



# LUNDS UNIVERSITET

## Ekonomihögskolan

*Institutionen för informatik*

---

# Big data, på gott och ont

GDPR som stöd för den enskilde individen

Kandidatuppsats 15 hp, kurs SYSK16 i Informationssystem

Författare: Gustav Stigestadh  
Felix Moberg

Handledare: Magnus Wärja

Examinatorer: Benjamin Weaver  
Umberto Fiaccadori

# Big data, på gott och ont: GDPR som stöd för den enskilde individen

Författare: Gustav Stigestadh, Felix Moberg

Utgivare: Inst. för informatik, Ekonomihögskolan, Lunds universitet

Framlagd: Vårterminen 2018

Dokumenttyp: Kandidatuppsats

Antal sidor: 102

Nyckelord: Big data, Big data-analys, Dataskyddsförordningen, GDPR, Personlig Integritet, PuL, Ostrukturerad data.

Sammanfattning (Max. 200 ord):

Big data-analys är ett kraftfullt verktyg för många organisationer idag. Genom att analysera stora mängder data kan organisationer se mönster, trender, dra slutsatser och förutspå saker som ännu inte skett. Men även om fördelarna med teknologin är många så kan Big data-analys kränka den personliga integriteten på många sätt. I samband med införandet av den nya Dataskyddsförordningen och de ökade kraven på den personliga integriteten undersöker vi hur detta påverkar Big data-analys. Vi undersöker även fördelar och nackdelar med teknologin, hur den kan kränka den personliga integriteten och vad GDPR ställer för krav på organisationer idag. I undersökningen har tre organisationer intervjuats som har god kunskap angående Big data-analys och insikt i lagändringen. Även Datainspektionen har intervjuats då vi vill ha en bred bild och få insikter från lagstiftningens perspektiv. Det visade sig att Big data-analys kommer att förändras. Med bland annat ökade krav på rättslig grund vid varje behandling leder detta till dyrare och mindre flexibla analyser samt en ökning av att anonymisera data, för att skydda personerna som behandlas. Även bättre kunskap inom organisationer angående vad som är en kränkning behövs, och tydligare etiska ramverk utöver GDPR-lagen.

## Innehållsförteckning

1	Introduktion.....	6
1.1	Problemformulering.....	7
1.2	Syfte.....	8
1.3	Avgränsningar .....	8
2	Litteraturgenomgång .....	9
2.1	Big data.....	9
2.1.1	Vad är Big data.....	9
2.1.2	Möjligheter med Big data.....	10
2.1.3	Analyser av Big data .....	11
2.1.4	Datastrukturer.....	14
2.1.5	Problematik med Big data .....	15
2.2	Skyddet av personuppgifter .....	16
2.2.1	Personlig integritet .....	16
2.2.2	GDPR – Dataskyddsförordningen.....	17
2.2.3	Personuppgifter .....	19
2.2.4	Anonymisering .....	19
2.2.5	Pseudonymisering .....	20
2.3	Sammanfattande modell av vår litteraturgenomgång .....	21
3	Metod.....	24
3.1	Intervjuer .....	25
3.1.1	Utformning av Intervjuguide.....	26
3.1.2	Bearbetning av data.....	26
3.1.3	Analys av data .....	26
3.1.4	Informanter.....	27
3.2	Kvalitet av data.....	28
3.2.1	Urval.....	28
3.2.2	Validitet och reliabilitet.....	29
3.3	Etik.....	30
4	Resultat .....	31
4.1	Resultat från intervjuer .....	31
4.1.1	Organisation 1: .....	31
4.1.2	Organisation 2: .....	35
4.1.3	Organisation 3: .....	38
4.1.4	Organisation 4: .....	41

---

5	Diskussion & Analys .....	44
5.1	Big data.....	44
5.1.1	Vad är Big data, definition .....	44
5.1.2	Möjligheter med Big data.....	44
5.1.3	Syfte vid datainsamling .....	45
5.1.4	Utmaningar med Big data .....	46
5.2	Skyddet av personuppgifter .....	47
5.2.1	Personlig integritet & Big data-analys .....	47
5.2.2	Big datas påverkan av GDPR.....	48
5.2.3	Anonymisering .....	49
5.2.4	Negativa aspekter kring GDPR .....	50
6	Slutsats .....	52
7	Förslag till vidare forskning.....	53
	Referenser.....	54
8	Bilagor.....	58
8.1	Intervjuguide.....	58
8.2	Transkribering Organisation 1 .....	59
8.3	Transkribering Organisation 2.....	71
8.4	Transkribering Organisation 3.....	82
8.5	Transkribering Organisation 4.....	91

## Figurer

Figur 1: The 3 V's of Big Data. (Russom, 2011, s. 6) .....	9
Figur 2: Taxonomi över data inom Big data (Sharda et al., 2014, s. 177) .....	14
Figur 3: Sammanfattande modell av litteraturgenomgången .....	21

## Tabeller

Tabell 1, Big data .....	22
Tabell 2, Skyddet av personuppgifter .....	23
Tabell 3: Informanter .....	27

# 1 Introduktion

Det tidigare konceptet Big data är inte längre ett koncept utan en verklighet, idag finns enormt stora ostrukturerade mängder data, som är oerhört värdefull för de som kan hantera och analysera den (Sharda, Delen, & Turban, 2014). Varje dag växer mängden data som produceras exponentiellt. Sensorer från en enda Boeing jetmotor producerar 20 terabyte data varje timme, Youtube producerar mer än en terabyte var 4:e minut och CERN partikelacceleratoren producerar otroliga en petabyte (1000 terabyte) data varje sekund (Sharda et al., 2014).

När vi använder våra mobiltelefoner, datorer, sociala medier, handlar i affären eller pratar med vår home pod samlas data in om oss, och insamlingen av data verkar inte minska (Kuner, Cate, Millard, & Svantesson, 2012). Varje like vi gör på Facebook sparas och genom våra sökningar via sökmotorer genereras direktreklam till oss (Grassegger & Krogerus, 2017). Genom att ha tillgång till stora mängder data från olika källor ges möjlighet att finna mönster och upptäcka hemligheter som tidigare varit gömda och okända (Kuner et al., 2012). Organisationer använder Big data-analys för flera syften och har blivit ett viktigt fundament för flera företags framgång (Deloitte, 2013). Organisationer använder sig bland annat utav Big data för att förbättra kundlojalitet, segmentera kunder, få konkurrenskraftiga fördelar och göra mer effektiva affärsbeslut (Deloitte, 2013). De senaste amerikanska valen har både republikanska och demokratiska kandidaterna använt sig av analyser av Big data; George W. Bush, Obama, Trump och Hillary Clinton för att nämna några (Markman, 2016). Där hade Big data-analys en stor roll för möjliggörandet av analysering på den amerikanska befolkningen (Markman, 2016). Även det senaste amerikanska valet 2016 lades mycket fokus på analyser av Big data för att skapa mer precisa kampanjer för olika områden, kön, ras eller till och med på individnivå (Grassegger & Krogerus, 2017).

Uppkomsten av Big data-analyser har ställt nya utmaningar för den personliga integriteten (Integritetskommittén, 2016). Skandalen med Facebook och Cambridge Analytica gjorde att många fick upp ögonen för Big data och dess problematik för den personliga integriteten. Enligt de senaste rapporterna rör det sig om känslig privat information från 87 miljoner människor som Facebook delat med sig av till Cambridge Analytica berättar Facebook själva i en bloggpost (Schroepfer, 2018). Detta bland annat, för att möjliggöra grundliga analyser av befolkningen och använda detta i bland annat den amerikanska valrörelsen (Miniter, 2018). Big data har med andra ord en stor inverkan på våra vardagliga liv även om användarna själva inte behöver vara särskilt medvetna om det (Integritetskommittén, 2016). Denna typ av teknologi, samt ett flertal typer av behandlingar ska vi i Europa vara skyddade mot i och med GDPR-lagen (General Data Protection Regulation) som börjar gälla den 25:e maj 2018 (Datainspektionen, 2017a).

GDPR-lagen skall hindra behandling av personuppgifter som kan kränka den personliga integriteten (Datainspektionen, 2017a). Den kommer ersätta de personuppgiftslagar som idag finns i EU:s medlemsstater och även gälla för alla länder utanför EU som hanterar personuppgifter tillhörande en EU-medborgare (Datainspektionen, 2017d).

Dock råder det delade meningar i samhället på vilken effekt GDPR kommer att få för individ och organisationer, EU-parlamentarikern Christofer Fjellner (M) menar på att de nya reglerna bygger på en föråldrad syn på hur den personliga integriteten bäst värnas och att den nya lagen kommer att hämma innovation och konkurrenskraft i Europa (Fjellner, 2018). Daniel Akenine, säkerhetschef på Microsoft, menar samtidigt på att det är en modern lagstiftning som lämpar sig väl i ett digitalt samhälle där vi individer kommer få en ljusare framtid i samband med utökade rättigheter i relation till företag som hanterar personuppgifter (Akenine, 2017).

## 1.1 Problemformulering

År 2016 utkom en rapport från Integritetskommittén, en svensk parlamentarisk kommitté med uppgift att kartlägga faktiska och potentiella risker för intrång i den personliga integriteten som kan uppstå med användning av IT i privat och offentlig verksamhet (Integritetskommittén, 2016). Kommittén fann bland annat att Big data för med sig allvarliga risker för den personliga integriteten.

*Faktorer som i detta sammanhang är av betydelse för risken för den personliga integriteten, är att Big data innebär att uppgifter hanteras för nya ändamål som inte är kända vid insamlingen. Det medför att den enskilde förlorar både kännedom om och inflytande över hur uppgifterna hanteras. Med Big data blir det särskilt tydligt att personuppgifter betraktas som en handelsvara med ett högt kommersiellt värde. Det medför en avsevärt ökad risk för spridning av uppgifterna till parter som inte är kända för den enskilde. Kommittén anser därför att Big data för med sig en allvarlig risk för den personliga integriteten.*

(Integritetskommittén, 2016, s. 115)

I Sverige är vi skyddade av personuppgiftslagen (PuL) som har till uppgift att skydda människors personliga integritet i samband med behandlingar av personuppgifter. Med behandling omfattar insamling, registrering, lagring, bearbetning med mera (Datainspektionen, 2007a). Den nya GDPR-lagen, som har en mer omfattande lagstadga samt högre och mer kännbara straff vid brott mot lagen än PuL tillträder den 25e maj 2018 (Datainspektionen, 2017a). I personuppgiftslagen finns en "missbruksregel" för ostrukturerad data som försvinner i samband med den nya GDPR-lagen (Datainspektionen, 2017c). Det innebär att samma regler gäller för alla typer av personuppgifter oavsett om det är ett personnummer i ett dokument, en ljudfil med en persons röst på, en bild på en person, filmer eller en illustration som föreställer en person (Datainspektionen, 2017c), allt detta klassas som en personuppgift (Datainspektionen, 2017c). Detta kommer att omfatta Big data som består till 95 % av ostrukturerad data (Gandomi & Haider, 2015). Den nya lagen medför en hel del administrativt arbete för hundratusentals företag framöver, om ett företag exempelvis vill publicera bilder på sociala medier eller på sin egen webbplats från det senaste företagseventet måste det finnas en rättslig grund till publiceringen (Datainspektionen, 2018). Det innebär att företaget måste vara väldigt tydliga med att bilder kommer tas under eventet och eventuellt publiceras offentligt (Datainspektionen, 2018).



Med den nya GDPR-lagen kommer EU-medborgare även få rätten att bli bortglömd och raderade ur organisationers databaser (Datainspektionen, 2017g). Även detta kan låta som en relativt simpel process men i verkligheten kan bli mycket kostsamt för organisationer på grund av det administrativa arbete som kommer att krävas. Dels måste organisationen radera all information de har om personen men också informera samtliga parter de lämnat ut information till att en radering skall ske (Datainspektionen, 2017g).

Organisationer som genomför Big data-analyser kan komma att genomgå stora förändringar i och med införandet av den nya lagen. Big data-analyser som fortfarande är ett relativt nytt fenomen har intensifierats de senaste åren och blivit en viktig förutsättning för flera företag (Columbus, 2017) men med de ökade kraven på personlig integritet kan detta komma att förändras (ICO, 2017). Den personliga integriteten är viktig i samhället idag, men till vilket pris kommer det att få för Big data-analyser samt organisationer som verkar i och mot EU-länder. Dessa frågor avser vi att få svar på med vår forskningsfråga:

*Hur inverkar större krav på personlig integritet Big data-analys ur ett organisationsperspektiv?*

## 1.2 Syfte

Syftet med denna uppsats är att undersöka hur GDPR-lagen kan komma att påverka Big data-analyser. Fokus kommer även ligga i möjligheter och problem med Big data kopplat till ifall eventuella kränkningar av personlig integritet kan komma att stävjas med i samband med att den nya GDPR-lagen träder i kraft.

## 1.3 Avgränsningar

- Vi själva kommer inte att försöka oss på egna tolkningar av GDPR (dataskyddsförordningen), utan kommer att använda oss av sekundärkällor, som exempelvis datainspektionen.
- Fokus kommer ligga på organisationer i Sverige då våra informanter i huvudsak arbetar på den svenska marknaden

## 2 Litteraturgenomgång

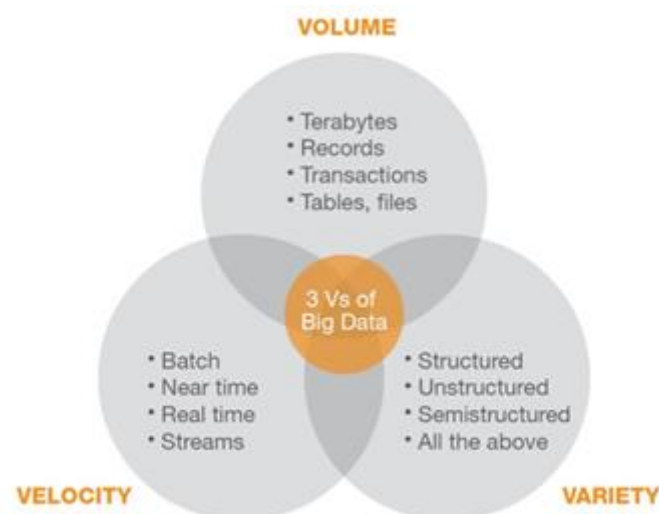
I detta kapitel går vi igenom tidigare forskning inom området, där litteraturgenomgången inleds med förklaring av Big data och övergår så småningom till GDPR och personlig integritet. Litteraturgenomgången ligger till grund för den empiriska undersökning som presenteras i kapitel 4. Detta kapitel avslutas med en sammanfattande tabell över vad som kommit fram i litteraturgenomgången.

### 2.1 Big data

#### 2.1.1 Vad är Big data

Trots att konceptet Big data har varit etablerat ett tag så finns det inte någon direkt vedertagen definition av begreppet (Gandomi & Haider, 2015). Bello-Orgaz, Jung & Camacho (2016) menar att Big data inte handlar om att lagra eller ha tillgång till stora mängder data utan målet med Big data är att analysera stora mängder data så att datan blir meningsfull och skapar värde. Enligt Russom (2011) har det blivit vanligt att förklara Big data genom att summera det i tre stycken V:n – *Volume*, *Variety*, *Velocity* (volym, variation, hastighet). Där man skall se de olika V:na som tre olika dimensioner av data (Rosyad, 2016).

Dessa V:n har även Gartner med i sin definition av Big data ”*Big data is a high-volume, high-velocity, and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation*” (Gartner, 2012).



Figur 1: The 3 V's of Big Data. (Russom, 2011, s. 6)

*Volume* refererar till magnituden av data som kommer från olika källor. Datavolymer som handlar om flera tera- och zetabytes, storlekar som ökas för varje dag som går (Gandomi & Haider, 2015). Enligt en undersökning som IBM gjorde år 2012 så menade mer än hälften av 1144 respondenter att datavolymer över 1 terabyte bör klassificeras som Big data eller stora datavolymer (Gandomi & Haider, 2015).

*Velocity* handlar om hastigheten på dataöverföringar, datainhämtning och den hastighet som data bearbetas (Rosyad, 2016). För flera företag och applikationer är hastigheten mer viktig än datavolym. Genom att kunna hantera realtidsinformation och data snabbt ger det företag en möjlighet att vara mer agila än sina konkurrenter (McAfee & Brynjolfsson, 2012).

*Variety* – Handlar om vilken struktur datan har, där det finns tre olika format; strukturerad, ostrukturerad och semistrukturerad (Gandomi & Haider, 2015). Strukturerad data refererar till data som är organiserad exempelvis tabelldata som finns i kalkylblad eller relationsdatabaser. Strukturerad data utgör ungefär 5% av all data (Gandomi & Haider, 2015). Ostrukturerad data handlar om text, bilder, ljud och har ökat explosionsartat i takt med sociala-mediers framfart. Semi-strukturerad data, exempelvis XML eller HTML, data som saknar formell struktur men som innehåller taggar för att separera dataelement (Charan, 2015).

Flera personer och organisationer har även adderat fler V:n som, *Value* och *Veracity* (värde och riktighet), för att få en ännu mer omfattande definition och förståelse om vad Big data är och innebär (Bello-Orgaz, Jung, & Camacho, 2016). Där value innebär att hämta värdefull information från stora uppsättningar data och där veracity hänvisar till att data och information som man hämtar och använder sig av är trovärdig och korrekt och kan användas vid beslutsfattning (Bello-Orgaz et al., 2016).

### 2.1.2 Möjligheter med Big data

Frasen "*kunskap är makt*" myntade den brittiska filosofen Francis Bacon för över 400 år sedan (Vickers, 1992). Big data handlar om att ju mer man vet om någonting, desto lättare blir det att få nya insikter, göra träffsäkrare förutsägelser och fatta bättre beslut (Berner, Graupner, & Maedche, 2014). Big data kräver kostnadseffektiva och innovativa former av informationshantering för förbättrad insyn och beslutsfattande (ICO, 2017). Genom att jämföra och analysera stora mängder data kan relationer som tidigare varit dolda att upptäckas. Dessa relationer ger organisationer större lärdom och bättre beslutsfattning (Marr, 2017).

Användningsområdena för Big data-analyser är många, det används bland annat inom forskning på sjukdomar och inom sjukvården. Forskare använder bland annat Big data-analyser för att kartlägga aktiviteten på över 100 miljarder neuroner i hopp om att finna svar på gåtorna kring Alzheimer och Parkinson (Michael & Miller, 2013). Inom sjukvården kan Big data-analyser vara ett verktyg för läkare och forskare. Mellan år 2000 - 2010 genomfördes en omfattande analys inom sjukvården i Storbritannien eftersom de hade väldigt låg överlevnadsnivå hos deras cancerpatienter. Genom att använda komplexa algoritmer som analyserade 118 miljoner journaler av 2 miljoner patienter kunde man se att nästan 25 % av alla fastställda cancerdiagnoser skedde i samband med ett akutbesök. Man kunde också se att de patienter som fått sin diagnos vid ett akutbesök hade lägre chans att överleva sin cancer (ICO, 2017). Detta gjorde att sjukhusen aktivt tog in patienter till läkarundersökningar som var i riskzonen utifrån de grundliga analyser man genomfört. Några år senare var

upptäcknings-nivån vid akutbesök nere på 20 % och tillsammans med genomförandet av kampanjer som riktar sig till riskgrupper fortsätter denna siffra att sjunka (ICO, 2017).

Andra användningsområden där Big data tillämpas är bland annat inom kundsegmentering. Informationen som företag samlar in låter dem profilera och segmentera deras kunder (IBE, 2016). Mestadels används informationen i marknadsföringssyfte men information tillåter även kartläggning av personliga attribut hos kunder som deras förmåga att handla vissa produkter eller tjänster och influera deras möjligheter att köpa vissa produkter (IBE, 2016).

Det finns en stor tendens att återförsäljare analyserar ditt köpbeteende och mönster för att räkna ut vad du gillar, vad du behöver och vilken rabatt som tilltalar dig mest (Duhigg, 2012). Ett numera klassiskt exempel på hur företag kan använda Big data skrev Duhigg (2012) om i New York Times. Den amerikanska detaljhandelskedjan Target kunde med hjälp av analys och algoritmer av kunddata och kunders köpbeteenden förutspå ifall en kvinna är gravid, detta utifrån vilka varor hon köpte. De lyckades med hjälp av analysen estimerade vilken månad de kunde vara gravida i och när de skulle föda. Denna information används sedan för att ge rabatt till bebis-relaterade varor och kuponger relaterade till i vilken graviditetsstadium kvinnan befann sig i. Metoder och analyser som detta exempel har blivit ifrågasatt som integritetskränkande och oetisk (Duhigg, 2012).

### 2.1.3 Analyser av Big data

Oavsett storlek eller typ av information är den ganska obetydlig om den inte används på rätt sätt, men genom att analysera den insamlade datan kan organisationer lära sig och se mönster av den (Sharda et al., 2014). Analyserna som genomförs kan visualisera data i olika typer av modeller, dessa modeller kan göra det möjligt att förutse saker som brottsaktivitet i vissa områden vid olika tider, sjukdomsspridningar, olika trender på marknaden, räkna ut flygbiljettspriser automatiskt och mycket mer (ICO, 2017; Sharda et al., 2014). Några olika typer på analyser av Big data är:

- **Användningen av algoritmer**  
När analys av data sker traditionellt sett så beslutas vad som skall hittas genom analysen, och sedan skapas en query för att ta fram den information som önskas. Big data-analyser startar sällan med en fördefinierad query, utan körs i en “upptäckningsfas” där ett stort antal algoritmer körs tillsammans på en stor mängd data (ICO, 2017). Detta för att hitta ett flertal likheter, samband och mönster. När ett önskat samband eller mönster upptäckts förändrar man algoritmerna i “användningsfasen” och man agerar på data istället för att tänka med data som i den tidigare fasen (ICO, 2017). Systemet lär sig sedan vilka kriterier som är relevanta genom de många analyser av data, detta är således en form av maskininlärning (ICO, 2017).
- **Opaciteten av bearbetning**  
De mest avancerade typer av maskininlärning kallas för djupinlärning som innebär att enorma mängder data passerar ett icke linjärt neuralt nätverk som klassificerar datan för varje lager den består av (ICO, 2017). Den stora komplexiteten genom denna typ av bearbetning gör att det kan uppstå en “svarta lådan”-effekt. Detta innebär att det kan vara svårt att förstå varför olika beslut tas av resultaten från djupinlärningen (ICO, 2017). Systemen kan ofta ta konstiga beslut i brist av mänsklig förståelse och beslut

kan i vissa fall vara irrationella (ICO, 2017). Detta är det som skiljer Big data-analyser från traditionella sätt att analysera data (ICO, 2017).

- **Tendensen att använda all data**  
Det är ofta nödvändigt att hitta statistiskt representativa urval eller slumpmässiga urval vid analys av data för forskning. Men Big data-analyser handlar mer om att analysera all den data som finns tillgänglig (ICO, 2017). Så istället för att genomföra en enklare enkät i en butik för att få en uppfattning om kundernas lojalitet, nöjdhet med mera (ICO, 2017). Så kan större analyser genomföras på alla som använder ett kundkort för att se olika mönster som vilka tider de går till butiken, handlar de endast när det är rea, vilka varor lämnas ofta tillbaka och mycket mer. Detta har gjort att det är lättare att lagra samt analysera en ständigt ökande mängd av data (ICO, 2017).
- **Återanvänd data**  
En annan stor fördel med att använda Big data-analyser är återanvändningen av data. I dessa analyser kan data som var insamlad för ett syfte användas för ett helt annat. Detta ger i sin tur många möjligheter för att hitta nya insikter och mönster mellan data som kan se helt diversifierad ut, men genom rätt analys faktiskt kan vara väldigt relevant för organisationen (ICO, 2017; Sharda et al., 2014). Genom att analysera geografisk data tillsammans med upplagda foton, tweets eller facebook-statusar i samband med fotbollsmatcher eller turistområden kan organisationer förstå var användarna kommer ifrån, var och hur länge de stannar, hur många de är, var dom handlar med mera (ICO, 2017). Detta kan sedan användas för att rikta reklam i deras egen hemstad för att öka turism eller försäljning (Sharda et al., 2014).
- **Textanalyser:**  
Data från sociala medier, bloggar, forum, email, dokument eller annan form av data som innehåller text räknas in som textuell data (Sharda et al., 2014). Genom att genomföra analyser av text producerad av människor kan organisationer dra slutsatser och ta beslut snabbare och bättre än förut (Bello-Organ et al., 2016). Detta beslutsstöd kan förutse hur börserna kommer röra sig utifrån analyser på text genererade från finansiella tidningar och liknande (Gandomi & Haider, 2015). Även personers åsikter analyseras kontinuerligt då detta är av stort intresse för många organisationer. Åsikter mot ett företag, en produkt eller ett event kan läsas av genom att analysera text skriven om dem och kan sedan användas för att förbättra produkten eller företaget (Gandomi & Haider, 2015).
- **Ljudanalyser:**  
Ljud som samlas in från t.ex. kundsupport eller hälsorådgivning räknas in som ljud-data (Gandomi & Haider, 2015). Detta kallas även för röst-data eftersom det i största utsträckning handlar intalat ljud (Gandomi & Haider, 2015). Detta kan vara användbart för kundservice via telefon där tusentals eller miljoner timmar av inspelade samtal kan analyseras för att förbättra kundupplevelsen, öka försäljning eller förstå hur kunder agerar i olika situationer (Gandomi & Haider, 2015).
- **Videoanalyser:**  
Att analysera videos har innan Big data-analyser varit svårt på grund av den enorma datamängd som produceras av videor. En sekund av högupplöst video motsvarar över 2000 sidor av skriven text (Gandomi & Haider, 2015; Sharda et al., 2014). Med hjälp

av smarta algoritmer som genomför videoanalyser på Big data kan övervakning ske automatiskt där systemet känner av om en person rör sig i ett förbjudet område, om ett rör har börjat läcka eller att någon försöker sabotera kameran (Gandomi & Haider, 2015). Systemet genomför automatiskt en åtgärd som att tillkalla personal, låta ett larm ljuda eller slå på lampor. Detta kan även användas av butiker som genom denna typ av analys kan få en demografisk bild av sina kunder (Gandomi & Haider, 2015). Hur gamla de är, vilken hudfärg, kön och liknande (Gandomi & Haider, 2015). Detta kan sedan sättas i perspektiv med hur många kunder butiken har vid vissa tider och vad de handlar (Gandomi & Haider, 2015). Den insamlade datan och resultatet av dessa analyser kan sedan användas för direktreklam, hur och var de ska placera produkter i butiken med mera (Gandomi & Haider, 2015).

- **Sociala medier analyser:**

Vid analys av sociala medier inkluderas; bloggar, mikrobloggar, sociala nätverk, wikis, media-delande medier (instagram, youtube m.fl.), forum, fråga-svar sidor, telefonapplikationer (Tinder m.fl.) med mera (Gandomi & Haider, 2015). All data som genereras av människor kategoriseras som användargenererad data, detta inkluderar text, videor, bilder med mera (Bello-Organ et al., 2016). Sedan analyseras relationer och interaktioner mellan olika nätverks entiteter t.ex. människor, företag, organisationer och produkter (Gandomi & Haider, 2015). Från ovanstående delas sedan analyser av sociala medier upp i två kategorier:

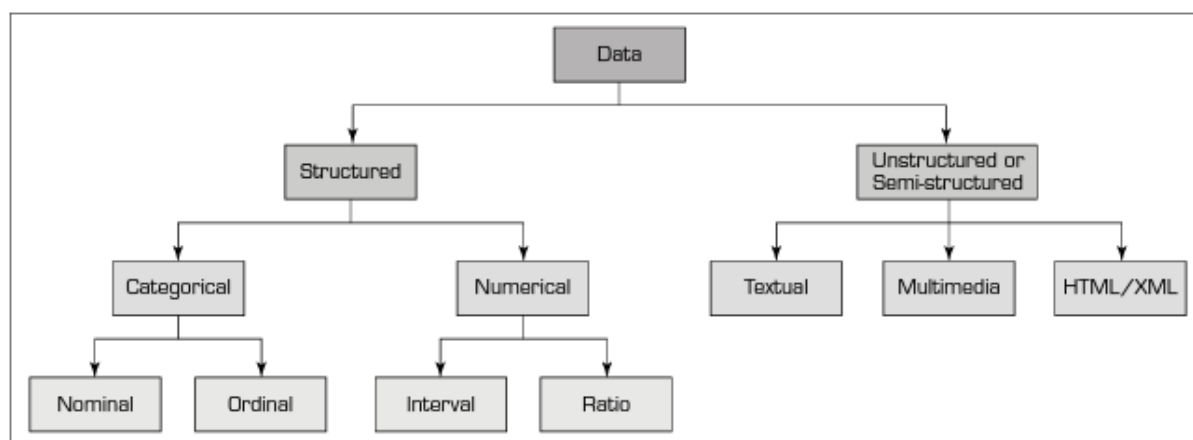
**Innehållsbaserad analys:**

Denna typ av analys fokuserar på data som producerats av användare. Det betyder att analys sker på bilder, videor, produktrecensioner, kundfeedback m.fl. Datat är ofta ostrukturerat och väldigt omfattande och därför kan en mix av text och video-analys kombineras för att möjliggöra slutsatser från denna typ av data (Gandomi & Haider, 2015).

**Strukturbaserad analys**

Genom att analysera relationen mellan olika entiteter inom sociala nätverk så kan kunskap extraheras från detta. Denna typ av analys kallas därför också för "sociala nätverks-analys". Strukturen av sociala nätverk är uppbyggd av noder som representerar deltagarna och deras relationer (Bello-Organ et al., 2016; Gandomi & Haider, 2015). Detta kan visualiseras i grafer för att identifiera olika befolkningsgrupper eller särskilda användare med stora sociala nätverk, men också se hur en tweet eller bild vidarebefordras och sprids mellan användare och som senare kan analyseras (Bello-Organ et al., 2016; Gandomi & Haider, 2015)

## 2.1.4 Datastrukturer



Figur 2: Taxonomi över data inom Big data (Sharda et al., 2014, s. 177)

### Strukturerad data

Strukturerad data är den data som kan klassificeras som kategorisk eller numerisk. Detta kan sedan delas in i nominal, ordinal, intervall eller ratio-data.

- Kategorisk data är en representation av olika klasser som används för att dela upp variabler i specifika grupper. Det kan vara variabler som ras, kön, åldersgrupper eller utbildningsnivåer (Sharda et al., 2014). Inom kategorisk data finns nominal data som innehåller mätningar av data som kan vara sann eller falsk, bra eller dålig, ja eller nej. Informationen kan t.ex. vara om en person är gift, frånskild eller singel. Ordinal data innehåller koder som är dedikerade till objekt eller event och kan beräknas samt kategoriseras som låg, medel eller hög (Sharda et al., 2014). Detta används t.ex. vid beräkning av kreditvärdighet. Det kan beskriva variabler som åldersgrupper, utbildningsnivå eller liknande (Sharda et al., 2014).
- Numerisk data ser till siffervärden och specifika variabler. Detta inkluderar åldrar, hushållsinkomst, resesträckor, antal barn i ett hushåll eller temperaturer (Sharda et al., 2014). Intervalldata är variabler som kan mätas på en intervallskala, såsom temperatur där skalan kan gå från 1/100 där 1 är fryspunkt och 100 är kokpunkt. Ratio-data beräknar variabler som ofta finns inom teknik och fysik. Variablerna kan vara massa, tid, längd, energi eller elektrisk laddning (Sharda et al., 2014).

### Ostrukturerad eller Semistrukturerad data

Data som behöver konverteras innan bearbetning räknas in i ostrukturerad eller semistrukturerad data. Big data består nästan uteslutande av ostrukturerad data. Så stor andel som 95% av all data inom Big data är ostrukturerad (Gandomi & Haider, 2015). Detta inkluderar text, bild, röst, videor, data från sensorer, sociala nätverk med mera. Denna typ av data måste konverteras till någon form av kategorisk eller numerisk representation för att det skall vara möjligt att processa dem genom olika typer av algoritmer (Sharda et al., 2014).

### 2.1.5 Problematik med Big data

Ett av de främsta problemen med Big data är dess potential på intrång på privatlivet och den personliga integriteten. (Richards & King, 2014). Varje gång en användare publicerar något på ett socialt nätverk, rör sig med sina mobiltelefoner i fickan eller genomför en sökning på Google så produceras metadata som flera organisationer tillskansar (Richards & King, 2014). En av de största utmaningarna företag står inför i samband med Big data är det etiska dilemmat mellan företags önskan att samla in och använda data för att förbättra deras kundservice och affärsverksamhet samtidigt som de har ett åtagande att respektera intressenternas integritet (IBE, 2016). När kunder använder vissa tjänster från företag så måste de lägga en stor tillit i hur organisationen använder uppgifterna och data, och ofta har de en liten insikt om hur informationen används, analyseras och samlas in (IBE, 2016).

Datinhämtning för analys kan skapa stora problem för den personliga integriteten (Integritetskommittén, 2016). Att skydda personlig identifierbar information blir allt svårare med större mängd data som ska processas snabbt (Terzi, Terzi, & Sagiroglu, 2015). En undersökning gjord av KPMG år 2016 angående folks inställning till att företag använder deras data, svarade 56 % av de tillfrågade att de var oroliga eller mycket oroliga för hur företag hanterar deras personliga data. Även ifall de flesta användare är uppmärksamma och medvetna om att deras information kan användas i andra syften än de själva tänkt eller säljas vidare till tredje part kände 84 % av de medverkade i studien att de inte har någon kontroll på hur deras personliga data används (KPMG, 2016). En annan rapport utförd av Boston Consulting Group (2013, s4) visar på samma mönster:

*“För 75 % av konsumenter i de flesta länder, oavsett ålder, fortsätter deras privata data vara det största problemet”.*

Även om de flesta användare är uppmärksamma och medvetna om att deras information kan användas i andra syften än de själva tänkt eller säljas vidare till tredje part känner 84 % av dem att de inte har någon kontroll hur deras data används (KPMG, 2016).

