AALBORG UNIVERSITET

BIG DATA, BAD DATA: KAN EU'S DATABESKYTTELSESFORORDNING IMØDEGÅ UDFORDRINGERNE FRA BIG DATA?

"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."

NICKLAS MØLLER MIKKELSEN
10. SEMESTER, POLITIK & ADMINISTRATION, AALBORG UNIVERSITET
Vejleder: Andrej Christian Lindholst

Abstract

This thesis concerns the challenges big data (BD) creates for the protection of personal data, known as data protection, in the European Union (EU) and whether the General Data Protection Regulation (GDPR), adopted by the European Parliament and the Council of the European Union in April 2016, is able to counter these BD challenges.

Data protection at EU level emerged on the political stage in the early seventies but not before 1990 did the European Commission (Commission) introduce a data protection package which, among other things, resulted in the adoption of the Data Protection Directive (DPR) in 1995. The rapid advance of technology has previously made national data protection laws obsolete which was one of the reasons why the Commission, in 2009, launched a reform work, within the area of data protection, that concluded in the adoption of the GDPR.

When the GDPR applies from 2018 it has been nine years in the making and in the meantime the technological development has continued with the emergence of, inter alia, BD. BD offers a considerable amount of benefits. But a darker side can also be identified which means that BD can be defined as a potential Weapon of Math Destruction which entails several challenges for data protection in the EU.

In continuation of these circumstances the thesis examined whether the GDPR is able to counter these challenges or whether the GDPR, even before it applies, is antiquated. To clarify this issue, the thesis first charted the challenges created by the drawbacks of BD. It was uncovered that the BD challenges for data protection are divided into two parts. On the one hand, BD creates or enhances phenomena such as dataveillance, secondary uses, profiling, weblining, the black box of BD, and data security threats that puts pressure on data protection as a right. On the other hand, BD undermines traditional data protection methods such as notice and consent, opt-out, and anonymization.

Given a number of theoretical views the thesis was able to devise a five-point model for the dismantling of the BD challenges. The model is comprised of five general strategies for the mitigation of the BD challenges: Privacy through accountability, measures against BD

predictions, the opening of the black box of BD, featurization of BD, and redefinition of the personal data concept.

The thesis, subsequently, used a document analytical content analysis of the GDPR to observe if the strategies, from the five-point model, could be identified on the basis of the reflection that the more of the content from the five-point model that could be detected in the GDPR, the stronger it would be to counter the BD challenges.

The analysis indicated that only two of the strategies, measures against BD predictions and featurization of BD, were incorporated in the GDPR while the privacy through accountability strategy only partially was present. Conversely, it was clear that the other two strategies, the opening of the black box of BD and the redefinition of the personal data concept, was nowhere to be found in the GDPR.

Even though the GDPR goes a long way in order to ensure data protection against BD the thesis concludes that the GDPR lacks the ability to counteract the BD challenges because a number of key data protection questions remains unanswered. The failure of the GDPR to integrate essential content from the five strategies means that the GDPR already looks set to be overtaken by the technological development even before it applies.

Indholdsfortegnelse

1. Indledning

Denne afhandling omhandler, hvilke udfordringer *big data* (BD) skaber for beskyttelsen af fysiske personer ved behandling af deres personoplysninger (databeskyttelse) i Den Europæiske Union (EU) og hvorvidt EU's databeskyttelsesforordning[1] (*forordningen*) er i stand til at imødegå disse.

1.1. Databeskyttelsesdirektivet

Den første databeskyttelseslovgivning på EU-niveau stammer fra den 24. oktober 1995, hvor Europa-Parlamentet (Parlamentet) og Rådet for Den Europæiske Union (Ministerrådet) vedtog databeskyttelsesdirektivet[2] (*direktivet*), som medlemsstaterne herefter skulle inkorporere i deres nationale lovgivning (European Parliament & the Council of the European Union, 1995:31).

Direktivet havde, jævnfør artikel 1(1) og 1(2), to formål: At sikre medlemsstaternes beskyttelse af fysiske personers grundlæggende rettigheder og friheder, særligt retten til privatlivets fred, i tilknytning til behandlingen af personoplysninger og forhindre at medlemsstaterne begrænsede eller forhindrede personoplysningernes frie bevægelighed.

Den nedenstående tabel giver et kort overblik over *direktivets* vigtigste indhold. Bestemmelserne i de nævnte artikler er ikke beskrevet udførligt og der henvises derfor til *direktivet* for en dybere indsigt.

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[2] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

Tabel 1: Vigtige bestemmelser i direktivet

Artikel	Bestemmelse
3	Anvendelsesområde *Direktivet* er gældende for behandling af personoplysninger, der foretages helt eller delvist automatisk samt ikke-automatisk behandling af personoplysninger, der efterfølgende vil blive indeholdt i et register.
6	Principper for personoplysningers kvalitet Personoplysninger skal: (1) behandles rimeligt og lovligt; (2) indsamles til specificerede og legitime formål. Senere databehandling må ej heller være uforenelig med formålene; (3) være tilstrækkelige, relevante og tilpas indskrænkede til kun at opfylde de angivne formål; (4) være korrekte og ajourførte; og (5) opbevares på en sådan måde, der kun tillader identifikation af den registrerede i det tidsrum, som er nødvendigt for opfyldelsen af behandlingens formål.
7	Principper for legitim databehandling Personoplysninger må kun behandles hvis: (1) den registrerede har givet samtykke; (2) behandlingen er nødvendig for opfyldelsen af en kontraktlig forpligtigelse; (3) behandlingen er nødvendig for at overholde en retlig forpligtigelse; (4) behandlingen er nødvendig for at beskytte den registreredes vitale interesser; (5) behandlingen er nødvendig for udførelsen af en opgave i samfundets interesse, som den dataansvarlige er blevet pålagt; og (6) behandlingen er nødvendig for, at den dataansvarlige kan forfølge en legitim interesse, medmindre den registreredes legitime interesser, grundlæggende rettigheder eller friheder overtrumfer dette.
8	Behandling af særlige kategorier af personoplysninger Behandling af personoplysninger omhandlende race eller etnicitet, politiske holdninger, religiøse eller filosofiske overbevisninger, fagforeningsmedlemskab samt helbred og seksuelle forhold forbydes.
10	Oplysningspligt når personoplysninger indsamles hos den registrerede Den dataansvarlige skal minimum stille følgende information til rådighed for den registrerede: (1) den dataansvarliges, eller dennes repræsentants, identitet; (2) formålet med databehandlingen; og (3) information om personoplysninger videregivet til tredjeparter, hvorvidt den registrerede er forpligtiget til at oplyse personoplysninger, konsekvenser ved nægtelse af at give disse oplysninger samt den registreredes ret til indsigt i og berigtigelse af personoplysninger.
11	Oplysningspligt når personoplysninger ikke indsamles hos den registrerede Den dataansvarlige skal, udover informationen i artikel 10, oplyse den registrerede om, hvilken type personoplysninger der er tale om.
12	Retten til indsigt Den registrerede har ret til at få: (1) oplyst om den pågældendes personoplysninger behandles, hvilke formål og kategorier af personoplysninger der ligger til grund for behandlingen, eventuelle modtagere af personoplysninger, meddelt hvor personoplysningerne stammer fra og hvilken logik der ligger bag den automatiserede databehandling; (2) berigtiget, slettet eller blokeret personoplysninger som ikke er blevet behandlet i overensstemmelse med *direktivet*; og (3) underrettet eventuelle tredjeparter om berigtigelser, sletninger eller blokeringer af personoplysninger.

Artikel	Bestemmelse
14	Indsigelsesretten Den registrerede har ret til: (1) at gøre indsigelse, på baggrund af vægtige legitime grunde, imod at den pågældendes personoplysninger behandles, med mindre andet fremgår af national lovgivning. Er indsigelsen berettiget, må den dataansvarlige ikke længere benytte disse personoplysninger i databehandlingen; og (2) gøre indsigelse imod behandling af personoplysninger med henblik på markedsføring.
15	Automatiserede beslutninger Enhver person har ret til ikke at være underlagt afgørelser, som udelukkende er baseret på automatisk databehandling, hvis disse afgørelser har retsvirkning eller påvirker personen i væsentlig grad.
17	Behandlingssikkerhed Den dataansvarlige skal implementere passende tekniske og organisatoriske foranstaltninger, under hensyntagen til det aktuelle tekniske niveau og omkostninger forbundet herved, til at beskytte personoplysninger imod tilfældig eller ulovlig tilintetgørelse eller hændelige tab, ændringer, uautoriseret offentliggørelse eller adgang samt beskytte imod andre former for ulovlig behandling.
18 & 19	Anmeldelsespligt Den dataansvarlige, eller dennes repræsentant, skal notificere tilsynsmyndigheden før en hel eller delvis automatiseret databehandling iværksættes. Notificeringen skal som minimum indeholde: (1) navn og adresse på den dataansvarlige eller dennes repræsentant; (2) databehandlingens formål; (3) en beskrivelse af de registrerede og de personoplysninger som udsættes for behandling; (4) mulige modtagere af personoplysningerne; (5) mulige overførsler af personoplysninger til tredjelande; og (6) en beskrivelse af hvilke foranstaltninger der søsættes for at sikre behandlingssikkerheden.
20	Forudgående kontrol Risikobetonede databehandlinger skal, inden deres igangsættelse, kontrolleres af tilsynsmyndigheden.
25	Overførsel af personoplysninger til tredjelande Overførsel af personoplysninger til tredjelande, med henblik på databehandling, må kun ske, hvis tredjelandet har et tilstrækkeligt databeskyttelsesniveau og *direktivets* bestemmelser overholdes.
27	Adfærdskodeks Medlemsstaterne og EU-Kommissionen skal opfordre til formuleringen af sektorbaserede adfærdskodekser, der skal være med til at sikre en korrekt implementering af *direktivet* i national lovgivning.
28	Tilsynsmyndighed Hver medlemsstat skal sørge for at én eller flere offentlige myndigheder udstyres med ansvaret for, uafhængigt, at monitorere overholdelsen af *direktivet*.

Den teknologiske udvikling på området for informations- og kommunikationsteknologi (IKT) har tidligere gjort national databeskyttelseslovgivning forældet (Mayer-Schönberger, 1998:221-233) og lidt over et årti efter vedtagelsen af den første databeskyttelseslovgivning på EU-niveau var der bred enighed om, at den samme skæbne nu var overgået *direktivet* (Hornung, 2012:65).

1.2. Databeskyttelsesforordningen

Herfor præsenterede EU-Kommissionen (Kommissionen) i januar 2012 et forslag til en omfattende reform af EU's databeskyttelsesregler (European Commission, 2012a:1). Reformen bestod af to dele; en forordning og et direktiv. Forordningen skulle erstatte *direktivet* fra 1995 og ville etablere en generel EU-databeskyttelseslovgivning, mens reformens direktiv omhandlede databeskyttelse i forbindelse med forebyggelse, efterforskning og retsforfølgning af overtrædelser af straffeloven (European Commission, 2012b:4). Det er reformens forordning, der vil være genstand for afhandlingens undersøgelse.

