019) Big Data in IoT. 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-7, doi: 10.1109/ICCCNT45670.2019.8944495.

Kilburn, T., Edwards D., Lanigan, M., Sumner, F. (1962) One-Level Storage System. IRE Transactions on Electronic Computers, (2), 223-235. https://doi.org/10.1109/TEC.1962.5219356

Kiran, G., Nalini, M. (2020). Enhanced security-aware technique and ontology data access control in cloud computing. International Journal of Communication Systems. https://doi.org/10.1002/dac.4554

Lee, G.Y., Yong, L. (2021). Security Management Suitable for Lifecycle of Personal Information in Multi-User IoT Environment. MDPI Journal.

Lemos, N. M. (2007). An Introduction to the Theory of Knowledge. Cambridge University Press. Cambridge, UK: Cambridge University Press, 2007. ISBN 9780521842136. https://search-ebscohost-com.ezproxy.jyu.fi/login.aspx?direct=true&db=nlebk&AN=304563&site=ehost-live.

Garcia, J., Marín-Torder, E., Masip-Bruin, X., Sinaeepourfard, A. (2016). Towards a Comprehensive Data LifeCycle Model for Big Data Environments. International Conference on Big Data Computing, Applications and Technologies. http://dx.doi.org/10.1145/3006299.3006311

Marr, B. (2017). Data Strategy How to profit from a world of big data, analytics and the internet of things. Kogan Page Limited. http://sutlib2.sut.ac.th/sut_contents/H173714.pdf.

Martín - de - Castro, G., Navas - López, J. E., López - Sáez, P., Alama - Salazar, E. (2005). Organizational capital as competitive advantage of the firm. Journal of Intellectual Capital. Vol. 7 No. 3, 2006 pp. 324-337. https://www.emerald.com/insight/content/doi/10.1108/14691930610681438/full/html.

Ong, Y.J., Qiao, M., Raphael, R., Routray, R. (2017). Context-Aware Data Loss Prevention for Cloud Storage Services. https://doi.org/10.1109/CLOUD.2017.58

O'Hanley, R., Tiller, S. (2014). Information Security Management Handbook, Volume 7. Taylor & Francis Group.

Ross, D. (2021). The Stanford Encyclopedia of Philosophy (Fall 2021 Edition). https://plato.stanford.edu/archives/fall2021/entries/game-theory/

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information Science. https://doi.org/10.1177/0165551506070706

Hyvärinen, M., Nikander, P., Ruusuvuori, J. (2010). Haastattelun analyysi. Vastapaino

Shannon, C. E. (1948). The synthesis of two-terminal switching circuits. The Bell System Technical Journal, 28(1), 59-98. https://ieeexplore-ieee-org.ezproxy.jyu.fi/stamp/stamp.jsp?tp=&arnumber=6771698&tag=1.

Shapiro, C. (1989). The Theory of Business Strategy. The RAND Journal of Economics, Spring, 1989, Vol. 20, No. 1 (Spring, 1989), pp. 125-137. https://www.jstor.org/stable/2555656

Sheldon, R. (2022). Data Lifecycle Management (DLM). https://www.techtarget.com/searchstorage/definition/data-life-cycle-management

Soomro, Z., Hussain, M., Ahmed. (2016). Information security management needs more holistic approach: A literature review. International Journal of Information Management Volume 36, Issue 2, April 2016, Pages 215-225. https://doi.org/10.1016/j.ijinfomgt.2015.11.009

Tuomi, J., Sarajärvi, A. (2018). Laadullinen tutkimus ja sisällönanalyysi. Kustannusosakeyhtiö Tammi.

Turkel, W., Muhammedi, S., Start, M. (2014). Grounding Digital History in the History of Computing. IEEE Annals of the History of Computing. https://doi.org/10.1109/MAHC.2014.21

Vaughan, S., (2019). Data. https://www.techtarget.com/searchdatamanagement/definition/data

Veritas Technologies (2016), "The Databerg Report: see what others don't: Identify the value, risk and cost of your data"

Yle (2020). Psykoterapiakeskus Vastaamon kiristäjä julkaisi yöllä lisää erittäin arkaluontoisia potilaskertomuksia. https://yle.fi/a/3-11606925.

Your Europe (2021). https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_fi.htm

Yujun, L. (16.12.2021). The Future is Now. https://www.tsmc.com/english/news-events/blog-article-20211216

LIITE 1 HAASTATTELURUNKO

- Millainen rooli sinulla on edustamassasi organisaatiossa, ja millä tavalla olet tiedon luokittelun kanssa tekemisissä?
- Millainen rooli datalla ja tiedolla on yrityksenne toiminnassa?
- Miten luokittelette omistamaanne dataa ja tietoa?
- Mikä on ollut tärkein syy sille, että dataa ja tietoa on alettu luokittelemaan?
- Millainen prosessi on käyty läpi, että on päädytty sellaiseen luokitteluun, kun on päädytty
- Millaisia hyötyjä tiedon luokittelulla on liiketoiminnalle?
- Millaisia hyötyjä tiedon luokittelulla on tiedon suojaamiselle?
- Miten tiedon luokittelua hallitaan ja ylläpidetään?
- Miten tiedon luokittelun ajantasaisuudesta huolehditaan sekä siitä, ettei luokittelematonta ja väärin suojattua dataa pääse syntymään?
- Miten vastuu tiedon luokittelusta jakautuu yrityksessänne?
- Entä vastuu tiedon suojaamisesta?
- Miten datan ikä/elinkaari vaikuttaa tiedon luokitteluun ja suojaamiseen?
- Millainen rooli erilaisilla teknologioilla on tiedon luokittelussa yrityksessänne?
- Miten tiedon luokitteluun liittyvä lainsäädäntö ja standardit ovat vaikuttavat tiedon luokitteluun?