Andra utmaningar med Big data är att det rör sig om stora och oorganiserade volymer data (Zoldan, 2013). Att organisera stora mängder data är oftast komplext och svårt. När Big data utnyttjas i beslutsfattning blir det därför inte alltid korrekt för att datan är ofullständig och saknar kontext (Zoldan, 2013). Trots detta tenderar organisationer att ha alldeles för stor tillit till data, för de besitter data i stora mängder och anser därför att den är användbar, vilket blir speciellt problematiskt om Big data algoritmerna tar fel slutsatser om exempelvis kunder och kundbeteende baserat på felaktig och ofullständig data (Zoldan, 2013). Utmaningarna kring oorganiserad data påvisades även i en undersökning av (IDG, 2016) angående organisationers användning av Big data, där menade 45 % av organisationerna att kunna hantera och få ut ett värde av all ostrukturerad data var en av deras största utmaningar de jobbade med angående Big data.

Vid dataanalys av stora datamängder blir det även lätt att finna samband och statistiska resultat, men det krävs en djup och omfattande förståelse för underliggande orsaker för att kunna få ut de bästa analyserna (Harford, 2014). I många fall vid Big data-analys ser man bara de statistiska korrelationerna utan att reflektera dess orsakssamband, vilket bidrar till att många analyser inte blir givande (Harford, 2014).



## 2.2 Skyddet av personuppgifter

*I detta avsnitt redovisar vi det skydd som finns för personuppgifter och behandlingar av dem i samband med införandet av den nya GDPR-lagen.*

### 2.2.1 Personlig integritet

Den ”digitala revolutionen” som har skett de senaste decennierna har fått en större effekt på den enskildas personliga integritet, detta på grund av att både offentliga förvaltningar och näringslivet har ändrat sättet att se på informationsbehandling, samt att analystekniker som Big data och data mining har blivit en realitet (Integritetskommittén, 2016).

Det finns ingen vedertagen definition av begreppet personlig integritet, utan betydelsen varierar utifrån kontext (Integritetskommittén, 2016). Integritetskommittén (2016), en svensk parlamentarisk kommitté, har förenklat begreppet personlig integritet till:

*”Allt som normalt framstår som angeläget att värna om för att den enskilde ska vara tillförsäkrad, rimlig, fredad, privat sfär”*

Skälet till att det inte finns en fullständig definition är bland annat för att rätten till en privat sfär inte är absolut, utan är en subjektiv uppfattning som avgörs av varje enskild person (Integritetskommittén, 2016).

Integritetskommittén har även lyft ett antal generella problem och utmaningar som tillkommer för den personliga integriteten. Bland annat har enskilda i stor utsträckning dåliga kunskaper och är omedvetna om hur och varför deras personuppgifter hanteras i olika sammanhang (Integritetskommittén, 2016). Det har även blivit svårare för den enskilde att påverka hur ens uppgifter lagras och sprids. Dels för att det har blivit en ökad vidareanvändning om personuppgifter, exempelvis genom Big data, men även för att den enskilde tenderar att använda utrustning som applikationer och sensorer som genererar och sprider personuppgifter (Integritetskommittén, 2016).

Enligt datainspektionen så är det inte möjligt att slå fast vad som är kränkning av den personliga integriteten, utan det måste bedömas för varje enskilt fall (Datainspektionen, 2008). Vad som är en kränkning ska inte bedömas endast utifrån vilka uppgifter som behandlas, utan man måste även ha i beaktning i vilket sammanhang uppgifterna förekommer, vilken spridning uppgifterna har fått eller riskerar att få, samt vad behandlingen av uppgifterna kan leda till (Datainspektionen, 2008). Datainspektionen menar också på att vad som är en kränkning för en viss person inte behöver upplevas som en kränkning för en annan person

*“Vad som kan vara en kränkning för en viss person eller i ett visst sammanhang inte behöver vara det för en annan person eller i ett annat sammanhang”*  
(Datainspektionen, 2008)

### 2.2.2 GDPR – Dataskyddsförordningen

General Data Protection Regulation (GDPR), eller dataskyddsförordningen, är den nya lag som skall slå fast hur och vem som får använda personuppgifter men även hur den insamlade informationen får behandlas för att knytas till en viss person (Danielsson, Nilsson, & Lindström, 2018). Den kommer ersätta den tidigare lagen om personuppgifter (PuL) som varit den lag haft i uppgift att skydda svenskarnas personliga integritet sedan 1998 (Datainspektionen, 2007a). Det huvudsakliga syftet med den nya lagen som träder i kraft 2018-05-25 är att skydda enskilda personers rättigheter och friheter, med ett särskilt fokus på personuppgifter (Datainspektionen, 2017a).

För att ha rätt till insamling och lagring av personlig data måste det finnas en rättslig grund för den personuppgiftsansvarige (Datainspektionen, 2017b). Vad som menas med rättslig grund är att personuppgiftsansvarig måste förhålla sig till ett antal rättsliga grunder, om det inte finns stöd i någon av de punkterna i lagen är inte behandlingen av data laglig. Lagen skiljer sig beroende på om det är en privat eller offentlig verksamhet (Datainspektionen, 2017b).

De grunder som riktar sig till privata företag, föreningar och andra organisationer är (Datainspektionen, 2017b):

1. **Samtycke:** Den registrerade har sagt ja till personuppgiftsbehandlingen. **Obs!** I många fall är det inte lämpligt eller kanske inte ens möjligt att stödja sig på den registrerades samtycke.
2. **Avtal:** Den registrerade har ett avtal eller ska ingå ett avtal med den personuppgiftsansvarige.
3. **Rättslig förpliktelse:** Det finns lagar eller regler som gör att den personuppgiftsansvarige måste behandla vissa personuppgifter i sin verksamhet.
4. **Intresseavvägning:** Den personuppgiftsansvarige får behandla personuppgifter utan den registrerades samtycke om den personuppgiftsansvariges intressen väger tyngre än den registrerades och om behandlingen är nödvändig för det aktuella ändamålet.

För offentlig verksamhet gäller (Datainspektionen, 2017b):

1. **Rättslig förpliktelse:** Det finns lagar eller regler som gör att den personuppgiftsansvarige måste behandla vissa personuppgifter i sin verksamhet.
2. **Myndighetsutövning och uppgift av allmänt intresse:** Den personuppgiftsansvarige måste behandla personuppgifter för att utföra sina myndighetsuppgifter eller för att utföra en uppgift av allmänt intresse.
3. **Avtal:** Den registrerade har ett avtal eller ska ingå ett avtal med den personuppgiftsansvarige.

Detta innebär att varje ändamål ett företag har för en behandling av personuppgifter måste vara kopplad till en rättslig grund. Det går inte att ändra rättslig grund under en pågående personuppgiftsbehandling och det är därför väldigt viktigt att rätt grund valts från början (Datainspektionen, 2017b). Ett exempel på detta kan vara att en behandling av personuppgifter påbörjas där samtycke valts som rättslig grund. När sedan personuppgiftsansvarig inte får in samtycke från samtliga personer och vill ändra rättslig grund går inte detta. Behandlingen måste därför avslutas direkt (Datainspektionen, 2017d).

Om ett företag eller organisation (detta gäller även offentliga verksamheter) bryter mot GDPR-lagen kan företaget komma att behöva betala (maximalt) böter på 20 miljoner Euro eller 4 % av årsomsättningen, det värde som är högst för företaget är det som blir boten enligt artikel 83, § 5 (Datainspektionen, 2017d).

Inom GDPR krävs det att det finns en personuppgiftsansvarig som ansvarar för bestämmelser över hur, när och varför personuppgifterna skall behandlas (Datainspektionen, 2017h). Detta kommer från den tidigare personuppgiftslagen (PuL) men inom GDPR kommer den personuppgiftsansvarige ha fler åtaganden (Datainspektionen, 2017h). En personuppgiftsansvarig är den organisation, förening, företag eller liknande som hanterar personuppgifter, men behöver således inte vara en fysisk person. Dock kan detta vara lämpligt eller krävbart på mindre organisationer eller enskilda firmor, då den som är personuppgiftsansvarig är en juridisk person (Datainspektionen, 2017h). Den personuppgiftsansvarige har ett ansvar över att GDPR följs, granska hur organisationens behandling av personuppgifter kan medföra olika integritetsrisker och genomföra åtgärder för detta. Olika sätt för att genomföra detta kan vara att personuppgiftsansvarig definierar de organisatoriska behandlingskraven i olika strategier, policys, certifieringar eller liknande som sedan alla inom organisationen måste följa (Datainspektionen, 2017h).

Personuppgiftsbiträde likt personuppgiftsansvarig har ursprung från Personuppgiftslagen. Inom GDPR kommer uppgiftsbiträdet vara den fysiska person, offentliga myndighet, juridiska person eller annat organ som behandlar personuppgifter för den personuppgiftsansvariges räkning, och ett uppgiftsbiträde finns utanför den egna organisationen (Datainspektionen, 2017i). För att sätta detta i perspektiv kan en personuppgiftsansvarig anlita ett personuppgiftsbiträde för att genomföra olika typer av behandlingar av personuppgifter, t.ex. lagring av personuppgifter. Uppgiftsbiträdet måste då följa de regler om hur behandlingen av personuppgifter som den personuppgiftsansvarige fastställt (Datainspektionen, 2017i).

Lagen gäller för alla, både privata och offentliga verksamheter, organisationer, företag, föreningar med mera förutom brottsbekämpande verksamheter (Datainspektionen, 2017a). Tillämpningsområdet för GDPR är hela EU samt EES-länderna, det innebär att alla som genomför någon sorts personuppgiftsbehandling med en EU-medborgare måste följa GDPR. Detta betyder att GDPR gäller för all typ av verksamhet oavsett om det är en förening, organisation, företag eller liknande någonstans världen som utför personuppgiftsbehandling med en EU-medborgare (Datainspektionen, 2017f). Missbruksregeln som tidigare inneburit att enklare personuppgifter i ostrukturerat material undantogs från Personuppgiftslagen (Datainspektionen, 2017c). Som personuppgift räknas nu således även personuppgifter i enklare ostrukturerad data, det innebär att bilder, ljudfiler med en persons röst, filmer med mera räknas in som personuppgifter och vid behandling av detta måste lagen följas (Datainspektionen, 2017c).

### 2.2.3 Personuppgifter

Artikel 4 i GDPR-lagen har förklaringen av personuppgifter som:

*”Varje upplysning som avser en identifierad eller identifierbar fysisk person (nedan kallad en registrerad), varvid en identifierbar fysisk person är en person som direkt eller indirekt kan identifieras särskilt med hänvisning till en identifierare som ett namn, ett identifikationsnummer, en lokaliseringssuppgift eller online identifikatorer eller en eller flera faktorer som är specifika för den fysiska personens fysiska, fysiologiska, genetiska, psykiska, ekonomiska, kulturella eller sociala identitet” (Datainspektionen, 2017d).*

Begreppet personuppgift begränsar sig således inte till namn eller personnummer, utan är uppgifter som kan kopplas direkt till en person. Enligt GDPR räknas bland annat bilder och ljudupptagningar om en fysisk in som personuppgifter, fastän namn inte nämns (Datainspektionen, 2017f). Även krypterade och kodade uppgifter är också personuppgifter, som exempelvis IP-adresser, förutsatt att den går att koppla till en specifik individ (Datainspektionen, 2017f).

### 2.2.4 Anonymisering

Anonymisering handlar om att avlägsna identifierbara attribut och information hos en person så att den inte blir identifierbar (Commissioner, 2016). Anonymiserad data kommer inte att påverkas av GDPR vilket åberopas i skäl 26 som förklarar att principerna om dataskydd inte skall gälla anonyma uppgifter som lämnas i ett sådant sätt att den registrerade inte ska kunna identifieras (Datainspektionen, 2017e).

Behovet av att anonymisera data kan enligt Raghunathan (2013) hänföras till följande faktorer:

- Behov av att skydda känslig data.
- Att reducera missbruk av personuppgifter och resulterande integritetsfrågor
- För att undkomma kostnader på grund av missbruk av personlig data och personuppgifter.
- Reducera risker som uppkommer genom operativa faktorer som outsourcing eller partnersamarbete
- Juridiska och överensstämelsekrav.

Det finns flera användningsområden för analys av anonymiserad data. Exempelvis är sjukvården är tvungna att skydda identiteten på sina patienter, men kan samtidigt behöva analysera och publicera patientresultat (ICO, 2012). Anonymisering hjälper organisationer att rätta sig efter dataskyddslagar samtidigt som de kan publicera och använda relevant information som inte kan kopplas till enskilda personer (ICO, 2012).

De senaste åren har det även visat sig att Big data och data mining har gjort det svårare att anonymisera information. Genom att kombinera stora datamängder och finna mönster så finns det möjlighet att göra anonymiserad information identifierbar och knutna till specifika individer (Tene, 2012). Detta påvisades bland annat år 2006 när Netflix släppte användares betygsättning på filmer, där användarna var anonymiserade. Genom att matcha den anonymiserade datan från Netflix med publik data från International Movie Database (IMDb) kunde två forskare på University of Texas av-anonymisera flera av de anonymiserade Netflix användarna (Narayanan & Shmatikov, 2008).

### 2.2.5 Pseudonymisering

Mellan anonym och personlig data finns en subkategori som kallas pseudonym data (Supriyadi, 2017). Pseudonym kommer ifrån grekiskans "pseudonumon" och betyder falskt namn (pseudo: falsk onuma: namn). Således är pseudonym ett fingerat namn, och inom IT innebär pseudonymisering att man maskerar den registrerade och dennes personuppgifter (Commissioner, 2016). Dataskyddsförordningens definition av pseudonymisering lyder:

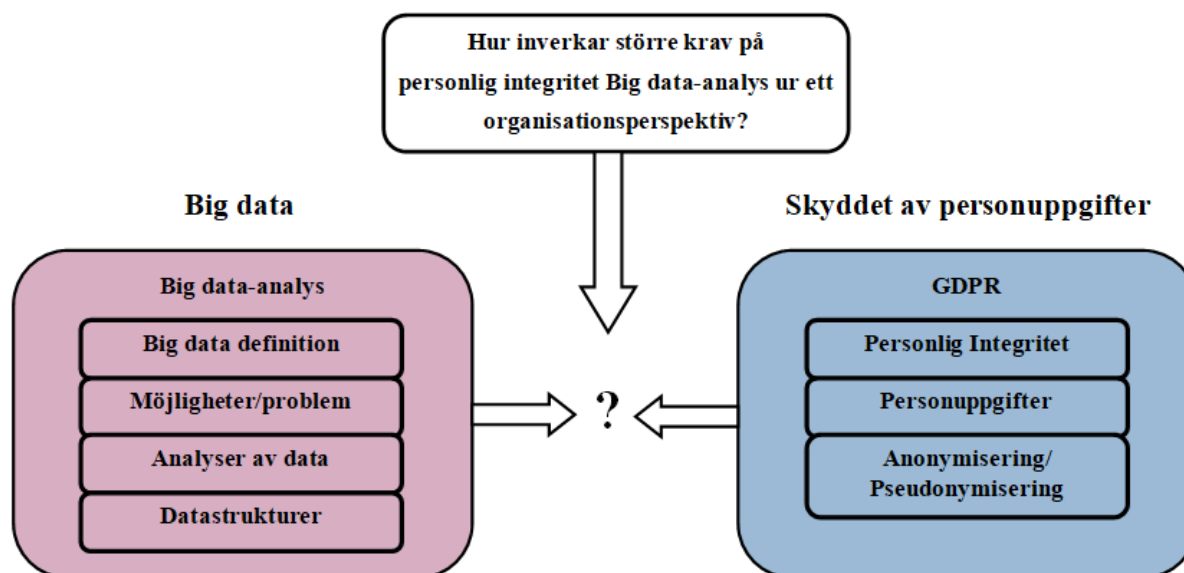
*"Behandling av personuppgifter på ett sätt som innebär att personuppgifterna inte längre kan tillskrivas en specifik registrerad utan att kompletterande uppgifter används, under förutsättning att dessa kompletterande uppgifter förvaras separat och är föremål för tekniska och organisatoriska åtgärder som säkerställer att personuppgifterna inte tillskrivs en identifierad eller identifierbar fysisk person"* (Datainspektionen, 2017d).

Pseudonymisering används bland annat inom medicinsk och samhällsvetenskaplig forskning där man byter ut namn, personnummer eller andra attribut som är bunden till en viss person (IDG, 2018). Dock är det fortfarande möjligt att identifiera en pseudonym person om man har tillgång till information som skall hållas dold och separerad från det undersökningsmaterial som behandlas (IDG, 2018). Eller genom att analysera bakomliggande eller relaterad data till den pseudonymiserade personen (Commissioner, 2016).

## 2.3 Sammanfattande modell av vår litteraturgenomgång

I vår litteraturstudie har vi observerat fördelar med Big data, de ekonomiska fördelarna är många och samtidigt kan det hjälpa organisationer med att få nya insikter (Sharda et al., 2014). Samtidigt som möjligheterna är många har vi även funnit en hel del sätt som Big data-analyser kan kränka den personliga integriteten (Richards & King, 2014). I och med den nya GDPR-lagen har vi också fastställt att Big data-analyser kommer att påverkas av lagändringen och då även ostrukturerad data kommer omfattas av lagen (Datainspektionen, 2017c). I vår sammanfattande modell, Tabell 1, belyser vi de mest betydelsefulla ämnena för varje område. Vi hänvisar till vem eller vilka författare som behandlat dem.

I Figur 3 sammanfattas vår litteraturgenomgång utifrån de viktigaste områdena för vår forskningsfråga. I den vänstra rutan omfattas samtliga underrubriker i huvudområdet *Big data-analys*. I den högra rutan omfattas samtliga underrubriker i *GDPR-lagen*. Dessa huvudområden har sedan stått till grund för vår Empiriska undersökning som skapar kapitel



Figur 3: Sammanfattande modell av litteraturgenomgången

Tabell 1, Big data

Område/tema	Sammanfattning/innehåll	Referens
Vad är Big data?	<p>Volym, variation, hastighet, värde, riktighet, datamängd över en terabyte.</p> <p>Strukturerad data, ostrukturerad data, semi-strukturerad data.</p>	<ul style="list-style-type: none"> <li>• (Russom, 2011)</li> <li>• (Bello-Orgaz et al., 2016)</li> <li>• (Charan, 2015)</li> <li>• (Gandomi &amp; Haider, 2015)</li> <li>• (Rosyad, 2016)</li> <li>• (Gartner, 2012)</li> <li>• (McAfee &amp; Brynjolfsson, 2012)</li> </ul>
Möjligheter med Big data?	<p>Finns stora möjligheter och användningsområden med Big data. Används bland annat inom olika typer av forskning, kundsegmentering och beslutsstöd.</p> <p>Genom att analysera Big data kan man finna relationer som tidigare varit okända</p>	<ul style="list-style-type: none"> <li>• (Michael &amp; Miller, 2013)</li> <li>• (Marr, 2017)</li> <li>• (IBE, 2016)</li> <li>• (Sharda et al., 2014)</li> <li>• (Duhigg, 2012)</li> <li>• (ICO, 2017)</li> <li>• (Vickers, 1992)</li> <li>• (Berner, Graupner, &amp; Maedche, 2014)</li> </ul>
Analys av Big data	<p>Finns flera olika typer av analyser. bland annat text-, video- och ljud- och sociala medier analyser.</p> <p>Visualiserar data i olika typer av modeller.</p>	<ul style="list-style-type: none"> <li>• (ICO, 2017)</li> <li>• (Bello-Orgaz et al., 2016)</li> <li>• (Sharda et al., 2014)</li> <li>• (Gandomi &amp; Haider, 2015)</li> </ul>
Datastrukturer	<p>Strukturerad data omfattar kategorisk och numerisk data.</p> <p>Ostrukturerad data inkluderar text, bild, röst, videor, data från sociala nätverk med mera.</p>	<ul style="list-style-type: none"> <li>• (Sharda et al., 2014)</li> <li>• (Gandomi &amp; Haider, 2015)</li> </ul>
Problem med Big data	<p>Finns en stor potential att kränka den personliga integriteten. Blir svårare att skydda personlig information i takt med att större mängd data processas.</p> <p>Personer har en låg medvetenhet och hög oro om hur dennes personliga data behandlas.</p>	<ul style="list-style-type: none"> <li>• (Richards &amp; King, 2014)</li> <li>• (KPMG, 2016)</li> <li>• (IBE, 2016)</li> <li>• (Terzi, et al, 2015)</li> <li>• (Zoldan, 2013)</li> <li>• (IDG, 2016)</li> <li>• (BCG 2013)</li> <li>• (Harford, 2014)</li> </ul>

Tabell 2, Skyddet av personuppgifter

Personlig integritet	Allt som normalt framstår som angeläget att värna om för att den enskilde ska vara tillförsäkrad, rimlig, fredad, privat sfär.	<ul style="list-style-type: none"> <li>• (Integritetskommittén, 2016)</li> <li>• (Datainspektionen, 2008)</li> </ul>
Dataskyddsförordningen - GDPR	<p>Lag som träder i kraft 2018-05-25, för att skydda enskilda personers rättigheter och friheter, med särskilt fokus på personuppgifter.</p> <p>Gäller för alla organisationer, företag, föreningar med mera förutom brottsbekämpande verksamheter.</p> <p>Missbruksregeln försvinner vilket innebär hårdare reglering vid behandling av ostrukturerat material.</p>	<ul style="list-style-type: none"> <li>• (Datainspektionen, 2017a)</li> <li>• (Datainspektionen, 2017b)</li> <li>• (Datainspektionen, 2017c)</li> <li>• (Datainspektionen, 2017d)</li> <li>• (Datainspektionen, 2017f)</li> <li>• (Datainspektionen, 2017h)</li> <li>• (Datainspektionen, 2017i)</li> <li>• (Danielsson, Nilsson, &amp; Lindström, 2018)</li> </ul>
Personuppgifter	<p>All slags information som direkt eller indirekt kan knytas till en fysisk person.</p> <p>Användning och hantering kan bli en integritetsfråga.</p> <p>Omfattas av GDPR.</p>	<ul style="list-style-type: none"> <li>• (Datainspektionen, 2007b)</li> <li>• (Datainspektionen, 2017d)</li> <li>• (Datainspektionen, 2017f)</li> </ul>
Anonymisering	<p>Identifierbara attribut och information för en person är borttagna. Ingen information ska kunna peka på en specifik person.</p> <p>Används för att skydda känslig data, reducera missbruk av personuppgifter och integritetsfrågor.</p> <p>Omfattas inte av GDPR.</p>	<ul style="list-style-type: none"> <li>• (Commissioner, 2016)</li> <li>• (Datainspektionen, 2017e)</li> <li>• (Raghunathan, 2013)</li> <li>• (ICO, 2012)</li> <li>• (Tene, 2012)</li> <li>• (Narayanan &amp; Shmatikov, 2008)</li> </ul>
Pseudonymisering	Pseudonymisering inom IT innebär att man maskerar personuppgifter, men som går att identifiera med kompletterande information	<ul style="list-style-type: none"> <li>• (Supriyadi, 2017)</li> <li>• (Commissioner, 2016)</li> <li>• (Datainspektionen, 2017d)</li> <li>• (IDG, 2018)</li> </ul>



### 3 Metod

*I den inledande delen av uppsatsen genomförs en litteraturstudie där vi undersökt tidigare forskning på området. Detta för att kartlägga vilka förutsättningar och kunskap som finns inom området och hur dessa kan komma att förändras i och med GDPR. I detta kapitel beskrivs och motiveras våra metodval i den undersökning vi genomfört som står som grund för vår insamlade empiriska data. I början av kapitlet beskriver vi vilka metodval som gjorts, hur vi genomfört våra intervjuer, vårt tillvägagångssätt att bearbeta den data vi fått in och analysen av vår data. Vi beskriver även våra informanter och varför vi valde just dem, vi avslutar kapitlet med kritik till oss själva och den etik vi följt för att uppnå god undersöknings-etik.*

---

Vid genomförandet av denna uppsats gjorde vi en avvägning att bestämma ifall vi skulle göra en kvantitativ eller kvalitativ ansats, vilket föll på det senare alternativet. Jacobsen (2002) menar på att en kvalitativ metod lämpar sig bäst när man vill skapa en större klarhet i vad som ligger i ett begrepp eller fenomen. Kvalitativ uppläggnings burkar vara avsedd för att få ut hur människor tolkar och förstår en given situation. Då vi delvis undersöker integritetsproblem med Big data, ett område där det inte finns några rätt eller fel, utan är upp till tolkningar för var och en fann vi denna metod mest lämplig för att driva vårt problemområde (Jacobsen, 2002).

En negativ aspekt med kvalitativ metod är att det är resurskrävande i form av tid och att det kan bli svårt att finna intervjupersoner som är villiga att ämna 1-2h. Vi har ändå valt att genomföra en kvalitativ undersökning med ett flertal intervjuer för att undersöka hur litteraturen skiljer sig från den verkliga världen (Jacobsen, 2002). Genom intervjuer med personer som arbetar inom området ämnar vi att undersöka om GDPR kan skydda användaren från analyser av Big data, och till vilket organisatoriskt pris. Vi valde en kvalitativ undersökning framför kvantitativ eftersom kvantitativa undersökningar kan standardiseras av oss som forskare och kan ses som deduktiva enligt Jacobsen (2002).

Litteraturstudien genomfördes innan vår empiriska data samlades in. Detta medför en risk av att vi själva har en bild av vad som är rätt eller fel vilket kan avspeglas i vår forskning. Men genom användning av många olika källor och via semistrukturerade intervjuer anser vi att detta kan undvikas i största möjliga utsträckning eftersom vi ständigt är öppna för nya perspektiv. Detta stöds av Jacobsen (2002) som anser att kvalitativa metoder är lämpligast när utveckling av nya teorier och hypoteser skall tas fram. Nackdelen med en kvalitativ ansats är den öppenhet som kan uppstå. Enligt Jacobsen (2002) kan den som forskar redan innan i ringa utsträckning bestämt i förväg vad hen ska leta efter. Och att de undersökta ofta är de som definierar vad som är korrekt eller inte (Jacobsen, 2002). Även detta är vi väl medvetna om men vi anser att det är viktigare att ta reda på detaljer om ämnet som gemene man förmodligen inte besitter. Därför har vi valt en kvalitativ undersökning med personer som arbetar inom ämnet och är synnerligen pålästa då detta ger oss möjlighet till djupare forskning men samtidigt ha en öppenhet för andra perspektiv (Jacobsen, 2002). Ämnet vi undersöker är även relativt nytt och mängden forskning på det är begränsat. För att kunna få relevanta svar och tankar kring vår frågeställning valde vi att försöka få intervjuer med personer som verkar inom området för Big data-analys eller är djupt insatta i GDPR. Enligt Jacobsen (2002) är den

kvalitativa ansatsen lämplig för oss då vi försöker skapa en större klarhet i ett oklart ämne samt få fram en nyanserad beskrivning av det.

Gällande den nya dataskyddsförordningen (GDPR) har vi använt oss av Datainspektionens tolkningar i absolut största utsträckning, som sekundärkälla. Vi anser oss själva inte besitta tillräcklig kunskap gällande lagtexter och tolkningar av dessa och därför använt oss av Datainspektionens tolkningar.

### 3.1 Intervjuer

Jacobsen (2002) menar på att den öppna intervjun lämpar sig bäst när relativt få enheter undersöks. En öppen intervju kännetecknas av att undersökare och uppgiftslämnare samtalar som i en vanlig dialog. Informationen kommer i form av ord, meningar och berättelser. De två vanligaste intervjuformerna är ansikte mot ansikte eller genom telefon (Jacobsen, 2002). Vi valde att försöka genomföra våra intervjuer ansikte mot ansikte då personer har en tendens att lättare tala om känsliga ämnen när de fysiskt sitter bredvid varandra. Då vissa menar på att personlig integritet kan vara ett känsligt ämne ville vi få en personlig kontakt för att det lättare skapar en intim och öppenhetig stämning (Jacobsen, 2002). Informanterna blir även mer uppriktiga och det skapas en mer personlig kontakt när man träffas fysiskt och samtalet blir mer givande än via telefon (Jacobsen, 2002).

Vårt absoluta mål med vår forskning var som sagt att genomföra samtliga intervjuer ansikte mot ansikte. På grund av våra begränsade resurser och att två av våra informanter inte befann sig i närområdet var vi tvungna att genomföra två intervjuer via telefon. Detta är enligt Jacobsen (2002) ett mer riskabelt tillvägagångssätt då informanten är svårare för oss att läsa av, få en personlig kontakt med och det är även lättare för informanten att undvika svåra frågor eller till och med ljuga. Vi förlorar även möjligheten att observera den vi intervjuar och hur hen uppträder vilket medför en viss förlust av känsla och dynamik i intervjun (Jacobsen, 2002). Vi anser dock att eftersom frågorna till den tillfrågade hålls på professionell nivå och inte är direkt utlämnande för informanten ser vi att det är en förhållandevis liten risk att hen vill ljuga för oss å företagets eller sina egna vägnar. Vi kommer därför använda oss av data från telefonintervjuerna på samma sätt som de fysiska intervjuerna.

Våra intervjuer har varken varit helt öppna eller helt stängda. Enligt Jacobsen (2002) skall en kvalitativ ansats ha så öppna intervjuer som möjligt vilket vi strävat efter. Detta för att möjliggöra för informanten att denna kan bistå med information som vi inte tänkt på tidigare. Men vi har även haft en viss struktur för att undvika ett för stort arbete vid bearbetningen av den insamlade informationen och för att hålla ett fokus under intervjun (Jacobsen, 2002).

Vi har innan intervjuerna skickat ut intervjufrågorna till informanterna för att ha en transparent och öppen ton till dem. Detta för att hålla det på en så professionell nivå som möjligt, ge informanterna en viss tid till förberedelse och se om de är rätt person i organisationen för oss att intervjuas. Då vårt ämne till viss del är avancerat och framförallt hur informanten och/eller dess företag kommer ställas inför förändringarna i och med Dataskyddsförordningen anser vi att de har rätt till en viss förberedelse. Detta för att vi skall få ut så mycket relevant data som möjligt från informanterna på den snäva tidsram vi har. Enligt Jacobsen (2002) så kan det dock vara ett problematiskt tillvägagångssätt då intressenten får kännedom om avsikten och kan uppge osanna eller inexakta upplysningar. Vi anser däremot att de flesta av våra frågor inte är så pass känsliga att detta ska ske. Vi kan

såklart aldrig vara helt säkra på att alla uppgifter vi fått är sanningsenliga och korrekta, något som är svårt oavsett tillvägagångssätt.

### 3.1.1 *Utformning av Intervjuguide*

Vi har format intervjuens struktur utifrån vår litteraturgenomgång där två teman kan räknas in: *Big data* och *skyddet av personuppgifter*. I dessa teman finns underkategorier som vad *Big data* är, användning av *Big data*, hur *Big data*-analyser kan komma att ändras, skyddet av personuppgifter och hur kränkningar på den personliga integriteten kan uppstå vid analyser samt skydd för den personliga integriteten. Intervjuens struktur finns i Bilaga 1 (Intervjuhandledning) men beroende på informant tillkom följdfrågor och svar på andra frågor vi inte tidigare tänkt på. Detta är ett resultat av att använda semistrukturerade intervjuer men något som vi ser som en styrka då vi fått mer intressant data än om vi enbart förhållit oss till en stängd intervjumetod.

### 3.1.2 *Bearbetning av data*

Samtliga intervjuer spelades in med två olika mobiltelefoner. Detta för att undvika diverse tekniska problem som kan uppstå (Jacobsen, 2002). Så snart som vi möjligtvis kunnat efter intervjuerna har vi transkriberat inspelningarna i sin helhet. Vi har valt att spela in intervjuerna för att ha möjlighet till så öppna samtal som möjligt då endast anteckning vid samtal kan hämma samtalet (Jacobsen, 2002). Däremot är transkribering en enormt tidsödande process då en intervju på en timme kan generera 10–15 sidor text (Bryman & Nilsson, 2011; Jacobsen, 2002). Men eftersom vår undersökning endast omfattar fyra intervjuer är detta enligt oss det bästa tillvägagångssättet. Vi har även antecknat under intervjuerna för att underlätta för oss själva när vi transkriberar eftersom det är svårt att hitta tillbaka till exakt rätt plats i inspelningen helt utan anteckningar (Jacobsen, 2002). Vi har inte antecknat informanternas reaktioner i samband med intervjuerna utan istället reagerat på dessa direkt i intervjun med ytterligare frågor från oss eller att vi bett dem förklara något igen. Vi har transkriberat hela intervjuerna men sorterat ut det som är relevant för vår undersökning då vi vill underlätta för oss själva vid bearbetning och för er läsare att enkelt hitta relevant data.

### 3.1.3 *Analys av data*

Samtliga intervjuer har inbringat en stor mängd data. För att göra detta tydligt har vi summerat den viktigaste informationen i överblickbara tabeller utifrån vår intervjuhandlednings teman. Detta tillvägagångssätt grundar sig från Jacobsen (2002) som poängterar vikten av att försöka hitta samband i den insamlade data vi har. Utifrån tabellerna kunde vi sedan urskilja mönster, likheter och olikheter, och resultatet av dessa ligger till grund för vårt resultatkapitel (4). Dessa tabeller valde vi sedan att inte ha med i uppsatsen då vi anser att en löpande text med god röd tråd är mer givande och lättare att ta sig an för läsaren.