Kommissionen nåede i december 2015 til enighed med Parlamentet og Ministerrådet om *forordningen*, som efterfølgende i april 2016 formelt blev vedtaget af de to organer med anvendelse fra den 25. maj 2018 (European Commission, 2015:1; European Parliament & the Council of the European Union, 2016:87-88).

Forordningens præambel peger på to grundlæggende udviklinger, som har skabt behovet for en opdatering af databeskyttelseslovgivningen i EU. Den hastige teknologiske udvikling og globaliseringen har øget omfanget af indsamlingen og delingen af personoplysninger markant og samtidig har den økonomiske og sociale integration i EU, på grund af det indre marked, ført til en substantiel stigning i den grænseoverskridende udveksling af personoplysninger. (European Parliament & the Council of the European Union, 2016:1-3)

EU ønsker fortsat at være med til at understøtte personoplysningernes frie bevægelighed for at opfylde målsætningen om etableringen af et fuldt fungerende digitalt indre marked. Det er dog vigtigt, at den frie bevægelighed for personoplysninger ikke sker på bekostning af databeskyttelsen, som er en grundlæggende rettighed i EU. (Ibid.)

Det er i denne balancegang, at *forordningen* indtager en nøglerolle på to måder. For det første er *direktivets* bestemmelser blevet implementeret og anvendt forskelligt i medlemsstaterne og har dermed ikke formået, i tilfredsstillende omfang, at fordre en harmonisering af den nationale databeskyttelseslovgivning. Dette er til hinder for den frie bevægelighed og kreeringen af et digitalt indre marked. For det andet har den divergerende implementering og anvendelse betydet, at databeskyttelsen er fragmenteret på tværs af medlemsstaterne og er blevet svækket. (Ibid.)

Sigtet med *forordningen* er at skabe denne manglende harmonisering, idet den i sin egenskab som EU-retsakt er direkte anvendelig i den enkelte medlemsstat. Hermed er forhåbningen, at *forordningen* vil bidrage til den fortsatte frie bevægelighed af personoplysninger og på samme tid skabe en ensrettet og stærk databeskyttelse i hele EU. (Ibid.)

De vigtigste bestemmelser i *forordningen* er angivet i tabellen på de følgende sider. Bestemmelserne i artiklerne er ikke beskrevet udførligt. Der henvises til selve *forordningen*, hvis et nærmere indblik ønskes.

Tabel 2: Vigtige bestemmelser i forordningen

Artikel	Bestemmelse
2 & 3	Anvendelsesområde *Forordningen* gælder for behandling af personoplysninger, der foregår helt eller delvist automatiseret samt ikke-automatiseret behandling af personoplysninger, der vil blive indeholdt i et register. *Forordningen* finder ligeledes anvendelse for dataansvarlige eller databehandlere etableret uden for EU, hvis deres behandling af personoplysninger vedrører en registreret bosiddende i EU og databehandlingen er relateret til: (1) udbud af varer eller services til registrerede i EU; og (2) overvågning af registreredes adfærd hvis denne adfærd udfolder sig i EU.
5	Principper for behandling af personoplysninger Personoplysninger skal baseres på principperne: (1) lovlighed, rimelighed og gennemsigtighed; (2) formålsbegrænsning; (3) dataminimering; (4) rigtighed; (5) opbevaringsbegrænsning; (6) integritet og fortrolighed; samt (7) ansvarlighed.
6	Lovlig databehandling Databehandling er kun lovlig, når ét af følgende forhold er gældende: (1) der er indhentet samtykke fra den registrerede; (2) behandlingen er nødvendig for at opfylde og vedligeholde en kontraktlig forpligtigelse; (3) behandlingen er nødvendig for at overholde en retlig forpligtigelse; (4) behandlingen er nødvendig for at beskytte den registreredes, eller en anden fysisk persons, vitale interesser; (5) behandlingen er nødvendig for at udføre en opgave i samfundets interesse, som kan være blevet pålagt den dataansvarlige; og (6) behandlingen er nødvendig for, at den dataansvarlige, eller en tredjepart, kan forfølge legitime interesser, så længe disse ikke trumfes af den registreredes grundlæggende rettigheder og friheder. Medlemsstaterne kan vedtage mere specifikke bestemmelser for at tilpasse appliceringen af *forordningen*. Når personoplysninger påtænkes at blive behandlet til et andet formål, end det de var indsamlet til i første omgang og behandlingen ikke er baseret på samtykke, EU-retten eller national lovgivning, skal den dataansvarlige vurdere, hvorvidt det nye databehandlingsformål er foreneligt med det oprindelige formål.
7	Betingelser for samtykke Er databehandlingen baseret på et samtykke, skal den dataansvarlige kunne påvise, at den registrerede har givet sit samtykke til behandlingen. Den registrerede har ret til at trække sit samtykke tilbage på hvilket som helst tidspunkt. Det skal være lige så let at trække sit samtykke tilbage, som det er at give sit samtykke.
9	Behandling af særlige kategorier af personoplysninger Behandling af personoplysninger om race eller etnisk oprindelse, politiske holdninger, religiøse eller filosofiske overbevisninger, fagforeningsmedlemskab, helbred, seksuelle forhold eller seksuelle orientering samt behandling af genetisk eller biometrisk data, med identificering af en fysisk person for øje, er forbudt.

Artikel	Bestemmelse
13	Oplysningspligt når personoplysninger indsamles hos den registrerede Indsamles personoplysninger fra en registreret skal den dataansvarlige give den registrerede følgende oplysninger: (1) identitet og kontaktoplysninger på den dataansvarlige eller dennes repræsentant; (2) kontaktoplysninger på den dataansvarliges *data protection officer* (DPO); (3) formålet med databehandlingen og dennes retsgrundlag; (4) den dataansvarliges, eller en tredjeparts, legitime interesser hvis databehandlingen er baseret på artikel 6(1.f); (5) eventuelle modtagere af personoplysningerne; (6) hvorvidt den dataansvarlige agter at overføre personoplysningerne til et tredjeland eller en international organisation. Herudover skal den dataansvarlige, for at sikre en rimelig og gennemsigtig databehandling, stille følgende oplysninger til rådighed for den registrerede: (1) tidsrummet for opbevaringen af personoplysningerne eller, hvis dette ikke er muligt, de kriterier der er med til at afgøre dette tidsrum; (2) retten til indsigt, berigtigelse, sletning, indsigelse og dataportabilitet; (3) retten til at tilbagetrække samtykker hvis databehandlingen er baseret på artiklerne 6(1.a) eller 9(2.a); (4) retten til at klage til en tilsynsmyndighed; (5) om indsamlingen af personoplysninger er baseret på et lovpligtigt eller et kontraktuelt grundlag, hvorvidt den registrerede er forpligtiget til at give personoplysningerne og hvilke konsekvenser der er forbundet med ikke at give disse personoplysninger; og (6) hvorvidt der forekommer automatiske afgørelser og hvis dette er tilfældet, oplysninger om logikken bag den automatiske afgørelse samt mulige konsekvenser ved denne form for databehandling. Påtænker den dataansvarlige at behandle personoplysninger til et andet formål, end det der blev givet i første omgang, skal den registrerede oplyses herom.
14	Oplysningspligt når personoplysninger ikke indsamles hos den registrerede Den dataansvarlige skal give den registrerede de samme oplysninger som i artikel 13 samt: (1) hvilke personoplysninger det drejer sig om; (2) den dataansvarliges, eller en tredjeparts, legitime interesser hvis databehandlingen grunder i artikel 6(1.f); og (3) hvilke kilder personoplysningerne stammer fra.
15	Retten til indsigt Den registrerede har ret til at få en bekræftelse på, om en dataansvarlig behandler den pågældendes personoplysninger. Er dette tilfældet, skal den registrerede have adgang til følgende information: (1) databehandlingens formål; (2) hvilke personoplysninger der er tale om; (3) eventuelle modtagere af personoplysningerne; (4) tidsrummet for opbevaringen af personoplysningerne eller, hvis dette ikke er muligt, de kriterier der er med til at afgøre dette tidsrum; (5) den registreredes ret til berigtigelse, sletning, begrænsning eller indsigelse imod databehandling; (6) retten til at klage til en tilsynsmyndighed; (7) hvor personoplysningerne stammer fra, hvis de ikke er indhentet fra den registrerede; (8) tilstedeværelsen af automatiske afgørelser og hvis ja, oplysninger om logikken bag de automatiske afgørelser samt mulige konsekvenser for den registrerede.
16	Retten til berigtigelse Den registrerede har ret til at få berigtiget forkerte eller unøjagtige personoplysninger om sig selv.

Artikel	Bestemmelse
17	Retten til sletning (*right to be forgotten*) Den registrerede har ret til at få slettet personoplysninger om sig selv hvis: (1) personoplysningerne ikke længere er nødvendige for at opnå det formål, de blev indsamlet til; (2) den registrerede trækker samtykket tilbage og der ikke ligger et andet retsgrundlag til grund for databehandlingen; (3) den registrerede gør indsigelse imod databehandlingen og der forefindes andre legitime grunde til databehandlingen, der går forud for den registreredes indsigelse; (4) behandlingen af personoplysninger er ulovlig; (5) personoplysningerne skal slettes, for at den dataansvarlige kan efterleve en retlig forpligtigelse i EU-retten eller national lovgivning; og (6) personoplysningerne er blevet indsamlet hos et barn i forbindelse med en informationssamfundstjeneste i relation til artikel 8(1).
18	Retten til begrænsning Den registrerede har ret til at få begrænset behandlingen af personoplysninger om sig selv, hos en dataansvarlig, hvis: (1) personoplysningernes rigtighed bestrides af den registrerede; (2) databehandlingen er ulovlig, men den registrerede ikke ønsker at få slettet personoplysningerne og derfor i stedet begærer, at de skal begrænses; (3) der ikke længere er behov for personoplysningerne i databehandlingen, men de fortsat er nødvendige for at kunne fastlægge, udføre eller forsvare retskrav; (4) den registrerede har gjort indsigelse imod databehandlingen. Begrænses behandlingen af personoplysninger må disse personoplysninger, med opbevaring som undtagelse, kun behandles med den registreredes samtykke, for at et retskrav kan fastlægges, udføres eller forsvares, for at beskytte en anden fysisk persons rettigheder eller for at beskytte vigtige samfundsinteresser.
20	Retten til dataportabilitet Den registrerede har ret til at få udleveret personoplysninger om sig selv af en dataansvarlig i et struktureret, almindelig anvendt og maskinlæsbart format samt at få disse personoplysninger transmitteret til en anden dataansvarlig når: (1) databehandlingen er baseret på samtykke eller på et kontraktligt forhold; og (2) databehandlingen foretages automatisk.
21	Retten til indsigelse Den registrerede har ret til at gøre indsigelse imod behandling af personoplysninger, der er baseret på artikel 6(1.e) og 6(1.f). Herefter må den dataansvarlige ikke længere behandle personoplysninger, medmindre den dataansvarlige kan redegøre for legitime grunde til databehandlingen, der ikke går forud for den registreredes interesser, grundlæggende rettigheder og friheder eller at databehandlingen er afgørende for at kunne fastlægge, udføre eller forsvare et retskrav. Bliver personoplysninger gjort til genstand for databehandling med henblik på direkte markedsføring, har den registrerede ret til at gøre indsigelse imod dette. Sker dette må personoplysningerne ikke længere anvendes til dette formål.
22	Automatiserede afgørelser Den registrerede har ret til ikke at være underlagt en afgørelse, der udelukkende er baseret på automatisk databehandling, hvis denne afgørelse har retsvirkning eller i udpræget grad påvirker den pågældende.