### 3.1.4 Informanter

Tabell 3: Informanter

<p><b>Informant 1:</b> Intervju 1</p> <p><b>Intervju genomförd:</b> 2018-04-12</p> <p><b>Längd:</b> 1:04:47</p>	<p><b>Roll:</b> Solution architect: Digital solutions</p> <p><b>Organisation - O1</b> Internationellt konsultbolag med 2000 anställda, över 2 miljarder SEK i omsättning.</p> <p><b>Relevant erfarenhet</b> Arbetat med Business Intelligence och analys av Big data i åtta år.</p>
<p><b>Informant 2:</b> Intervju 2</p> <p><b>Intervju genomförd:</b> 2018-04-17</p> <p><b>Längd:</b> 1:02:01</p>	<p><b>Roll:</b> Big data- och data science konsult</p> <p><b>Organisation - O2</b> Internationellt IT-bolag med över 350 000 anställda, över 560 miljarder SEK i omsättning.</p> <p><b>Relevant erfarenhet</b> Arbetat med Big data-analys i tre år, skrev även sin masteruppsats om Big data-analyser.</p>
<p><b>Informant 3:</b> Intervju 3</p> <p><b>Intervju genomförd:</b> 2018-04-20</p> <p><b>Längd:</b> 59:53</p>	<p><b>Roll:</b> Senior Manager inom Digital Applied Intelligence</p> <p><b>Organisation - O3</b> Internationellt IT/konsultbolag med över 380 000 anställda, över 270 miljarder SEK i omsättning.</p> <p><b>Relevant erfarenhet</b> Arbetat med dataanalys i många år redan innan det blev ett vedertaget begrepp i Sverige. Arbetat med Big data-analys sedan 8 år tillbaka.</p>
<p><b>Informant 4:</b> Intervju 4</p> <p><b>Intervju genomförd:</b> 2018-04-24</p> <p><b>Längd:</b> 53:36</p>	<p><b>Roll:</b> Säkerhetsspecialist</p> <p><b>Organisation - O4 <i>Datainspektionen</i></b> Datainspektionen, Statlig förvaltningsmyndighet sedan 1970, har i huvudsaklig uppgift att skydda människors personliga integritet vid behandling av personuppgifter.</p> <p><b>Relevant erfarenhet:</b> Expert inom GDPR och lagens förhållning till Big data-analyser, personuppgiftsbehandling och mycket mer.</p>

## 3.2 Kvalitet av data

Vi har haft ett genomgående upplägg för samtliga intervjuer för att inte påverka informanterna från vår sida. Endast en person av oss har den som ställer frågor och den andra har varit sekreterare. Detta gäller även under telefonintervjuerna då ett sammanhängande upplägg är viktigt enligt Jacobsen (2002). Vårt mål har varit att skapa ett så snarligt scenario som möjligt och förutsättningarna för informanterna skall vara så lika som det bara går. Då våra intervjuer har haft ett semistrukturerat upplägg och vi är öppna för eventuella förändringar i intervjun har informanten alltid fått stor plats och på så sätt kunnat påverka intervjun till viss del (Jacobsen, 2002). Vi har dock alltid fallit tillbaka till vår intervjuguide när informanten svävat från ämnet i för stor utsträckning. Samtliga frågor är grundade på vår litteraturstudie och den sammanfattande litteraturmodellen, detta för att vi vill hålla hög validitet på frågorna (Jacobsen, 2002). Vi har haft två inspelningsapparater för varje intervju då vi dels försäkras oss mot ett eventuellt tekniskt bekymmer, och vid tillfällen där det är svårt att tyda vad som sägs vid transkriberingen har den andra ljudinspelningen varit till stor hjälp.

### 3.2.1 Urval

För en första kontakt med organisationer skapade vi en email-mall som skickades ut till över 20 företag där vi kort förklarade vilka vi var, vad vi ville undersöka och varför. Vid ett positivt svar skickades intervjufrågorna för att säkerställa att de var rätt person för oss att intervjua. Om de inte var rätt person i organisationen blev vi hänvisade till den person som hade rätt kunskap, och dessa är våra informanter. Vi har valt de informanter som redovisas i tabell 3 eftersom de alla har kunskap inom vårt forskningsområde samt insikter i hur analyser genomförs på Big data idag. Samtliga av våra intervjupersoner har haft många projekt och kunder som genomfört analyser av Big data. De har också kunskap angående GDPR och hur lagen påverkar en organisation inför en eventuell omställning inför lagändringen. Vi fokuserade på att göra urvalet utifrån ett informationsperspektiv för att få så mycket värdefull data som möjligt ur varje intervju (Jacobsen, 2002). Men eftersom vi aldrig kan vara säkra på detta innan genomförd intervju hade vi även bredd och variation i åtanke som enligt Jacobsen (2002) säkerställer att vi har större chans att få den kompetens som är relevant för vår undersökning. Vi hade som mål att få tag på organisationer som genomför Big data-analyser för sin egen räkning, men då inga av dessa ville ställa upp på intervju fick vi fokusera på organisationer som hjälper sina kunder med detta. Det innebär att de tillfrågade inte aktivt genomför analyser på sina kunder, utan hjälper sina kunder att analysera deras kunder. Vi ser dock inte detta som något enbart negativt då resultaten vi fått kan ha en mer objektiv och ärlig bild än om vi intervjuat organisationer som genomför analyser för sin egen räkning, då de kan tycka frågorna är mer obekväma och därför ter sig tala mer osanning (Jacobsen, 2002). Vi valde också att intervjua Datainspektionen då vi tror de besitter intressant kunskap om vårt ämne och GDPR-lagen. Eftersom de övriga informanter vi intervjuat kommer förhålla sig till lagen anser vi att Datainspektionen är viktig för vår undersökning för att få en bredare och tydligare bild från lagstiftningens perspektiv. De informanter vi genomfört intervjuer med hade vi inte fått tag på om vi inte utgått från bredd och variations-perspektivet då de fallit ur innan vår urvalsprocess (Jacobsen, 2002).

### 3.2.2 Validitet och reliabilitet

Vid undersökningar skall målet vara att försöka minska och förhindra problem som kan påverka validiteten (giltighet) och reliabiliteten (tillförlitlighet) i undersökningen, även fast vår undersökning har varit av den kvalitativa ansatsen måste också vi sträva efter detta (Jacobsen, 2002). Validitet kan sedan vidareutvecklas till två underkategorier, intern validitet och extern validitet. Den interna validiteten innebär att se till vår egen forskning och om vi anser oss fått tag i den kunskap vi ville ha (Jacobsen, 2002). Den externa validiteten betyder om det resultat vi fått fram kan appliceras och överföras till andra sammanhang. Reliabiliteten i undersökningen innebär att vi måste ställa oss frågan om informationen vi tillförskaffat oss går att lita på och är korrekt (Jacobsen, 2002).

Vi är medvetna om att endast fyra informanter kan anses förhållandevis få i sammanhanget men trots förfrågningar till över 20 företag var dessa de enda som hade tid och möjlighet. Vår forskning hade kunnat stärkas på ett bättre sätt vid fler svar och genomförda intervjuer. För vår forskningsfråga och den tidsram vi har anser vi dock att fyra informanter bör räcka för diskussion och slutats då de flesta informanter kom med liknande svar och alla skall ses som lika viktiga (Jacobsen, 2002).

Vår intervjuguide (Bilaga 1) är grundad i den sammanfattande modellen från litteraturstudien och är väl förankrad i vår forskningsfråga. Intervjuerna har genomförts under så lika förhållanden som möjligt för samtliga informanter. Vi har även undvikit att påverka informanterna i absolut största utsträckning. Dock kan våra frågor vara lite väl stängda, något vi upptäckte efter vi genomfört våra intervjuer. Vårt mål var att ha öppna frågor som låter informanterna svara så brett som möjligt och att vi påverkar intervjuobjekten så lite som möjligt (Jacobsen, 2002). Vid närmare granskning inser vi att våra frågor är ganska stängda. Vår fråga från intervjuguiden (bilaga 1) som lyder "*Hur tror du att GDPR kommer att påverka Big data och analys?*" skulle istället kunnat ha ställts som "*tror du att GDPR kommer påverka Big data och analys av Big data?*". Detta medför en viss påverkan från oss på intervjuobjektet där vi leder intervjuobjektet. Vi hoppas dock att den data vi fått fram ändå håller liknande validitet som vid än mer öppna frågor. Detta eftersom vi genom en tidigare fråga säkerställt att intervjuobjektet t.ex. anser att det finns en problematik angående den personliga integriteten vid analys av Big data, för att sedan ställa frågan hur intervjuobjektet anser att detta sker.

Vårt mål att säkerställa den externa validiteten var att våra intervjuobjekt hade en hög kompetens och bred erfarenhet inom branschen och ämnet. Vi tolkade själv inte den nya lagen utan höll oss till Datainspektionens tolkningar och försökte hålla oss så brett som möjligt inom den (Jacobsen, 2002). Vi försökte också ha så breda frågor som möjligt där vi även hade frågor som inte direkt är kopplade till ämnet för att upprätthålla en flexibilitet för eventuella nya vinklar (Jacobsen, 2002). De intervjuobjekt vi haft har många olika typer kunder som genomför analys av Big data idag. Detta eftersom vi vill få en så bred bild som möjligt av hur lagen kan komma att påverka olika typer av branscher och organisationer.

Reliabiliteten säkerställdes genom att vi skickade ut frågorna i förväg och höll en öppen ton mot våra intervjuobjekt, detta för att bygga upp en tillförlitlighet mellan oss och dem (Jacobsen, 2002), samt att de har tid att förbereda sig för att ge oss så korrekta svar med mycket information som möjligt. Eftersom våra intervjuobjekt inte heller genomförde data-analys för sin egen räkning hoppas vi att det gör dem mindre benägna att tala osanning och ge

felaktig information för att rädda sig själva. Att GDPR-lagen inte är implementerad ännu kan säkerligen ge en viss påverkan på vårt resultat av den insamlade datan. Våra intervjuobjekt och vi har så långt det går generaliserat vad de ökande kraven från lagen kommer att innebära. Vissa av intervjuobjekten hade redan implementerat lösningar åt sina kunder som nu är redo för lagändringen medan andra ännu inte gjort det.

### 3.3 Etik

För att uppnå en god undersöknings-etik har vi följt de etiska aspekter mellan forskare och undersökt som Jacobsen (2002) argumenterar för. Enligt Jacobsen (2002) uppstår dessa etiska dilemman först och främst när vi vill dölja avsikten med vår undersökning för den undersökta. Detta har vi inte gjort utan redan från början varit väldigt öppna med våra frågor mot våra informanter. Vi har även skickat ut samtliga intervjufrågor till dem i förväg för att följa principen angående "Full information" som enligt Jacobsen (2002) gör att först när en tillfrågad till fullo har information angående undersökningen kan avgöra om de vill ställa upp eller inte. Vi har också följt principen informerat samtycke som innebär att den som undersöks frivilligt skall delta i undersökningen (Jacobsen, 2002). Samtliga av våra informanter ställde upp frivilligt och hade alla rätt till anonymitet. I vår vetenskap har inga påtryckningar skett från oss eller informanternas överordnade för att ställa upp för vår undersökning. Vi frågade samtliga i början av intervjun om det var okej att vi använde deras namn och organisations namn i uppsatsen. Samtliga informanter svarade ja men vi vill ändå säkerställa deras rätt till privatliv och professionella roll genom att pseudonymisera dem (Jacobsen, 2002). Den information som kan användas för att av-pseudonymisera dem har enbart författarna tillgång till och den insamlade informationen används enbart till denna uppsats.

## 4 Resultat

*Resultatet från vår empiriska studie har inbringat mycket data. I avsnitten nedan redovisar vi hur våra informanter ställer sig till de intervjufrågor vi format utifrån vår litteraturgenomgång. Vi har sammanställt informationen i löptext för lättare översikt av det insamlade data. Samtliga intervjuer finns transkriberade och kommer att hänvisas till i löptexten i detta kapitel. Vid transkriberingen har vi delat upp intervjuerna till stycken för att underlätta hänvisning. När vi refererar så kommer vi referera till vilken intervju, 1 till 4, och sedan till vilket stycke, t.ex. (i1, 114), detta innebär då intervju nummer 1, stycke 114. Detta följer sedan vår ordning av organisation/informant. Så (i1, 114) betyder organisation 1, informant 1 och intervju 1, stycke 114. De teman som avsnitten förhåller kommer från vår intervjuguide i bilaga 1 som grundar sig på vår litteraturgenomgång.*

---

### 4.1 Resultat från intervjuer

*Vi har valt att redovisa vårt resultat organisation för organisation. Detta eftersom den data vi fått från Empiriska studie är relativt stor och en sammanslagning av all data skulle bli svårare för läsaren att följa.*

#### 4.1.1 Organisation 1:

##### 4.1.1.1 Big data

**Definition:**

Termen Big data som används i marknadsföring är inte riktigt Big data för informant 1, utan Big data för informanten är när datavolymer och strukturen inte stämmer in i vanliga transaktionsbaserade datasystem (i1,6). Informanten menar på att det inte Big data bara för att det är stora datavolymer utan Big data skall innehålla mer ostrukturerad data.

**Möjligheter:**

De största möjligheterna med Big data är när man har en frågeställning och får svar på den i sin data, och att utan frågeställning är inte Big data värt mycket. Samtidigt så menar han på att man inte har alla frågor när man samlar in data men att Big data kan ge svar på saker man inte hade tänkt på när man samlar in information (i1, 10).

**Utmaningar:**

De största utmaningarna vid Big data är att se det ekonomiska värdet med all data (i1, 16). Det är inte dyrt att lagra data men investeringsbeslutet är svårt att ta för att lagra och analysera information (i1, 18).



**Big datas betydelse:**

Informanten anser inte att Big data har den betydelse som den skulle kunna ha. Det finns flera olika områden där Big data har en stor potential men där han inte sett att det används korrekt, bland annat menar han på att produktion och kvalitetsmätning av produkter skulle kunna bli mycket bättre med hjälp av Big data-analys (i1, 30).

*”En annan sak som jag var inblandad i var ett företag som har maskiner och där kan man mäta vibrationerna på maskinerna i tillverkningen, när de vibrationerna ökar så börjar maskinerna behöva servas eller bli dåliga. Datan av vibrationerna de lagrades, men de gör ingenting utav den. De har inte riktigt förmåga att förädla den informationen på rätt sätt så de kunde fånga det i tid och göra arbete.” (i1, 30)*

**Användning:**

Informantens företag använder inte Big data för eget syfte, utan endast för kunders räkning, vilket är olika scenarier. Ett exempel informanten tar upp är hur de hjälpt en kund med en lösning att de kan se temperaturer för varje enskild lägenhet i ett lägenhetskomples. På så vis kan de lättare få återkoppling till hur värmesystemen är inställda till vilken temperatur det finns i lägenheten. Tidigare kunde det ta ett par år att jobba sig igenom hela beståndet, men med hjälp av dagens teknik kan de få Realtidsanalyser. (i1, 32) Informanten har inga nuvarande projekt som berör personuppgifter, men har haft ett projekt med ett postorderföretag som besatt ett kundregister på flera hundratusen. Där de skickade olika typer av kampanjer och standardutskick ett par gånger per år. Han menar på att integritetsbiten i uppgifterna var ganska basala då de mestadels jobbade med kön och ålder. (i1, 34)

**Syfte vid datainsamling:**

Informanten tror att det alltid finns ett tydligt syfte för deras kunder när de samlar in data. I och med att kunderna är strikta på hur mycket pengar de vill lägga på att utveckla saker och ting så börjar det nästan alltid med ett syfte, de vill inte besitta information som de ser kostar pengar (i1, 12).

Informanten är inte närvarande vid insamling av data utan hen kommer oftast in i ett skede när all data redan finns på något media, på någon disk av något slag (i1, 20). Men datan kan komma in i flera olika former, det kan vara från sensorer, besöksräknare eller ”riktig” information från ett affärssystem. Informanten tror även att data köps in från tredje part. (i1, 26).

#### 4.1.1.2 Big data och skyddet av personuppgifter

##### **Personlig integritet & Big data-analys:**

Informanten menar på att det kan finnas en etisk problematik med Big data, men att problemen inte är så stora som framställs i media:

*“Därför att människor som interagerar med företag gör det för att de är intresserade av företagets produkter och då tror jag att det finns en allmän acceptans att man blir kartlagd av företaget för att företaget ska hitta saker och ting till en som är bättre, som bättre passar, så man slipper att se dåliga annonser till exempel.” (i1, 36)*

Hen menar att det finns en problematik med handeln av personuppgifter:

*“När man delar med sig av informationen till någon annan för att dem ska kunna göra något annat med det då har man gjort en datainsamling av en individ för ett syfte och sen så har man ur ett egoistiskt perspektiv delat med sig det till någon annan som vill använda det för ett annat syfte.” (i1, 36).*

Informanten menar på att så länge uppgifterna används för det självklara inhämtade ändamålet så upplever individerna det som mindre känsligt. Hen menar också att Big data-analys gör det möjligt på nya sätt att inkräkta på den personliga integriteten, där han drar upp lägenhets exemplet igen, för där kan man se om personer duschar ofta, vilket kan bli en integritetsfråga. Det kan bli väldigt tydligt när man mäter fuktigheten i lägenheten hur ofta de duschar. (i1, 36). Informanten menar på att det kan finnas en slarvig syn hos företagen gällande personlig information, som skulle kunna hamna i fel händer beroende på att den är slarvigt hanterad (i1, 38).

##### **Big datas påverkan av GDPR:**

Informanten tror inte att Big data-analys inte kommer att påverkas märkvärdigt. För att de som gör Big data-analys med individknutna uppgifter, har redan jobbat med de frågorna i PuL, och de som jobbar med dataanalys där det inte är individ spårbar information kommer inte drabbas av GDPR (i1, 52). Han menar på att det är andra processer som kommer att påverkas mer under GDPR som datahantering, att de blir mycket striktare och dyrare (i1, 54).

För den personliga integriteten tror inte informanten att det kommer göra någon skillnad med den nya lagen (i1, 66). Han menar på att som lagen är utformad så gör den ett otroligt litet värde för konsumenten, men den ger en väldigt stor börda för företagen (i1, 64):

*“GDPR kommer att sätta grus i maskineriet för Europa. Det är en katastrof-lagstiftning för Europas skull.” (i1, 64)*

Informanten fortsätter med att han inte tror att företag som Spotify och andra världsledande IT-koncept kommer kunna skapas i Europa i framtiden. Tröskeln kommer vara för hög när det kommer till att hantera kundinformationen så företagen behöver större investering redan från start än idag (i1, 68). Han tror att nya idéer kan komma att strypas med den nya lagen, han tar ett exempel från en hemsida för scouterna i Tyskland som var tvungen att stänga ner då den innehöll namn och epost på medlemmar, den drevs av scouterna men stängdes ner på grund av informationen den innehöll, och att det är den lagstiftningen vi nu får till Sverige (i1, 68).

*“Det är en katastroflagstiftning för den europeiska konkurrenskraften inom IT-området.” (i1, 68)*

**Anonymisering:**

Inom BI-ledet så tror informanten att de flesta kommer att stryka all individ spårbar information som adresser, namn, personnummer. Informanten tror att det kan bli en del anonymisering men att det mestadels kommer bli att istället för att anonymisera information så kommer företag se till att det inte finns någon individ spårbar information överhuvudtaget. De flesta kommer att stryka all individspårbar information som adresser, namn, personnummer, i alla fall inom Business Intelligence ledet (i1, 44).

**Negativa aspekter kring GDPR:**

Informanten anser dock att någon sorts lagstiftning behövs för att skydda den personliga integriteten men att GDPR går alldeles för långt för hur företagen skall agera, och att den ger för lite nytta för konsumenten gentemot bördan för företagen (i1, 70). Han tror även att allting inom IT kommer bli dyrare i Europa på grund av de nya lagen (i1, 72). Informanten anser också att det finns en obalans hos befolkningen, att personer är mer rädda över hur staten använder deras uppgifter än hur företag använder dem, och fortsätter med att detta kommer fortsätta likadant eftersom brottsbekämpning eller förebyggande av brott finns undantaget i GDPR-lagen (i1, 87). Han fortsätter med att lagen framförallt kommer påverka de små företagen då stora företag har tillräckligt med jurister som kan hjälpa dem (i1, 115), och han tror att det just är de stora företagen kommer kunna navigera sig igenom lagen (i1, 115).

När de själva är ute på uppdrag åt sina kunder får de aldrig spara data på sina egna maskiner utan är väldigt noga med att informationen aldrig får lämna kundens miljöer (i1, 109). Hen fortsätter med att förklara att som konsult idag kan det vara ganska tungrovt med tanke på alla inloggnings, brandväggar, VPN-lösningar med mera. Men att hen tycker det är rätt med så hög säkerhet idag, men att den största faran sällan kommer utifrån utan inifrån ett företag (i1, 107).

Informanten tror inte att GDPR kommer sätta stopp för företag som Cambridge analytica eller andra företag som använder data på ett ifrågasättande sätt. För att de som vill göra analys som inte är laglig eller etiskt accepterbar kommer göra det ändå (i1, 78).

## 4.1.2 Organisation 2:

### 4.1.2.1 Big data

**Definition:**

När det kommer till att definiera Big data följer Informanten den teoretiska definitionen 3 V:n, Volume, Variety, Velocity. Hen påpekar att Volymen nästan alltid är given, i alla fall för större bolag. Med variety blir det lätt problem med den ostrukturerade datan, och att när datan blir analyserad, det är då den blir värdefull. Velocity gäller rent allmänt för flera saker som ska göras i "realtime" (i2, 8).

**Möjligheter:**

De största möjligheterna Big data-analys ger är enligt informanten att användningsområden ökar eftersom ostrukturerad och svårtydd data nu blir användbar (i2, 10). Man kan också använda Big data-analyser för forskning med mera menar informanten, men att det just är styrkan i att kunna använda all den mängd data som finns och få ut konkret information av den som är den stora möjligheten med Big data-analys (i2, 10).

**Utmaningar:**

De största utmaningarna är enligt informanten att datan just är ostrukturerad, det är svårt att veta vilken data som skall analyseras för att få ut rätt resultat, samtidigt som ostrukturerad data i sig kan vara väldigt komplex (i2, 20).

**Betydelse:**

Enligt Informanten är Big data-analys nästan en förutsättning idag (i2, 22):

*“Jag skulle säga att Big data nästan är en förutsättning nu. Nu sitter vartenda företag med någon form av Big data i den klassiska definitionen och det innebär att det är bara så det är”*

**Användning:**

Informantens organisation genomför idag analyser åt sina egna kunder, det kan bland annat innefatta att åskådliggöra data och förutspå saker ur den (i2, 24).

**Syfte vid datainsamling:**

Informanten kan inte svara på om det alltid finns ett tydligt syfte med datainsamlingen då organisationen är så stor, så hen vill inte uttala sig om det, men att det alltid ska vara tydligt vid en insamling av data (i2, 28). Men att informantens organisation har många processer för om de hittar känslig data på fel ställe hos en kund. Organisationen har haft råd att vara den som står på integritets sidan enligt Informanten och att de har metoder för att automatiskt anonymisera uppgifter i samband med dataanalys (i2, 30). De har även etiska riktlinjer kring databehandling som de måste följa (i2, 81).

#### 4.1.2.2 Big data och skyddet av personuppgifter

##### **Personlig integritet & Big data-analys:**

Informanten påpekar att det finns en problematik med Big data angående den personliga integriteten, men att det även gör det för alla typer av data. Dock menar informanten att problematiken är större med Big data eftersom ju större mängd data man har desto lättare är det att få en kombination av data som blir identifierbar (i2, 44). Informanten menar att ju större mängd och typ av data desto större fokus på etik måste man ha, att man kan t.ex. kombinera en e-postadress med en GPS-position och då snabbt kan dra slutsatser om vem personen är (i2, 44).

*“Så absolut, ju mer data man samlar in eller ju fler typer av data man samlar in desto större etiskt fokus måste man ha. “(i2,44)*

Informanten anser att det finns flera olika sätt den personliga integriteten kan kränkas på i samband med Big data-analys men att det kan vara mer eller mindre specifikt (i2, 46). Han menar att det kan vara allt ifrån att sprida reklam på ett oetiskt sätt till att granska cookies och se en persons surfhistorik. Informanten anser också att med den nya tekniken kan man samla in mycket mer data än men att man alltid måste ha i åtanke att datan inte alltid är korrekt eller okej att använda (i2, 46).

*“Och även om vi då med ny teknik kan samla in mycket mer data än vad vi borde kunna så innebär inte det att den datan alltid är okej. Att det finns tillgänglig data som man inte hade fått av en person om man hade frågat. “(i2, 46)*

Informanten anser även att om någon har tillräckligt mörka syften så går det alltid att använda på ett eller annat sätt (i2, 46). Att de som är villiga att begå brott spelar det inte så stor roll vad för regler som finns (i2, 74). Informanten berättar att många idag lämnar ifrån sig information utan att faktiskt veta vad det kan användas till som i t.ex. en undersökning där användaren kan vinna ett pris och lämnar sin e-postadress och att det blir problematiskt då det är svårt att kolla hur detta används för användaren (i2, 48). I avseende på Facebook-skandalen påpekar Informanten att de företag som är ärliga och kan hantera känslig information på ett bra sätt kommer vara vinnarna i slutändan (i2, 48).

*“Vi lämnar ifrån oss information utan att faktiskt kunna veta om någon kommer att kränka vår integritet, och där tror jag att företag som kan hantera det på ett bra sätt kommer komma vinnande ur den här stormen” (i2, 48)*

##### **Big datas påverkan av GDPR:**

Hen tror att GDPR-lagen kommer främst göra folk medvetna om vad data egentligen är. Att människor mer funderar på vad de skriver på för avtal i samband med de tjänster som de använder (i2, 50). Hen tror också att människor kommer göra mer medvetna och aktiva val när det kommer till avtal och produkter. GDPR-lagen kommer sätta stopp för att användaravtal kan låsa kunden för att kunden inte accepterar ett visst villkor och att människor kommer bli mer medvetna om vad de kan säga nej till (i2, 75). Hen fortsätter med att vissa företag kanske inte kan ta bort all data eftersom det kan komma att behöva användas av polis och myndigheter, fortsätter med att hen tror att det kan vara ganska lätt att hävda sig ha rätt till data-lagringen då de har en kundklubb eller liknande. Hen påstår att även om reglerna kan verka väldigt tydliga i GDPR-lagen återstår det att se vad som gäller i

verkligheten och vad som faktiskt kommer behöva göras (i2, 50). Enligt Informanten kommer högre krav ställas på var data sparas, även geografiskt.

*“Man måste veta var datan finns, man måste allokera datakapacitet till ett visst område kanske. Det gör ju att man inte till exempel kan koka ihop data och köra allting på ett och samma datacenter någonstans i USA liksom.”* (i2, 70)

Hen fortsätter att om Big data-analys varit enkelt tidigare kommer det bli krångligare, men att det är en sådan sak som bara ska lösas (i2, 70). Att applikationer kanske sammansluter data från annat håll eller att molntjänster kommer användas på fel sätt (i2, 72).

Informanten tror inte att sådana kränkningar som skett i samband med Facebook & Cambridge analytica inte kommer upphöra, eftersom deras plattform är öppen för deras utvecklare finns redan där ett stort integritetsproblem (i2, 52). Men att hen hoppas på att GDPR-lagen kommer öka medvetenheten för utvecklare så att man inte designar system som är dåliga från början. Hen fortsätter med att förhoppningsvis gör lagen att inte mer data är nödvändigt samlas in och behandlas (i2, 52).

Informanten är positivt inställd till lagändringen och hoppas att mycket gott kommer utav den, men att det är för tidigt att säga exakt hur påverkningen kommer bli (i2, 52). Informanten tycker att GDPR-lagen är ett bra steg framåt på vissa områden, t.ex. användaravtal som blir tydligare och kunden kan säga nej till mer saker. Men att tidigare lagstiftning inte har förhindrat stöder av t.ex. lösenordsdatabaser men att en ny lagändring kanske gör detta ännu säkrare (i2, 79).

### **Anonymisering:**

Gällande anonymiserad och pseudonymiserad data tror informanten att detta kommer att bli mer och mer vanligt i samband med GDPR-lagen. Hen tror att om aktörer kan göra hela koncept som är anonymiserade kan detta leda till mer professionalitet (i2, 54).

### **Negativa aspekter av GDPR:**

När det kommer till de negativa aspekterna av GDPR-lagen påpekar Informanten att hen tror att det kommer leda till högre omkostnader för företag. Kanske i viss mån även en omöjlig administrativ situation (i2, 56). Informanten tar ett exempel om hen ber Röda Korset ta bort all data om hen, så måste de garantera att all data tas bort. Även pappersprotokoll från en lokalförening för 10 år sedan. Informanten fortsätter (i2, 5):

*“Det kan kosta hur mycket pengar som helst, och bolag kan inte behöva gå i konkurs bara för att någon begär att bli av med all sin data.”*

Informanten hoppas på att, efter införandet av lagen, att en konsensus om vad som är rimligt att göra och inte kommer tas fram. Informanten fortsätter med att hen är rädd att hederliga aktörer kommer drabbas negativt. Informanten tycker inte GDPR-lagen känns vattentät men att ingen lag gör det för den delen, dock kan just GDPR-lagen få stora konsekvenser för något vi inte tänkt på just nu utan inser först i framtiden (i2, 56).

### 4.1.3 Organisation 3:

#### 4.1.1.3 Big data

**Definition:**

Informant 3 är inte en stor supporter av begreppet Big data, men kände igen att man brukar se Big data genom ett par V:n, som hen dock inte finner återspegla verkligheten.

*”Om det inte finns variety i datan så skulle det inte vara Big data, det finns inte en sådan binär grej” (i3, 20).*

**Möjligheter:**

Informanten menar på att de Big data-teknikerna har möjliggjort mer affärsrelevant-, affärsnära-analys av väldigt granulär data. Det är där informanten anser att det stora värdet ligger i, inte att det genereras mer data, utan att datan är mer tillgängliggjord för företagen (i3, 24).

**Utmaningar:**

Ett stort problem som informanten har sett är att flera av hans kunder har haft ett alltför stort tekniskt fokus på implementeringen, och glömt bort det huvudsakliga syftet. Företag fokuserar för mycket på datainsamling, att alla ska få in mer och mer data.

*”Alla datakällor ska in och det blir någon form av pokemonjakt att man ska ha alla datakällor i sin datalake. Man tappar fokus på slutsyftet” (i3, 30).*

Informanten menar på att det finns stora affärsvärden att hämta genom att analysera data, men att det lätt blir fokus på fel saker för att det är en ny teknik och kompetensen inte alltid finns där, vilket blir en stor utmaning för företag (i3, 30).

*“Det är problemet med Big data, jag förknippar det väldigt mycket med folk som lyckas med investeringar med saker de inte har någon aning att få värde av. För att de saknar den analytiska kompetensen. Många förväxlar att om man har data eller analytiska modeller som är dålig, att slutresultatet blir bättre om man vrider på mängden data. Det vill jag starkt motsäga mig ifrån” (i3, 8)*

Informanten har sett att företag har sparat fel data och sparat data om kunder som egentligen inte skulle vara sparade för att de inte är kunder längre. Han menar på att det inte är av anledningen att man vill missbruka uppgifterna, utan att det varit tekniskt svårt att åtgärda det. (i3, 36). Informanten anser att GDPR är bra, för att företagen kämpar mer med att ha korrekt data i och med att det blir hårda konsekvenser ifall man inte följer lagarna (i3, 36).

**Betydelse:**

Big data är ett viktigt affärsområde för informantens organisation, han har sett att det finns mycket stora investeringar hos deras kunder inom Big data-området. Han menar på att för hela marknaden finns det bolag vars affärsmodell bygger väldigt nära Big data och analys som exempelvis Spotify och Netflix (i3, 40). För de mer traditionella företagen tror informanten att de ser mer Big data som ett hot, eller att de måste använda Big data för annars kommer de bli iväg sprungna (i3, 40).

**Användning:**

Informantens bolag gör allt inom Big data. De jobbar dels med teknisk implementation som handlar om integration av lösningar, där man samlar in datan, skapar en plattform och lagrar datan. Till att skapa en miljö att samla in data och sedermera analysera datan ifrån.

Informanten tillhör gruppen som analyserar data, de hjälper kunder att analysera deras data för att skapa bättre insikter och driva ut det till affärsvärde (i3, 42).

**Syfte vid datainsamling:**

Vid de projekt som informanten har arbetat med så har han alltid sett att det funnits ett tydligt syfte med att samla in datan. Samtidigt är mycket av datan som samlas in mer granulär, vilket gör den mer känslig. Så även om det finns ett syfte med att samla in datan, så kan man lättare missbruka den, och där tror informanten att företag i en mycket större utsträckning nu än tidigare måste fundera på hur de vill använda datan och till vilka syften, att man inte går helt ”bananas” på datan och skapar konstiga användar-kampanjer eller individuella anpassningar som kan slå fel (i3, 37).

**4.1.2.3 Big data och skyddet av personuppgifter****Personlig integritet & Big data-analys:**

Informanten menar på att det existerar en etisk problematik med Big data, men han menar samtidigt att det gör det inom alla branscher. Men han menar på att med en större tillgång till data och mer tekniska och funktionella möjligheter att analysera datan så behövs det etiska riktlinjer för hur man ska jobba med sin data (i3, 48). Hen anser att Big data-miljöerna faktiskt kan vara bättre uppbyggda för att anpassas efter GDPR-lagen än gamla Business Intelligence miljöer (i3, 36). Hen har sett att det finns sparad data om personer som inte längre är kunder hos företaget, och att det mer är på grund av den tekniska utmaningen att få ur kunderna ur systemen än ett medvetet val (i3, 36). Hen fortsätter med att hen tror GDPR kan få bukt med sådana saker eftersom konsekvenserna är stora om inte lagen följs (i3, 36).

Informanten anser också att Big data-analys och dess data är mer känslig än annan data och därför finns ett större behov att diskutera och fundera på hur man vill använda datan (i3, 36).

*“Med en större tillgång till data och mer tekniska möjligheter, funktionella möjligheter att analysera data, sätta ihop data på olika sätt så ökar behovet att ha etiska riktlinjer, eller ihopsatta regelverk för hur man ska jobba med sin data.” (i3, 48)*

Informanten har arbetat med kunder som har plattformar dedikerade för ett visst syfte, där systemet automatiskt anonymiserade data. Där säkras kundens integritet samtidigt som det affärsmässiga värdet bibehålls, och att vi kommer se fler sådana lösningar (i3, 48).



**Big data påverkan av GDPR:**

Informanten förmodar att andra områden kommer slå hårdare än Big data med GDPR. De största utmaningarna tror hen ligger i att man smidigt ska kunna ta bort data i de gamla systemen (i3, 60). Informanten tror att GDPR kommer att försvåra analys, för att det kommer vara svårt för personer eller legala avdelningar hos företag att förstå arbetssättet när man jobbar med analys, vilket kommer att påverka saker som faktiskt är bra för kunden.

*”När man till exempel analyserar två varianter av en produkt och man vill förstå vilken produkt som skapar mest nöjda kunder, då tror jag att har man inte tillåtelse att analysera eller samla in data för en sådan sak som ändå syftar på att göra sin produkt bättre, jag tror att det i vissa lägen kommer att försvåra analyser i slutändan som faktiskt är bra för kunden.”*  
(i3, 64)

Här ser informanten en problematik med GDPR, att vissa tror att GDPR bara kommer göra det bättre för kunderna, vilket det på vissa sätt kommer att bli, men att det kommer att försvåra analys som faktiskt är bra för slutkunderna (i3, 64). Det kommer även bli mer kostsamt och tidskrävande att göra analys.

*“Istället för att göra sina tio olika analyser har man kanske bara råd att göra fem eller tre”*  
(i3, 66).

Informanten är tveksam om att GDPR kommer att ha en effekt på företag likt Cambridge analytica som använder och analyserar stora datamängder på ett ”skumt” sätt.

*“Jag tror att även med GDPR att man får legala restriktioner, så tror jag att det fortfarande kommer att finnas massa med case där det är etiskt rätt eller inte.”* (i3, 62)

Hen talar om hur användningen av ansiktsgenkänning genomförs på delar av befolkningen i Kina och att Informanten ser detta som mycket integritetskränkande. Men att på den europeiska marknaden tror Informanten att GDPR kommer ha effekt men att det alltid kommer finnas de som ser affärsmöjligheter eller andra drivande krafter som genomför analyser som vare sig är legalt eller etiskt försvarbara (i3, 62).

**Anonymisering:**

Informanten har sett kunder som har specifika plattformar som är dedikerade för ett syfte, där man har anonymiserat datan så att man kan göra den analys man vill, utan att göra intrång på den personliga integriteten, men att det affärsmässiga värdet bibehålls (i3, 48). Informanten tror att efter GDPR kommer anonymiserad data att öka, att den blir mer GDPR kompatibel (i3, 54). Men också att man kommer jobba mer med lösningar så att man inte kan titta på individer i grupper av människor, utan att man vid analys på något sätt täcker över alla unika mönster eller svar som man kan härleda till en viss person, vilket han även tror kommer att skydda individerna i slutändan (i3, 60).

**Negativa aspekter kring GDPR:**

Informanten menar på att GDPR kommer ge effekter som blir mer kostsamt och tidskrävande för företagen, vilket inkluderar dataanalys (i3, 66).

#### 4.1.4 Organisation 4:

##### 4.1.1.4 Big data

**Definition:**

För informant 4 är det inte intressant huruvida någonting är Big data eller inte, utan för datainspektionen är det mer intressant att behandling av personuppgifter sker i enlighet med gällande rätt. Men nämner att Big data kan vara stora datavolymer som man kan använda på olika sätt (i4, 4)

**Möjligheter:**

Informanten anser att det finns enorma fördelar med Big data, det kan vara allt inom vård och forskning där man använder Big data som direkt tillför människan och mänskligheten. Hen menar på att det är användningen som avgör om det är på gott eller ont (i4, 7). Informanten ser det som att det är en ganska ung och omogen teknik men som besitter enorm potential (i4, 15).

**Utmaningar:**

Ur Informantens organisationsperspektiv är de största utmaningarna med Big data är behandlingen av personuppgifter, att förhålla sig enligt ansvarsskyldigheten till gällande lagstiftning.

*”För det första får man bara behandla personuppgifter för berättigade ändamål, och det kan jag säga så här långt att det finns ingen som har behandlat en personuppgift som inte har ansett ha haft ett berättigat ändamål, men det är inte riktigt samma sak utan här ska det rent objektivt sätt vara ett berättigat ändamål” (i4, 9).*

Sen när det gäller särskilda kategorier av uppgifter, således känsliga personuppgifter, så är utgångspunkten ett förbud, och då måste man hitta ett applicerbart undantag från det här förbudet, att man får behandla känsliga personuppgifter enligt Informanten (i4, 11). Hen fortsätter med att detta gäller oavsett om tekniken beskrivs som Big data, AI, eller någonting sådant, och att regelverket är generellt (i4, 11). Informanten fortsätter med att de som ligger längst fram i krigsfronten kanske inte är de som ligger längst fram i insikt i och förhållande till gällande regelverk, även om de naturligtvis borde det och har en skyldighet att följa gällande regelverk (i4, 11). Till exempel då att känsliga personuppgifter behandlas på ett sätt som innebär en kränkning av den registrerade personliga integriteten (i4, 11). Hen fortsätter med att mena på att tillsammans med de möjligheter som finns idag med så stora datamängder och så vidare kan en brist uppstå vid behandling av personuppgifter utan stöd eller i strid med ett förbud (i4, 13).

*“Det ligger väl i entreprenörens natur att inte låta sig begränsas” (i4, 13)*

#### 4.1.2.3 Big data och skyddet av personuppgifter

##### **Personlig integritet & Big data-analys:**

Informanten ser att det finns både sociala och etiska frågor att ställa med användningen av Big data, men att det kommer att stjälpas i samband med det nya regelverk som kommer till Europa som är betydligt ordrikare och således mer beskrivande och förklarande på vad som är okej och inte okej (i4, 23).

*”Jag brukar säga så här att väldigt många av de ”nyheter” som kommer med GDPR egentligen inte är nyheter, men man måste vara väldigt duktig på att läsa PuL för att läsa ut att det är just det där som krävs. Det nya regelverket är i många hänseenden ett förtydligande om vad som gäller idag” (i4, 23).*

I hur informanten ser att den personliga integriteten kan kränkas i samband med Big data-analys så beror det på vad analysen avser, vilka uppgifter det är som behandlas, behandlingens ändamål, sammanhang, omfattning och att man har rättsligt stöd. Kunskapsbrist gällande regler är en stor risk som informanten anser kan innebära kränkning av den personliga integriteten (i4, 27).

Informanten berättar att all behandling av personuppgifter är ett intrång inom de registrerades personliga integritet (i4, 29). Och intrånget får inte vara otillbörligt för då är det en kränkning, otillbörlighet i detta fall är att behandlingen inte sker i enlighet med GDPR-lagen (i4, 29). Hen fortsätter med att den personuppgiftsansvarige kanske inte förstår vilka skyldigheter den har, och vilka rättigheter den registrerade har när de bestämmer ändamålen för analysen (i4, 29). Och Informanten fortsätter med att inga praktiska åtgärder kan ersätta kravet på rättslig grund (i4, 29).

*”Det jag har varit med om ofta är att man missar lite grann, för att man är så inne på att ändamålet är så berättigat, det här är så viktigt, särskilt forskarna, det här är viktigt för mänskligheten liksom. Vissa har gjort det på ett jättesäkert sätt och på rätt och vis, men ni saknar rättslig grund” (i4 31)*

##### **Big datas påverkan av GDPR:**

Hur GDPR kommer att påverka Big data-analys så tror och informanten hoppas att medvetenheten hos företag och personer kommer att bli större, men det betyder inte nödvändigtvis betyda att man förstår vad regelverket kräver av en (i4, 41).

När vi frågar Informanten om hur konkreta konsekvenser för organisationer som behandlar personuppgifter i och med införandet av den nya lagen svarar Informanten (i4, 67):

*“Att man hela tiden har ett risktänk i förhållande till den behandling som man ska utföra, har man skaffat helt rättsligt stöd för att utföra den och bör även reda ut andra alternativ för den här behandlingen. “*

Informanten belyser vikten av att lagen följs, samtliga organisationer som behandlar personuppgifter är skyldiga att samarbeta med till skyddsmyndigheten, vilket innebär att de när som helst kan få krav att visa hur de behandlat personuppgifter. De tar även in tips och har en omvärldsbevakning (i4, 79). Informanten fortsätter med att det finns otroligt många sätt

som gör att datainspektionen får upp ögonen för en personuppgiftsbehandling som är tveksam eller problematisk (i4, 79). Vid behandling av personuppgifter kommer det även krävas konsekvensbedömningar som alltid skall produceras separat enligt Informanten (i4, 75).

*“Och då om datainspektionen kommer och säger “detta var en högriskbehandling, kan vi få titta på er konsekvensbedömning” då ska den bara vara att lägga fram.” (i4, 77)*

### **Anonymisering:**

På grund av den ökade tydligheten av förordningen så tror informanten att Big data-analyser kommer användas mer av anonymiserad och pseudonymiserad data, för att man i största möjliga mån ska undvika direkta personuppgifter. Är det inte personuppgifter eller om det rör sig om helt anonymiserad data så är inte GDPR tillämpligt. Informanten menar dock på att det visat sig vara svårt att anonymisera stora datamängder, vilket bland annat Big data har påvisat historiskt.

*”Så det har visat sig svårare att anonymisera uppgifter än vad man tidigare trott. För sådana här funktioner, om vi pratar nu Big data, kan ju också användas för att åter-identifiera uppgifter” (i4, 47).*

Samtidigt påpekar informanten att även ifall man använder anonymiserad data så måste det finnas en rättslig grund för analyserna (i4, 45). Informanten menar att kombinationer av data kan vara direkt utpekande, även ifall det inte står ens namn eller personnummer. För vad som är personuppgift beror på vilket sammanhang uppgifterna är i (i4, 51).

*”En vanlig fråga jag får är en IP-adress en personuppgift? Då brukar jag säga är blå en personuppgift? Ja, om det står jämte mitt namn och datainspektionen så är blå en personuppgift, eftersom den säger någonting om mig. Det kanske är fel, jag kanske är brunögd, men det är fortfarande en uppgift som är hänförlig till mig.” (i4, 49)*

### **Negativa aspekter kring GDPR:**

Vilka negativa konsekvenser som informanten ser att GDPR kan ge är att på grund av okunskap så har vissa företag vidtagit väldigt drastiska åtgärder, förstört en massa uppgifter som inte hade behövts förstöras vilket hen menar kan skada en verksamhet (i4, 59). Informanten fortsätter med att säga att folk hetsar upp sig i onödan och att det inte finns några rättsliga hinder, utan bara rättsliga förutsättningar (i4, 55).

## 5 Diskussion & Analys

*I detta kapitel jämförs och analyseras den insamlade data från vår empiriska studie mot vår data från vår litteraturgenomgång. Vi belyser likheter, olikheter och intressanta fynd. Kapitlets struktur utgår från de två stora teman vi fastställt i litteraturgenomgången för att hålla en struktur för läsaren och en röd tråd genom uppsatsen.*

---

### 5.1 Big data

#### 5.1.1 Vad är Big data, definition

Trots att begreppet Big data har varit etablerat några år så har vi märkt att det kan ligga en förvirring i begreppet och att det råder delade uppfattningar hur folk definierar det. De tre V:na som både Russom (2011) och Gartner (2012) har med i sin definition är den mest etablerade beskrivningen vi funnit, trots detta finner vi i vår intervjustudie att begreppet är svårdefinierat och är lätt att tolkas olika. Big data för informant 1 är när när datavolymer och strukturen inte stämmer in i vanliga transaktionsbaserade datasystem och att den ska innehålla ostrukturerad data. Informant 2 hade stött på de tre V:na när han gjorde sin masteruppsats inom Big data och menar på att de 3 V:na stämmer väldigt väl in i praktiken. Informant 3 tyckte inte om begreppet Big data, hen följde dock de V:n som finns men att dessa inte riktigt återspeglar verkligheten. Det är framförallt att datan måste vara varierande som Informant 3 tyckte var lite väl binärt. Informant 4 anser att utifrån sitt perspektiv som representant för datainspektionen spelar det ingen roll vad det kallas så länge lagen följs och behandling av personuppgifter sker i enlighet med gällande rätt. Men att Big data innefattar stora datavolymer som man kan använda på olika sätt höll Informant 4 med om vilket samtliga informanter också påpekat.

Bello-Orgaz et al. (2016) anser också att Big data inte handlar om enbart lagring och stora datamängder utan att målet med Big data är att analysera den för att skapa ett värde. Så med andra ord finns olika syn på begreppet och utifrån vår empiriska studie har vi funnit att det således finns en viss förvirring angående innebörd och definition. Dock instämmer samtliga Informanter med litteraturen att Big data innebär stora datavolymer med stora mängder ostrukturerad data som ofta är oanvändbart utan ett syfte för analysering och behandling.

#### 5.1.2 Möjligheter med Big data

Att Big data har kommit att få en stor betydelse för både organisationer och den enskilde individen är både informanterna och vår litteratur till stor del överens om. Med all data som genereras idag och alla möjliga olika tillämpningsområden inom Big data så har det blivit ett viktigt affärsområde. Big data har möjliggjort att finna nya relationer i data som tidigare varit dolda, vilket kan bidra till större lärdom och bättre beslutsfattning (Marr, 2017). Detta ser även Informant 2 som den stora styrkan med Big data, att kunna använda stora mängder data och få ut konkret information utav den. Informant 2 menar även på att möjligheterna med Big data är att ostrukturerad och svårtydd data blir lättare användbar med Big data-tekniker. Informant 1 är även inne på samma spår där han menar på att de största möjligheterna ligger i att få svar

på sin frågeställning med hjälp av all data, och man kan få svar på saker man inte hade tänkt på vid datainsamlingen, detta diskuteras vidare i punkt 5.2.3. Informant 3 menar på att tillgängliggörandet av all data, det är där den största potentialen finns. Genom att tillgängliggöra data så är det lättare för företag att hitta insikter som påverkar affärs- och operativa beslut.

Informant 2 menar på att Big data har i princip kommit att bli en förutsättning för vissa företag och affärsområden, för att de ska kunna vara konkurrenskraftiga så måste de ta tillvara och finna mönster i all data. Deloitte (2013) är inne på samma princip, att Big data har blivit ett viktigt fundament för flera företag, där användningen kan vara för att hitta nya mönster, eller för att ta bättre affärsrelaterade beslut. Informant 4 anser att Big data inom vård och forskning har bidragit direkt till människan och mänskligheten. Informant 3 förklarade att för hela affärsmarknaden har det kommit företag vars affärsmodell grundar sig i stora datamängder och bygger på Big data-analys, vilket Cambridge analytica är ett tydligt exempel på.

De företag vi intervjuat arbetar med Big data och Big data-analys, där de arbetar med att hjälpa sina kunder att genomföra analyser, förutom Informant 4 som arbetar på datainspektionen och arbetar inte direkt med Big data men är insatt i ämnet. Informant 1 har bland annat jobbat med Big data för att kartlägga temperaturer i olika lägenheter. Informant 2 & 3 har arbetat med olika projekt, där det ofta innefattar att åskådliggöra data och förutspå saker från den för att skapa bättre insikter och driva ut det till ett affärsvärde. Det verkar finnas få områden där det finns tillgång och genereras stora mängder data som man inte kan applicera värdeskapande analys ifrån. Informant 1 menar bland annat att Big data skulle kunna ha en mycket större betydelse i samhället och för företag än vad den har idag, och menar bland annat att inom produktion och kvalitetsmätning finns det en mycket större potential att tillämpa analys av stora datamängder för att få ut bättre resultat.

### 5.1.3 Syfte vid datainsamling

Samtliga informanter menar på att datainsamling alltid börjar med ett visst syfte, i alla fall under de projekt som de arbetat med. Samtidigt kan data som varit insamlat för ett syfte, användas för något annat vilket kan bidra till att finna nya insikter och mönster Sharda et. al, (2014). Informant 1 förklarar i ett exempel på hur han har hållit på med Geocaching och GPS i flera år och sparade GPS tracksen från varje tur han var ute på. Senare visade sig att han kunde geo-tagga alla bilder, något som inte gick när han samlade in datan, och därmed fick han en ny användning för datan, och han menar just på att Big data ger möjlighet till nya insikter som man inte tänkt på när man samlar in datan. Detta går emot principen att insamling och behandling av data alltid har ett tydligt syfte, och visar enligt oss på en problematik som finns med Big data-analyser. Vi anser att det kan bli ett hot för den personliga integriteten om data som samlats in för ett specifikt syfte används för något helt annat. Detta går i linje med Integritetskommittén (2016) som säger att det finns en problematik att uppgifter hanteras för nya ändamål som inte är kända vid insamlingen. Här ser vi även ett mönster, att allt blir problematiskt så länge det rör personuppgifter, och som informant 4 menar på så kan abstrakta saker vara personuppgifter så länge de finns i ett sammanhang som gör att det kan knytas till en viss person. Frågan är hur man vet vad som är en personuppgift eller inte, vad som helst verkar kunna bli en personuppgift så länge det kan knytas till en viss person. Eftersom den nya lagen så gäller samma lag för ostrukturerad som strukturerade personuppgifter.

Informant 3 menar på att även om det finns ett sunt syfte med Big data-analys så kan den missbrukas till andra syften, och av denna anledning anser vi, i enighet med Informant 3 att företag sannerligen bör överväga hur datan skall användas och till vilket syfte innan genomförandet av en analys. Han menar vidare på att innan så reglerade det sig självt när datan och tekniken inte fanns där. I vissa fall har lagarna och regler inte upprätthållits å jour till den snabba tekniska utvecklingen, och där tror vi GDPR kan hjälpa till eftersom det krävs en rättslig grund från den personuppgiftsansvarige (Datainspektionen, 2017b) samt att lagen inkluderar ostrukturerad data (Datainspektionen, 2017c) vilket vi tror är två starka riktlinjer för företagen att förhålla sig till. Vi tror att Big data-analyser kommer bli mer begränsade och reglerade efter införandet av GDPR-lagen, detta fortsätter att diskuteras i punkt 5.3.2.

#### 5.1.4 Utmaningar med Big data

Mycket av litteraturen kring problematik och utmaningar med Big data utmynnar i dess potential på intrång på den personliga integriteten. Enligt IBE (2016) är en av de främsta utmaningarna som företag står inför som använder Big data-analys är dess önskan att förbättra deras kund- och affärsverksamhet, samtidigt som det kan betyda potentiellt intrång på den personliga integriteten. När vi ställde frågan till våra informanter om vilka utmaningar de ser med Big data, kom det svar som inte alltid var knutna till den personliga integriteten. Informant 1 menar på att en av de största utmaningarna med Big data är att kunna se och få ut ett ekonomiskt värde. Att lagra data är nödvändigtvis inte en stor kostnad, men att analysera data kan bli en relativt kostsam process vilket kräver investeringsbeslut som kan vara svårt att få, och där menar informant 1 att ett tydligt syfte är ett väsentligt för att få ett positivt investeringsbeslut. Samtidigt såg informant 3 att syftet ofta glöms bort för att det blir för stort fokus på datainsamlingen. Informant 3 menar även på att en stor utmaning inom Big data är att det är ett relativt nytt fenomen och därmed är kompetensen inte alltid hög, vilket bidrar till att det lätt fokuseras på fel saker vid analysen. Att ha rätt kompetens menar (Harford, 2014) är av stor vikt vid analys, för att kunna få ut de bästa analyserna behövs en djup och omfattande förståelse för att kunna förstå samband och få ut de bästa analyserna.

En stor utmaning som informant 2 ser med Big data är att stora mängder av datan oftast brukar vara ostrukturerad och därmed kan vara komplex. Det kan bli svårt att veta vilken data som skall analyseras för att få ut rätt resultat och är datan ofullständig eller saknas i en kontext så ger den inget värde, vilket Zoldan (2013) menar på att flera organisationer inte inser vikten i utan har en övertro på datan bara för att man besitter stora mängder av den. Vilket även informant 3 fann detta problematiskt, som menar på att det finns en tendens att vissa bara vill samla in data, då de tror att ju större mängd man besitter ju mer värde kommer det att skapa, vilket i vissa fall kan vara sant ifall man lyckas få ett värde ur det. Denna utmaning stöds även av (IDG, 2016) undersökning om användning av Big data, där nästan hälften av organisationerna såg den ostrukturerade datan som en av de största utmaningarna.

Informant 3 återkommer en del i intervjun till att det krävs en analytisk kompetens för att Big data-analys ska bli värdefullt. Han har sett att folk tror att slutresultatet blir bättre bara för att man vrider och vänder på datan, och det är just detta som Harford (2014) nämner som ett problem inom Big data-analys, att man letar bara efter statistiska korrelationer utan att reflektera innebörden. Informant 3 menar på att många saknar den analytiska kompetensen, vilket gör att flera analyser inte ger ett värdeskapande resultat, för är man analytiskt lagd då vill man titta på datan på ett analytiskt sätt och förstå vad den säger (i3, 26).

## 5.2 Skyddet av personuppgifter

### 5.2.1 Personlig integritet & Big data-analys

I vår studie har vi funnit att samtliga informanter anser att det finns en problematik för den personliga integriteten i Big data-analys. Detta går i samma linje som litteraturen. Enligt Integritetskommittén (2016) har de analystekniker som uppkommit i samband med Big data och även andra typer av informationsbehandling fått en större effekt på den personliga integriteten. Det finns ingen direkt definition av vad personlig integritet är eftersom rätten till en privat sfär inte är absolut utan en subjektiv bedömning (Integritetskommittén, 2016). Den personliga integriteten kan pressas då enskilda individer i stor utsträckning har dåliga kunskaper om hur deras personuppgifter behandlas i olika sammanhang. Ytterligare försämringar för individen gällande påverkan hur personuppgifter lagras, sprids och används uppstår genom bland annat Big data-analys (Integritetskommittén, 2016). Detta går i linje med åsikterna från Informant 2, att många personer idag lämnar ifrån sig information utan att veta vad det kan användas till. Informant 1 anser istället att de människor som interagerar med företag gör det för intresset för deras produkter, och att därför bör de inte känna sig kränkta i lika stor utsträckning då den allmänna acceptansen för kartläggning är högre. Här ser vi en problematik eftersom det är upp till varje individ att avgöra vad som är en kränkning eller ej och även om acceptansen är högre för vissa, så betyder inte det att det inte är ett intrång på den personliga integriteten. Vidare kan detta bli problematiskt om individen inte vet hur eller varför datan används, i vilket syfte ens data används, som stöds av Integritetskommittén (2016).

Vi tror att balansen mellan personlig integritet och Big data-analys kan vara ganska så stram eftersom det många gånger är oklart hur datan samlats in och till vilket syfte (Integritetskommittén, 2016). Informant 1 säger även att spridningen av information och att den används i ett annat syfte än det ursprungliga är problematiskt. Så länge uppgifterna används för det självklara inhämtade ändamålet så upplevs det som mindre känsligt enligt Informant 1. Men hen anser samtidigt att Big data-analys är värdefullt även vid behandling av data än den ursprungliga insamlingens syfte som tas upp i punkt 5.2.3. Detta ser vi som problematiskt för den personliga integriteten eftersom viljan att återanvända data som är insamlat för ett annat syfte finns hos organisationer idag. Men Informant 1 anser samtidigt att Big data-analys gör det möjligt att kränka den personliga integriteten på nya sätt, som att se onormalt hög vattenförbrukning i en lägenhet. Om detta är en kränkning är upp till den analyserade att avgöra, ifall det inte finns tydliga riktlinjer hos organisationen så är det analytikerns sunda förnuft som gäller. Det finns ingen definition av vad som är en kränkning eller inte då detta är subjektivt och från fall till fall (Datainspektionen, 2008) och detta håller Informanterna med om. Detta är problematiskt då både synen på Big data och definitionen vad kränkning är inte är tydlig. Det kan i sin tur leda till att företag själva gör dessa definitioner, och då definierar vad kränkande användning av Big data är utifrån sin egen vinning, och inte ur ett konsumentperspektiv samtidigt som nästan allt kan ses som en kränkande behandling. Informant 2 är lite fundersam om hur en Big data-analys kan vara kränkande men ger flera exempel som bland annat granska cookies om en persons surfhistorik till att sprida reklam som en individ inte önskat. Informant 1 fortsätter med att generellt sett finns en "slarvig inställning" hos företag idag gällande personlig information och att den då kan hamna i fel händer. Informant 4 anser också att det finns en kunskapsbrist gällande regler och därför höjs risken att den personliga integriteten kränks.



Från vår studie har vi funnit att problematik för den personliga integriteten finns för all typ av data men att den är större med Big data-analys eftersom ju större mängd data desto lättare att kombinera den på ett kränkande och identifierande sätt. I enlighet med litteraturen påpekar Informant 2 och 3 detta, det stöds av Integritetskommittén (2016) som menar att datainhämtning för analys kan skapa stora problem för den personliga integriteten. Informant 2 fortsätter med att det krävs ett större fokus på den personliga integriteten ju mer mängd och typer av data som samlas in och behandlas. Informant 3 håller med Informant 2 inom detta:

*“Med en större tillgång till data och mer tekniska och funktionella möjligheter att analysera data och sätta ihop data på olika sätt så ökar behovet att ha etiska riktlinjer, eller ihopsatta regelverk för hur man ska jobba med sin data. “(i3, 48)*

Vidare har vi funnit att samtliga förutom Informant 1 är positivt inställd till den nya lagstiftning som är på ingång. Informant 1 tycker att någon sorts lagstiftning behövs för den personliga integriteten men att GDPR går för långt. De övriga 3 är mer positiva till GDPR och anser i linje med litteraturen att Big data-analys och annan databehandling bör göras säkrare för individen och hoppas på en smartare design av systemen redan från början.

### 5.2.2 Big datas påverkan av GDPR

I vår studie har vi kunnat se att Big data-analys kommer att påverkas av GDPR-lagen bland annat för den nya lagen nu inkluderar personuppgifter i ostrukturerad data (Datainspektionen, 2017c). Detta står i enlighet med den empiriska studie vi genomfört. Informant 1 tror dock inte att lagen kommer förändra just Big data-analys särskilt mycket då många som genomfört analyser följt den tidigare lagstiftningen. Detta stöds även av Informant 2 och 3 där Informant 2 mer tror att GDPR kommer göra människor uppmärksamma och mer medvetna på vilka val de gör i samband med olika användaravtal, vilket indirekt kan komma att påverka Big data-analys då personer kommer vara mer medvetna om vilken information de lämnar ut och behandlas. Informant 3 hävdar att GDPR kommer “slå hårdare” på andra områden än Big data-analys, detta eftersom de äldre systemen kommer vara svårare att anpassa än de modernare analys-systemen.

Dock anser Informant 3 att GDPR kommer påverka Big data-analys negativt på andra sätt. För det första tror hen att GDPR kommer försvåra analyser, sedan tror hen att det kommer bli dyrare eftersom förarbetet inför en analys kommer bli mer omfattande med ett inkluderande av legala avdelningar med mera.

*”När man till exempel analyserar två varianter av en produkt och man vill förstå vilken produkt som skapar mest nöjda kunder, då tror jag att har man inte tillåtelse att analysera eller samla in data för en sådan sak som ändå syftar på att göra sin produkt bättre, jag tror att det i vissa lägen kommer att försvåra analyser i slutändan som faktiskt är bra för kunden.”*  
(i3, 64)

Hen fortsätter med att det finns en problematik med GDPR, där Informanten tror att analyser som är bra för kunden kommer hämmas, och att det kommer bli mer tidskrävande och kostsamt att genomföra analyser. Dock är detta enligt oss svårt att definiera, på vilket sätt något är bra för företaget eller slutkunden kan bero på synvinkel. Informant 3 menar att detta kommer drabba slutkundens upplevelse och resultat, istället för att göra tio olika analyser kanske organisationen endast har råd att göra fem eller tre säger Informant 3. Detta går i linje med vad Informant 1 säger om att lagen är utformad på det viset att den ger en väldigt stor börda för företagen men ett litet värde för konsumenten.

“Som den är utformad, regelverket, så ger den ett otroligt litet värde för konsumenten, men den ger en väldigt stort börda för företagen som hanterar informationen” (i1, 68)

Informant 2 är inne på samma spår som ovan, genom att högre krav kommer ställas på var data sparas och att organisationer nu måste kontrollera att inte flera applikationer sammanstrålar data från olika håll. Informanten fortsätter med att man nu inte kommer kunna “koka ihop data” och köra på ett och samma datacenter i t.ex. USA. Hen menar att om Big data-analys setts som enkelt tidigare kommer det nu bli krångligare men att det bara måste lösas. Informant 2 tror även att GDPR kommer påverka och öka medvetenheten hos utvecklare som gör att systemen designas på ett smartare sätt och att inte mer data än nödvändigt samlas in och behandlas.

Informant 4 går i samma linje som Informant 2 och hoppas att medvetenheten och företag och personer kommer bli större, men att det inte betyder att man kommer förstå vad regelverket kräver av en ändå. Informant 4 anser att den största förändringen för Big data-analys i samband med GDPR-lagen kommer vara att man alltid måste ha ett risktänk i förhållande till den behandling som ska utföras. En annan förändring är att det nu även kommer krävas konsekvensbedömningar som alltid skall kunna uppvisas om datainspektionen gör en utredning på en organisation. Dessa ska produceras separat vid en behandling av personuppgifter och skall kunna uppvisas direkt när datainspektionen gör en utredning enligt Informant 4.

### 5.2.3 Anonymisering

Enligt IBE (2016) har företag idag stora dilemman i förhållande med önskan att samla in och använda data för att förbättra organisationer gentemot respekterande av intressentens integritet. I vår studie har vi sett att organisationer kommer att använda sig av anonymiserad data i större utsträckning än tidigare. Informant 2 & 3 har arbetat i miljöer som automatiskt skyddar de analyserades integritet genom anonymisering. Så det affärsmässiga värdet bibehålls tillsammans med skydd av den personliga integriteten. Och båda tror att detta kommer bli mer vanligt.

Informant 1 tror att företag kommer undvika att behandla personuppgifter i största utsträckning och helt enkelt stryka all individ spårbar information i samband med behandlingen. Detta tror inte vi kommer ske i samma utsträckning som Informant 1 då Big data-analyser genomförs med en oerhörd mängd olika typer av data. Och som Datainspektionen (2017c) säger är en personuppgift allt ifrån ett personnummer till en IP-adress, att helt undvika behandling av personuppgifter efter införandet av GDPR-lagen tror vi kommer vara komplext, om inte omöjligt. I och med att det även inkluderar ostrukturerad data som bilder, inspelat ljud med mera. Informant 2 och 3 tror som sagt att anonymisering

kommer att öka, men också att lösningar som är skyddande redan från början kommer att tas fram. Informant 3 tror att gruppanalyser kommer bli vanligare och att man vid analys täcker över alla unika mönster eller svar som kan härleda till en viss person för att skydda individen.

Dock påpekar Informant 4 att anonymisering inte är så lätt som man tidigare trott, och att det fortfarande måste finnas en rättslig grund för analyseringen som stöds av litteraturen (Datainspektionen, 2017b). Informant 4 fortsätter:

*”Det har visat sig svårare att anonymisera uppgifter än vad man tidigare trott. För sådana här funktioner, om vi pratar nu Big data, kan ju också användas för att åter-identifiera uppgifter” (i4, 47).*

Vi tror att detta absolut är en risk för den personliga integriteten, att detta är möjligt har påvisats tidigare genom bland annat det exempel som lyfts i litteraturgenomgången. Genom kombinationer och matchning av olika datatyper av-anonymiserades den “anonyma” Netflix-datan (Narayanan & Shmatikov, 2008). Att möjligheten finns är försvårande men i hur stor utsträckning det kommer att ske är svårt att säga. Informant 4 fortsätter med att kombinationer av data kan vara utpekande även om inte ens namn eller personnummer står där.

*”En vanlig fråga jag får är en IP-adress en personuppgift? Då brukar jag säga är blå en personuppgift? Ja, om det står jämte mitt namn och datainspektionen så är blå en personuppgift, eftersom den säger någonting om mig. Det kanske är fel, jag kanske är brunögd, men det är fortfarande en uppgift som är hänförlig till mig.” (i4, 49)*

Så har man som syfte att vilja försöka identifiera vissa personer, och datan stödjer det så kommer det förmodligen vara möjligt, även efter GDPR. Vi tror att Big data-analyser kommer att förändras efter införandet av GDPR-lagen. Hur dessa förändringar kommer att se ut återstår att se, vi tror dock att färre analyser kommer genomföras på grund av högre kostnader, mer anonymisering och gruppanalyser, högre krav på rättslig grund vid behandling av data och flexibiliteten för organisationer kommer att hämmas. Förhoppningsvis till vinst för den personliga integriteten.

#### 5.2.4 Negativa aspekter kring GDPR

*Nedanstående har inte en direkt koppling till vår forskningsfråga men vi finner ändå detta intressant att rapportera.*

Samtliga Informanter är överens om att GDPR är en försvårande och kostsam lag att förhålla sig till, dock anser Informant 4 att detta mer beror på okunskap hos vissa organisationer som vidtagit alldeles för drastiska åtgärder. Informant 2 anser att lagen kommer leda till höga omkostnader för företag och att i vissa fall kommer ge en omöjlig administrativ situation för dem.

*“Det kan kosta hur mycket pengar som helst, och bolag kan inte behöva gå i konkurs bara för att någon begär att bli av med all sin data.” (i2, 5)*

Detta stöds av Informant 1 och 3. Informant 3 anser även att Big data-analyser kommer bli mer kostsamma och tidskrävande. Informant 1 anser också att befolkningen mer är rädd för

övervakning från staten än av företag. Och att detta kommer fortsätta även med GDPR eftersom det finns undantag för brottsförebyggande eller brottsbekämpande organisationer i lagen. (detta är svårt för oss att utläsa i lagen och har svårt att ta ställning till, att de har undantag är givet men frågan är till vilken utsträckning) Informant 1 tror även att lagen mest kommer påverka små företag som inte har den juridiska kraft som större företag har, och att de enkelt kommer navigera sig igenom den ändå.

*“GDPR kommer att sätta grus i maskineriet för Europa. Det är en katastrof-lagstiftning för Europas skull.” (i1, 64)*

Informant 1 fortsätter med att världsledande IT-koncept som Spotify och liknande förmodligen inte kommer uppstå i Europa i framtiden. Hen tror att tröskeln för hantering av kundinformation är för hög och alltför mycket investeringar kommer att krävas innan företagen är lagligt körbara. Och att trots den höga säkerheten som finns idag är den största faran fortfarande internt hos företagen, i form av informationsläckor eller liknande.