Artikel	Bestemmelse
24	Den dataansvarliges ansvar Den dataansvarlige skal, under hensyntagen til databehandlingens karakter, omfang, kontekst, formål og mulige risici for fysiske personers rettigheder og friheder, implementere passende tekniske og organisatoriske foranstaltninger for at sikre, og for at være i stand til at påvise, at databehandlingen udføres i overensstemmelse med *forordningen*.
25	Databeskyttelse gennem *privacy by design* og *privacy by default* Den dataansvarlige skal, under hensyntagen til det aktuelle tekniske niveau, implementeringsomkostninger samt databehandlingens karakter, omfang, kontekst, formål og mulige risici for fysiske personers rettigheder og friheder, implementere passende tekniske og organisatoriske foranstaltninger, som er designet til at integrere databeskyttelsesprincipper, f.eks. dataminimering, i databehandlingen for at sikre at der leves op til *forordningens* krav og registreredes rettigheder beskyttes. Tillige skal den dataansvarlige implementere passende tekniske og organisatoriske foranstaltninger for at sikre, gennem en teknologis standardindstillinger, at der kun indsamles de personoplysninger der er nødvendige for at opnå databehandlingens formål.
28	Databehandler Skal databehandlingen foretages på vegne af den dataansvarlige, benytter denne, kun databehandlere der kan garantere, at de vil implementere passende tekniske og organisatoriske foranstaltninger, således at deres databehandling lever op til *forordningens* krav og sikrer registreredes rettigheder.
30	Journalisering af behandlingsaktiviteter Dataansvarlige, eller dennes repræsentant, skal føre protokol over de behandlingsaktiviteter, som den dataansvarlige selv har ansvaret for og behandlingsaktiviteter, som en databehandler foretager på vegne af den dataansvarlige.
32	Behandlingssikkerhed Den dataansvarlige og databehandleren skal, under hensyntagen til det aktuelle tekniske niveau, implementeringsomkostninger samt databehandlingens karakter, omfang, kontekst, formål samt risiciene for fysiske personers rettigheder og friheder, implementere passende tekniske og organisatoriske foranstaltninger, der sikrer et tilfredsstillende sikkerhedsniveau.

Artikel	Bestemmelse
33 & 34	Notificering af tilsynsmyndigheden og den registrerede ved brud på behandlingssikkerheden Ved brud på behandlingssikkerheden skal den dataansvarlige, senest 72 timer efter at være blevet bekendt med sikkerhedsbruddet, orientere tilsynsmyndigheden, medmindre sikkerhedsbruddet ikke indebærer risici for fysiske personers rettigheder og friheder. Er det sandsynligt at sikkerhedsbruddet vil medføre risici for fysiske personers rettigheder og friheder, skal den dataansvarlige ligeledes orientere de berørte registrerede.
35 & 36	Konsekvensanalyse og forudgående konsultering Når en dataansvarlig påtænker at igangsætte en databehandling, der baserer sig på ny teknologi og sandsynligvis vil medføre høje risici for fysiske personers rettigheder og friheder, skal den dataansvarlige udfærdige en konsekvensanalyse af databehandlingens konsekvenser for databeskyttelsen. Den dataansvarlige skal konsultere tilsynsmyndigheden, hvis konsekvensanalysen viser, at databehandlingen indebærer høje risici for fysiske personers rettigheder og friheder ved manglende implementering af tekniske og organisatoriske foranstaltninger. Er tilsynsmyndigheden af den opfattelse, at databehandlingen er i modstrid med *forordningen*, særligt hvis den dataansvarlige ikke på tilfredsstillende vis har identificeret risiciene eller søgt at begrænse dem, skal tilsynsmyndigheden rådgive, jf. dens beføjelser i artikel 58, den dataansvarlige.
37 & 38	*Data Protection Officer* En dataansvarlig, eller databehandler, skal udpege en DPO hvis: (1) databehandlingen foretages af en offentlig myndighed med undtagelse af domstole; (2) den dataansvarliges, eller databehandlerens, hovedaktiviteter udgøres af databehandling, der kræver kontinuerlig og systematisk monitorering af registrerede i stort omfang; eller (3) den dataansvarliges, eller databehandlerens, hovedaktiviteter består af databehandling af særlige kategorier af personoplysninger eller personoplysninger vedrørende straffedomme og lovovertrædelser. En DPO skal involveres tilstrækkeligt og rettidigt i alle databeskyttelsesspørgsmål og skal i sin rolle kunne fungere uafhængigt og uden at modtage instrukser fra andre. Herudover må en DPO ikke blive afskediget eller straffet på grund af sin opgaveudførelse.
40	Adfærdskodeks Medlemsstaterne, tilsynsmyndighederne, *European Data Protection Board* (EDPB) og Kommissionen skal opfordre til etableringen af adfærdskodekser, som kan understøtte en korrekt anvendelse af *forordningen*, der tager hensyn til specifikke sektorer samt små og mellemstore virksomheders særlige behov.

Artikel	Bestemmelse
42	Certificering Medlemsstaterne, tilsynsmyndighederne, EDPB og Kommissionen skal opfordre til, især på EU-niveau, til etableringen af databeskyttelsescertificeringsmekanismer og databeskyttelsesmærkninger med det formål, at dataansvarlige og databehandlere kan demonstrere, at deres databehandlingsaktiviteter overholder *forordningen*. Certificeringen skal være frivillig og gennemføres på en gennemsigtig måde.
45, 46, 48 & 49	Overførsel af personoplysninger til tredjelande eller internationale organisationer Overførsel af personoplysninger til databehandling i et tredjeland eller en international organisation kan foregå, hvis Kommissionen har besluttet at tredjelandet, eller et område eller en specifik sektor i det pågældende tredjeland, eller den internationale organisation har et tilfredsstillende databeskyttelsesniveau. Hvis der ikke forefindes en sådan beslutning fra Kommissionen må en dataansvarlig eller databehandler kun overføre personoplysninger til et tredjeland eller en international organisation, hvis den dataansvarlige eller databehandleren har sikret de fornødne foranstaltninger og garanteret tilgængeligheden af rettigheder og stærke retsmidler for registrerede. En dom afsagt af en domstol eller en afgørelse truffet af en administrativ myndighed i et tredjeland som påkræver, at en dataansvarlig eller databehandler skal udlevere personoplysninger er kun gyldig, hvis den bygger på en international aftale mellem tredjelandet og EU eller en medlemsstat. Forefindes der ikke en Kommissionsafgørelse om et tiltrækkeligt databeskyttelsesniveau i et tredjeland eller en international organisation eller fornødne foranstaltninger og garantier kan overførsel af personoplysninger kun finde sted hvis: (1) den registrerede har givet et udtrykkeligt og velinformeret samtykke; (2) overførslen er nødvendig for vedligeholdelsen og opfyldelsen af en kontraktlig forpligtigelse mellem den registrerede og den dataansvarlige; (3) overførslen er nødvendig for vedligeholdelsen og opfyldelsen af en kontraktlig forpligtigelse mellem den dataansvarlige og en anden fysisk eller juridisk person i den registreredes interesse; (4) overførslen har en vigtig samfundsinteresse; (5) overførslen er nødvendig for fastlæggelsen, udførelsen eller forsvaret af et retskrav; (6) overførslen er nødvendig for at beskytte den registreredes eller en anden fysisk persons vitale interesser, hvis den registrerede ikke er fysisk eller juridisk i stand til at afgive sit samtykke; og (7) overførslen finder sted fra et offentligt tilgængeligt register.

Artikel	Bestemmelse
51	Tilsynsmyndighed Hver medlemsstat skal sørge for, at én eller flere uafhængige offentlige myndigheder er ansvarlige for monitoreringen af anvendelsen af *forordningen* for at beskytte fysiske personers grundlæggende rettigheder og friheder samt fremme personoplysningers frie bevægelighed i EU.
68, 69, 70 & 71	European Data Protection Board *Forordningen* opretter EDPB som et uafhængigt EU-organ. EDPB skal sikre en ensartet og stringent efterlevelse af *forordningen* på tværs af medlemsstaterne. I dette øjemed skal EDBP udøve tilsyn, rådgive Kommissionen, udstede anbefalinger, henstillinger og retningslinjer, fremme *best practice* inden for databeskyttelse, øge samarbejdet mellem de nationale tilsynsmyndigheder og udarbejde en årlig rapport, hvori der redegøres for tilstanden på databeskyttelsesområdet i EU.
83	Administrative bøder Tilsynsmyndigheden kan, ved overtrædelse af *forordningen*, sanktionere med administrative bøder. De administrative bøder kan være op til 20 mio. € og hvis der er tale om en virksomhed, kan bøden udgøre op til 4% af dennes årlige globale omsætning, såfremt dette beløb er højere end de 20 mio. €.

Det er afgørende, at *forordningen* formår at sikre en ensrettet og stærk databeskyttelse, da dette er indskrevet som en grundlæggende rettighed i EU, hvilket fremgår af artikel 8 i EU's Charter om Grundlæggende Rettigheder: *"Everyone has the right to the protection of personal data concerning him or her"* (European Union, 2010:10).

Denne grundlæggende rettighed er ydermere blevet traktatsikret gennem inkorporeringen i Traktaten om Den Europæiske Union (TEU) og Traktaten om Den Europæiske Unions Funktionsmåde (TEUF) i henholdsvis artikel 6 og artikel 16 (European Union, 2010:19-55).