## 6 Slutsats

*I detta kapitel presenterar vi de viktigaste upptäckterna baserat på vår empiri och litteraturgenomgång. Vi ämnar besvara vår forskningsfråga och uppfylla uppsatsens syfte. Vi presenterar de förändringar vi funnit som Big data-analyser står inför samt hur den personliga integriteten påverkas.*

---

Vår studie syftar på att besvara vår forskningsfråga som är:

*Hur inverkar större krav på personlig integritet Big data-analys ur ett organisationsperspektiv?*

Vi har genom vår studie funnit att Big data-analyser är en svårdefinierad framgångsfaktor med hög betydelse för flera organisationer idag. I och med införandet av den nya GDPR-lagen kan vi finna utifrån vår undersökning att Big data-analyser kommer genomgå förändringar. Den nya lagen omfattar många datatyper, även personuppgifter i ostrukturerad data och därmed kommer inkludera Big data. Vi har funnit att en av de förändringar Big data-analyser står inför är att det kommer krävas högre krav på rättsligt grund för behandlingen. Detta innebär att färre analyser kommer att genomföras än tidigare då all behandling måste ha stöd i lagen, och insamlad data ej får användas i annat än det ursprungliga syftet. Flexibiliteten kommer att reduceras och det kommer även bli svårare i vissa fall att få igenom beslut för att utföra analyser.

Vi har även funnit att grupp-analyser och anonymisering av data kommer att öka i samband med Big data-analys. Det är problematiskt att se detta som en problemfri lösning. I och med alla mängder av data och unika mönster som Big data-teknologin kan finna, ser vi en problematik med att lösa många kommande problem vid Big data-analys genom anonymisering. Då det påvisats att anonymiserad data kan av-anonymiseras.

I vår studie har det även framkommit åtskilliga sätt den personliga integriteten kan kränkas på i samband med Big data-analys. Eftersom det kan vara otydligt för vad som är en kränkning på den personliga integriteten finns en risk för att detta ses utifrån ett organisationsperspektiv, där den egna vinningen kan väga tyngre än den personliga integriteten. Vi uppfattar dock att GDPR-lagen är ett förstärkande skydd för den personliga integriteten men att högre kunskap inom organisationer, tydligare etiska ramverk och ett starkt förhållningssätt till lagstiftningen kommer krävas för att kränkningar på den personliga integriteten skall upphöra.

Konsekvenserna av GDPR-lagen är att den personliga integriteten kommer att få högre skydd, men till priset av att Big data-analyser blir svårare att genomföra, dyrare samt att vissa kommer att upphöra som hade kunnat gynna kunden och individen i slutändan. Dock med all data som användare lämnar ut och genereras från deras handlingar online så kommer användare fortfarande behöva lägga en stor tillit att organisationer använder dem till goda syften.

## 7 Förslag till vidare forskning

Eftersom denna uppsats bygger på en lag som ännu inte kommit i kraft så kan det vara av intresse att utvärdera om ett par år hur Big data-analyser har förändrats och påverkats av GDPR. Vi finner även ett intresse att undersöka hur GDPR kommit att påverka Big data-analyser ur ett individperspektiv, eftersom vi i vår slutsats kom fram till att GDPR kommer att försvåra genomförandet av Big data-analyser för organisationer, som i vissa fall kan anses vara positiv för kunden och individen i slutändan. Vidare finner vi det intressant att forskning görs för att få en uppfattning på kostnader för organisationer i omställning till den nya lagen. Vi anser också att vidare forskning bör göras på organisationernas egna definitioner av vad som anses kränkande på den personliga integriteten och jämföra med användares uppfattning.

## Referenser

### Tryckta referenser:

- Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *An International Journal on Information Fusion*, 28, 45–59.
- Berner, M., Graupner, E., & Maedche, A. (2014). The Information Panopticon in the Big Data Era. *Journal of Organization Design*, 3(1), 14.
- Bryman, A., & Nilsson, B. (2011). *Samhällsvetenskapliga metoder*.
- Charan, A. (2015). *Marketing Analytics: A Practitioner's Guide to Marketing Analytics and Research Methods*. World Scientific Publishing Company.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.
- Jacobsen, D. I. (2002). *Vad, hur och varför?: om metodval i företagsekonomi och andra samhällsvetenskapliga ämnen*. Studentlitteratur.
- Kuner, C., Cate, F. H., Millard, C., & Svantesson, D. J. B. (2012). The challenge of “big data” for data protection. *International Data Privacy Law*, 2(2), 47–49.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 90(10), 60–66, 68, 128.
- Michael, K., & Miller, K. W. (2013). Big Data: New Opportunities and New Challenges. *Computer*, 46(6), 22–24
- Narayanan, A., & Shmatikov, V. (2008). Robust De-anonymization of Large Sparse Datasets. In 2008 IEEE Symposium on Security and Privacy (sp 2008).
- Raghunathan, B. (2013). *The Complete Book of Data Anonymization*.
- Richards, N. M., & King, J. (2014). *Big Data Ethics*.
- Rosyad, A. (2016). *BIG DATA: Edisi Onno W Purbo*. Springer India.
- Sharda, R., Delen, D., & Turban, E. (2014). *Business Intelligence: A Managerial Perspective on Analytics*. Pearson.
- Supriyadi, D. (2017). *Personal and Non-Personal Data in the Context of Big Data*.
- Tene, P. (2012). Privacy in the Age of Big Data: A Time for Big Decisions Symposium Issue. *Stanford Law Review*.
- Terzi, D. S., Terzi, R., & Sagioglu, S. (2015). A survey on security and privacy issues in big data. In 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST) (pp. 202–207).

Vickers, B. (1992). Francis Bacon and the Progress of Knowledge. *Journal of the History of Ideas*, 53(3), 495.

### **Rapporter:**

Deloitte. (2013). *The Analytics Advantage*.

IBE. (2016). *Business Ethics and Big Data*. Institute of business ethics.

ICO. (2012). *Anonymisation: managing data protection risk*.

ICO. (2017). *Big data, artificial intelligence, machine learning and data protection*. ICO.

IDG. (2016). *Data & Analytics Research*.

Integritetskommittén. (2016). *Hur står det till med den personliga integriteten? - En kartläggning av Integritetskommittén*. Integritetskommittén, Regeringskansliet.

KPMG. (2016). *Crossing the line: Staying on the right side of consumer privacy*.

Russom, P. (2011). *TWDI Best Practices Report: Big Data Analytics*. Renton, Washington: TWDI - The Data Warehousing Institute.

### **Artiklar**

Akenine, D. (2017, June 1). "Välkommen GDPR! Det är dags att tala om fördelarna med den nya dataskyddslagen."

<https://computersweden.idg.se/2.2683/1.683614/gdpr-valkommen-debatt-2018-04-28>

Columbus, L. (2017, December 25). *53% Of Companies Are Adopting Big Data Analytics*. Forbes.

<https://www.forbes.com/sites/louiscolombus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/> 2018-04-06

Danielsson, L., Nilsson, S., & Lindström, K. (2018, February 14). *GDPR: Här är allt du behöver veta om EU:s nya dataskyddsregler*.

<https://cio.idg.se/2.1782/1.674864/gdpr-konsekvenser-utvecklare> 2018-04-03

Duhigg, C. (2012). *How Companies Learn Your Secrets*. The New York Times.

<https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html> 2018-04-07

Fjellner, C. (2018, March 2). "GDPR kan öka företagens it-kostnader med 40 procent".

<https://www.nyteknik.se/opinion/gdpr-kan-oka-foretagens-it-kostnader-med-40-procent-6902062> 2018-05-03

Grassegger, H., & Krogerus, M. (2017, January 28). *The Data That Turned the World Upside Down*. [https://motherboard.vice.com/en\\_us/article/mg9vvn/how-our-likes-helped-trump-win](https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win) 2018-04-04

Harford, T. (2014, March 28). *Big data: are we making a big mistake?*

<https://www.ft.com/content/21a6e7d8-b479-11e3-a09a-00144feabdc0> 2018-04-11



Markman, J. (2016, August 8). Big Data And The 2016 Election.

<https://www.forbes.com/sites/jonmarkman/2016/08/08/big-data-and-the-2016-election/>  
2018-04-17

Marr, B. (2017, March 14). The Complete Beginner's Guide To Big Data In 2017.

<https://www.forbes.com/sites/bernardmarr/2017/03/14/the-complete-beginners-guide-to-big-data-in-2017/> 2018-04-04

Minitzer, F. (2018, April 9). Facebook's Surprising List Of Security "Fixes"

<https://www.forbes.com/sites/frankminiter/2018/04/09/facebooks-surprising-list-of-security-fixes/> 2018-04-17

Schroepfer, M. (2018). An Update on Our Plans to Restrict Data Access on Facebook | Facebook Newsroom.

<https://newsroom.fb.com/news/2018/04/restricting-data-access/> 2018-05-07

Zoldan, A. (2013, May 10). More Data, More Problems: Is Big Data Always Right?

<https://www.wired.com/insights/2013/05/more-data-more-problems-is-big-data-always-right/>  
2018-04-11

### **Hemsidor:**

Commissioner, D. P. (2016, September 9). Anonymisation and pseudonymisation - Data Protection Commissioner - Ireland.

<https://www.dataprotection.ie/docs/Anonymisation-and-pseudonymisation/1594.htm> 2018-04-11

Datainspektionen. (2007a, August 28). Personuppgiftslagen - Datainspektionen.

<http://www.datainspektionen.se/lagar-och-regler/personuppgiftslagen/> 2018-04-03

Datainspektionen. (2007b, October 16). Ordlista - Datainspektionen.

<http://www.datainspektionen.se/ordlista/> 2018-04-11

Datainspektionen. (2008, February 4). Vad menas med en kränkning av den personliga integriteten? - Datainspektionen.

<http://www.datainspektionen.se/fragor-och-svar/personuppgiftslagen/vad-menas-med-en-krankning-av-den-personliga-integriteten1/> 2018-05-07

Datainspektionen. (2017a, January 26). Dataskyddsförordningens syfte - Datainspektionen.

<http://www.datainspektionen.se/dataskyddsreformen/dataskyddsförordningen/introduktion-till-dataskyddsförordningen/dataskyddsförordningens-syfte/> 2018-04-03

Datainspektionen. (2017b, January 26). Rättslig grund för personuppgiftsbehandling - Datainspektionen.

<http://www.datainspektionen.se/dataskyddsreformen/dataskyddsförordningen/rattslig-grund-for-personuppgiftsbehandling/> 2018-04-27

- Datainspektionen. (2017c, January 26). Samma regler för alla uppgifter - Missbruksregeln försvinner - Datainspektionen.  
<http://www.datainspektionen.se/dataskyddsreformen/dataskyddsforordningen/missbruksregeln-upphor/> 2018-04-23
- Datainspektionen. (2017d, March 1). Dataskyddsförordningen - texten i sin helhet - Datainspektionen.  
<http://www.datainspektionen.se/dataskyddsreformen/dataskyddsforordningen/forordningstexten/> 2018-04-03
- Datainspektionen. (2017e, March 3). Dataskyddsförordningens beaktandesatser - Skäl - Datainspektionen.  
<http://www.datainspektionen.se/dataskyddsreformen/dataskyddsforordningen/forordningstexten/beaktandesatserna/> 2018-04-11
- Datainspektionen. (2017f, March 7). Tillämpningsområde - Datainspektionen.  
<http://www.datainspektionen.se/dataskyddsreformen/dataskyddsforordningen/tillampningsomrade/> 2018-04-03
- Datainspektionen. (2017g, March 8). Rätt till radering - Datainspektionen.  
<http://www.datainspektionen.se/dataskyddsreformen/dataskyddsforordningen/de-registrerades-rattigheter/ratt-till-radering/> 2018-04-24
- Datainspektionen. (2017h, March 9). Personuppgiftsansvarig - Datainspektionen.  
<http://www.datainspektionen.se/dataskyddsreformen/dataskyddsforordningen/skyldigheter-for-de-som-behandlar-personuppgifter/personuppgiftsansvarig/> 2018-04-03
- Datainspektionen. (2017i, March 9). Personuppgiftsbiträde - Datainspektionen.  
<http://www.datainspektionen.se/dataskyddsreformen/dataskyddsforordningen/skyldigheter-for-de-som-behandlar-personuppgifter/personuppgiftsbitrade-och-bitradesavtal/> 2018-04-03
- Datainspektionen. (2018, March 9). Publicera bilder, filmer och ljud på internet - Datainspektionen.  
<http://www.datainspektionen.se/dataskyddsreformen/dataskyddsforordningen/missbruksregeln-upphor/publicera-bilder-filmer-och-ljud/> 2018-05-8
- Gartner. (2012, May 25). What Is Big Data? - Gartner IT Glossary - Big Data.  
<https://www.gartner.com/it-glossary/big-data> 2018-04-24
- IDG. (2018, April 16). IDG:s ordlista, 2017.  
<https://it-ord.idg.se/ord/pseudonymisering/> 2018-04-11

## 8 Bilagor

### 8.1 Intervjuguide

#### Introduktion:

1. Är det okej att vi spelar in denna intervju?
2. Kan vi nämna ert och/eller företagets namn i uppsatsen?
3. Kan du berätta lite om dig själv och din roll på företaget? Hur länge har du jobbat med Big data?
4. Hur många projekt eller år har företaget jobbat med Big data?

#### Big Data & Användning av Big data:

1. Hur skulle du kort definiera begreppet Big data (eller hur definierar företaget begreppet)?
2. Vilka ser du som de största möjligheterna med Big data?
3. Vilka ser du som de största utmaningarna med Big data?
4. Vilken betydelse har Big data idag?
5. Hur använder ni Big data idag?
6. Vilken typ av analyser genomför ni eller era kunder? (text, video, sociala medier etc)
7. Vilka utmaningar ser ni vid analys av stora datamängder?
8. Har ni alltid ett tydligt syfte med er datainsamling? Hur samlas data in?
9. Hur vet ni att datan är kvalitativ och korrekt?

#### Big data kopplat till GDPR (Skyddet av personuppgifter)

1. Anser ni att det finns några etiska problem med Big data och Big data analys?
2. Hur anser du att den personliga integriteten kan kränkas i samband med dataanalys?
3. Har ni några etiska riktlinjer i samband med Big data och Big data analys (utöver GDPR)?
4. Hur tror du att GDPR kommer att påverka Big data och analys?
5. Tror du att GDPR kan hindra de analyser av Big data som är kränkande för den personliga integriteten?
6. Hur tror du att användningen av Pseudonymiserad & anonymiserad data kommer användas i samband med Big data efter GDPR?
7. Ser ni något negativt med GDPR? Vilka negativa konsekvenser kommer det att ge?

#### Avslutning:

20. Finns det något annat ni vill berätta?

## 8.2 Transkribering Organisation 1

### Organisation 1.

Intervju 1 – Informant 1 – Organisation 1

Intervjuare – Gustav Stigestad (GS)

Sekreterare – Felix Moberg

#### Start.

1. **GS:** Hur länge har du jobbat med BI?

2. **Informant 1:** Heltid sedan 2010, och ytterligare deltid i tre år till.

3. **GS:** Vet du hur många år ert företag har jobbat med BI eller Big data?

4. **Informant 1:** I Sverige sedan 2007.

5. **GS:** Hur skulle du definiera big data?

6. **Informant 1:** Det är inte en helt självklar fråga. Det som används i marknadsföring som big data är inte big data för mig. För big data för mig är när data volymerna och strukturen inte riktigt stämmer in i vanliga transaktionsbaserade datasystem. Ett exempel som vi kallar för big data: Vi har en kund som har lägenheter som de mäter temperatur i och sen använder dem det för att styra sina värmesystem, men inte en hård koppling tillbaka utan tittar på siffrorna och sen gör de en bedömning och styr värmesystemen utifrån det. Och det kallar vi Big data ur ett marknadsföringsperspektiv och det är tveksamt om det är big data även om datavolymerna är stora. Det är ganska strukturerad information och finns i några olika databaser eller molntjänst. Så ja det är ett gränsfall där, men det betraktas som big data när man tittar på det, men jag är inte så säker om jag skulle kalla det big data. För jag tycker big data är ytterligare ett snäpp av volymer och som är ännu mer ostrukturerade till exempel klick på en webbtjänst eller hur saker och ting hänger ihop. Lite mer ostrukturerad information.

7. **GS:** Vilka ser du som de största möjligheterna som finns med big data?

8. **Informant 1:** De största möjligheterna är nog egentligen när man har en frågeställning och får svar på den i sin data. Och utan frågeställning är Big data inte värt så mycket. Sen ska man inte bortse från att man inte har alla frågor på bordet när man gör datainsamlingen.

9. **GS:** Hur menar du då?

10. **Informant 1:** Ett exempel är att jag har hållit på med Geocatching och gps i många år och jag har alltid sparat mina gps tracks från det, från varje tur jag varit ute på, plötsligt för några år sedan så gick det att börja geotagga bilderna i efterhand. Så alla foton jag hade tagit tidigare kunde jag geotagga med hjälp av mina gamla GPS spår. Det gick inte när jag samlade in spåren, eller det kanske gick men jag visste inte att man kunde göra det för det var så länge sedan. Och det är en sån sak, det är en aha upplevelse, som jag får från mina ostrukturerade GPS spår som finns i tusentals GPX filer och som finns i en disk och låtit dom ligga där och plötsligt kunde jag använda dem för att geotagga bilder i efterhand för att veta precis att denna

bilden är tagen där. Och det är precis detta som Big data ger möjligheter till sånt som man inte har tänkt på när man har samlat in informationen. Men egentligen tillhör det undantaget för den mesta informationen som man samlar in, har man inte en frågeställning så kommer man inte samla in tillräckligt hög detaljgrad på det.

**11. GS:** Har ni alltid ett syfte när ni samlar in data?

**12. Informant 1:** Ja det kan man säga. Man kan säga att kunderna är så pass strikta med hur mycket pengar de vill lägga på att utveckla saker och ting att det börjar nästan alltid med ett syfte. Sen får man också ta hänsyn till att det ofta finns en konflikt mellan IT-avdelningen och verksamheten hos företagen. Där det är IT-avdelningen som sitter på servrar, resurser, datakapacitet och regelverk hur saker och ting får lagras medan verksamheten har andra behov och det är IT-avdelningen som kommer ta investeringsbeslutet hur mycket information som får lagras och liknande. Så den konflikten är nästan alltid verkligen i företagen idag. Och det gör ju att IT bromsar, de vill inte ha information som de ser kommer kosta pengar.

**13. GS:** Hur mäter ni att informationen är korrekt och användbar?

**14. Informant 1:** Det kan man nästan säga är ett kapitel för sig, men där är ingen skillnad från traditionell BI lösning.

Det handlar om att man gör stickprov på transaktionsnivå eller på rad nivå. Först försöker man få alla sånna som man kan hitta i stickprov rätt och när man får rätt på dem så går man vidare och vänder på frågeställningen och försöker titta på aggregerad information, medelvärde, snitt, antal värden osv. Och när de stämmer så börjar man titta på många poster, eller många saker, så det är liksom en expanderande verifieringsprocess och där är ingen skillnad med vilken typ av information man jobbar med.

Men sen har man problemet när man exempelvis jobbar med temperaturer i lägenheter, vad är rätt? Du kan inte veta om termometern visar rätt eller inte. Om det bor fem personer i lägenheter eller en person eller om de alltid har fönstret öppet det ser du inte direkt på siffrorna utan du kan se det på lite mönster. Till exempel man ser en kurva på hur temperaturen ändras i lägenheter så kan man lätt se att denna person, öppnar vädringsfönster när den åker hemifrån och så står det öppet hela dan och så stängs det när personen kommer hem. Så den typen av saker ser man ju, men det går ju inte egentligen att säga om det är rätt på dem siffrorna eller inte. Så det som det här företaget har gjort då är att de har ett antal mobila mätare, små mätare som har speciella koder som de kan de lägga ut i en lägenhet så mäter dem temperatur och fuktighet tillsammans med den fasta mätaren som sitter installerad. Och då kan de jämföra siffrorna. Och då måste de ha ett verktyg för att titta på informationen så de kan se det på ett enkelt sätt på driftskontoret utan att behöva knappa i en massa databaser. Och på så sätt kan de jämföra och se ifall det är fel på någon mätare.

**15. GS:** Men vad är de största utmaningarna med Big data?

**16. Informant 1:** Jag skulle nog säga att de största utmaningarna är ekonomiskt. Att se det ekonomiska värdet med datavolymerna.

**17. GS:** Det är en kostsam process med Big data?

**18. Informant 1:** Lagringsmässigt är inte datavolymer dyrt men investeringsbeslutet är svårt att ta för att lagra och analysera information. Och det kommer återigen att man behöver ha ett

syfte med den informationen man samlar in. Man har en frågeställning som gör att man vill samla in det.

**19. GS:** Hur samlar ni in data?

**20. Informant 1:** Det är väldigt varierande. Jag kommer oftast in i ett skede när informationen redan finns på något media, den finns på disk av något slag. Och det är olika beroende på vad det är för typ av lösning. Om det är en exempelvis en butiksräknare som räknar på hur många som kommer in genom dörren så finns det speciella lösningar för det som levererar exempelvis XML filmer in till nåt system som sedan någon läser in i någon databas och då kanske det bli tillgängligt till analys. Så det går oftast i flera steg.

**21. GS:** Men är det en sensor då som räknar?

**22. Informant 1:** I butiksräknaren tror jag att det finns många varianter, det finns bland annat räknare som räknar antalet personer som passerar på gatan utanför, för då vill man kunna se hur stor andel som går in. Och antalet som passerar utanför är en bra indikator på hur bra butiksläge det är. Det är en parameter, sen så räknar man hur många som går igenom dörren. Sen kombinerar man ofta det med verklig information i kassan, dvs hur många avslut har du gjort i kassan och då får man något som i butiksvärlden kallas "Commercial rate" dvs hur många av de som gick in i butiken fick du ett kassa avslut på. Och det är jätteviktiga nyckeltal som används i butiksvärlden. Och då använder man olika källor, man kan ha en räknare från ett företag som räknar på hur många som passerar utanför butiken, en räknare från ett annat företag som räknar på hur många som passerar dörren in till butiken och en räknare i kassan, som är mer riktig information i affärssystemet. Och då ska man försöka kombinera ihop de här, man försöker gå ner på detaljnivå och när man går ner på timmar och klockslag bryter det ofta ihop för att man inte har rätt tidszon eller har upplösning på siffrorna som gör att man inte kan gå ner på lägre nivå.

**23. GS:** Det ska vara samma filformat?

**24. Informant 1:** Ja precis och de som tillhandahåller insamlingen är nyckelspelare att få informationen i ett format som sedan går att använda vidare.

**25. GS:** Köps data in från tredjepart?

**26. Informant 1:** Det gör man nog också en del, om man kan.

**27. GS:** Vilken betydelse anser du att Big data har idag?

**28. Informant 1:** Jag skulle nog säga att den inte har så stor betydelse som den skulle kunna ha.

**29. GS:** Det finns mycket större potential?

**30. Informant 1:** Mycket större, framförallt i produktion och kvalitetsmätning av produkter skulle kunna bli mycket bättre. Ett exempel, ett företag som tillverkat en apparat som de släpper ut på konsumentmarknaden, som är uppkopplad tillbaka så man kan veta hur den används. Det finns ingen information om serienummer och batcher och sånt på själva devicen

i den informationen, så de kan inte veta vilken av alla sina devicer som sitter ute hos en viss kund. Det gör att de har ingen möjlighet att göra en kvalitetsuppföljning, skulle de ha en kvalitetsproblem i sin tillverkning så vet de inte att det är den här serien som är drabbad. Det handlar mer om att tänka igenom sin strategi från början så att man utnyttjar den potentialen som finns. Det hade varit så himla lätt att de hade serienummer på själva tillverkade enheten och sen skulle dem när de gjorde tillverkningsprocessen haft serienummer eller ett batch nummer på alla ingående komponenter, så att de visste precis vad som satt i den hära. Det hade bara varit en fråga att hålla register över den, och sen kunna koppla den informationen så den går tillbaka och det gör dem inte och det är nog inte många som är där. Det är typiskt en sådan sak som man skulle kunna använda det för kvalitetsuppföljning att se hur sakerna fungerar och när de inte fungerar kunna identifiera dem sakerna. En annan sak som jag var inblandad i var ett företag som har maskiner och där kan man mäta vibrationerna på maskinerna i tillverkningen, när de vibrationerna ökar så börjar maskinerna behöva servas eller bli dåliga. Datan av vibrationerna de lagrades, men de gör ingenting utav den. De har inte riktigt förmåga att förädla den informationen på rätt sätt så de kunde fånga det i tid och göra arbete.

**31. GS:** Men hur använder ni big data idag?

**32. Informant 1:** Vi gör ju inte för oss själva, utan vi gör det bara för våra kunders räkning. Och det är ju lite olika scenarier. Vi kan ta lägenhetsexemplet. Där har vi hjälpt kunden en lösning att de kan se temperaturerna för varje enskild lägenhet, visualiserat i hur huset har det fördelat. Och på det viset kan de lättare få återkoppling för hur värmesystemen är inställda till hur det ser ut i lägenheten.

Så tidigare kunde de kanske jobba sig igenom hela beståndet på ett eller ett par år, alltså att de justerade alla värmesystem sen fick de ut och mäta och koppla tillbaka stoppa in det i Excel och räkna ut normaltemperatur, de fick vänta tills det var -5 eller +10 grader osv, för att se att det stämmer för hela kedjan. Nu får de den informationen tillbaka redan nästa dag, så nu har de kunnat jobba hela igenom hela beståndet av värmesystemet för det ändras och kopplas om och så vidare så de stämmer inte riktigt. Så de har kunnat jobba genom hela det systemet på bara någon månad och på det viset har de kunnat komma närmre sin måltemperatur i alla lägenheterna och därmed har de kunnat sänka snitt temperaturen för oftast är det så att de får hålla lite för varmt istället. Och har de en måltemperatur på till exempel 21 grader så är det onödigt om det är 24. Och de har de kunnat göra och därmed sparat jättemycket pengar på att kunna få denna återkoppling snabbare och kunna ställa in det under samma säsong.

**33. GS:** Har ni några projekt överhuvudtaget som har med personuppgifter, analysera kunder och så?

**34. Informant 1:** Nja, det skulle jag inte säga att jag har aktivt just nu. Jag har haft ett projekt på ett företag som sålde på postorder och hade ett kundregister. Och där hade de ett stort kundregister på några hundratusen medlemmar som de gjorde utskick till. Och där skickade de ut olika typer av kampanjer på dem när de fyllde år till exempel eller standard utskick flera gånger om året och så vidare. Och sen hade de så många olika kampanjer som de skickade ut så när de här kunderna kom in i butik eller i sin e-butik och använde kampanjkoden som fanns på kampanjen så kunde de spåra tillbaks till precis vilken kampanj det var som hade gett resultat. Det använde dem mycket, men man kan säga att den integritetsbiten där som säger hur varje enskild person handlar, där var de ganska så basala. De jobbade med kön och ålder. Eller jag är osäker om de jobbade med kön men de jobbade med ålder. Så de såg att denna personen exempelvis är mellan 40 och 50 och då är det den här kampanjen som passar till

dem här, och den här personen är mellan 50-60 då är det den här kampanjen. Så de hade inte mer fragmentering och uppföljning egentligen. Den uppföljningen där för att se vad som skall ut till en viss användargrupp, det gjorde dem förmodligen lite analyser på, men inte på ett strukturerat sätt. Utan de gjorde mer stickprov. Fungerade den här kampanjen? Den informationen gick tillbaks till marknadsavdelningen som avgjorde om kampanjen gick bra eller dåligt och sen satt det i fingertopparna på dem i marknadsavdelningen hur de skulle lägga upp kampanjerna, snarare än ett databehandlingsperspektiv. För det enda databehandlingen gjorde var att se att den här kampanjen resulterade i så och så många köp inom en viss tid i det att kampanjen gick. Så att dom hade ett stort kundregister men de använde inte det för en individanpassad marknadsföring mer än för en ålderskategorisering var det som styrde dem.

**35. GS:** Anser du att det finns några etiska problem med Big data och Big data analys?

**36. Informant 1:** Ja, men det måste man absolut säga. Men generellt sätt så är problemen inte så stora som framställs i media anser jag. Därför att människor som interagerar med företag gör det för att de är intresserade av företagets produkter och då tror jag att det finns en allmän acceptans att man blir kartlagd av företaget för att företaget ska hitta saker och ting till en som är bättre, som bättre passar, så man slipper att se dåliga annonser till exempel. Och den medvetenheten tror jag är ganska stor hos folk. Sen finns det ju andra exempel där man handlar med informationen om människor då plötsligt kommer man in på ett annat område, när man delar med sig av informationen till någon annan för att dem ska kunna göra något annat med det då har man gjort en datainsamling av en individ för ett syfte och sen så har man ur ett egoistiskt perspektiv delat med sig det till någon annan som vill använda det för ett annat syfte. Det tror jag folk har betydligt svårare för att acceptera. Så länge det används för det som liksom är självklart är inhämtat för så tror jag det är mindre känsligt. Men samtidigt som den här lägenhetsinformationen med temperaturer och fuktighet i lägenheter, där ser man till exempel om personer duschar hela tiden, det kan på ett vis bli en integritetsfråga. Men det syns jättetydligt när man mäter fuktigheten i lägenheten. Någon som duschar 5 gånger om dagen eller 10 gånger eller något sånt, eller vattenförbrukning samma sak, om de går på toa hela tiden och den typen. Och det är klar det är information som inte ska spridas hur som helst, den måste man hantera på ett sätt som även om en så enkel information om vattenförbrukning eller temperaturer och fuktighet i en lägenhet känns som helt ofarlig så är den inte det. All information bör man hantera med varsamhet.

**37. GS:** Ser du några andra exempel där den personliga integriteten kan kränkas med Big data eller Big data analys?

**38. Informant 1:** Jag har inte stött på den typen av frågor egentligen mer än en liten slarvig syn hos företagen med personlig information. Som då skulle kunna hamna i fel händer, beroende på att den är slarvigt hanterad. Inom företagen har man en hög lojalitet tröskel mot sitt företag och man tror och litar på sina kollegor. Men det finns ju utomstående faktorer som kan vilja göra intrång och komma åt informationen och man är nog lite naiva för den frågeställningen och då kan de hamna i orätta händer. Sen ett annat exempel har jag jobbat med... Nu ska vi se om jag kan beskriva det... Det handlar om en lösning som motsvarar Tellstick och de bitarna, det vill säga att man kan ha byggnadsövervakning med fjärrkontroll till lampor och så vidare som man kan styra med telefonen.

**39. GS:** Ja just



40. **Informant 1:** Den typen av lösning som är en webbaserad tjänst och där finns känslig information i den att det dem jobbar med är bara konsumentprodukter men i samma system finns information om larm och då blir det plötsligt superkänsligt, för då vet man om det är larmat hemma hos någon. Men den informationen har de varit noga med att den inte ska kunna spridas, men sen har de då vetskap om GDPR så de vet om att efter maj månad så kan de inte ha den här typen av information. Men de har varit tvungna att ta med sig mycket information som är individknuten eller som går att spåra tillbaka bara för att kunna felsöka systemet. Och det har de kanske inte tänkt på redan från början att de behöver bygga in en felsökning som är av anonymiserad för annars kan man inte följa upp det på ett vettigt sätt. Och det är en sådan sak som jag tror att många kommer drabbas av nu, ”nu stänger vi ner det informationsflödet, nu kan vi inte felsöka det längre”.

41. **GS:** Men hur är det. Tror du att företag kommer att anonymisera personerna eller pseudonymisera personerna och sånt nu när GDPR börjar gälla?

42. **Informant 1:** Jag tror att de flesta kommer nog att stryka all information i BI-ledet som är individspårbar, som adresser, namn, personnummer och telefonnummer och liknande att man inte tar med det i sina BI-system. Och generellt sätt är det inte mycket sådan information i BI-lösningar heller, man jobbar mer med siffror och man jobbar med företagskunder kanske. De jobbar man med namnen på men inte vilka människor som sitter bakom. De är sällan jätteintressanta. De som kanske har behov av det är butikskedjor som verkligen vill göra en riktad marknadskampanj mot ett visst kundsegment som de redan har relation till, men de har alla de här processerna för GDPR redan, eller god förståelse att de måste ha dem. För de vet om att de har känslig kundinformation, medan övriga företag känner att ”det var synd att vi inte kunde följa upp det längre” då bara stryker vi det.

Så att deras mognad på att känna till det är inte så katastrofal heller som det skrivs om i tidningarna, de vet om vad det handlar om. Vi har inte fått någonstans om att de har börjat gå igenom sina BI-system och då är det 1-1,5 månad kvar. Så de kommer förmodligen hända här i början i början av juni att ”oj vi måste ta bort lite information” och då handla det bara om att ta bort informationen. Blanka bort informationen och ta bort fält.

43. **GS:** Så anonymisering och den biten är inget du tror kommer...

44. **Informant 1:** Jo, jag tror att det kommer en del sånt, men jag tror man tar nog ett helhetsgrepp. Så att istället för att göra en anonymisering av informationen så ser man bara till att där finns ingenting som är individspårbart och om man behöver gå tillbaks till individspårbar information då gör man förmodligen det i ett annat register.

45. **GS:** Okej.

46. **Informant 1:** Så att man har informationen i ett skyddat register som är klassat så att vem som har accessat det och så vidare. Men det är det som är problemet med BI-lösning att informationen hamnar där och när det väl hamnar där då kan du inte veta vem som accessat den. Och vill du styra över vem som accessar den så får du en mycket svårare lösning, då är det inte ett normalt BI-system längre, då är det någon annan lösning. Så du kan i princip inte ha någon individspårbar information i ett generellt BI-system. För att få lov att ha individspårbar information så behöver man veta om vem som accessar den och vem som har

rätt att accessa den och så fort den ligger ickekrypterad i en databas så finns det alldeles för många som kan komma åt informationen.