1.3. Forældet inden anvendelse?

Arbejdet med at udarbejde *forordningen* kan spores helt tilbage til maj 2009, hvor Kommissionen igangsatte en revision af EU's lovgivning på databeskyttelsesområdet (European Commission, 2010). Dermed havde *forordningen* været undervejs i syv år inden dens vedtagelse og vil på tidspunktet for dens anvendelsesdato allerede have ni år på bagen. I forlængelse af dette opstår problemstillingen om, hvorvidt den hastige teknologiske udvikling i mellemtiden har overhalet *forordningen* indenom og endnu engang gjort databeskyttelseslovgivning utidssvarende.

Den teknologiske udvikling på IKT-området er, i de seneste år, virkelig accelereret, hvilket har ført til fremkomsten af BD. Grundlaget for BD er den eksplosion i generering af data og information, der er foregået det seneste årti og som udelukkende forventes at stige eksponentielt fremadrettet. (Mayer-Schönberger & Cukier, 2014:6; Kitchin, 2014:69)

Brancheforeningen TechAmerica definerer BD som:

> *"(...) large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information."* (TechAmerica, 2012:10)

Google Flu Trends er et skoleeksempel på udfoldelsen af BD. Tilbage i 2009 præsenterede Google deres såkaldte *Google Flu Trends*, der kunne forudsige influenzaudbrud i USA én til to uger før de amerikanske sundhedsmyndigheder. Dette skyldtes at Google, på baggrund af en korrelationsanalyse af tidligere influenzaudbrud og 50 mio. af de mest indtastede søgeforespørgsler på deres søgemaskine fra 2003 til 2008, fandt frem til, at hvis en række bestemte søgeforespørgsler figurerede blandt de mest søgte, var der en stor sandsynlighed for, at et influenzaudbrud var under opsejling. (Ginsberg et al., 2009: 1012-1014)

Aktører der formår at anvende og beherske BD kan opnå en række fordele og gevinster såsom bedre beslutningstagning, mere innovation, reduktion af omkostninger og større effektivitet (Manyika et al., 2011:61-83). Der kan dog også identificeres en skyggeside ved BD. Cathy O'Neil, en tidligere dataanalytiker, betegner, på baggrund af denne skyggeside, BD som et *Weapon of Math Destruction* (O'Neil, 2016:3).

Skyggesiden relaterer sig særligt til de mange fordele og gevinster, som BD stiller i udsigt til dens brugere. Fordelene og gevinsterne betyder, at der fremover vil være endnu flere der indsamler, anvender og gemmer et væld af data og information, herunder personoplysninger som kan misbruges. Den stigende anvendelse af personoplysninger, på både gængse men også nye måder blandt flere og flere aktører i samfundet, sætter ikke bare databeskyttelsen under pres men også selve privatlivet. (Mayer-Schönberger & Cukier, 2014:151)

En hændelse, vedrørende den amerikanske internetudbyder AOL, kan eksemplificere det BD-skabte pres på databeskyttelsen. I 2006 frigav AOL, til brug i forskningsprojekter, en anonymiseret database med over 20 mio. søgeforespørgsler fra omkring 650.000 kunder. På trods af anonymiseringen lykkedes det for journalister at identificere kunde nummer 4.417.749 som den 62-årige Thelma Arnold fra Lilburn i USA. (Barbaro & Zeller Jr., 2006)

BD er et af de seneste skud på stammen fra den teknologiske udvikling på IKT-området og der er bred enighed om, at der er behov for at nytænke reglerne på databeskyttelsesområdet (Kitchin, 2014:170). Blot at opdatere eller revidere de nuværende databeskyttelsesregler er ikke nok, hvilket skyldes at BD ændrer selve spillereglerne på området:

"If the threat is simply larger, then the laws and rules that protect privacy may still work in the big-data age; all we need to do is redouble our existing efforts. On the other hand, if the problem changes, we may need new solutions. Unfortunately, the problem has been transformed." (Mayer-Schönberger & Cukier, 2014:151)

Spørgsmålet er derfor, om *forordningen* er nytænkende nok til at kunne imødegå udfordringerne fra BD og sikre en stærk databeskyttelse?

1.4. Problemformulering

Med det ovenstående in mente arbejdes der i afhandlingen ud fra følgende problemformulering:

Hvilke udfordringer giver BD for databeskyttelsen i EU og kan forordningen modvirke disse?

1.5. Afhandlingens videre forløb

Det følgende kapitel omhandler det anvendte forskningsdesign og undersøgelsesmetode. Forskningsdesignet karakteriseres som værende et beskrivende casestudie. Derefter foretages en operationalisering af casestudiets analyse. Det efterfølgende afsnit redegør for

undersøgelsesmetodens dokumentanalyse, hvor der blandt andet knyttes kommentarer til *forordningens* repræsentativitet og mening, samt indsamlingen og udvælgelsen af empiri.

Kapitel tre udfolder de empiriske overvejelser, der først opridser udviklingen af databeskyttelsesområdet i EU og herefter udspecificeres BD-begrebet med gennemgang af blandt andet *de tre V'er*, grundlaget for BD og dets datakilder.

I det fjerde kapitel udrulles teorifeltet bestående af to dele. Den første del kortlægger de udfordringer, som BD's todelte skyggeside giver for databeskyttelsen. I anden del identificeres strategier til at afbøde de udfordringer, som BD's skyggeside giver for databeskyttelsen. Strategierne opsummeres afslutningsvist i en fempunktsmodel, som vil udgøre omdrejningspunktet for den efterfølgende analyse.

Femte kapitel udgøres af analysen der afklarer, hvorvidt fempunktsmodellens strategier kan identificeres i *forordningen*. Herefter diskuteres det, på baggrund af analysens resultater, om *forordningen* er i stand til at afbøde BD-udfordringerne for databeskyttelsen.

Slutteligt opsummeres og afrundes afhandlingen gennem konklusionen i kapitel seks.

2. Forskningsdesign og undersøgelsesmetode

I indeværende kapitel omtales det anvendte forskningsdesign og undersøgelsesmetode, som danner rammen for den senere analyse.

2.1. Forskningsdesignet

2.1.1. Databeskyttelsesforordningen som case

Afhandlingen søger gennem et casestudie at afgøre, hvorvidt *forordningen* er opbygget på en sådan måde, at den kan afbøde BD-udfordringerne og derved sikre databeskyttelsen. Til dette formål bliver *forordningen* anvendt som case.

2.1.2. Det beskrivende eller forklarende forskningsdesign

I forlængelse af dette er det væsentligt at få fastlagt, om casestudiet kan karakteriseres som værende beskrivende eller forklarende.

Generelt kan forskellen på det beskrivende og forklarende forskningsdesign siges at komme til syne gennem undersøgelsesspørgsmålets formulering. Beskrivende forskningsdesign stiller spørgsmålet "hvad" der foregår, mens forklarende forskningsdesign omvendt fokuserer på "hvorfor" det foregår (de Vaus, 2001:1).

Skarpt optrukket vil det beskrivende forskningsdesign for eksempel koncentrere sig om at beskrive, hvilke BD-elementer der skaber udfordringer for databeskyttelsen. Omvendt vil et forklarende forskningsdesign søge at forklare, hvorfor disse udfordringer er et problem for databeskyttelsen.

Afhandlingens forskningsdesign må siges at tilhøre den beskrivende gren, idet undersøgelsesspørgsmålet søger at "beskrive", hvorvidt *forordningen* kan modvirke BD-udfordringerne for databeskyttelsen.

To øvrige elementer diskvalificerer forskningsdesignet fra at kunne beskrives som forklarende. For det første er det forklarende forskningsdesign enten teoritestende eller teoribyggende (de Vaus, 2001:5-8). Da denne afhandling, på baggrund af observationer fra *forordningen*, hverken forsøger at verificere eller falsificere teoretiske overvejelser eller

forsøger at udvikle teoretiske generaliseringer på databeskyttelsesområdet gennem induktion, er det tydeligt, at der ikke er tale om et forklarende forskningsdesign. For det andet er *forordningen* endnu ikke trådt i kraft, hvilket vanskeliggør et forklarende forskningsdesign, da retrospektive og prospektive observationer samt vurderinger i vidt omfang er udelukket (de Vaus, 2001:227-228).

2.1.3. Operationalisering af det beskrivende casestudie

Casestudiet omhandler en analyse af, hvorvidt *forordningen* er sammensat således, at BD-udfordringerne for databeskyttelsen kan afbødes.

Til dette formål anvendes en idiografisk forklaringsmåde hvor databeskyttelsen i EU og BD forklares grundigt. Herved gives en komplet og dybdegående indsigt i de kontekster, der udgør genstandsfeltet for casestudiet, hvilket baner vejen for en høj intern validitet. (de Vaus, 2001:233-234)

For at casestudiet bliver i stand til at afgøre, om *forordningen* har mulighed for at opretholde databeskyttelsen overfor BD-udfordringerne, er det nødvendigt med en kortlægning af, hvilke databeskyttelsesudfordringer BD enten forstærker eller skaber.

Herfor foretages der i teorikapitlet først en teoretisk diskussion, der fastlægger, hvilke udfordringer BD giver for databeskyttelsen. Kortlægningen af BD-udfordringerne afgrænser hvilke teoretiske løsninger, der er relevante at indlemme i den senere analyse. Dette skyldes at kortlægningen udstikker retningen for, hvor der i den teoretiske litteratur skal indhentes løsninger.

Herefter oprulles teoretiske anskuelser, der opstiller strategier for hvordan BD-udfordringerne afbødes. Disse strategier, og deres instrumenter, fremstilles i en fempunktsmodel.

Analysen af *forordningens* evne til at modvirke BD-udfordringerne for databeskyttelsen tager afsæt i denne model. Modellen anvendes til at analysere *forordningens* elementer ud fra tesen om, at jo mere af modellens indhold der kan iagttages i *forordningen*, jo stærkere er denne til at afbøde BD-udfordringerne.

2.2. Undersøgelsesmetoden

2.2.1. Dokumentanalyse

Undersøgelsen af hvorvidt fempunktsmodellens indhold kan genfindes i *forordningen*, sker gennem en dokumentanalyse af den lovgivningsmæssige retsakt indeholdende *forordningen*. Retsakten er offentliggjort i EU-Tidende L 119, årgang 59 fra den 4. maj 2016 på siderne 1-88. Der kan opnås direkte adgang til denne retsakt gennem EU's internetbaserede databank *EUR-Lex*.

Dokumentanalysen anvender en kvalitativ indholdsanalyse, hvor *forordningen* analyseres med henblik på at identificere elementer, der kan relateres til fempunktsmodellens strategier (Bryman, 2012:557). Indholdsanalysen er dermed hypotetisk-deduktiv, idet modellens strategier fungerer som på forhånd teoretisk fastlagte indikatorer, der skal kunne lokaliseres i *forordningen*, således at det kan vurderes, hvorvidt *forordningen* kan modvirke BD-udfordringerne. (Lynggaard, 2010:144).

I dokumentanalyser er vigtige fokusområder autenticitet, troværdighed, repræsentativitet og mening (Lynggaard, 2010:147). Som et officielt EU-dokument, offentliggjort i EU-Tidende, forefindes der ikke nævneværdige autenticitets- eller troværdighedsproblemstillinger. Derimod kan der knyttes to mindre kommentarer til repræsentativiteten og meningen med *forordningen* som dokument.

Med hensyn til repræsentativiteten skal EU's *forordning* ikke opfattes som en universelt dækkende tilgang til databeskyttelse. For eksempel adskiller den amerikanske tilgang til databeskyttelse sig fra den europæiske på afgørende punkter.

Divergensen mellem den amerikanske og europæiske tilgang ses allerede i selve navngivningen af området: I USA kaldes dette for privatlivspolitik, mens det i EU går under navnet databeskyttelse. Herudover er det muligt i USA at handle med eller sælge retten til sine personoplysninger, mens databeskyttelse i EU anses for at være en umistelig rettighed. Et tredje eksempel på forskelligartetheden er at fagforeningsmedlemskab i EU regnes for en særlig sensitiv personoplysning, mens dette ikke gør sig gældende i USA. (Minelli et al., 2012:158-161)

I relation til meningen tages der udgangspunkt i den engelske udgave af *forordningen* på trods af at der findes en dansk version. Dette valg er baseret på to ting. For det første anvendes der empiri fra EU, hvor en dansk udgave ikke er tilgængelig, hvorfor den engelske udgave er blevet valgt. For kontinuitetens skyld er der derfor udelukkende anvendt EU-empiri på engelsk. Dertil har der, for det andet, tidligere været afvigelser i formuleringer i EU-databeskyttelseslovgivning blandt de forskellige sprogudgaver (Fuster, 2014:142). Derfor benyttes den engelske variant af *forordningen*.

2.2.2. Indsamling og udvælgelse af empiri

Foruden *forordningen* benyttes en mængde empiri til redegørelsen for databeskyttelsen i EU og BD samt den teoretiske diskussion af BD-udfordringer og løsninger til at modvirke disse.

Empiriren er indsamlet gennem den såkaldte sneboldmetode, hvor der med udgangspunkt i "moderempiri" forfølges referencer til andet empiri, indtil der er opbygget et dybt og omfattende empiriapparat (Lynggaard, 2010:141).

På BD-siden udgøres "moderempirien" af Viktor Mayer-Schönberger og Kenneth Cukiers *Big Data: A Revolution That Will Transform How We Live, Work, and Think* og Rob Kitchins *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*, mens databeskyttelsessiden tager afsæt i Gloria González Fusters *The Emergence of Personal Data Protection as a Fundamental Right of the EU* samt Den Europæiske Unions Agentur for Grundlæggende Rettigheder og Europarådets *Håndbog om europæisk databeskyttelse*.

3. Empiriske overvejelser

Det følgende kapitel er inddelt i to dele. Første del belyser udviklingen på databeskyttelsesområdet i EU, mens den anden del definerer BD som begreb.

3.1. Udviklingen af databeskyttelse i EU

Udviklingen af databeskyttelsesområdet i EU kan spores tilbage til begyndelsen af 1970'erne med Kommissionen og Parlamentet som hovedaktører (Fuster, 2014:112).

Elektronisk databehandling

I 1970'erne identificerede Kommissionen den fremvoksende elektroniske databehandling (EDB) som en fremtidig kerneindustri, der skulle satses på i Europa. En udfordring var dog, at EDB-industrien var stærkt domineret af amerikanske virksomheder. På daværende tidspunkt sad IBM for eksempel på mere end halvdelen af det globale marked. (Commission of the European Communities, 1972:1-2)

For at styrke europæiske virksomheders konkurrenceevne og vækstmuligheder på EDB-området præsenterede Kommissionen derfor et forslag til en fælles EF[3]-politik på området for Ministerrådet i 1973. Heri omtalte Kommissionen ligeledes behovet for fælles tiltag til at beskytte borgerne i relation til den stigende etablering af databaser:

> *"When police, and tax, and medical records, and the files of hire purchase companies concerning individuals are held in data banks, the rules of access to this information become vital."* (Commission of the European Communities, 1973:13)

Kommissionen opfordrede til en større debat i EF-regi angående denne fælles beskyttelse i relation til EDB og databaser (Ibid.). Opfordringen fra Kommissionen fandt dog hverken genklang hos Ministerrådet eller Parlamentet (Fuster, 2014:113).

[3] EU gik på daværende tidspunkt under navnet Det Europæiske Fællesskab (EF).

Den første interesse i Europa-Parlamentet

Debatten fik dog i marts 1974 en saltvandsindsprøjtning, da Pierre Bernard Cousté, medlem af Parlamentet (MEP), under et plenarmøde bragte emnet på banen:

"(...) I believe that this is the first time that this matter has been raised in this House (...) in our Community there is a real and growing awareness of the dangers to the individual freedom of our citizens resulting from the use of computers on personal data." (European Parliament, 1974:34)

Cousté argumenterede for at Ministerrådet, i samarbejde med Kommissionen, skulle fastsætte fælles principper og regler frem for at overlade beskyttelsen af borgernes privatliv til den enkelte medlemsstat. Nødvendigheden af sådanne fælles principper og regler skyldtes, ifølge Cousté, faren for, at personoplysninger kunne falde i hænderne på ukendte aktører, der kunne misbruge personoplysningerne og risikoen for at de indsamlede personoplysninger var falske, forældede, mangelfulde eller ukorrekte, hvilket kunne føre til uretfærdig behandling af borgerne. (Ibid.)

Ministerrådets repræsentant på plenarmødet anerkendte behovet for at beskytte borgerne imod de trusler, som den stigende anvendelse af EDB i samfundet skabte. Repræsentanten slog dog fast at Ministerrådet, uden de fornødne konsultationer med interessenter, endnu ikke var i stand til at lægge sig fast på en bestemt holdning og politik. (European Parliament, 1974:35)

Debatten på plenarmødet startede dermed for alvor diskussionen i EF om beskyttelse af borgernes privatliv i en tid, hvor EDB og databaser fyldte mere og mere. Der var bred enighed om behovet for denne beskyttelse af borgerne, men debatten afdækkede en klar skillelinje om, hvordan denne beskyttelse skulle sikres. På den ene side var der dem som mente, at beskyttelsen skulle opnås gennem fælles EF-lovgivning mens der, på den anden side, var dem der mente, at dette skulle ske gennem national lovgivning som herefter skulle harmoneres. (European Parliament, 1974:34-37)

Europa-Parlamentets udredning fra 1975

I forlængelse af denne vitaliserede diskussion og interesse opfordrede den daværende EF-kommissær for industri og entreprenørskab, Altiero Spinelli, Parlamentet til at nedsætte en komité til nærmere udredning af spørgsmålet. Dette skyldtes at Kommissionen anså spørgsmålet for at være af konstitutionel karakter og derfor ville det være mest hensigtsmæssigt, at Parlamentet tog teten. (Fuster, 2014:114)

Parlamentet uddelegerede ansvaret for udredningen til dets retsudvalg i oktober 1974. Udredningen, der blev forelagt Parlamentet i februar 1975, konstaterede:

"(...) that a directive on 'individual freedom and data processing' is an urgent necessity, not only to ensure that Community citizens enjoy maximum protection against abuses or failures of data processing procedures, but also to avoid the development of conflicting national legislation." (European Parliament, 1975a:5)

Et andet vigtigt element i udredningen var understregningen af, at fokus for øjeblikket ikke skulle være på at etablere en EF-ret angående privatliv, men i stedet rettes imod:

"(...) examining the more restricted field of intrusion on privacy by misuse of automatic data processing techniques, of identifying the dangers and studying the ways in which they may be minimized." (European Parliament, 1975a:7)

Europa-Parlamentets tre resolutioner

På baggrund af udredningen vedtog Parlamentet kort tid efter en resolution[4], hvori der blev givet opbakning til udredningens anbefaling om et direktiv om individuel frihed og EDB. Herudover nedsatte resolutionen et særligt udvalg til nærmere undersøgelse af sagen og tilskyndede ligeledes Kommissionen til, på baggrund af det særlige udvalgs anbefalinger, at forberede et direktiv, der kunne beskytte den enkelte borger imod misbrug i forbindelse med behandlingen af personoplysninger i elektroniske databaser. (European Parliament, 1975b:48)

[4] Resolution of the European Parliament of 21 February 1975 on the protection of the rights of the individual in the face of developing technical progress in the field of automatic data processing.

Resolutionen affødte ingen nævneværdig reaktion fra Kommissionen eller Ministerrådet, hvilket resulterede i, at Parlamentet i april 1976 vedtog endnu en resolution[5] (Kuster, 2014:116). I denne resolution udtrykte Parlamentet, at det var opsat på at intensivere dets aktiviteter på området og opfordrede endnu engang Kommissionen til at udarbejde fælles EF-lovgivning for blandt andet at undgå modstridende national lovgivning blandt medlemsstaterne. Derudover blev Parlamentets retsudvalg endnu engang udstyret med opgaven om at udfærdige en rapport vedrørende hvilke tiltag og redskaber, der kunne varetage borgernes rettigheder over for udbredelsen af EDB og databaser. (European Parliament, 1976:27)

Retsudvalget nedsatte et specialudvalg til at udføre den pålagte opgave. I maj 1979 præsenterede specialudvalget dets arbejde i den såkaldte *Bayerl-rapport*. Rapporten gennemgik og evaluerede udviklingen på området for beskyttelse af borgerne i relation til anvendelsen af EDB i medlemsstaterne samt i udvalgte lande fra Europarådet og Organisationen for Økonomisk Samarbejde og Udvikling (OECD). Rapporten noterede blandt andet hvordan den østrigske databeskyttelseslov fra 1978 skænkede de østrigske borgere en forfatningssikret: *"(…) right of personal data secrecy."* (Kuster, 2014:117)

Samtidig med at det nedsatte specialudvalg var undervejs med *Bayerl-rapporten*, besluttede Ministerrådet, at Kommissionen skulle igangsætte en række studier angående EDB og IKT. Et af studierne skulle behandle de konsekvenser, som EDB og IKT havde for datasikkerheden og –fortroligheden samt hvordan borgerne kunne beskyttes. (Council of the European Communities, 1977:25)

Parlamentet vedtog i 1979 en tredje resolution[6] på baggrund af *Bayerl-rapporten*. Her blev resolutionens anvendelsesområde udvidet fra udelukkende kun at omhandle EDB til at vedrøre databehandling generelt. Herudover indeholdte resolutionen synspunkterne om, at de teknologiske fremskridt skal være til for mennesket og ikke omvendt og at sikringen

[5] Resolution of the European Parliament of 8 April 1976 on the protection of the rights of the individual in the face of developing technical progress in the field of automatic data processing.