**47. GS:** Allright! Har ni några etiska riktlinjer i samband med data analys eller big data?

**48. Informant 1:** Ja, vi har jobbat rätt mycket med det och det handlar ju om att vi ska hjälpa kunderna kan man säga, till att de ska kunna hantera sin information på rätt sätt. Det kom till och med ut någon intern presentation här, bara för någon dag sedan, hur vi ska kunna göra det bättre med att hjälpa kunderna och hur vi själva ska agera, vilka papper som ska vara ifyllda när man gör olika roller. Så de bitarna har vi ganska väl på plats tycker jag.

**49. GS:** Okej

**50. Informant 1:** Och generellt sätt har vi sagt till kunderna, åtminstone sen ett år tillbaks, att så fort vi ser individknuten data i systemen så har vi berättat för kunderna att detta måste ni ta bort innan GDPR går igenom.

**51. GS:** Hur tror du GDPR kommer påverka Big data analys?

**52. Informant 1:** Jag tror inte det kommer påverka dataanalysen speciellt mycket. För jag tror att de som gör dataanalys med individanknutna uppgifter, de har redan jobbat med de här frågorna i PUL. Och de som jobbar med annan dataanalys där det inte är individspårbar information där spelar det ingen roll. Så just ur ett dataanalys perspektiv tror jag inte påverkningen blir så stor.

**53. GS:** Okej!

**54. Informant 1:** Det är nog de andra processerna runt GDPR som är hysteriskt kostsamma och det är egentligen inte dataanalysen utan datahanteringen som de har i sina affärssystem och så vidare för man måste ha kundinformation för att kunna exekvera orders och så vidare. Hanteringen i dem systemen är mycket värre och dyrare. Men där hade nog mina kollegor sagt något annat. ”Det är det värsta som finns”.

**55. GS:** Men kan du inte berätta lite mer då hur du tror datahanteringen kommer förändras med GDPR?

**56. Informant 1:** Jag skulle ju säga om du tittar på hur data hanteras idag så är det många system som inte har förmågan att glömma kunder. Det finns ingen tydlig struktur i systemen för att kunna få en samlad bild över all kundinformation. Det finns en otydlighet i vilken information du måste spara i förhållande till skatteverket på fakturaunderlag och vilken information du får lov att ta bort.

Det är ju jätte centrala frågor som man måste ta med och samma sak med ett garantiåtagande, hur länge får du spara information om kunden om du har ett garantiåtagande? Det kan hända att du måste ha tillräckligt med saker på ditt kvitto för att kunden ska kunna göra en komplett garanti med serienummer och sånt för att du har varit tvungen att ta bort informationen för kunden har sagt ”ta bort mig ur systemet”.

**57. GS:** Just det!

58. **Informant 1:** Så alla de frågorna är mycket svårare att reda ut och få en klarhet i. Och generellt sett tror jag inte kundklubbarna har kommit fram till att göra det men de kommer att behöva nya godkännande avtal från alla användarna som är med i kundklubbar och de avtalen kommer att säga att man får lov att spara informationen och så länge man är med i kundklubben och när man går ur kundklubben så försvinner man ur registret. Men det får också till följd om man har gått ur och man har glömt en så kommer man inte kunna få hjälp av förmåner, man kommer inte kunna titta på gamla ordrar och dylikt.

59. **GS:** För att man blir helt borttagen då?

60. **Informant 1:** Man blir helt borttagen. Eller man blir förmodligen inte helt borttagen för det går inte att göra i affärssystem men man blir anonymiserad att man inte kan hitta den kunden igen men det kundnumret kanske ligger kvar men det finns ingen koppling till mig. Mitt namn är borta och telefonnummer och så vidare.

Men om man tittar på gamla underlag. Det finns många system tror jag som inte har sparat den informationen på ett strukturerat sätt utan till exempel på PDF och lagrat in den på disk. Vad gör man med dem? De måste man nästan ta bort, det går inte att söka igenom vilken information som hör till en viss kund.

61. **GS:** Nää och man kan inte stryka i PDFerna och så.

62. **Informant 1:** Nej. Så det är mycket såna frågor som kostar mycket pengar för företag.

63. **GS:** Men känner du att GDPR liksom kommer sätta käppar i hjulen på något vis?

64. **Informant 1:** Definitivt. GDPR kommer att sätta grus i maskineriet för Europa. Det är en katastrof lagstiftning för Europas skull.

65. **GS:** Men om man tänker på just den personliga integriteten och sådant?

66. **Informant 1:** Men det gör ju ingen skillnad

67. **GS:** Den kommer inte göra någon skillnad tror du?

68. **Informant 1:** Nej det tror jag inte kommer göra någon skillnad alls. Utan som den är utformad, regelverket, så ger den ett otroligt litet värde för konsumenten, men den ger en väldigt stort börda för företagen som hanterar informationen och den sätter begränsningar i saker och ting som man kan göra. Så Europa kommer inte att vara IT-ledande som det mångt och mycket har varit. Det kommer inte att komma några nya stora företag som har uppstått i Sverige som Spotify och andra. Det kommer inte komma sådana företag i Europa i den utsträckningen därför att tröskeln för att våga göra någonting när man hanterar information om kunder kommer vara så stor att man behöver stora investerare bakom sig. Så de små startupsen som sen köps upp och blir något bättre och som växer de kommer att strypas i stor utsträckning redan tidigt på grund av detta.

Jag jobbade med en tysk websida privat för ett par år sedan, eller jag hade en svensk website och kompisar i Tyskland som körde en likadan website i Tyskland och de kunde inte köra den i privat regi utan de var tvungna att köra den under scouterna i Tyskland, så scouterna var huvudman för hemsidan för att de kunde inte driva en website privat som innehöll

personuppgifter, även om det bara innehöll namn och epost adress ingenting annat. Men av juridiska skäl kunde de inte köra den, även fast det bara var en hobby website. Och det är den lagstiftningen vi har fått till Sverige nu. Så det är en katastroflagstiftning för den Europeiska konkurrenskraften inom IT-området.

**69. GS:** Flera hävdar på att det är den personliga integriteten som är det största etiska bekymret med Big data, vad skulle annars vara bättre att skydda den personliga integriteten än GDPR, ser du några andra lösningar?

**70. Informant 1:** Nej men på sätt och vis kan jag säga att någon sorts lagstiftning behövs. Men det jag kan tycka med GDPR är att den går alldeles för långt till att sätta strikta regelverk för hur företagen ska agera och stora bötesbelopp och som inte är tillräckligt tydligt definierade om vad som skall göras, som är tillräckligt kraftiga för att göra att saker och ting inte kan utvecklas, så det är en väldigt restriktiv lagstiftning som ger väldigt lite nytta för konsumenten i förhållande till restriktionerna som lagstiftningen ger. Det är väl min syn på GDPR.

**71. GS:** Ja men vill du utveckla mer med katastrof för Europa?

**72. Informant 1:** Som jag ser det så lägger den en ganska kraftig sordin på utvecklingsarbete inom IT-området i Europa, för allting kommer bli dyrare. Det blir liksom inte lika lätt att göra någonting nytt. Så fort man stiger utanför den befintliga ramen där man redan finns, eller där andra redan finns, för att hitta på någon ny tjänst så kommer man att fastna i ett juridiskt träsk. Vågar vi göra detta? Eller vad behöver vi? Behöver vi ha ett jättelångt godkännande dokument för alla användarna där de verkligen kryssar i på sju ställen att de godkänner detta och detta. Och håller det sen i domstol? Eller kommer vi att drabbas av miljonböter för att vi har missat en liten kryss ruta? Det kommer inte att finnas den typen av initiativförmåga utan att det finns jättestora investerare bakom.

**73. GS:** Aha okej.

**74. Informant 1:** Om man tittar historiskt, antalet IT företag i Sverige som har uppstått och kommit in i nya branscher framför allt i informationssamhället, per invånare om man jämför det med Tyskland så har Sverige varit fantastiskt ledande, medan Tyskland har varit bra på sina traditionella industrier.

Men det är den tyska lagstiftningen vi har fått i Europa nu.

**75. GS:** Men tror du att GDPR kommer sätta stopp för företag som Cambridge analytica och sådana bolag?

**76. Informant 1:** Det tror jag inte. Det hjälper inte.

**77. GS:** men på vilket sätt kommer de kunna fortsätta då?

**78. Informant 1:** De som vill göra analys som inte är laglig så att säga, de gör ju det ändå.

**79. GS:** De kör ända in i kaklet bara?

**80. Informant 1:** Som exempelvis Ryssland idag, som jag hörde på radion i morse. Då har

MSB (myndigheten för samhällsskydd och beredskap) gått ut att främmande makt samlar in information om de som har interagerat med sensationella nyheter på Facebook. Det vill säga då får de ett jättebra register över personer som har varit lättpåverkade av en sensationell nyhet. Det är hur lätt som helst att samla in, du behöver inte vara Facebook för det, du behöver bara vara Cambridge Analytica. Det är bara att släppa den typen av sensationella nyheter så får du del av den informationen. Namn och profiluppgifter, det är lite jobbigare att samla in om du inte gör någon scriptmotor som gör det åt dig. Så du är inte beroende av att det är Facebook som gör det och säljer informationen till Cambridge analytics. Du kan skapa den informationen och samla in den själv. De bryr sig inte, för de kommer använda detta i ett syfte som dem gör. Faller helt utanför GDPR, så det kommer göra en väldigt liten skillnad för konsumenten.

**81. GS:** Det känns som att vi har fått svar på de flesta vi har kommit fram till så att säga, men jag tycker vi kan fortsätta prata lite.

**82. FM:** Ja det var väldigt intressant detta.

**83. Informant 1:** Det blir lite Big data över det som är ostrukturerat.

**84. GS:** För det är lite den bilden vi har fått också ju mer man forskat på det här att. Vet inte riktigt hur mycket GDPR kommer sätta stopp, för det ni gör och så, det känns inte så jättekränkande. Men det är ju den stora kränkningen som vi är intresserade av, och den vet jag inte hur GDPR kommer att påverka.

**85. Informant 1:** Jag vet inte var man hittar den stora kränkningen heller. På en laglig verksamhet, vem kränker som man kan stoppa med en lagstiftning? För det sker utanför lagens ramverk.

**86. GS:** Så det är frågan hur GDPR kan sätta åt dem där.

**87. Informant 1:** Det finns en obalans också, jag upplever att folk inte speciellt bekymrade om att företag använder information om dem för sin marknadsföring. Folk är mycket mer bekymrade över att staten använder information om dem, för sin kontroll över medborgarna och så fort det handlar om brottsbekämpning eller förebyggande av brott så finns det undantag i GDPR för det. Så där finns liksom ingen skillnad, staten får fortsatt göra precis som de vill så länge de kan hävda att det är förebyggande av brottsbekämpning. Så den biten som folk upplever som mest integritetskränkande idag, den adresserar inte GDPR över huvud taget.

**88. GS:** Videoövervakning och sådana grejer?

**89. Informant 1:** Videoövervakning eller hur skatteverket får titta på Facebook för att se om man har delat någonting där som de kan använda i sin övervakning.

**90. GS:** Men om man ska bena ut dem, vad är de mest negativa konsekvenserna med GDPR?

**91. Informant 1:** Det är administrativ börda för företagen och riskanalysen som gör att man stryper ny idéutveckling. Och det är nästan ny idéutveckling som är det värsta. För då kommer andra länder som inte har så strikta regler att styra utvecklingen. Jag kan inte

motsvarigheten för GDPR i USA och Sydkorea till exempel, men om de länderna jobbar med att inte ha så strikta regler, så kommer de att ta ledningen i informationssamhället framöver. För det är nog Sydkorea, USA och Europa som är de ledande i informationssamhället skulle jag säga idag.

Men jag vet i alla fall att Storbritannien har minst lika stränga regler som EU, så det är inte lönt att flytta dit.

**92. GS:** Haha nä precis! Är det något annat du känner du vill berätta om?

**93. Informant 1:** Nää men min sammanfattning är att jag kan tycka att företagen är lite naiva i sin säkerhetshantering eller hantering av informationen, men de har samtidigt en god insikt i att inte ha någon information som är känslig. Men sen kanske naiva att det finns information som trots allt är känslig.

**94. GS:** Som inte de anser är känslig, men andra kan uppfatta som känslig?

**95. Informant 1:** Mm.

**96. GS:** Men hur tror du det här med samtycke med GDPR, man måste ju ha samtycke, som du sa när man ska kryssa i, hur tror du det kommer att se ut? Alltså när man går in i en webshop kommer det bli att man är tvungen att acceptera, eller hur kommer det fungera i praktiken tror du?

**97. Informant 1:** Det tror jag. Och sen så kommer de att behöva lagra samtycket, datum, klockslag. Sen kommer det nog behövas version på samtycket så man vet vilken version av samtycket som gäller, för det kommer komma nya sådana.

**98. GS:** Men kommer inte det påverka dataanalys och sånt tror du om det blir mindre samtycke att man får hantera deras data?

**99. Informant 1:** Ja fast samtidigt, du kan inte avtala bort allting. Om jag förstått det rätt så får du bara samtycka till det som står beskrivet i samtycket. Så du kan inte ha generellt att "jag får lov att använda informationen" till vilken databehandling jag vill. Så det betyder att det är väldigt svårt att göra en datainsamling för att använda vid senare tillfälle, för att någonting som jag vill gräva fram och ha den individknuten. Jag tror det är väldigt svårt att skriva den typen av samtycke.

**100. GS:** Okej!

**101. Informant 1:** Jag tror inte ens de försöker göra det utan de ser det nog mer som att skapa en profil över användaren, exempelvis vilka varor den här kunden handlar för att kunna ge ett bättre erbjudande. Det är lite självreglerande också, för jag pratade med några som jobbade med butiksdata och i princip är det så att de vet tillräckligt mycket om dig, vad du handlat tidigare och så vidare. Så om du går in i en butik, om du skulle bli identifierad när du går in i butiken, så skulle de kunna ge riktade erbjudanden till dig på skärmar när du är där i butiken. Men detta skulle folk uppleva som integritetskränkande. Möjligen skulle de kunna komma dit hän, att man är uppkopplad på wifi i butiken och får meddelande i telefonen, men det är ett minfält.

**102. GS:** Precis!

**103. Informant 1:** Där företagen själva inser att den här informationen tycker folk är integritetskränkande så vi kan inte använda den för att göra det som vi kanske hade velat. Sen är den kanske inte speciellt integritetskränkande, att jag har ett visst mode. Det är inte speciellt känslig information.

**104. GS:** Nää. Det är väldigt upp till individen också, vissa vill ju ha att när man går in på Ica så är det anpassat efter dom. Det är ett minfält såklart.

**105. Informant 1:** Så det är inte helt självklart. Men jag tror som sagt att den största faran är informationsläckor. Och det är så jag har tolkat från företag att de är mest rädda för när de hanterar persondata.

**106. GS:** Alltså säkerhetsaspekter?

**107. Informant 1:** Ja, de är väldigt noga med hur de släpper in och brandväggar och så vidare, de har flera multifaktorinloggning och så vidare. De är strikta. Och det gör ju att som utomstående konsult som jag som jobbar med 15-20 olika kunder blir det ganska tungrott, att ha olika VPN lösningar och inloggning för alla. Så klart det blir tungrott men de har bedömt det som nödvändigt och det tycker jag nog de gör rätt i. Sen ska man nog inte bortse att den största faran nog inte är utomstående utan det är inside, och där är de nog lite naiva fortsatt.

**108. GS:** Ja, det var det vi hade, tusen tack för att du ville ställa upp på denna intervju!

## 8.3 Transkribering Organisation 2

### Organisation 2:

Intervju 2 – Informant 2 – Organisation 2

Intervjuare – Gustav Stigestadh (GS)

Sekreterare – Felix Moberg

#### Start:

1. **GS:** Berätta lite om dig själv, vad du gör och så.

2. **Informant 2:** Jag är i grund och botten civilingenjör i industriell ekonomi, men riktade mig mer mot IT. Skrev exjobb här, om Big data faktiskt, jag kom på det faktiskt att jag har en liten bok om det. Där analyserade vi hur investeringar i customer analytics inom Big data området, hur det faktiskt leder till bättre kunskap om sin kund, om man kan förutspå vad kunden vill och så vidare.. Och hur de här investeringarna faktiskt ser ut. Det ledde då till jobb här, och det var 2015, snart varit här i 3år och jag är konsult och vad det hetat har bytt lite namn över tiden men jag är någonstans i gränslandet mellan Big data analytics och data science. Så det rör sig om att hantera data på olika sätt och göra predictions, och analysera data. Samtidigt som det finns ett spann av management consulting. Så jag måste också vara den som måste prata med kunden kring GDPR när det gäller våra projekt. Eller få kunden att hitta vilka problem vi ska angripa och så där.

3. **GS:** Så ganska bred roll ändå

4. **Informant 2:** Ja, det får man väl säga. Det är väl kanske dilemmat om man pluggar brett. Men jag får arbeta med många olika saker.

5. **GS:** Och hur många år har du jobbat med Big data? företaget har ju..

6. **Informant 2:** Ja de var ju nästan de som uppfann, eller nä inte riktigt. Men för min del är det 3 år.

7. **GS:** Hur skulle du definiera big data?

8. **Informant 2:** Ja, jag är lite färgad utav akademien har gjort det. Och absolut handlar det om de stora datamängderna, sen finns det ju de här 3V:na. Volume, variety, velocity. Och jag måste säga att även fast det var teori för mig när jag skrev mitt exjobb så tycker jag det stämmer väldigt väl. När man pratar med större bolag om det här så volymen vet man, den finns alltid där, det är alltid stora datamängder. Variety innebär då att det är oftast så att den strukturerade datan har man någorlunda koll på, men den ostrukturerade datan är det som liksom är problemet egentligen. Så samtidigt då att man ska hantera en stor mängd ostrukturerad data så ska man också kunna verifiera att man kan få ut någon form av information av den datan som finns. Och där är mycket problem. Sen velocity gäller rent allmänt för att många saker ska göras i det fantastiska buzzwordet "realtime" och det gäller att man har en infrastruktur som kan hantera det. Och velocity kan innebära att man måste kunna arbeta med iterationer väldigt snabbt, säg att ett företag får in data så vill man göra en kampanj på det. Då måste det gå fort annars tappar det värde i sig självt nästan direkt. Så för mig är det att Volymen är nästan given, men just att det handlar om olika typer av data och att



man måste göra analys på all data för att den ska bli värdefull. För strukturerad data har vi haft hur länge som helst.

**9. GS:** Vilka är då de största möjligheterna enligt dig med Big data?

**10. Informant 2:** De största möjligheterna är nog att vi kan.. ja hur ska man formulera det här. Man får actionable information, vad det nu ska vara på svenska ur information som egentligen bara har funnits men man inte ansett vara tillgänglig tidigare. Så även om man samlat på sig data under många år och det finns enorma bibliotek med data så är det inte förens den senare tiden som man liksom börjar köra data mining på det här på ett strukturerat sätt, och att man faktiskt kan få ut något som är användbart. Det är lite som man gör jämförelsen med forskning till exempel, det forskas på extremt många olika saker och man kan debattera om det är liksom nyttigt eller inte nyttigt vissa saker. Men när man har den här mängden forskning så handlar det om att kunna ta reda på vad är det som faktiskt är verifierbart, vad är det vi vet är sant? eller någorlunda sant och hur kan vi arbeta med det? Och där tycker jag att det finns massa ostrukturerad data och vi börjar nu kunna på ett strukturerat sätt använda det på ett nyttigt sätt för oss. Och det innebär bara att mängden data som finns och sättet som vi kan jobba med det nu gör att vi får konkret information helt enkelt.

**11. GS:** Har du något exempel från när du har arbetat med det, alltså vad för typ av information?

**12. Informant 2:** Ja, jag har till exempel ett projekt jag jobbade med ett företag inom möbelbranschen, där man hade väldigt mycket försäljningsdata, och ja, problemet var bara att man inte egentligen såg några mönster i det här. Man kunde se att en bokhylla till exempel var såld så här många gånger och samma personer köpte samma lampa. Men det gav inte så mycket mervärde. Däremot när man började titta i datan och se vilka fler attribut hade datan, så kunde man då börja länka de här produkterna som hade sålts samtidigt på helt andra premisser. Inte bara lampa och stol, utan kanske vilka färger, i vilken del av varuhuset, vilken typ utav product range. De kunde göra all den här informationen av varje produkt åskådlig, åskådlig göra den. Då kunde man då dra slutsatser att exempelvis den här lådan passar väldigt bra med den här bokhyllan trots att de inte är byggda för varandra.

**13. GS:** Är det ni som drar slutsatser då eller?

**14. Informant 2:** Det beror lite på, man kan säga att som data scientist så börjar man då med att hitta den här formen utav möjliga connections i datan. Det visade sig också till exempel att med varje produkt var det hög korrelation att man köpte värmeljus också. Det visade sig att alla köper värmeljus, därför finns det en connection. Men när det gäller specifika varor så dyker det här just upp och då kan man, vår roll kanske är mer att sortera bort bruset och då kan man jobba som konsult med kunden och ta reda på vad kan ligga bakom det här. Kanske då intervjua eller ställa frågor. Vissa saker kanske är lite lättare att dra slutsatser kring, exempelvis om det handlar om någonting där man förstår domänen. Men i många fall så är det kunden som har bäst koll.

**15. GS:** Man lär sig liksom av kundernas beteenden.

16. **Informant 2:** I det här fallet menar jag kunden, alltså som företaget. De har jättebra koll på det de gör, men vi kan ge dem ett bättre underlag att förstå vad de gör.

17. **GS:** Aa precis

18. **Informant 2:** Men till exempel skulle det röra tillverkningsprocessen, eller något sånt där, man ser att något går sönder vid en viss temperatur. Då kan vi hitta den temperaturen, då kan vi dra en slutsats utifrån det. Men många gånger handlar det om att hitta de mönster som kan vara intressanta, och sen i samråd med den som äger datan egentligen, ta reda på vad man ska göra.

19. **GS:** Vilka ser du som de största utmaningarna och problemen med Big data?

20. **Informant 2:** Det är just det att ostrukturerad data är just ostrukturerad. Och rör det sig om stora mängder, eller nej det kanske inte är mängden i Big data som är problemet egentligen utan det handlar om att kunna veta med viss sannolikhet att man gör rätt sak. För gör man fel sak så kommer man ha ett datacenter som går i en vecka och jobbar med någonting som är fel, och då har man bränt ganska mycket både pengar och tid. Så det handlar fortfarande att veta hur ska man analysera vilken data och hur säkra kan vi vara på att det vi får ut är rätt.

Utmaningen är att ostrukturerad data kan vara väldigt komplext.

21. **GS:** Vilken betydelse ser du Big data har idag, generellt i världen och branschen?

22. **Informant 2:** Jag skulle säga att Big data nästan är en förutsättning nu. Det var ett sånt buzzword kring typ 2013-14, där började det här "booma". Alla pratade om att med big data kommer vi kunna göra saker. Men nu sitter varenda företag med någon form av big data i den klassiska definitionen och det innebär att det är bara så det är. Nu så är det så att med den infrastrukturen vi har nu så kan vi behandla den här datan. Jag tror bara att vi har börjat kolla på vilken typ av data vi har.. Att vi har börjat tänka mer kring vad som är strukturerat/ostrukturerat. Att man har en bättre governance på plats, att man vet hur man ska jobba med datan. Ett av de kanske största problemen som är mest övergående det är att man har dålig master data management. Det innebär att dubbla kundregister finns för en och samma kund, att man har registrerat sig på olika epost-adresser men man vet att det är samma person, och om man ska göra ett reklamutskick så kanske den här personen inte får någonting, för att det är konflikter och sådant. Så governance och master data management är egentligen det som är följd problemen som har kommit ur utav att datamängden bara ökar och ökar. Så mycket av det jobbet vi jobbar med när det gäller just den här typen, när det handlar om datakvalitet till exempel så är det just då att man inte vet, eller man inte haft en tillräckligt tydlig styrning på plats från början, utan det har bara växt. Ett företag på 10 personer har blivit 200 och då har det bara sparat ur liksom. Big data är någonting som bara är.

23. **GS:** Och hur använder ni Big data?

24. **Informant 2:** Vi jobbar ju med kunder av alla möjliga olika slag då, och alla har stora datamängder. Många gånger så kanske man börjar göra exploration på ett litet dataset, sen förstorar man upp det här bara när man har gjort en liten pilot först då. Men när det gäller företag som mest kanske vill bygga data warehouses eller data lakes och åskådliggöra den

data man har, då finns det ju ganska stora ramverk som beskriver just så här bör man hantera data för att man liksom ska ha någonting man kan jobba med helt enkelt.

Använder är svårt att säga men vi, men hur vi kommer i kontakt med den är att antingen göra den åskådlig eller så handlar det om att dra analyser och sånt. Att det handlar om stora datamängder, det är nästan det som är vår styrka för vi har infrastrukturen som kan jobba med den.

När det gäller data warehouse eller någonting sånt där så är det många gånger att man vill göra transformationer av datan och då kan det bli väldigt överväldigande om det är fyra petabytes som ska köras igenom och då har vi en infrastruktur som klarar av det. Där kommer vår styrka in delvis då. När det handlar om att förutspå saker ur data så handlar det mer om att vi måste använda våra data science skills, men då tar vi också hjälp av att om det är stora datamängder så kan vi även lösa det där. Då är det lite bonus på toppen.

**25. GS:** När ni analyserar stora datamängder åt kunder och så, vad är det för utmaningar där? Som du sa så är det Governance och master data management och så?

**26. Informant 2:** Ja, det är nog dem. Jag tror de är de största. Sen är det så att det är nya utmaningar för varje nytt uppdrag, när det handlar om att ta reda på, säg att det är data från en biltillverkare och att man vill veta vilken bil som kommer gå sönder om fyra år. Då kan det vara ett svårt uppdrag i sig, men där är inte utmaningen datamängden utan det handlar om vilken typ av data, eller hur den är strukturerad eller ostrukturerad. I och för sig om mycket datamängder håller mycket ostrukturerad data, då ökar utmaningen. Ju mer variety av data ju jobbigare blir det i princip.

**27. GS:** Jag vet inte om just ni samlar in så mycket data, men era kunder har de alltid ett tydlig syfte med datainsamlingen?

**28. Informant 2:** Jag har ingen aning om vi har till exempel har Facebook som kunder och så där, där har det ju varit väldigt otydligt exakt vad det används till. Det som alltid ska vara tydligt hos oss, det är ju att vi då som data processer i GDPR processen ska ha ett tydligt direktiv om det skulle vara så att vi ska behandla någons data, då måste vi veta vad det är för premisser det rör sig om och hittar vi då någonting som verkar strida mot lagen så måste vi alltid uppmärksamma det. Jag skulle säga att just det här så är det ganska tydligt att alla vi på XXX går igenom en.. vad heter det nu igen.

**29. GS:** GDPR utbildning?

**30. Informant 2:** Ja dels det, och sen så i våra business combat guidelines. Det är för att överhuvudtaget ha en anställning så måste man gå igenom det. Det förklarar då vilken typ av integritet man måste hålla för att utföra jobb på XXX överhuvudtaget. Och där finns det ganska stora regler kring just då vad är etiskt eller inte etiskt. Är det inte etiskt kan man alltid embraca det till en manager, och så finns det en hel process för det där. Jag skulle säga att om vi skulle jobba med ett företag som har en dold agenda hur man samlar in data, så skulle det ganska fort fångas upp i processerna. Det jag exponeras mest för det är att man får ut data från en kund, som man ska försöka analysera eller få en överblick över och vad jag kan säga så har det alltid varit relevant data. Just på insamling, i och för sig XXX har ju ganska många patent och har uppfunnit flera algoritmer som ska göra det också. Det som är intressant är att många av de tools vi säljer då, till exempel för markandsföring, eller för analys. Så finns det automatisk anonymisering av data. Vissa kan man inte ens välja på, alltså samlar du in data

för marknadsföring så kommer den vara presenterbar bara på ett anonymiserat sätt. Det har kanske varit lite av ett krav för att de här sakerna ska vara top of the line. Där har jag i alla fall blivit positivt överraskad över att XXX har liksom den typen av integritet för att då inte stänga ute framtida kunder. Samtidigt jobbar vi med säkerhetstjänster och då är det på ett helt annat sätt, då är det lagar och så. Men när det gäller just för företag i den privata sfären så skulle jag säga att XXX har haft råd att vara den som står på integritetssidan. Med det här verktyget så jobbar vi med ett etiskt hållbart sätt kan man säga. Det kanske är en lite lyxig position att vara konsult i det här, för vi kan vara de som säger nej till saker medans ett företag är helt baserat på att samla in data om sina kunder rent tekniskt sett borde vara intresserade av att samla in allt som finns. Och då tror jag GDPR kommer i det här fallet kommer komma in och vara den naturliga bromsen för datainsamling.

**31. GS:** Hur vet ni att datan är korrekt och kvalitativ?

**32. Informant 2:** Ja, det är många gånger det dyker upp skräpdata hos kunder, så är det ju. När vi jobbar med datakvalitet och korrekt data, data governance egentligen, så har vi byggt ett ganska stort ramverk för att jobba med det här, och vi har experter som är bra på att göra om egentligen hela kundens, ja, allt i från system till tänk kring hur man ska jobba med det. Och det är ett ganska vanligt uppdrag hos XXX att jobba med. Där finns det då ett 200 sidors dokument man kan gå igenom eller som vi jobbar med som konsulter på företaget. Men jag tänkte att vi kunde dra de här dimensionerna då som vi har satt fram som de viktiga. Det är sju dimensioner, så det är lite fler än dem som de brukar vara. Jag har en tabell som jag kan skicka till er, men för diskussionens skull så tänkte jag ta dem direkt.

**33.GS:** Perfekt!

**34. Informant 2:** Vi sätter de här sju dimensionerna för att då kunna både ha konversationen med en kund kring, vad är datakvalitet och vad man faktiskt ska mäta, kvantifierbart och kvalitetsmässigt.

Första är om datan är *valid*. Det handlar om ett dataelement har alla features för att vara acceptabelt. Till exempel i ett kundregister så ska en data entry(?) innehålla både personnummer, e-postadresser och namn och så vidare.

Nästa är att den är *unique*. Varje datanummer ska vara unikt, inga duplikat.

Tredje är *complete* och det handlar om då att ett dataelement alltid behövs, och alltid behövs baserat på ett condition från ett annat dataelement. Dett innebär då att ett telefonnummer kommer aldrig vara viktigt om det inte är så att ett dataelement, en person, behöver telefonnumret. Skulle man kunna säga ganska enkelt. Eller tillexempel två produkter som beror på varandra måste vara länkade på något sätt.

Fjärde är *consistent*. Det är ganska välkopplat till GDPR, men det är att det inte finns någon variation i dataelementet eller viss typ. Personnummer ska vara av samma längd, är ett exempel på ett superenkelt fall. Det blir kanske lite mer komplext när det handlar om andra variabler.

Sen nästa är *timely*, den ska innehålla den senaste och mest uppdaterade informationen och här ser vi stora problem inom master data management. Att datan inte är up-to-date hos många av våra kunder. Vanligt klassiskt är att man har fyllt i en adress någon gång och sen har man flyttat.

**35. GS:** Aa precis

36. **Informant 2:** Efter det har vi *accurate* och det kan man tänka är väldigt likt att det ska vara valid och sådant där. Men jag skulle säga att det kanske handlar mer om hierarkin i datan. Enkelt sagt att det inte är en röra, utan att man lätt kan sortera in datan, speciellt inom domäner som kan vara land, väldigt viktigt inom GDPR då att man vet vilken data som hör till vilka kunder och vilket land och om de befinner sig i GDPR zonen. Den sista är *precise* och det handlar om att datan är ett på ett sådant sätt att man använder det enbart för sitt intended purpose. Det innebär också att man måste förstå vad det är för data som finns och vad den ska användas till och vad man har för karaktäristika egentligen. Man kan ju också se att de här som vi har använt ett tag då mappar ganska väl mot GDPR som också säger att man måste veta vad den här datan ska användas till och sånt där. Har man de här sju dimensionerna in place så är det ganska lätt att börja titta på GDPR.

37. **GS:** Är detta ett välkänt ramverk?

38. **Informant 2:** Det här är ju det som XXX använder officiellt, och det finns flera variationer av det är, vissa använder bara tre och då är det lite luddigare. Vi har sett att man man behöver nästan utöka det till sju för att kunna peka på vissa saker, när det handlar om till exempel vilket syfte man har med datan, då är det just en dimension som behandlar det här.

39. **GS:** Vilket bra svar!

40. **Informant 2:** Det finns många som har tänkt till på XXX som jobbar just det här.

41. **GS:** Och uppfyller man de här kraven så är datan korrekt?

42. **Informant 2:** Ja, och kvalitén är hög. Man skulle kunna dra in i kvalitén hur snabbt man behandlar data och sånt där, men det är kanske mer på kravnivå. Men de här sju dimensionerna skulle jag säga är om korrekthet och kvalitet.

43. **GS:** Då går vi över till etik och GDPR. Anser du att det finns några etiska problem med Big data och big data analys?

44. **Informant 2:** Ja, det gör det ju. Det finns ju problem med all form av data egentligen, men speciellt om det handlar om data som gör den identifierbar och ju större mängd man har desto lättare är det att få en kombination som gör det identifierbart. Låt säga att en epostadress behöver i sig självt inte vara en identifierare, även om man behandlar det så, men kombinerar man e-postadress med GPS-position och så vidare, då har man ganska fort personal information. Hos oss går de under de hårdaste kraven av integritet. Så absolut, ju mer data man samlar in eller ju fler typer av data man samlar in desto större etiskt fokus måste man ha.

45. **GS:** Och anser du att den personliga integriteten kan kränkas i samband med dataanalys?

46. **Informant 2:** Det kan vara mer eller mindre specifikt. Säg att det finns oetiska sätt att sprida reklam på till exempel, som möjliggörs av att man har mer data än man borde. Det kan även vara så luddigt att man kanske bara vet vilka sidor en person surfar på genom en cookie och det kan jag tycka är personlig integritet man kränker. Så även om det inte egentligen inte

skadar mig på samma sätt som om man hade sålt samma data till någon som hade ett helt annat syfte.