[6] Resolution of the European Parliament of 8 May 1979 on the protection of the rights of the individual in the face of technical developments in data processing.

af individuelle rettigheder, i relation til de teknologiske fremskridt på databehandlingsområdet, er en essentiel del af udviklingen af demokratiske samfund. (European Parliament, 1979a:34-35)

I resolutionen anmodede Parlamentet endnu engang Kommissionen til at forberede et forslag til et direktiv, der ville give borgerne en stærk beskyttelse af deres personoplysninger gennem blandt andet harmonisering af medlemsstaternes nationale lovgivning efter højeste fællesnævner. For at sikre den mest solide beskyttelse af personoplysninger opfordrede resolutionen ydermere medlemsstaterne til at tilslutte sig den databeskyttelseskonvention, som Europarådet var på vej med samt at det skulle undersøges, hvorvidt EF, som helhed, kunne tiltræde denne kommende konvention. (European Parliament, 1979a:35-36)

Til forskel fra dens to forgængere opstillede resolutionen en række specifikke principper for, hvad en kommende EF-lovgivning skulle basere sig på. Disse principper fastlagde at: (1) personoplysninger skulle være indhentet lovligt; (2) behandlingen af sensitive personoplysninger krævede samtykke fra den registrerede; (3) personoplysninger måtte kun registreres og videregives inden for de angivne formål; (4) personoplysningerne skulle være præcise og nødvendige for databasens endemål; og (5) ulovlige, unøjagtige og forældede personoplysninger skulle slettes. (European Parliament, 1979a:37)

Principperne udstyrede tillige borgerne med rettigheder, der sikrede dem: (1) ret til indsigt i anvendelsen af deres personoplysninger; (2) ret til at få fjernet personoplysninger der ikke levede op til de førnævnte principper; (3) ret til at få korrigeret urigtige personoplysninger; og (4) ret til at få vurderet lovligheden af registrerede personoplysninger hos en medlemsstats datamyndighed. (European Parliament, 1979a:37-38)

Med hensyn til datamyndigheder ønskede Parlamentet slutteligt i resolutionen, at der blev oprettet et fælles EF-datakontrolorgan til at føre tilsyn med beskyttelsen af personoplysninger i medlemsstaterne (European Parliament, 1979a:38).

Efter vedtagelsen af den tredje resolution fastholdte Parlamentet presset på Kommissionen på området (Fuster, 2014:120). Dette kom blandt andet til syne i september 1979 under en debat i Parlamentet, kaldet *Databeskyttelse i EF*, anført af spørgeren MEP Jochen van Aerssen:

"(...) we are moving from and industrialized society into a modern computerized society, in which data processing is an essential element (...) We are not afraid of the computer age (...) On the other hand, we also know that modern computer technology enables us to compose a kind of 'X-ray' picture of an individual using the data collected by the computer, and it is here that the great danger lies." (European Parliament, 1979b:19)

van Aerssen påpegede at en række medlemsstater enten havde eller var i gang med at vedtage national lovgivning på området, men dette var ikke fyldestgørende på grund af databehandlingens grænseoverskridende karakter. Herfor blev opfordringen om et EF-direktiv gentaget for at sikre en harmoniseret og kompatibel lovgivning. (Ibid.)

Kommissionens repræsentant under debatten, dens daværende næstformand Lorenzo Natali, slog fast, at Kommissionen anerkendte behovet for en international løsning og at et EF-direktiv kunne være en vej frem, men at det var Kommissions holdning at afvente resultatet af forhandlingerne om en databeskyttelseskonvention i Europarådet. (European Parliament, 1979b:20).

Konvention 108

I januar 1981 nåede Europarådet til enighed om en databeskyttelseskonvention[7], kendt som Konvention 108, der dermed blev lagt frem til ratificering af Europarådets medlemmer. Formålet med konventionen var at:

"(...) secure in the territory of each Party for every individual, whatever his nationality or residence, respects for his rights and fundamental freedoms, and in particular his right to privacy, with regard to automatic data processing of personal data relating to him (...)" (Council of Europe, 1981:1)

[7] Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data.

Reaktionen fra Kommissionen kom prompte med anbefalingen om, at medlemsstaterne hurtigst muligt skulle ratificere konventionen. Dette skyldtes Kommissionens vurdering om, at konventionen var velegnet til at skabe en stærk og ensartet databeskyttelse i Europa. Kommissionen gjorde det dog også klart, at såfremt ikke alle medlemsstater tilsluttede sig konventionen, ville den forbeholde sig retten til at foreslå et decideret EF-instrument for Ministerrådet. (Commission of the European Communities, 1981:1)

I marts 1982 udstedte Parlamentet endnu en resolution[8], der, i tråd med Kommissionen, forholdte sig positivt til konventionen og opfordrede de resterende medlemsstater, som endnu ikke var tiltrådt, til at ratificere denne. Samtidig blev ønsket om, at selve EF skulle tilslutte sig som part gentaget (European Parliament, 1982:40-41).

På ét område gik Parlamentets resolution dog imod Kommissionens udmelding fra 1981. Resolutionen understregede fortsat nødvendigheden for deciderede EF-databeskyttelsesregler og opfordrede endnu engang Kommissionen til at fremlægge et forslag til et EF-direktiv. Herudover blev det for første gang nævnt i en resolution, at det skulle undersøges, hvorvidt det var muligt og ønskværdigt at gøre databeskyttelse til en menneskerettighed gennem inkorporering i Den Europæiske Menneskerettighedskonvention[9]. (Ibid.)

Databeskyttelsespakken fra 1990

Ministerrådet, der indtil videre i databeskyttelsesspørgsmålet havde holdt sig mest i baggrunden, udviste i april 1984 interesse for at finde frem til potentielle harmoniseringsmuligheder og redskaber på området i EF gennem beslutningen om at finansiere undersøgelser til afdækning af dette (Council of the European Communities, 1984:31).

[8] Resolution of the European Parliament of 9 March 1982 on the protection of the rights of the individual in the face of technical developments in data processing.

[9] Convention for the Protection of Human Rights and Fundamental Freedoms.

Hermed kunne der op gennem 1980'erne identificeres et skifte, hvor Ministerrådet blev mere positiv over for udviklingen af fælles lovgivning for databeskyttelse (Fuster, 2014:121-122).

Dette kulminerede i 1990 med Kommissionens fremlæggelse af en databeskyttelsespakke bestående af: (1) et generelt direktiv omhandlende databeskyttelse; (2) en resolution der udvidede det generelle direktivs anvendelsesområde, så dette også omhandlede, de personoplysninger som ikke faldt ind under fælleskabsretten i medlemsstater; (3) en erklæring om at institutioner og organer i EF også skulle efterleve det generelle direktiv; (4) et direktiv omhandlende databeskyttelse og privatliv inden for telesektoren; (5) en anbefaling om at EF, som helhed, skulle tiltræde Konvention 108; og (6) et forslag til en Ministerrådsafgørelse om informationssikkerhed. (Commission of the European Communities, 1990a:6-9)

Grundlaget for fremlæggelsen af databeskyttelsespakken var, ifølge Kommissionen, for det første, at de meget forskelligartede tilgange til databeskyttelse i medlemsstaterne bragte den europæiske integration i fare, for det andet ikke sikrede en stærk nok databeskyttelse og for det tredje at Konvention 108 ikke havde haft den ønskede effekt på harmoniseringen af medlemsstaternes lovgivning på området. (Commission of the European Communities, 1990a:14-18)

Databeskyttelsespakkens sigte var på den ene side introduktionen af:

"(...) measures to ensure the protection of individuals in relation to the processing of personal data and to enhance the security of information processing in the context, notably, of the development of open telecommunications networks." (Commission of the European Communities, 1990a:2)

På den anden side skulle fælles EF-regler fjerne hindringer for færdiggørelsen af det indre marked og lette den grænseoverskridende udveksling af data for at fremme databehandlingsindustrien og databaserede services (Commission of the European Communities, 1990a:4).

I oktober 1992 fremlagde Kommissionen et revideret forslag til det generelle databeskyttelsesdirektiv (Commission of the European Communities, 1992:30). På trods af modstand fra navnligt Storbritannien og Irland, der ikke brød sig om harmonisering på området, lykkedes det for stats- og regeringscheferne i Det Europæiske Råd at nå til enighed om en fælles udtalelse, hvori de inviterede Ministerrådet og Parlamentet til at vedtage det reviderede databeskyttelsesdirektiv; det som i afhandlingen er benævnt *direktivet* (Fuster, 2014:128-129).

Særligt den britiske accept af *direktivet* blev sikret, da det var lykkedes Tyskland, der på daværende tidspunkt havde formandsskabet for EU, sammen med Kommissionen at forhandle en række kompromisser på plads i form af for eksempel særregler for behandlingen af oplysninger på sundhedsområdet (Ibid.).

Omtrent 25 år efter at databeskyttelse for første gang kom på dagsordenen i EU, var vejen dermed banet for EU-lovgivning på området, da *direktivet* i oktober 1995 formelt blev vedtaget (European Parliament & the Council of the European Union, 1995:31-38).

Databeskyttelsespakken fra 1990 indeholdte, som før nævnt, et forslag til et direktiv omhandlende databeskyttelse og privatliv inden for telesektoren. Dette direktiv[10] blev endeligt vedtaget i december 1997 (European Parliament & the Council of the European Union, 1997:1).

Direktivets karakteristik som et direktiv betød, at medlemsstaterne skulle indarbejde dets bestemmelser i den nationale lovgivning, men dette var ikke gældende for EU's institutioner og organer (Fuster, 2014:143). Herfor bestod databeskyttelsespakken fra 1990 også af et forslag til en erklæring om, at institutioner og organer i EU skulle rette sig efter *direktivet*.

[10] Directive 97/66/EC of the European Parliament and of the Council of 15 December 1997 concerning the processing of personal data and the protection of privacy in the telecommunications sector.

Dette spørgsmål blev løst med Amsterdamtraktatens[11] underskrivelse i oktober 1997, hvor artikel 286 blev tilføjet Traktaten om Oprettelse af Det Europæiske Fællesskab[12], der slog fast at:

> *"(...) Community acts on the protection of individuals with regard to the processing of personal data and the free movement of such data shall apply to the institutions and bodies set up by, or on the basis of, this Treaty."* (European Union, 1997:294)

For at leve op til dette fremlagde Kommissionen i september 1999 et forslag til en forordning, der skulle implementere artikel 286 (European Commission, 1999:24). Denne forordning[13] blev i december 2000 vedtaget af Parlamentet og Ministerrådet (European Parliament & the Council of the European Union, 2000:1).

For at føre tilsyn med at EU's institutioner og organer efterlevede forordningen oprettedes en uafhængig tilsynsmyndighed kaldet European Data Protection Supervisor (EDPS) (European Parliament & the Council of the European Union, 2000:4).

Databeskyttelse som en fundamental rettighed

Indførelsen af databeskyttelse som en fundamental EU-rettighed blev opnået i december 2000, da Kommissionen, Ministerrådet og Parlamentet sammen bekendtgjorde ikrafttrædelsen af EU's Charter om Grundlæggende Rettigheder, der slog fast, hvilke rettigheder, friheder og principper som skulle anerkendes og respekteres i EU (European Union, 2000:8).

Artikel 7 og 8 i Chartret kan knyttes til databeskyttelse. Førstnævnte fastslår at alle har ret til respekt for deres privatliv og familieliv, sit hjem og sin kommunikation, mens sidstnævnte decideret nævner databeskyttelse:

---

[11] Treaty of Amsterdam amending the Treaty on European Union, the Treaties establishing the European Communities and certain related acts.

[12] Treaty establishing the European Community.

[13] Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data.

"Everyone has the right to the protection of personal data concerning him or her (…) Such data much be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified (…) Compliance with these rules shall be subject to control by an independent authority." (European Union, 2000:10)

På trods af EU-institutionernes fælles proklamering af Chartret var det ikke juridisk bindende og der skulle gå ni år, før det opnåede denne status med ikrafttrædelsen af Lissabontraktaten[14] den 1. december 2009 (Fuster, 2014:213-230).

Med Lissabontraktaten reformeredes EU's to grundlæggende traktater: TEU og Traktaten om Oprettelse af Det Europæiske Fællesskab. Sidstnævnte traktat blev omdannet til den føromtalte TEUF. (Ibid.)

I de konsoliderede udgaver af TEU og TEUF nævnes databeskyttelse som en rettighed. Artikel 6 i TEU rummer tilkendegivelsen af, at EU anerkender Chartrets rettigheder, friheder og principper og at disse skal have samme juridiske status som TEU og TEUF. Samtidig fastlægger artiklen, at EU skal tiltræde Den Europæiske Menneskerettighedskonvention som juridisk person. (European Union, 2010:19)

Artikel 16 i TEUF, der teknisk set erstatter den førnævnte artikel 286, fastsætter, at enhver har ret til databeskyttelse. Derudover giver artikel 16 Parlamentet og Ministerrådet kompetence til at vedtage ny databeskyttelseslovgivning både inden for EU's institutioner og organer samt i medlemsstaterne, når denne ligger inden for EU-rettens anvendelsesområde. Der er dog i artikel 16 tilføjet, at databeskyttelseslovgivning skal tage hensyn til regler om den frie bevægelighed af personoplysninger i EU. (European Union, 2010:55)

[14] Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community.

Lovgivning på området for databeskyttelse efter 2000

I 2002 blev Direktiv 97/66/EC erstattet med Direktiv 2002/58/EC[15], idet der var behov for at tilpasse det oprindelige direktiv, i lyset af udviklingen på markedet og de teknologiske fremskridt på området for elektroniske kommunikationstjenester, for at sikre en fortsat databeskyttelse og privatlivets fred. Det nye direktiv introducerede for eksempel begrebet lokaliseringsdata, som ikke var med i det oprindelige direktiv. (European Parliament & the Council of the European Union, 2002:37-43).

Direktiv 2002/58/EC indeholdte derudover en bestemmelse om at medlemsstaterne, på baggrund af hensyn til den nationale sikkerhed, forsvaret, den offentlige sektor eller forebyggelse, efterforskning, afsløring og retsforfølgning i straffesager, kunne tilsidesætte en række af direktivets bestemmelser. Herunder lovgive om tilfælde hvor det er muligt at lagre personoplysninger i en begrænset periode (European Parliament & the Council of the European Union, 2002:46).

Medlemsstaternes mulighed for at lovgive om situationer hvor bestemmelserne i Direktiv 2002/58/EC kunne tilsidesættes, skabte en stor divergens blandt medlemsstaternes nationale lovgivning. For at sikre en harmonisering på dette punkt, og forbedre muligheden for at udnytte de store fordele og gevinster som elektronisk kommunikationsdata kunne have for forebyggelse, efterforskning, afsløring og retsforfølgning af kriminalitet og strafbare handlinger, blev Direktiv 2006/24/EC[16] vedtaget i marts 2006. (European Parliament & the Council of the European Union, 2006:54)

Debatten om databeskyttelse i EU var i årene efter 2000 under stor indflydelse af terrorangrebene den 11. september 2001 i USA (Fuster, 2014:220), hvilket resulterede i

[15] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

[16] Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC.

Ministerrådets rammeafgørelse 2008/977/JHA[17]. Rammeafgørelsen forsøgte at balancere målsætningen om at opretholde EU som et sikkert, retfærdigt og frit område med forpligtigelserne inden for databeskyttelse.

Rammeafgørelsen identificerede at medlemsstaternes politisamarbejde og retlige samarbejde, hvor udveksling af blandt andet ersonoplysninger på tværs af grænserne, var et afgørende element i at sikre et højt tryghedsniveau, men for fortsat at leve op til borgernes grundlæggende rettigheder, herunder databeskyttelse, var der behov for nye regler. Dette skyldtes at *direktivet* ikke var gældende i forbindelse med en aktivitet, der ikke er dækket af fællesskabsretten såsom den offentlige sikkerhed, forsvaret, statens sikkerhed og statens aktiviteter på det strafferetlige område. (Council of the European Union, 2008:60)

Herfor var rammeafgørelsens målsætning, at:

> *"(...) ensure a high level of protection of the fundamental rights and freedoms of natural persons, and in particular their right to privacy, with respect to the processing of personal data in the framework of police and judicial cooperation in criminal matters (...) while guaranteeing a high level of public safety."* (Council of the European Union, 2008:64)

Efter rammeafgørelsen var den næste udvikling på databeskyttelsesområdet, Kommissionens præsentation af en strategi for moderniseringen af EU's databeskyttelsesregler der i 2016 mundede ud i vedtagelsen af *forordningen*.

3.2. Hvad er big data?

I 2003 var datalogen Oren Etzioni på vej til sin brors bryllup med fly fra Seattle til Los Angeles. Etzioni havde bestilt flybilletten over internettet flere måneder i forvejen med den forventning om, at jo tidligere flybilletten blev bestilt, jo billigere ville den være. Ombord på flyet fandt Etzioni ud af, at passageren på sædet ved siden af havde købt sin flybillet

[17] Council Framework Decision 2008/977/JHA of 27 November 2008 on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters.

langt senere men stadig betalt en betydeligt mindre pris. (Mayer-Schönberger & Cukier, 2014:3)

På grund af denne oplevelse besluttede Etzioni sig for at udvikle en metode, så folk kunne forudsige, hvornår en pris på en flybillet, som de så *online*, var en god handel. Etzioni udviklede en model, baseret på 12.000 prisobservationer over en periode på 41 dage fra et rejsewebsite, der gav dens brugere fine besparelser:

> *"The model had no understanding of why, only what. That is, it didn't know any of the variables that go into airline pricing decisions (…) It based its prediction on what it did know: probabilities gleaned from the data about other flights."* (Mayer-Schönberger & Cukier, 2014:4)

For at virke havde modellen behov for masser af data og for at forbedre dens præstationer fik Etzioni adgang til en stor flyreservationsdatabase. Hermed behandlede modellen pludseligt omkring 200 mia. fortegnelser over billetpriser, hvilket satte den i stand til at lave forudsigelser for de fleste flyruter i USA. (Mayer-Schönberger & Cukier, 2014:5)

I 2012 ramte modellen rigtigt i 75% af tilfældene og sparede dens brugere for gennemsnitligt $50 pr. billet. Omkring modellen skabte Etzioni virksomheden Farecast. (Ibid.)

Farecast er indbegrebet af en virksomhed, der benytter sig af BD. Overordnet set drejer BD sig om data i råt, ustruktureret, semistruktureret eller struktureret format der er så stort og omfattende, at analysen af dette data ikke kan gennemføres ved hjælp af gængse analysemetoder og –værktøjer (Zikopoulos et al., 2012:1).

Nye kraftige teknologiske databearbejdningsredskaber, der kan håndtere denne store mængde af kompleks information, er i stedet nødvendig at benytte, for at få adgang til den "guldmine" som BD udgør (Mayer-Schönberger & Cukier, 2014:6; Zikopoulos et al., 2012:1).

Victor Mayer-Schönberger og Kenneth Cukier definerer således BD som:

"(...) things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value (...)." (Mayer-Schönberger & Cukier, 2014:6)

3.2.1. De tre V'er

Men hvad består BD nærmere af? Til at afklare dette spørgsmål er det relevant at kaste et blik på de såkaldte *tre V'er*, der er centrale karakteristika for BD. De *tre V'er* udgøres af: *Volume, variety* og *velocity* (Zikopoulos et al., 2012:5).

Volume

I det seneste årti er der sket en kraftig stigning i genereringen og behandlingen af data. En stigning der forventes at fortsætte og vokse eksponentielt fremover. (Kitchin, 2014:69)

At der i dag er tale om en overflod af data og information, på tværs af samfundet, tydeliggøres hvis størrelsen på den samlede mængde data i verden overvejes: IBM har anslået at 90% af alt data i verden er blevet genereret inden for de seneste to år (IBM, 2017); i 2013 oplyste den daværende EU-kommissær for den digitale agenda, Neelie Kroes, at der dagligt, i verden, blev kreeret 1,7 mio. mia. databytes i minuttet (Rial, 2013); på globalt plan blev det estimeret at der i 2007 fandtes 295 exabyte data, 295 mia. gigabytes, svarende til 404 mia. CD-ROM'er. Blev disse stablet ovenpå hinanden, ville de kunne nå månen og et godt stykke forbi (Hilbert & López, 2011:62); og i 2020 forventes den samlede mængde data at udgøre 35 zettabytes eller 35.000 mia. gigabytes (Zikopoulos et al., 2012:5).

For at forstå hvor denne enorme mængde data stammer fra, kan en række eksempler overvejes: I 2010 indsamlede *Sloan Digital Sky Survey*, gennem dets teleskop i New Mexico, på et par uger mere data end der hidtil havde været indsamlet i hele astronomiens historie (The Economist, 2010a); *Large Hadron Collider*, en af Counseil Européen pour la Recherche Nucléaires (CERN) partikelacceleratorer, skaber 40 terabytes data i sekundet lig med 40.000 gigabytes (The Economist, 2010b); ca. 114 mia. e-mails, 24 mia. tekstbeskeder og 12 mia. telefonopkald genereres dagligt globalt (Strohm & Homan, 2013); Google håndterer hver dag 24 petabytes data omkring 24 mio. gigabytes (Davenport et al., 2012:43); Facebook behandler dagligt 2,5 mia. indholdselementer, 2,7 mia. "synes godt om"-tilkendegivelser og 300 mio. billedoverførsler (Constine, 2012); og den britiske

supermarkedskæde Tesco genererer 1,5 mia. dataelementer hver måned (Manyika et al., 2011:21).