Men det är lite svårt att svara på. Det handlar just om att vi måste nog se skillnad på, även om den finns, information som man inte vill ge ut, som man i samtal med någon inte hade gett ut. Och även om vi då med ny teknik kan samla in mycket mer data än vad vi borde kunna så innebär inte det att den datan alltid är okej. Att det finns tillgänglig data som man inte hade fått utav en person om man hade frågat.

**47. GS:** Men du var och nosade på det innan, att man just sätter data i samband med en e-postadress med en kanske faktiskt adress. Är det ett sätt att kränka, alltså i samband med dataanalys.

**48. Informant 2:** Ja, ehm... Visst alltid beror lite på hur man lämnat ifrån sig det. Om någon har tillräckligt mörka syften så går det alltid att använda det på något sätt. Säg att man i en undersökning vill du vinna ett pris så lämna din e-postadress och den här datan behandlas anonymt och så vidare. Men om de har min e-postadress så kan de kontakta mig baserat på vad jag svarat, förmodligen, om de inte har byggt in ett system som automatiskt slumpar ut id. Men det kan jag inte veta heller. Vi lämnar ifrån oss information utan att faktiskt kunna veta om någon kommer att kränka vår integritet, och där tror jag att företag som kan hantera det på ett bra sätt kommer komma vinnande ur den här stormen som har dragit igång då, kanske lite oberättigat kring Facebook eftersom de är inte de enda som gör det här, och de har gjort det under en lång tid som vi har vetat om. Så varför är det ett problem just nu när det alltid varit ett problem.

Men ja, hmm, svaret är kanske att den kan kränkas direkt om det är någon som tar kontakt trots att man tror att man har lämnat något anonymt, eller så kan den kränkas i form av att man sätter samband flera olika typer av data och använder det på ett sett som är drabbar en personligt. Svaret är absolut ja, att det är.. .

**49. GS:** Hur tror du GDPR kommer att påverka Big data och Big data analys?

**50. Informant 2:** Jag tror först och främst att det gör folk medvetna om vad data egentligen är, nu pratar jag inte om IT-konsulter i allmänhet utan alla personer. Man börjar tänka på, okej vad har jag egentligen skrivit under. Skämten har ju gått kring att man säkert skriver under sin egen dödsdom, när man till exempel skriver under iTunes avtal. Men det är ingen som har brytt sig om det tidigare, och nu kommer det då faktiskt lägga fram att alla måste ta ett beslut när man blir medlem i en kundklubb till exempel, och man kommer då få reda på att, okej ni kommer använda min e-postadress för att skicka mail till mig, då väljer jag att inte godta det här. Jag tror det kommer göra folk mer medvetna om att man faktiskt kan ta aktiva val när det gäller det här och det tror jag kanske är den största grejen. Sen så återstår det att se vad som kommer att hända med företag, det finns en del frågetecken att man inte som företag kan ta bort all data för det kanske kan komma att behövas användas utav polis och myndigheter till exempel. Samtidigt som det kan vara ganska lätt att säga att vi behöver den här datan för att personen är kund hos oss, hen är med i vår kundklubb och då kan man inte heller ta bort den. Det känns även om att reglerna kan verka väldigt tydliga inom GDPR, så kommer det att återstå att se vad praxis faktiskt kommer säga och vad man faktiskt kommer behövas göra.

**51. GS:** Tror du att den kommer stävja Cambridge analytica och sådana...?

52. **Informant 2:** Nja. Nu är jag inte jättebra påläst om just Cambridge analytica men om det var olagligt så kommer de inte att stoppa dem ändå.

Å andra sidan, Facebook hade alltså en öppen plattform för att dela information med sina developers, och redan där finns det ett enormt problem. 36:51

Om man då ser det ur personlig integritet, eller vad man då har skrivit över till Facebook när man är aktiv där. Samtidigt som de då samlar in information om sådana som inte är kunder vilket är ett ännu större problem. Men jag tror att GDPR kommer att öka medvetenheten som gör att man inte designar system som från början inte är dåliga. Det kanske faktiskt kommer leda till att man inte samlar in och behandlar mer data än vad man behöver och om jag inte vill vara med i en kundklubb så är jag inte det, och då minskar min risk, eller exponering mot risk ifrån cyberangrepp och liknande. Jag tror att mycket bra kommer komma utav det här och så får vi väl se nu då i sommar om det är så att lagen blir så oerhört svår att följa till exempel, eller om det är så att de måste stämma en stor del av företagen för att man inte kan eller inte vet eller hamnar i tvistemål med andra myndigheter till exempel.

53. **GS:** Tror du att användning av anonymiserad och pseudonymiserad kommer att ökas i samband med dataanalys efter då att GDPR.. Ni använder mycket anonymiserad data sa du.

54. **Informant 2:** Ja många gånger, ur ett rent behandlings är det bättre att ha personlig data även om det förknippar man med många fler regler för vad man kan göra med den. Men anonymiserad data är väl egentligen lite på man räknar med. Så jag skulle nog säga, ja det kommer nog öka och det kommer nog fungera bra för företag som faktiskt jobbar seriöst med det här. Samtidigt då som det öppnar en marknad för att faktiskt anonymisera data så tror jag att det kommer nog gå hand i hand. Mer anonymiserad data och då även aktörer som kan göra hela konceptet kommer att gynnas och det tror jag är bra. Det tror jag kommer leda till mer professionalitet.

55. **GS:** Vad ser du för negativt med GDPR? Alltså vilka negativa konsekvenser tror du att det kommer att ge?

56. **Informant 2:** Jag tror absolut att det kan leda till högre omkostnader för företag. Kanske i viss mån en omöjlig administration situation. Det klassiska exemplet är väl Röda korset eller liknande organisation om man ber dem då att ta bort all min data så måste de garantera att all data tas bort, som även gäller pappersprotokoll från en lokalförening för 10 år sedan. Det kan kosta hur mycket pengar som helst, och bolag kan inte behöva gå i konkurs bara för att någon begär att bli av med all sin data. Så jag misstänker att det kommer att dyka upp några sådana här fall, där det händer konstiga saker, men då hoppas jag att administrationen av GDPR kan komma till något konsensus, vad är rimligt att göra och vad är inte rimligt. Sen är frågan om det finns hederliga aktörer som kommer att drabbas negativt utav det, jag kan inte komma på något klockrent exempel men jag misstänker att data som inte har varit timely och precise i början kanske blir det efter ett tag. Det kanske blir så att man har data i ett kundregister som man inser att man har efter ett tag. Om den hade raderats från början kanske man hade missat (hör inte vad han säger41:36) . Och det rör sig enbart om vinst och omsättning för ett specifikt företag, men jag tänker att det kommer ändras sättet man jobbar med det här, jobbar med data överhuvudtaget. Sen finns det någon klausul grej med GDPR om det är livsavgörande, så kan man frånga GDPR, så det kommer nog inte drabba sjukvården där det hade kunnat vara ett jättestort problem annars.

Det är svårt att säga, det känns inte helt vattentätt, det gör det inte. Men det gör nog ingen lag i och för sig, men här kan det komma få stora konsekvenser över någonting vi inte tänkt på nu som blir ett jättestort problem i framtiden.

**57. GS:** Men, ni på XXX kanske inte samlar in jättemycket, men när ni arbetar med företag och de har personlig data som används in fel, vad sa ni var ert tillvägagångssätt då, vem flaggar ni för?

**58. Informant 2:** Det beror lite på hur datan kommer i kontakt med oss. Men i mitt perspektiv om jag skulle jobba med en kund och skulle göra en prediktiv analys på någonting och skulle se att datan skulle innehålla någonting som är djupt oetiskt, då handlar det om att sätta igång en process. Just det här handlar då att höja en flagga till antingen den som är projektledare eller project executive, och de har all utbildning hur man ska ta hand om det här då. Säg att jag bara är enskild konsult ska jag alltid säga till den som är chef helt enkelt. XXX skulle aldrig utsättas sig för någon sån risk för en kunds skull. Det har man skrivit papper på från början, i alla avtal har man redan klargjort om en kund gör någonting olagligt ska inte vi ha någonting med det att göra.

**59. GS:** Ni liksom separerar er från...

**60. Informant 2:** Man kan ju säga om kunden begår ett brott så är det något som åligger oss plikt i att göra någonting åt det eller anmäla.

Det finns även funktioner internt på XXX att man kan lyfta en sådan fråga till ett råd som är internationellt, så att det garanterat inte blir något eftergift. Jag kan till exempel inte få sparken för ett fel min chef gör.

**61. GS:** Då känns det att ni har väldigt bra stadgar hur ni förhåller er till det.

**62. Informant 2:** Ja, det finns mycket att läsa på där om man vill kika på omfattningen utav det. Det är just rotat i att varje person som har gått igenom utbildningen i business conduct guidelines och då signerar på att man agerar på det här etiska sättet, samtidigt som alla chefer har en extra utbildning och att det finns ett säkert sätt att rapportera olagligheter eller fel. Det är någonting som man kan känna det egna bolaget, absolut, men det är också för att skydda den egna individen. Om inte den enskilda anställda känner sig säker i att göra rätt sak, då vill man för det första inte jobba där och det kommer inte rapporteras av ren lättja liksom. Jag tycker det är en viktig kultur man kan välja att odla och det är bra för alla parter.

**63. GS:** I och med GDPR kommer att samtycke krävas, det kommer att bli formulär att läsa igenom och acceptera och så, hur tror du det kommer att säkerställas? Hur kommer era kunder få in samtycken från sina kunder rent praktiskt?

**64. Informant 2:** Jag tror delvis att det kommer att handla om design, hur man utformar appar och programvara. Det måste vara på något sätt tillgängligt, det är A och O, man måste kunna se vad det man ger sig in på. Och där tror jag absolut att ux-design kommer spela en liten roll ändå. Sen utöver det så handlar det för företag att veta vad man ska använda datan till, och då kanske man ska ha en tydlig strategi. Har man en tydligare strategi så är det lättare att förmedla den till kunden och så kan man få ett ja eller nej.

**65. GS:** Aa.



66. **Informant 2:** Så jag tror att det handlar ganska mycket om de mjuka värdena i det faktiskt, inte bara att det ska vara en tydlig checkbox utan även i "hur har vi kontakt med kunden" och "hur leder vi kunden igenom, t.ex. köpet".

Jag har själv upplevt att man ser att många avtal nu har dykt upp på nytt igen, de är lite annorlunda utformade men jag kan fortfarande inte säga att någon har varit tydlig med att "så här vill vi använda din data" är det okej?

67. **GS:** Nää precis

68. **Informant 2:** Nu har det inte riktigt gått igenom ännu, men många har ju redan uppdaterat sina avtal. Och jag tror att om det här bara leder till ännu längre EU-lag(?48:27 ) så hjälper det oss ändå inte. Frågan är hur tydligt kommer tydligt vara.

69. **GS:** Har du något annat, eller allmän åsikt om GDPR som du tänkt på?

70. **Informant 2:** Jo det kan ju ställa högre krav på hur och var man sparar data, säg att kundregister kommer att bli väldigt uppdelade på just geografiskt position, man måste veta vilken data, så inte den här datan finns. Det känns ju att med det kommer då utformningen av infrastrukturen för Big data analytics kommer att ändras. Man måste veta var datan finns, man måste allokera datakapacitet till ett visst område kanske. Det gör ju att man inte till exempel kan koka ihop data och köra allting på ett och samma datacenter någonstans i USA liksom. Då måste det vara clearat att man får flytta datan och så där. Jag tror att det kommer få vissa konsekvenser, att ja om det varit enkelt innan så kommer det bli lite krångligare. Men som sagt, det är en sådan grej som bara ska lösas.

71. **GS:** Aa

72. **Informant 2:** Men under början kan det absolut vara så att det har glömts bort kanske. Man inte inser att den här appen faktiskt mergar data någon annanstans, eller någon molntjänst någonstans som kommer användas på fel sätt. Och har man inte koll på det kan det bli stora dataproblem. Så innan alla har fått koll på sina white papers, och hur deras tjänster fungerar då kan det vara ett litet tidshål i början, där det inte utförs på rätt sätt ändå.

73. **GS:** Men tror du att GDPR... den kommer inte stoppa dataanalyser så klart men tror du att de kommer att stoppa de mer kränkande analyserna? Eller blir det så att de som vill göra det kommer att göra det så att säga?

74. **Informant 2:** Alltså, de som är villiga att begå ett brott spelar det inte så stor roll vad vi har regler för.

75. **Informant 2:** Jag tror att, om det blir som man tänker sig, så kommer det nog att göra att folk kommer att kunna säga nej till de sakerna vi egentligen tycker är kränkande, men som vi har på något sätt blundat för, för att man bara har accepterat och gått vidare med en tjänst. Så det på sättet ja. Människor blir mer medvetna om vad man kan säga nej till, då tror jag att vi kan minska en del utav det. Som jag förstår det, så är den allmänna opinionen är att man inte tycker om att folk har ens data.

76. **GS:** Nää precis!

77. **Informant 2:** Om man lätt kan säga nej till det.. För nu har det inte riktigt gått, för om man inte godtar de här användaravtalen får man inte använda produkten, och det sätter GDPR stopp för, att man fortfarande kommer att kunna använda produkten men det kommer inte finnas ett avtal som man måste godkänna för att använda den. På det sättet får man inte liksom låsa in kunden.

Och det tycker jag är ett bra steg framåt. Gemene man är inte jurist som är specialiserad på avtal, men jag tror att det går att göra det tillgängligt att förstå vad det är som händer och vad man egentligen vill gå med på. Det tror jag är en bra utveckling rent allmänt.

78. **GS:** Vet du hur det kommer skilja sig från PuL?

79. **Informant 2:** Ja, GDPR är först och främst fler regler. Vi har haft ett relativt bra skydd i Sverige tidigare. Men som jag tycker så handlar det mer om att visa upp syftet och vara mer öppen, vilket jag tror leder till att vem som helst kan mer förstå vad som händer och säga ja och nej. I och för sig säger väl PuL att data som ska förvaras på ett sätt som gör att man inte lätt kan bli av med den. Men det finns så mycket lösenordsdatabaser som inte är krypterade som stjäls ändå. Så det känns som att det har inte hjälp ändå. Om man nu gör en ny iteration så kanske det sätter igång en ny våg utav, ja mer uppmärksamhet som kan leda till något bra.

80. **GS:** Men bra.. Men visst hade XXX etiska riktlinjer kring Big data?

81. **Informant 2:** Absolut! Och jag skulle säga att den är kanske främst fokuserad på vilken typ av data det handlar om. Och det gäller främst om personlig data. Handlar det om väderdata så är det inte så allvarligt, men gäller det personlig data så finns det guidelines som finns publikt som finns på hemsidan som man kan läsa. Vill man göra business med XXX så kan man också se vilka åtagande, eller extra steg, tar vi för att bemöts gå dels lagar och regler men även etiska standarder som vi har satt upp då.

82. **GS:** Bra! Är det något annat du tänkt på eller så?

83. Nä det var nog det mesta tycker jag. Jag skulle nog säga att det finns ett behov på marknaden för experter, även om ett företag är stort och funnits i branschen länge så innebär inte det att governance av data har hängt med. Det blir mer och mer fokus på det, men det kommer ta tid innan alla bolag har en så pass bra governance att det är självruddande. Ökat fokus är jättebra, men jag tror att det fortfarande finns ett ökat utrymme för experter. Företag som är experter på det här, att agera och faktiskt tillföra mycket nytta.

84. **GS:** Tusen tack för detta. Vi är väl klara skulle jag anse!

## 8.4 Transkribering Organisation 3

### Organisation 3:

Intervju 3 – Informant 3 – Organisation 3

Intervjuare – Gustav Stigestadh (GS)

Sekreterare – Felix Moberg

#### Start.

1. **GS:** Hur länge har du jobbat med Big data?

2. **Informant 3:** Om jag får ta mig friheten och kringgå lite frågorna så hoppar jag direkt på något som stör mig lite i kanten. Jag har jobbat med dataanalys länge, jag studerade machine learning och artificiell intelligens. När jag pluggade tekniskt fysik, så var jag ett utbytesår i USA, Colorado, där var de väldigt duktiga med artificiell intelligens, så jag höll på med machine learning och computer vision och AI under ett år där. Det är väl 15 år sedan nu då. Jag hade möjlighet att fortsätta att forska, jag hade det alternativet för jag tyckte det var väldigt kul med AI. I framförallt Sverige var det ett ganska smalt område, det fanns inte mycket expertis i det och framförallt fanns det inga kommersiella applikationer eller kommersiella aktörer som applicerade det, utan det var en teoretisk nivå, hemmasnickeri mest.

3. **GS:** Aha

4. **Informant 3:** Eller på akademisk nivå.

5. **GS:** När var det här ungefär?

6. **Informant 3:** Jag var klar från USA 2004, kom hem så funderade jag lite på vilken väg jag skulle gå och då valde jag att ta en svängom tillbaka och läsa mer ekonomi och bli managementkonsult. Då var jag en person som jag tror många uppfattade som väldigt kvantitativ driven då jag gillar kvantitativ analys. Hade nån slags kandidat i företagsekonomi och valde att bli managementkonsult. Kom till företaget 2006, sen efter lite stapplande första år på företaget, så började jag inse att det fanns vissa projekt på företaget som passade mig lite bättre och det var när det fanns data och dataanalys att göras.

7. **GS:** Okej!

8. **Informant 3:** Sen 2009 kanske så startade vi en egen gruppering inom vårt globala bolag som jobbade med analys, analytics enhet. Analytics började segla upp som ett "buzzword" var det då, men kom att bli ganska standard efter ett par år. Sen dess har jag jobbat med kvantitativ analys. Sen tror jag att någonstans runt 2010, 2011 så började Big data explodera på allvar, det var ännu större buzzword och framförallt de bakom drivande krafterna bakom Big data var mycket större än den som var för analytics. För det var mycket mer hårdvara, mjukvara, analytics hade folk hållit på med längre. Mjukvaran SAS hade hållit på i 30, 40 år med statistisk analys i en företagskontext, och de var väldigt duktiga. Men sen kom big data vågen och det blev väldigt mycket mer pengar i det, man skulle köpa stora parker av hårdvara bara för att kunna hantera större mängder data. Jag gillar utveckling, men jag ogillar

när utveckling inte drivs av affärsmässighet då. Det är problemet med Big data, jag förknippar det väldigt mycket med folk som lyckas med investeringar med saker de inte har någon aning att få värde av. För att de saknar den analytiska kompetensen. Många förväxlar att om man har data eller analytiska modeller som är dålig, att slutresultatet blir bättre om man vrider på mängden data. Det vill jag starkt motsäga mig ifrån. Sen finns det dock väldigt intressanta applikationer av Big data, det är bara att kolla Google, Spotify, Netflix där man använder den här Big data teknologin för att kunna leverera en tjänst på en stor skala.

Men många vill ju samla in data, bara för att man tror att ju större mängd data ju mer värde kommer det att skapa utan att riktigt ha koll. Det är en risk i det.

Men sen finns det många fördelar, nu börjar man bli lite mer klarsynt i affärsvärlden med Big data. Jag jobbar i ett projekt idag där man har väldigt stora datamängder, det är signaldata, det är... det är väldigt stora data mängder helt enkelt och det finns ett affärsvärde i det.

## 9. GS: Aa

10. **Informant 3:** Men det har tagit ett tag att komma till ett sådant projekt. För att svara på den första frågan så har jag jobbat med dataanalys i 15år. Jag har jobbat med Big data till och från men det har inte varit min drivkraft att jobba med big data, utan jag jobbar med data analys.

11. **GS:** Har du någon koll hur ert företag har jobbat med Big data?

12. **Informant 3:** Jag har inte stenkoll, men vi är ett bolag med 400 000 globalt så det är lite svårt att uttala sig exakt. Men på en global skala så har vi varit med väldigt tidigt och har haft pionjärer inom Big data. Vi får inte tala om vilka kunder vi har, men alla de stora tech bolagen har nog varit våra kunder. Jag skulle nog säga att vi har jobbat med Big data sedan 2008-2009 skulle jag tro. Men avancerad analys, beroende på hur man ställer frågan, så har vi haft personer som kan SAS och jobbat med SAS i över 20 år skulle jag säga.

13. **GS:** Men om du skulle vilja definiera Big data, hur skulle du definiera det kort?

14. **Informant 3:** Jag är inte ett så jättestort fan av begreppet Big data, men man brukar väl se det som fyra eller fem V.

15. **GS:** Aa tre, fyra eller fem är väl de vanligaste.

16. **Informant 3:** Velocity, variety, volym. Vad är det mer då?

17. **GS:** Value är en

18. **Informant 3:** Ja, det måste vara påklarat på slutet.

19. **GS:** Aa exakt. Sen finns det veracity också. Men ni brukar köra efter de begreppen också?

20. **Informant 3:** Jag kanske inte tycker att de återspeglar riktigt verkligheten. Att det ska finnas variety i datan, om det inte finns variety i datan så skulle det inte vara Big data, det finns inte en sådan binär grej. Utan det är väl lite min syn på det, det blir lite meningslös genom någon modell så är inte detta Big data, om det kommer in flera terabyte data om dagen, och de kommer med en viss hastighet, då är det ju fortfarande Big data problem. Sen kanske

man inte behöver använda alla tekniska finesser som krävs för att hantera en variation av datan som kommer. Jag kan tycka att data volume och velocity som..

**21. GS:** Det är störst fokus på det?

**22. Informant 3:** Aa det är störst fokus på det. Det är oftast det är så. Men exempel variety, om det kommer en dag 10 000 events och en annan dag 20 000 events är det då variety? Jag skulle säga nej. Så funkas alla datasystem, i vanliga BI så har ju haft variety också. Man behöver kapacitet att behöva det, men det är mer att man behöver bandbredden. Men variety är mer att ibland kommer det en bild, ibland kommer det en video ibland kommer det ljudfiler. Den känns lite konstruerad, den situationen har jag aldrig upplevt att det är så.

**23. GS:** Men vilka är de största möjligheterna tycker du, med att analysera Big data?

**24. Informant 3:** Nä men de största möjligheterna med Big data. De nya teknikerna har möjliggjort mer affärsrelevant, affärsnära, analys av väldigt granulär data. Och det är där värdet ligger främst tycker jag. Inte att datan är mycket större nödvändigtvis utan att den är mer tillgängliggjord för företagen. I form av de initiativ man gör, för man tänker om. De största problemen med BI - när man konstruerar business intelligence så glömmer man fullständigt bort Intelligence, att det bara skulle vara en dum rapport maskin. Business intelligence, är misslyckat, för mig är analytics vågen snarare än Big data, analytics vågen är snarare ett sätt att reparera de misstag man gjorde med Business intelligence. Business Intelligence är mer och mer ett skällsord och har varit det senaste åtta åren typ.

**25. GS:** Aha Okej?

**26. Informant 3:** För Business intelligence... Jag tror man talar mer och mer om rapportering och liksom destruktiv analys till och med. Men BI är konstruerat för att lätt att få ut, typ hur mycket sålde vi förra månaden, till controllers som ska ha det i årsredovisning och rapportera det till marknaden och så. Men om man däremot från affären vill ha reda på hur man kan använda den här datan för att förbättra vårt sätt vi agerar mot våra kunder, till exempel. Vilka kunder säger upp sina abonnemang hos oss? För controllern vill bara visa att tusen abonnemang sas upp under februari, det måste de redovisa i sin månadsrapport. Men är man mer analytisk lagd, då vill man titta på datan och på ett analytiskt sätt förstå varför de säger upp, eller vilka de är. Då behöver man datan på ett annat sätt, Business intelligence miljön är inte byggd för det där.

**27. GS:** Okej.

**28. Informant 3:** Och det är det för mig, att hela vågen att man bygger sådana data lakes, det gör att man kan hitta, eller tillgängliggöra väldigt granulär data för verksamheten. Och förhoppningsvis kan verksamheten hitta mer insikter som påverkar affärsbeslut och operativa beslut. Så det är de största potentialen och det som Big data levererat - att tillgängliggöra data.

**29. GS:** Och vad finns det för utmaningar med det? Vad finns det för utmaningar för Big data?

**30. Informant 3:** Som ni hört, kanske för att jag läste företagsekonomi, att jag blev en ganska affärscentrerad person. Men med ett väldigt analytiskt sinne skulle jag säga. Men jag ser ju

flera kunder som har en allt för tekniskt fokus på sina implementeringar. Att det blir för fokuserat på att få in mer och mer data. Alla datakällor ska in och det blir nån form av pokemonjakt att man ska ha alla datakällor i sin datalake. Man tappar fokus på slutsyftet. Jag ser ofta att det blir en väldigt stor IT komponent, att det blir ett IT-projekt.

Det finns oerhörda affärsvärden för olika affärsenheter att få genom att analysera data.

Det är väl största utmaningen, för att det är en ny teknik, ny kompetens som företag måste skaffa sig och det blir lätt fokus på fel saker.

Men typ den här velocity grejen då, om man inte är Netflix eller Spotify, de har en affärsmodell att när man trycker på play i sin telefon ska det kännas som att sången finns på min telefon. Ni kanske är uppväxta med Spotify nästan, hur gamla är ni?

**31. GS:** Vi är 91or båda två.

**32. Informant 3:** Aa, men då är det nästan så att ni växte upp med spotify, inte riktigt.

**33. GS:** Mja nästan.

**34. Informant 3:** Man laddade ner musiken tidigare på sin telefon eller ipod eller var det va. Men Spotify har revolutionerat, för det känns likadant fast man inte har datan på telefonen, där är Big data teknik som verkligen behövs trimmas i detalj för att kunna hantera stora mängder data, sen finns det även massa leverans komplexitet i det där som kanske är annat än Big data. Men dom behöver tänka på hur fort går det, och hela den grejen. Men till exempel stora företaget X som försöker få lite insikt i sina kunder, ett traditionellt producerade bolaget, eller en bank eller en telekomoperatör eller liknande. Om de vill förstå sina kunder lite bättre, då är det sällan så att affären har.. liksom deras främsta drivkraft är att datan som finns i den här data laken ska vara helt up to date på några minuter eller om nån säger att att det tar en dag eller ett par dagar innan datan kommer, de är helt fine med det. De vill kunna titta på, åh här finns all grandulär data i ett halvår eller två, jättebra! De behöver hitta statistiska modeller för att förklara saker som sker. Då spelar det inte så stor roll att det tar en timme, en dag eller en vecka för datan att komma in. På det sättet kan jag tycka att de blir lite för tekniskt drivna, sådana här projekt.

**35. GS:** Okej! Ser du några utmaningar rent etiskt som kan uppstå? Du sa att det nästan är pokemonjakt, man tar all data som kan tänkas liksom. Har du någon gång kommit i kontakt med att vi har "de här uppgifterna, de ska vi inte ha i det projektet" eller så där? Eller skulle du säga det också är en utmaning med Big data, eller skulle du säga att övriga saker är utmaningar med Big data?

..

**36. Informant 3:** Om man tar det från ett GDPR perspektiv, eller PuL. Då tror jag egentligen att Big data miljöerna är bättre förberedda för det än gamla legacy systemen och BI-systemen. För BI-system då är det också många lager emellan, problemet att man sparar fel data och sparar data om kunder man inte borde, har jag sett i BI-system länge, och tyvärr hos kunder, att man sparar data om sådana som inte längre är kunder. Det har jag tyvärr sett. Det är inte av anledningen att man vill missbruka det utan för att det har varit tekniskt svårt att åtgärda det. Men det är bra med GDPR, jag tror det är det alla kämpar med, sen är det att rätta till det för att man får väldigt hårda konsekvenser om man inte följer lagarna.

Men det är mer huruvida man sparar saker man inte borde, tex sådana som inte är kunder. Det finns väl lite specialtilfällen där man får göra det en stund. Just att man har processer att ta bort de efter ett halvår, det är det jag tror jag GDPR har fått med sig.

Sen när det gäller Big data och huruvida man kan missbruka den, Ja den här datan är mer känslig och jag tror definitivt ett större behov att diskutera och fundera på hur man vill använda datan.

**37. GS:** Men brukar det alltid finnas syften med datainsamlingen och dataanalysen?

**38. Informant 3:** Det är svårt att säga, det finns lite risk att mitt svar blir representativ för verkligheten och jag vet inte om det är det. Det är inte så att man har gjort hundra olika projekt i det här området. Men av de 5-10 projekten jag gjort så Ja, även innan GDPR har man haft ett syfte att samla in datan, Men mycket av datan man samlar in och framförallt de nya datakällorna och att den blir väldigt mer granulär, det innebär också att datan blir mycket känsligare. Så även om det finns ett syfte att samla in den så kan man missbruka den. Absolut. Och därmed tror jag att företag i en mycket större utsträckning nu än tidigare måste fundera på hur de vill använda datan, och vilka syften. Att man inte går helt bananas på datan och skapar konstiga användarkampanjer eller individuella anpassningar från det, det kan slå fel. Det reglerar sig självt sen förut när datan och tekniken inte fanns.

**39. GS:** Men vad skulle du säga overall, vilken betydelse är Big data idag? För några år sedan som du sa var det ett Buzzword. Vilken betydelse har Big data idag för era kunder och för er?

**40. Informant 3:** För oss är det ett viktigt affärsområde. Det finns mycket investeringar hos våra kunder i det här området. Vi jobbar både med dataanalys och tekniska implementeringar och big data lösningar.

För marknaden skulle jag säga att det finns bolag som Klarna och iZettle. Klarna är kanske ett bra exempel, hela deras affärsmodell bygger på analys och att man har en stor mängd data och att man kan göra det snabbt. Hela deras lösning bygger på nära den här Big data världen och analysvärlden. Spotify och netflix också, finns fler exempel också.

Så för vissa företag är det det som kunnat gjort att de kunnat skapa och bli en framgång, hela produkten, som deras företag bygger kring.

Sen de traditionella spelarna. Hur mycket betyder Big data för en traditionell bank eller liknande? Mer och mer, delvis då att man är rädd för att bli disrupted(?). iTunes då som blev väldigt marginaliserat av Netflix och framförallt Spotify.

Jag tror många av de traditionella företagen ser det mer som ett hot, eller att de måste göra det för annars kommer det någon annan göra det åt dom och förstöra allting. Det blir mer ett försvars sätt. Det känns som det är någonting de måste göra och de hade nog hellre sett att det inte kom. Bankerna hade nog tyckt det var mer bekvämt om det inte kom, med mer digitala konkurrenter och så.

**41. GS:** Haha, ja precis. Men vilken typ av analyser genomför ni för era kunder? Alltså är det allt som du sa, video, textfiler?

**42. Informant 3:** Ja, vi är ett stort bolag, vi gör allt. Vi jobbar väl dels med tekniskt implementation som handlar om integration av lösningar. Där man mer försöker samla in datan, skapa en plattform i fysiska burkar och lagra det. Skapa en miljö att samla in data och sedermera analysera data från. Sen jobbar vi, den gruppen jag tillhör, med analys av data, där vi hjälper kunder analysera data, att skapa insikt och driva ut det till affärsvärde. Jag vet inte riktigt, IBM till exempel är ett väldigt stort bolag, de har väl mer en teknisk approach till det hela. De har watson och sådana grejer.

Jag vill tro att vårt bolag är mer breda än andra spelare i området, vi kan köra allt och göra allt.

**43. GS:** Men hur vet ni att datan är korrekt och kvalitativ när ni utför analyser och sådant?

**44. Informant 3:** Det är en bra poäng, men jag tror att ofta när man sitter och analyserar data, om man har en data lake, de som har ingestat datan 37:33 ska ha gjort en viss kvalitetstest när de integrerat den, men sen också att de har monitorering så att de vet att datan flödar. Men det är svårt, ju större datavolym det är och ju mer komplex data det är desto svårare att säga huruvida att det saknas data.

När vi jobbar med analys så tittar man på datan och vrider och vänder på den. Man kollar, stämmer det, verkar det rimligt och så vidare, sen kanske man går in i en annan mode och försöker göra själva affärsanalysen. Men först handlar det om att man känner sig trygg med datan.

**45. GS:** Har ni något ramverk, att datan ska uppfylla något krav så att ni vet att den är korrekt?

**46. Informant 3:** Bra fråga, det kan vara så att vi har det. Det kanske finns ramverk för de som jobbar med injektion, de som tar in datan som håller på med det. Jag vet faktiskt inte.

**47. GS:** Men vilka anser du är de största etiska problemen med Big data och Big data analys? Eller om det finns några överhuvudtaget.

**48. Informant 3:** För mig är det lite som att säga finns det några etiska problem med metalltillverkning eller nåt. Det är lite för bred pensel att kunna säga. Absolut, det finns. Det finns tillfällen där det sker etiska övertramp med Big data, det är bara att titta på Cambridga analytica och se vad som gjordes där. Jag tror att, som jag var inne på förut, med en större tillgång till data och mer tekniska möjligheter, funktionella möjligheter att analysera data, sätta ihop data på olika sätt så ökar behovet att ha etiska riktlinjer, eller ihopsatta regelverk för hur man ska jobba med sin data. Jag har sett kunder som har specifika plattformar som är dedikerade för ett syfte, där man anonymiserar data på ett speciellt sätt som gör att man kan göra den analys man vill, och egentligen det affärsmässiga värdet bibehålls. Men att man ändå säkrar kundernas integritet. Det finns inte möjlighet i den analysmiljön att missbruka det. Den blir väldigt säker på det sättet då, för den går inte att missbruka analys av fel saker.

**49. GS:** Ni använder inga etiska riktlinjer med Big data analys? Alltså inga skrivna? Alltså utöver GDPR då tänker vi.

**50. Informant 3:** Nej vi har inte det. Eller ja och nej. För det första, integritet är en våra sex core values. Integritet har alltid varit viktigt för oss. Vi kan inte prata om våra uppdrag eller kunder, men våra uppdrag är känsliga och får inte prata med vilka kunder vi jobbar med. Oavsett om man jobbar med dataanalys eller strategi, oavsett vad det är, så har vi en väldigt djupt rotad kultur att inte prata med våra kundengagemang. Det är en viktig del när vi jobbar med analys och big data projekt, att man kan se saker man se saker man inte borde. Man kan se integritetskränkande saker, men det är ingenting vi kan ta med oss och prata med andra om. Ibland jobbar vi med att drifva ett företags finanssystem eller SAP system, vi är inne och kör, vi kan driva vissa saker som är finansiellt känsliga eller personlighetskänslig information. Vi har en intern skyddsprocess som säkrar att vi inte förlorar eller läcker viktig information om våra kunder.



**51. GS:** Men innefattar det också era kunders kunder då?

**52. Informant 3:** Ja exakt! Det gör vi på alla projekt. Man märker att vissa saker är mer viktiga på vissa geografier till exempel religion är någonting som vi i Sverige inte ser som det största faran att missbruka, men i andra länder kanske det är värre.

**53. GS:** Ja men precis. Men då har ni lite riktlinjer och så ju. Men hur tror du GDPR kommer att påverka Big data och Big data analys? och hur har det påverkat om du till och med har exempel på det?

**54. Informant 3:** Ja, jag tror att man kommer att jobba med mer som det här exemplet som jag berättade om att man anonymiserade data. Att den blir då GDPR kompatibel. Det är väl de implementeringar jag sett redan då. Att man jobbar mer med att anonymisera data.

**55. GS:** Pseudonymiserad data är det någonting ni arbetar med också?

**56. Informant 3:** Yes!

**57. GS:** Så de två kommer ökas i samband med GDPR?

**58. Informant 3:** Ja det tror jag. Sen också mer spärrar, man sparar oftast inte data för att gå in i individnivå utan det finns kundtjänstsystem för att kolla vilka ärenden har den här kunden haft etc.

**59. GS:** Att man gör lite mer gruppanalyser då?

**60. Informant 3:** Ja precis, och inte det att man manuellt grupperar dom utan mer att om du gör en analys att du på något sätt täcker över alla unika mönster eller svar som du kan härleda till en person. Det tror jag också är ett sätt att skydda individerna på i slutändan. Jag tror man kommer få se mer många saker kring analys, det är ytterligare ett steg att skydda slutkunderna. Som jag sa förut så tror jag att GDPR i Big data miljö kommer ha viss påverkan. Framförallt att man inte sparar icke kunder, sen tror jag att största utmaningen kring GDPR ligger i de gamla systemen. Man måste få sätt att få ta bort data där och så vidare. Så GDPR slår nog lite hårdare på andra områden än big data faktiskt.

**61. GS:** Tror du att GDPR kommer att hindra de kränkande analyserna som har skett med exempelvis Cambridge Analytica? Trots du det kommer att stävjas, företag som använder Big data på ett skumt sätt liksom?

**62. Informant 3:** GDPR är ju ett europeiskt initiativ, man kan jämföra med hur det är i Kina där det är helt andra synsätt på integritet och vad man kan göra med data. Där man jobbar med ansiktsgenkänning med kameror som finns runt om på staden, där man verkligen kan integritetskränka människor. Det känns det som att där håller det på att bli mer och mer saker som vi finner kränkande. Det är läskigt tycker jag. Men på den europeiska marknaden tror jag den kan ha effekt, men sen tror jag också att det kommer finnas de som ser möjligheter, affärsmöjligheter, eller andra drivande krafter som gör saker som vare sig är legalt eller etiskt försvarbart. Det är väl det som är det intressanta, jag tror att även med GDPR att man får

legala restriktioner, så tror jag att det fortfarande kommer att finnas massa med case där det är etiskt rätt eller inte.

**63. GS:** Ser du några negativa konsekvenser GDPR kommer att ge ur ett Big data perspektiv?

**64. Informant 3:** Ja, jag kan tycka att det försvårar ibland analys, det verkar vara ganska svårt att få en person eller legala avdelningen för ett företag att förstå arbetssättet när man jobbar med analys. När man till exempel analyserar två varianter av en produkt och man vill förstå vilken produkt som skapar mest nöjda kunder, då tror jag att har man inte tillåtelse att analysera eller samla in data för en sådan sak som ändå syftar på att göra sin produkt bättre, jag tror att det i vissa lägen kommer att försvåra analyser i slutändan som faktiskt är bra för kunden. Men det här är hela problemet, att vissa tror att det bara kommer bli bättre och bättre för kunderna, och på vissa vis har det även blivit så, men det även när det inte blivit så och det har missbrukats, och det är väl det GDPR ska råda bot på. Men ja absolut, jag tror att det kommer att försvåra analys som faktiskt är bra för slutkunderna. Det kommer göra det stökigare, i vissa lägen kommer man inte vara helt datadriven när man tar sina beslut.

**65. GS:** Men förutom då att det försvårar analys, finns det andra grejer med GDPR som kan vara negativt? Det sätter käppar i hjulen men, finns det nåt annat? Att det kommer bli mycket krångligare eller kostsamt för er eller era kunder?

**66. Informant 3:** Ja men absolut, det är lite ihopkopplat med det. Mer kostsamt och tidskrävande och därmed också att det inte blir gjort på samma sätt, istället för att göra sina tio olika analyser har man kanske bara råd att göra fem eller tre. Ja absolut det kommer att ge effekter, det kommer bli svårare.

**67. GS:** Aa, har du någon insikt i hur era kunder kommer att.. För nu med GDPR kommer det krävas samtycke för att samla in data, vet du hur det kommer att säkerställas det samtycket rent praktiskt? Eller era kunder kanske redan idag har implementerat det?

**68. Informant 3:** Hmm, ja jag vet inte. Jag har fått lite mejl från Ica, Åhlens och telia. Jag får lite olika GDPR information från de här. Ica och Åhlens var väl mer information tror jag, där behövde jag inte logga in.

**69. GS:** Tycker du de har haft tydlig information, vad de ska göra med datan de samlar in? Att de har ett tydligt syfte och så?

**70. Informant 3:** Både ja och nej, det är på en nivå för att vi gör det här för att förbättra tjänsten. Men sen jag som gör analyser att förbättra tjänsten innebär då att man ska rannsaka data från väldigt många olika håll. Men det som slog mig från Telia var att det inte kändes att man hade så mycket val, den var väl konstruerad på ett sätt som skulle optimera eller maximera antalet som säger ja. Det fanns inte ett tydligt nej, det fanns ett ja eller stäng ner rutan. De hade inte ja eller nej, de hade bara Ja eller inget svar.

**71. GS:** Haha okej. Men är det ungefär så det kommer bli för era kunder också? Om du inte vet så är det lugnt, då är det bara att passa på den.

**72. Informant 3:** Nä men jag vet att kommunikation från en abonnemangsaffär som Telia, så

varje gång man skickar ut mail eller fysiska brev som säger att det här är dina nya villkor. Dels är det jäkligt dyrt för oftast så måste man skicka brev till en hel massa människor, och vissa kanske man kan maila, men varje gång man gör en sådan sak så är det en kostsam sak för ett gäng kunder inte är så aktiva och måste påminnas, och så kanske de påminns att de måste säga upp sina abonnemang för de inte använder dem..

Man måste uppdatera villkoren och göra det på ett så smart sätt som möjligt, som inte driver kunduppsägningar.

**73. GS:** Är det något annat som du har tänkt på som vi inte har tagit upp som du skulle vilja snacka om eller så?

**74. Informant 3:** Nej, inte något sådant där. Men kul diskussion ändå.

**75. GS:** Ja, det är ett sanslöst intressant ämne, ju mer man läser ju mer.. Det finns så många olika vinklar på det. Så det är jätteintressant.

**76. Informant 3:** Hoppas ni fick ut det ni hade hoppats på.

**77. GS:** Ja verkligen, tusen tack!

## 8.5 Transkribering Organisation 4

### Organisation 4.

Intervju 4 – Informant 4 – Organisation 4

Intervjuare – Gustav Stigestadh (GS)

Sekreterare – Felix Moberg

#### Start.

**1. GS:** Vill du berätta lite om dig själv, hur länge du har varit på datainspektionen och hur länge du har kollat på Big data och data analyser och så?

**2. Informant 4:** Jag jobbar som IT-säkerhetsspecialist på datainspektionen, på enheten för myndigheter, vård och utbildning. Jag har jobbat här snart i 14 år och granskar skyddet för personuppgifter och behandlingar av personuppgifter. När det kommer till Big data, termen är hyfsat ny men företeelsen är liksom inte helt ny. Det började med data warehousing och business intelligence och sånt där. Men det har blivit allt mer avancerat och som jag förstår det så räknas Big data i det konceptet väger man in uppgifter som inte lagras enligt den gamla Data warehousing modellen. Där pratar vi ändå om hantering av stora datavolymer, lagrade datavolymer, medan man lika gärna kan ägna sig åt strömmade data när det gäller den mera moderna hanteringen.

**3. GS:** Ja precis. Men om du skulle definiera Big data, har du någon egen definition som du har tänkt på? Just det du snackade om med stora datavolymer?

**4. Informant 4:** Alltså det finns ingen juridisk definition av Big data, det är ett marknadsföringskoncept med mera. Det som är intressant för oss är inte huruvida någonting är eller inte är Big data, eller AI eller machine learning. Det som är intressant för oss är behandling av personuppgifter, och att det sker i enlighet med gällande rätt. Sen vad man kallar det, det är mer en fråga om typ och smak. Men Big data, ja, stora volymer och det kan vara allting från realtidsräkningar till mera maskininlärning och sådant där som man använder det till.

**6. GS:** Jag vet att frågan kanske blir lite konstig för att ni är en myndighet, men ser du några möjligheter med användandet av Big data eller analys med Big data?

**7. Informant 4:** Jo men absolut, självklart är det så. Det är klart att det finns enorma fördelar med det, det kan vara allt inom vård och forskning som direkt tillför människan och mänskligheten. Det är lite så där, finns det några fördelar med att använda IT, klart det finns! Det är vad man använder det till som avgör om det är gott eller ont om man säger så.

**8. GS:** Jo men precis, men vilka är de största utmaningarna tror du med användandet av Big data?

**9. Informant 4:** Från vårt perspektiv blir lite utmaningen, de som vill använda sig utav sådana här tekniker för behandling av personuppgifter, att förhålla sig enligt

ansvarsskyldigheten till gällande lagstiftning. För det finns några av de här grundläggande kraven när man behandlar personuppgifter. För det första får man bara behandla personuppgifter för berättigade ändamål, och det kan jag säga så här långt att det finns ingen som har behandlat en personuppgift som inte har ansett ha haft ett berättigat ändamål, men det är inte riktigt samma sak utan här ska det rent objektivt sätt vara ett berättigat ändamål. Sen får man inte behandla fler uppgifter än som behövs för det berättigade ändamålet, man får inte lagra den en längre tid än vad som behövs för det berättigade ändamålet. Sen får man inte behandla de här uppgifterna för ett ändamål som är inkompatibelt med det ursprungliga ändamålet. Det kan ställa till det lite grand för det här. Nu är det så att man måste hitta en rättslig grund för det berättigade ändamålet och där finns det ett antal olika varianter. Man kan säga att dataskyddet går egentligen ut på att man ska veta vad man gör och varför, och sen ska man se till att det är lagligt.

**10 GS:** Aa.

**11. Informant 4:** Har man den första delen så har man kommit väldigt långt.

Sen när det gäller särskilda kategorier av uppgifter, alltså känsliga personuppgifter. Så är utgångspunkten ett förbud, och då måste man hitta ett applicerbart undantag från det här förbudet, att man får behandla känsliga personuppgifter. Detta gäller ju oavsett som gäller inom ramen för någonting som skulle låta sig beskrivas som Big data, AI, eller någonting sånt. Regelverket är generellt. De som ligger längst fram i krigsfronten kanske inte inte ligger längst fram i insikt i och förhållande till gällande regelverk, även om de naturligtvis borde det och har en skyldighet att följa gällande regelverk. Till exempel då är att känsliga personuppgifter behandlas på ett sätt som innebär en kränkning av den registrerade personliga integritet.

**12. GS:** Precis, så det svåraste är egentligen de som samlar in ska ha ett tydligt syfte och hålla sig till reglerna så att säga.

**13. Informant 4:** När möjligheterna blir så stora, som de är idag, vi har ju stora datasamlingar, lagliga upprättade samlingar inom exempelvis sjukvården och så vidare. Där är väl liksom entreprenörstänket, det ligger väl i entreprenörens natur att inte låta sig begränsas. Där uppstår det då en brist att man behandlar personuppgifter utan ordentligt stöd eller i strid med ett förbud.

**14. GS:** Men vad tror du, vilken betydelse har Big data idag, om man tänker allmänt på företag, sjukhus och er så att säga? Är det en viktig beståndsdel i dagens informationsklimat?

**15. Informant 4:** Det vet jag väl inte riktigt, för det kan komma sig över att jag granskar myndigheter och vård och forskning. Men det dyker upp lite blänkare, som beslutsstöd inom sjukvården till exempel. Men jag ser det som att det är en ganska ung och omogen teknik. Som naturligtvis har mycket stor potential.

**16. GS:** Fick du frågorna av oss?

**17. Informant 4:** Aa, de ligger någonstans här.

**18. GS:** Nä, för de är skrivna utifrån en mall så att säga, de är mer riktade till företag så vi kan hoppa över användningen av Big data.

19. **Informant 4:** Vi använder inte Big data idag.

20. **GS:** Nää precis, så vi hoppar till det som vi finner är mest intressant då egentligen, och det är Big data och GDPR då.

Så anser ni att det finns några etiska problem med Big data och Big data analyser?

21. **Informant 4:** Jo men det finns både sociala och etiska frågor i alla fall. Naturligtvis lagliga frågor som vi arbetar med. De legala förutsättningarna. Där står vi just nu inför ett regelverk skifte av en ganska betydande magnitud, och där är ju förordningen betydligt tydligare i. I förhållande till nuvarande är den ordrikare och därigenom mera beskrivande och förklarande.

22 **GS:** Jo precis, den är framförallt gällande för samtliga EU stater så att det...

23. **Informant 4:** Ja precis, den gäller exakt lika i alla länder, och det är ju en nyhet, och så finns det då lite ytterligare saker. Jag brukar säga så här att väldigt många av de "nyheter" som kommer med GDPR egentligen inte är nyheter, men man måste vara väldigt duktig på att läsa PuL för att läsa ut att det är just det där som krävs. Det nya regelverket är i många hänseenden ett förtydligande om vad som gäller idag. Som det här med privacy by design, inbyggt dataskydd till exempel, man måste vara bra på att förstå och tolka innebörden av de grundläggande kraven i kombination med kraven på att vidta lämpliga säkerhetsåtgärder i PuL för att förstå att det är privacy by design som man ska ha som utvecklingsparadigm liksom. På så vis blir det enklare, tror jag, för de flesta att förstå vad som förväntas av dem när GDPR börjar tillämpas.

24. **GS:** Men vad finns de för etiska frågor i samband med Big data analys?

25. **Informant 4:** De etiska dimensionerna som jag ser tydligast då är just kring forskningsområdet när det gäller känsliga personuppgifter, och där finns det ett helt regelverk kring det också. När det gäller forskning på människor har vi skyddsåtgärder med etikprövningar och etikprövningsnämnderna som är ett annat regelverk för att hantera de etiska aspekterna. Men vi ger inga riktlinjer utifrån våra regelverk. Men jag menar det är bara att ta det senaste från Cambridge Analytica för att se att det finns väldigt viktiga sociala frågor kring det här också.

26. **GS:** Ja men precis, hur tror du att den personliga integriteten skulle kunna kränkas i samband med Big data analys?

27. **Informant 4:** Det beror ju på vad analysen avser, vilka uppgifter det är som behandlas. Behandlingens ändamål, sammanhang och omfattning, som det står i GDPR. Att förhålla sig till GDPR, egentligen, om man är väldigt bra på de här frågorna räcker det med de här principerna som finns i artikel 5, de är liksom som budorden. Där utöver ska man se till att man har rättsligt stöd och sen finns det specialregler för olika kategorier av uppgifter. Sen så har den registrerade ett antal rättigheter och sen så har de som initierat behandlingen av personuppgifter ett antal skyldigheter, lite grann så kan man se på det hela. Vi får då väldigt många frågor om framförallt vilka skyldigheter har vi nu och hur ska vi tolka det här? Då måste man se på de här skyldigheterna som konkretiseringar av principerna. Man måste tolka de här skyldigheterna med de här principerna framför ögonen. På ett väldigt generellt plan kan

man säga att en tolkning som är bra för den registrerade, det är nog en rätt tolkning. En tolkning som inte är det, är ingen bra tolkning. Syftet med detta och de här skyldigheterna, de här stammarna liksom utifrån de här principerna, utifrån korrekthet, transparens, öppenhet, säkerhet, konfidentialitet, etc.

Så att kunskapsbrist gällande regler är ju en stor risk som kan innebära kränkningar av den personliga integriteten.

## 28. GS: Mm

29. **Informant 4:** Om man då ska förklara vad som är kränkande av den personliga integriteten vid behandling av personuppgifter så är det så här att all behandling av personuppgifter är ett intrång inom de registrerades personliga integritet. Och det här intrånget får inte vara otillbörligt, för då är det en kränkning. Och otillbörligt är det här intrånget om det inte sker i enlighet med dataskyddsreglerna. Den risk som jag pekat på här nu då, att den personuppgiftsansvarige som det heter inom vår terminologi, de be bestämmer ändamålen med den här analysen inte förstår vilka skyldigheter den har och rättigheter den registrerade har. Att oavsett hur bra man är på privacy by design och oavsett hur bra man är på att skydda de här uppgifterna och att man kodar de så att de inte finns i direkt utpekande form etc. Inga praktiska åtgärder kan ersätta kravet på en rättslig grund.

## 30. GS: Nää precis.

31. **Informant 4:** Det jag har varit med om ofta är att man missar lite grann, för att man är så inne på att ändamålet är så berättigat ,det här är så viktigt, särskilt forskarna, det här är viktigt för mänskligheten liksom. Vissa har gjort det på ett jättesäkert sätt på rätt och vis, men ni saknar rättslig grund. För det här som ni gör kanske grundas på ett samtycke, men ni har inte samlat in samtycke från alla 200 000 till exempel. Där finns då det kompletterande etiska regelverket också.

32. **GS:** Aa juste. Men när man ska ha en sådan rättslig grund då?

33. **Informant 4:** Det går att finna den i artikel 6.

34. **GS:** Är det ofta att så att folk är duktiga på det tekniska men slopar efter på att de ska ha en rättslig grund för sin undersökning eller insamling eller så?

35. **Informant 4:** Ja så kan det vara.

36. **GS:** Men hur tror du GDPR i övrigt kommer att påverka Big dataanalyser, förutom att man måste ha de här kraven?

37. **Informant 4:** Ja alltså hela den här inför nu, när GDPR ska börjas tillämpas nu den 25e maj, det har skett på ett sådant sätt att man kan hoppas att ganska många har blivit medvetna om de här regelverket. Jag menar mina barns storkusin som är franchisetagande bilmeck har haft GDPR ångest, förstår du vad jag menar?

38. **GS:** Ja

39. **Informant 4:** Det säger någonting av graden av penetration för en sådan här fråga.

Därmed inte sagt att man förstår vad regelverket innebär, men jag tror att det är väldigt många.. Och då är det dels så att de inte har använt sig av avancerade personuppgiftsbehandlings.

**40. GS:** Nä men precis.

**41. Informant 4:** Förhoppningen, i alla fall från min sida är att medvetenheten ska ha höjts, men därmed inte sagt att man förstår vad regelverket kräver av en.

**42. GS:** Jag vet inte hur det funkar för er, men har ni någon hot line för företag att ringa in på för att fråga sånt här?

**43. Informant 4:** Ja det har vi, man kan både ringa och mejla.

**44. GS:** Aha okej, det är ju bra.

Men tror du att nu med GDPR att företag som gör stora Big data analyser och så kommer att använda sig mer utav anonymiserad och pseudonymiserad data?

**45. Informant 4:** Ja alltså på grund av den ökande tydligheten av förordningen. Men återigen det är ingenting som ersätter kraven på rättslig grund. Det är den där käpphästen som kommer tillbaka hela tiden va. För att man ska i största möjliga mån undvika direkta personuppgifter. Är det inte personuppgifter så är GDPR inte tillämpligt, om det är helt anonymiserade uppgifter. Nu har det visat sig att det är ganska svårt att anonymisera stora datamängder, jag vet inte, för några år sedan var det en comeback om metropolit som aktualiserades igen, för det var några som med mycket hög sannolikhet återidentifierat personerna i det här registret. Det här var en stor dataskandal, ni kan googla på det, som involverar ett mycket stort antal människor och uppgifter.

**46. GS:** Vad hette skandalen sa du?

**47. Informant 4:** Metropolit. Det var det som ledde fram till att vi fick en datalag 1973. Så det var ett tag sedan. Så det har visat sig svårare att anonymisera uppgifter än vad man tidigare trott. För sådana här funktioner, om vi pratar nu Big data, kan ju också användas för att åter identifiera uppgifter, och faktum att det är väldigt många uppgifter för varje registrerad så att säga.

**48. GS:** Ja och en personuppgift är väldigt brett också, det är allt ifrån en röstinspelning till ett personnummer.

**49. Informant 4:** Ja! En vanlig fråga; är ip-adress en personuppgift? Då brukar jag säga är blå en personuppgift? Ja, om det står jämte mitt namn och datainspektionen så är blå en personuppgift, eftersom den säger någonting om mig. Det kanske är fel, jag kanske är brunögd, men det är fortfarande en uppgift som är hänförlig till mig.

**50. GS:** Ja precis, så det är svårt att anonymisera allt och få ur det att man inte kan sätta det i samband.

**51. Informant 4:** Ja särskilt i stora datamängder, kombinationen av data kan vara direkt



utpekande till mig i alla fall, även om det inte står mitt namn eller personnummer, eller var jag jobbar eller någonting sådär.

**52. GS:** Ja, men ansvaret hos företaget är väl absolut anser väl ni.

**53. Informant 4:** Ja det är så regelverket är.

**54. GS:** Men om vi tänker någonting negativt med GDPR, finns det något negativt tror du? Något affärsmässigt, eller att det sätter käppar i hjulen eller något annat?

**55. Informant 4:** Man kan säga så här det här är lite tillspetsat och alla jurister håller inte med om det här. Men folk är så upphetsade över rättsliga hinder, men det finns inga rättsliga hinder. Det finns bara rättsliga förutsättningar. Sen finns det sådana här exempel om sekretesslagen kan hindra utlämnande, men då är det ju ett övervägt syfte som har passerat riksdagen. Då har vi demokratiskt gemensamt beslutat att det där ska vi inte göra.

**56. GS:** Nä men precis.

**57. Informant 4:** Och då blir det ju en politisk uppfattning om det är rätt eller fel om vi har bestämt att man inte ska göra så. Det är inget perspektiv vi kan ha som tillsynsmyndighet, vi är ju tjänstemän. Vi ska vara för att citera: "*Regeringens lydiga redskap till förverkligande av statens politik*".

**58. GS:** Haha ja precis, men tror du att det kan komma att få några negativa konsekvenser?

**59. Informant 4:** Jag har ju hört inför att det tillämpas att det är många som har, återigen på grund av okunnighet, vidtagit väldigt drastiska åtgärder och kommit till en massa felslut och kanske förstört en massa uppgifter som de inte kanske hade behövt förstöra och sånt kan naturligtvis skada en verksamhet. Men återigen kommer det tillbaks till att man måste veta vad man gör och varför och se till att det är lagligt då.

**60. GS:** Tror du att det är ett problem att folk har svårt att tolka lagen eller så? Alltså att okunskapen är såpass stor trots att man har haft två år på sig att förbereda sig?

**61. Informant 4:** Ja alltså, som sagt, det här har inte varit en hemlighet under två år i alla fall. Sen så är det en fråga om organisationens mognad. Verksamheter som redan är väldigt reglerade som finanssektorn till exempel, de har tonvis med regelverk som reglerar deras verksamhet och som ställer mycket detaljerade krav. För dem är detta just another regulation.

**62. GS:** Ja och småföretagare som bilmeckare som får panik.

**63. Informant 4:** Ja och det är synd, men det har ju varit ett väldigt ett starkt fokus på frågorna och det är väldigt många som har sagt väldigt mycket om de här administrativa funktionsavgifterna som är en nyhet i förordningen.

**64. GS:** Men har ni märkt någonting med frågor från företag och sådant, det här med samtycke alltså hur företag ska lyckas säkerställa samtycke när man ska samla in data och så?

65. **Informant 4:** Nu är ju samtycke bara en rättslig grund av många. Visst är det så att förordningarna för GDPR, den tydliggör lite grann och man kan säga att den skärper upp lite grann om möjligheterna med samtycke, den begränsar lite vad man kan samtycka till, men det är inte samma sak att alla i alla lägen måste be om den registrerades samtycke. Den rättsliga grunden kan ju finnas någon annanstans. En intresseavvägning eller allmänt intresse, ja artikel 6. Så det är en vanlig missuppfattning, att man helt plötsligt och skaffa samtycken till allting. Däremot så hänger det ihop med någonting annat, att man har informationskyldighet, att man måste informera om den personuppgiftsbehandling man utför. Det är inte samma sak som att alla måste skaffa samtycke.

66. **GS:** Vilka konkreta faktiska konsekvenser får det för dem som behandlar personuppgifter?

67. **Informant 4:** De här skyldighetspaketet som jag sa innan där säkerhet är en och informationsplikten är en och så vidare. Så finns de här med the risk based approach som det kallas. Att man hela tiden har ett risktänk i förhållande till den behandling som man ska utföra, har man skaffat helt rättsligt stöd för att utföra den och bör även reda ut andra alternativ för den här behandlingen.

68. **GS:** Ok.

69. **Informant 4:** Och där kanske riskerna är ganska små för Bengans bilmeck, medans de är lite större för karolinska institutet.

70. **GS:** Senast idag fick jag från Facebook faktiskt och de samlar in samtycken, att man ska klicka i att det är okej att de genomför ansiktsgenkännings analyser.

71. **Informant 4:** Ja och där är det till skillnad från de här Big data som vi pratade om, eller om vi nämnde forskningen till exempel, i Facebooks fall så är den registrerade samma som kunden eller användaren, så är det inte nödvändigtvis med en myndighet. Så därför gäller då naturligtvis andra regler för myndigheter där regeringen ska samtycka. När någon vänder sig direkt till dig med en vara eller tjänst och därför måste registrera uppgifter om dig, då ser informationskyldigheter lite annorlunda ut och jag menar, de kan stödja det på ett samtycke, självklart, det är väl kanske att föredra. Men då finns det vissa krav att uppfylla för att samtycket ska vara vinstfritt. Där är det alltid den personuppgiftsansvarige och i ditt fall Facebook som har bevisbördan.

72. **GS:** Men vi börjar nästan snart bli klara, men GDPR är väl en anmälningsbaserad lag så att säga..

73. **Informant 4:** Vad sa du?

74. **GS:** Att den är anmälningsbaserad, så ni kommer inte att söka aktivt efter företag som bryter mot lagen väl?

75. **Informant 4:** Nej det är precis tvärt om faktiskt, inte om det att vi kommer aktivt att söka för det kommer vi göra naturligtvis, men en av de stora förändringarna i GDPR §36 närmare bestämt, men i Sverige har vi inte tyckt att det är en särskild vettig verksamhet för

datainspektionen att sitta och ta emot sådana här anmälningar, så Sverige har utnyttjat alla de möjligheter som finns i direktivet som ligger i grund till personuppgiftslagen, vi använder oss av alla möjligheter till undantag till den här anmälningsplikten som direktivet medger. Så vi har en miniimplementation, i mina snart 14 har jag varit med en enda gång där ett tillsyns objekt faktiskt har kunnat konstaterat brytat mot den här anmälningsplikten. Det var antipiratbyrån så det var många som tyckte det var väldigt roligt. Det här var 10 år sedan. En annan bakgrund till GDPR är att andra länder, t.ex Frankrike har gått åt helt motsatt håll, man maxiimplementerar den här anmälningsskyldigheten. Så att stora delar av den franska datainspektionens verksamhet bygger på den här anmälningsskyldigheten. Men så har vi aldrig jobbat av naturliga skäl, för att vi har minimerat anmälningsskyldigheten i den Svenska lagstiftningen. Det har det som har varit problemet, regelverket har haft lite olika tillämpningar och har varit väldigt svårt för bolag som har varit verksamma i hela EU, att behöva förhålla sig till olika lagar som ser lite olika ut och utformade efter olika rättstraditioner och sånt där. Det är därför man har gjort GDPR. Då har man tagit bort det här anmälningsskyldigheten och istället för att kompensera så har man lagt till det här kravet med konsekvensbedömningar i artikel 35 och läser man skälen som motiverar artikel 35, en allmän anmälningsskyldighet har inte alltid lett till det bästa personuppgiftsskyddet men har ibland inneburit en stor administrativ börda.

Det är väldigt snirkligt formulerat det här. För då har man resonerat så här att alla kan inte behöva anmäla allting, men om det handlar om någon som gör en behandling där det finns en hög risk för att lagstiftningen inte följs, att det finns en hög risk för den registrerades fri och rättigheter i och med den här behandlingen, då är man skyldig att göra en sedvanlig risk- och sårbarhetsanalys med den som dokumenteras på ett särskilt sätt.

## 76. GS: Okej

77. **Informant 4:** Och då om datainspektionen kommer och säger "detta var en högriskbehandling, kan vi få titta på er konsekvensbedömning" då ska den bara vara att lägga fram.

Så nej GDPR är inte en anmälningsskyldighets reglering, utan tvärtom.

78. **GS:** Men hur fungerar det i praktiken då? Alltså när kommer man se att någon har brutit mot lagen, är det att ni får in ett tips eller är det ni som upptäcker det, eller hur fungerar det?

79. **Informant 4:** Det finns alla sätt. Jag menar alla som behandlar personuppgifter, är skyldiga att samarbeta med tillskyddsmyndigheten i GDPR och jag menar, vi får ju naturligtvis in tips, vi har en omvärldsbevakning, vi tar del av media. Det finns hur många sätt som helst som kan göra att det kommer till datainspektionens kännedom att det finns tveksamma eller problematiska personuppgiftsbehandlingar. Sen har vi hela vår tillsynserfarenhet men det är det som förordningen understryker där, tidigare har det varit så här att det har varit först när datainspektionen har börjat ställa frågor som man själv börjar utreda vad man gör och varför och hur det skulle kunna vara lagligt. Men med ansvarsskyldigheten så ska det kunna levereras när vi ringer och frågar.

## 80. GS: Okej.

81. **Informant 4:** Så den här ansvarsskyldigheten i artikel 5,2 den innebär just det, att man har en skyldighet att veta vad man gör och varför. Man ska kunna visa att det man gör är lagligt

och på vilket sätt det är lagligt och dessutom om det är högriskbehandling ska det vara en fullödig konsekvensbedömning att visa upp.

**82. GS:** Okej så då hade jag missuppfattat det faktiskt, jag trodde att GDPR var anmälningsbaserad men den är baserad på alla möjliga situationer. T.ex om det blir en ny skandal som Cambridge analytica, då hugger väl ni direkt så att säga?

**83. Informant 4:** Haha, ja alltså det är så här att det här gäller exakt lika i hela Europa och om det rör en aktör som bara finns i Sverige, då är det naturligen datainspektionen som så att säga hugger på det. Men om det är en gränsöverskridande behandling då innehåller förordningen, utöver det jag sa om principerna, reglerna kring rättslig grund, de sista delen om förordningen handlar just om vem som ska göra vad och på vilket sätt. Den reglerar de europeiska dataskyddsmyndigheternas samarbete. Om det är någon som har sitt huvudsäte i Paris men även har filialer över halva Europa, då är det ändå så att den franska dataskyddsmyndigheten som ska hantera detta.

**84. GS:** Men jag ska inte haka upp mig på det, men om vi i Sverige har gått ifrån så långt vi kan med det anmälningsbaserade och Frankrike tvärtom, vad innebär det i praktiken att vara ett väldigt anmälningsbaserat lagstadga?

**85. Informant 4:** När man behandlar personuppgifter på ett sätt som träffas av reglernas omfattning, det handlar inte det privata undantaget, det handlar inte om yttrandefrihet, det handlar inte om journalistiska ändamål etc. Då uppstår ett personuppgiftsansvar, med det ansvaret så följer ett antal skyldigheter. I det nuvarande regelverket är en av de skyldigheterna en anmälningsplikt. Men vi har minimerat den skyldigheten så långt direktivet medger, medan i andra länder har gjort andra implementationer av det här. Det är så här att redan när du planerar en behandling av personuppgifter så inträder det här personuppgiftsansvaret och det finns en del skyldigheter som slår till redan innan du börjar behandla uppgifterna, t.ex samla in dem. Det är det här med exempel privacy by design, att man vidtar åtgärder för att säkerställa att uppgifterna behandlas i några andra ändamål, att man inte samlar in fler uppgifter än vad man behöver för ändamålet etc.

Den som besvarar frågorna varför gör vi det här och hur gör vi det här, enligt regelverket åläggs dem massa skyldigheter som dem självständigt och på eget ansvar ska säkerställa att de följer, särskilt nu med GDPR och ansvarsskyldigheten att man verkligen måste kunna visa att man följer lagen och varför. Det är den personuppgiftsansvarige som självständigt och på eget ansvar ser till att man följer alla reglerna, annars så kan man råka ut för oss eller så kan man bli dragen in till tinget och bli krävd på skadestånd. Det är inte bara datainspektionen som övervakar det här regelverket utan den registrerade har ett antal rättigheter, och om den registrerade tycker att de här rättigheterna de tillgodoses inte, då kan de begära skadestånd och det har ingenting med oss att göra, det är helt fristående från vår verksamhet.

**86. GS:** Ja men precis. Men är det något annat som du har tänkt på eller vill berätta?

**87. Informant 4:** Kring Big data?

**88. GS:** Big data och GDPR egentligen då.

**89. Informant 4:** Nä men man måste förstå vad regelverket kräver av en om man vill använda sig utav sådana här tekniker och det innebär en behandling av personuppgifter. Och som jag

har sagt flera gånger och kan upprepa det en sista gång, man ska veta vad man gör och varför och kunna visa att det är lagligt.

**90. GS:** Ja men det är bra avslut tycker jag.

**91. Informant 4:** Ja det är det väl! Har du fått kött på benen?

**92. GS:** Ja verkligen, tack så jättemycket!