Variety

Den store mængde data der genereres stammer fra vidt forskellige datakilder, som de ovenstående eksempler bevidner. Overordnet set kan der opstilles syv former for data: Internetdata, primært forskningsdata, sekundært forskningsdata, lokalitetsdata, billeddata, forsyningskædedata og enhedsdata (Minelli et al., 2012:10).

Dette betyder at data ikke længere kun optræder i det traditionelle strukturerede format, men nu ligeledes udgøres af rådata samt semi- eller ustruktureret data. Herudover bliver disse forskelligartede dataformater oftere sammenkædet og kombineret. (Zikopoulos et al., 2012:7-8)

Evnen til at udnytte og kombinere disse nye dataformater skyldes den teknologiske udvikling på computer- og databaseområdet. Relationelle databaser, som *Structured Query Language* (SQL), havde kapaciteten til at beherske struktureret data, der let kunne indekseres og placeres i bestemte felter; for eksempel en persons navn i ét felt, personens adresse i et andet og personens telefonnummer i et tredje (Driscoll, 2012:16-23).

Men rå, ustruktureret eller semistruktureret data er umuligt at indeksere og placere i bestemte felter, hvilket betød at disse dataformater ofte: *"(...) was either ignored or, at best, used inefficiently"* (Open Data Center Alliance, 2012:7).

Introduktionen af *Not Only Structured Query Language* (NoSQL) har dog gjort det muligt at bearbejde disse forskelligartede rå, ustrukturerede og semistrukturerede data, idet NoSQL er mindre kompleks samt mere fleksibel, udvidelig og behersker dynamisk og divergerende data (Ibid.; Driscoll, 2012:22-26).

Velocity

Genereringen af de utrolige mængder data i forskelligartede formater, *volume* og *variety*, sker i dag med en enorm hastighed. Men *velocity* handler dog ikke kun om hvor hurtigt

data genereres, oplagres og behandles. *Velocity* omhandler ligeledes den nye analysetankegang, som det *tredje V* fordrer. (Zikopoulos, 2012:8)

Tidligere var undersøgelser baseret på tidsafgrænsede dataindsamlinger, men BD har muliggjort undersøgelser, hvor data indsamles kontinuerligt og i realtid (Kitchin, 2014:76) gennem for eksempel *Radio-frequency identification*-sensorer (RFID) og andre smarte apparater, der konstant indsamler og transmitterer data i informationsstrømme (Zikopoulos, 2012:8)

Den nye analysetankegang medfører, at undersøgelsers databehandling baseres på data der er: *"(...) still in motion, not just after it is at rest"* (Zikopoulos, 2012:9). Et eksempel på forskellen mellem undersøgelser forankret i statisk data kontra flydende data er som følger:

> *"In traditional processing, you can think of running queries against relatively static data: for example, the query 'Show me all people living in the New Jersey flood zone' would result in a single result (...) With streams computing, you can execute a process similar to a continuous query that identifies people who are currently 'in the New Jersey flood zones', but you get continuously updated results, because location information from GPS data is refreshed in real time."* (Ibid.)

Foruden de *tre V'er* kan karakteristikken af BD suppleres med yderligere seks karakteristika: *Exhaustivity*, *resolutionality* og *indexicality*, *relationality* samt *flexibility* og *scalability* (Kitchin, 2014:68).

Exhaustivity

Tidligere var undersøgelser begrænset af manglende evne til at indsamle, strukturere, oplagre og behandle data, som prøveudtagninger og stikprøvekontroller blandt andet var en løsning på (Mayer-Schönberger & Cukier, 2014:19-20).

De teknologiske fremskridt har dog i vidt omfang fjernet disse begrænsninger gennem tilvejebringelsen af omkostningslet og fuldt ud tilstrækkelig indsamlings- og databehandlingskapacitet. Dette har medført et skifte, på mange undersøgelsesområder,

væk fra kun at indsamle et udsnit af data om et fænomen til at inddrive så meget som muligt eller, hvis realiserbart, alt data; det vil sige *n=alt*. (Mayer-Schönberger & Cukier, 2014:26-27)

Resolutionality og indexicality

Ved siden af muligheden for *n=alt* er BD ligeledes meget mere detaljeret og indekserbar (Kitchin, 2014:73). Detaljerigdommen kan blandt andet ses i udviklingen på området for satellitbilleder af Jordens overflade. I 1980'erne havde de bedste satellitbilleder af Jorden, der var tilgængelige for offentligheden, en opløsning, hvor hver pixel svarede til 30 x 30 meter. I dag svarer hver pixel, på Google Earth, cirka til 2,5 X 2,5 meter, hvilket skaber grobund for en mere finkornet analyse. (Ibid.)

Et andet eksempel på den stigende detaljerigdom kan hentes fra folketællinger. Indtil 2011 opdelte det irske Central Statistics Office (CSO) folketællingerne efter 3.409 valgdistrikter, men herefter valgte CSO at offentliggøre folketællinger baseret på 17.000 såkaldte *Small Areas*. (Gleeson et al., 2009:7-9)

At BD er mere indekserbar skyldes den omfattende applicering af identifikationskoder i samfundet, der både hæftes på det enkelte individ, for eksempel personnumre og kreditkortnumre, og materielle objekter i form af unikke identifikatorer såsom stregkoder, *smart labels* og RFID-chips. (Dodge & Kitchin, 2005:852-875)

Relationality

Den store mængde data findes, som sagt, i mange forskellige former. En syvende karakteristik ved BD er dog, at disse uensartede dataformater, på grund af den teknologiske udvikling på databehandlingsområdet, i høj grad er kompatible, hvilket betyder, at de forskellige data kan kombineres og forbindes på et utal af måder, der har potentiale til at afsløre nye mønstre og skabe ny indsigt. (Boyd & Crawford, 2011:1-2)

Den omfattende kombinerings- og sammenkædningsevne ved BD blev for eksempel udnyttet under det amerikanske præsidentvalg i 2012. Her skabte præsident Barack Obamas kampagnestab en kæmpe database indeholdende varierende data fra blandt andet meningsmålinger, pengeindsamlinger, forbrugeroplysninger, kampagnemedarbejdere,

sociale medier og vælgerprofiler i svingstater. Disse data blev blandt andet brugt til at styrke kampagnens pengeindsamling, give bedre og mere målrettede TV-annoncer samt etablere detaljerede vælgerprofiler, hvilket gav grobund for en mere effektiv kontakt til og påvirkning af vælgerne (Scherer, 2012).

Flexibility og scalability

Undersøgelser der foretages på baggrund af prøveudtagninger eller stikprøvekontroller er afhængige af at være repræsentative, for blandt andet at deres undersøgelsesresultater kan generaliseres. Behovet for repræsentativitet bevirker, at disse undersøgelser bliver rigide i deres undersøgelsesdesign og –metode når selve dataindsamlingen påbegyndes.
Eksempelvis er det:

"(...) essential that every person captured by the census fills in exactly the same form to ensure that the data are comparable across the whole population (...) Likewise, in scientific experiments and environmental studies to enable comparison and replication, the research design is usually inflexible once initiated." (Kitchin, 2014:77-78)

Denne rigiditet er BD-undersøgelser ikke præget af, hvilket blandt andet. skyldes brugen af NoSQL-databaser, der løbende kan udvides med nye datafelter og håndtere foranderlige datainputs på løbende basis (Kitchin, 2014:78).

Sidst men ikke mindst karakteriseres BD af *scalability* der omhandler, at BD-systemer automatisk kan skalere kapaciteten og ydeevnen alt efter hvor meget data systemet tilføres. Dette er et vigtigt karaktertræk, idet systemet ellers ville bryde sammen på grund af BD-karakteristikaene *volume* og *velocity*. (Marz & Warren, 2015:6)

Twitter har for eksempel behov for denne *scalability*, idet antallet af *tweets* kan fluktuere voldsomt i antal; helt op til titusindvis i sekundet ved store globale begivenheder (Kitchin, 2014:78).

3.2.2. Grundlaget for big data

Fremkomsten af BD skyldes en række teknologiske landvindinger siden 2. Verdenskrig på IKT-området: Kraftigere og billigere computere; internettets udvikling; udbredelsen af allestedsnærværende computerteknologi; dominansen af indekserbar og maskinlæsbart data; samt datalagringens hastigt stigende kapacitet og styrtdykkende omkostninger. (Kitchin, 2014:80-81)

Kraftigere og billigere computere

De første computere i 1950'erne og 60'erne var bekostelige og havde begrænset computerkraft, der betød, at de ikke kunne løse alt for komplekse og store opgaver (Kitchin, 2014:81).

Men den teknologiske udvikling bevirkede, at computerkraften over de næste årtier steg betydeligt, mens bekosteligheden faldt voldsomt. Fra 1940 og frem til 2001 er computerkraften årligt vokset gennemsnitligt med 55% (Nordhaus, 2001:28) mens:

> *"(...) we have seen a 1.000.000.000.000.000-fold decrease in the cost of computation in the last 100 years (...)."* (Computing Research Association, 2003:1)

Væksten i computerkraft er fortsat i samme tempo efter 2001, hvor nye apparater som bærbare computere, *smartphones* og *tablets* ligeledes har set dagens lys (Kitchin, 2014:81-82). Disse har en klart kraftigere ydeevne end datidens supercomputere dog til en brøkdel af størrelsen. For eksempel indeholder den gennemsnitlige *smartphone* i dag en *Central Processing Unit* (CPU), der er ti gange så kraftig som den amerikanske Cray-1 supercomputer fra 1976. (Townsend, 2013:xiii)

Nutidens store computerkraft har muliggjort den databehandlingskapacitet, som BD kræver og det fortsat faldende omkostningsniveau har gjort anvendelsen af BD tilgængelig for de fleste aktører (Kitchin, 2014:82).

Internettets udvikling

Lige så afgørende som computerens opfindelse og kontinuerlige evolution har været for BD, lige så skelsættende har internettets indtog og indtrængen i de fleste dele af samfundet

været (Ibid.). Allerede tilbage i 1969 lykkedes det at skabe den første netværksforbindelse mellem to computere gennem det såkaldte *Advanced Research Projects Agency's Network* (ARPANET); internettets forløber (O'Neill, 1995:76-79). Sidenhed har udviklingen på området taget fart og kulminerede i 1990'erne, da en række internetsystemer opstod, herunder *World Wide Web* (WWW) der blev altdominerende grundet dens enkelthed (Gillies & Cailliau, 2000:11).

Siden årtusindeskiftet er internettet blevet suppleret med moderne IKT som *Near Field Communication* (NFC), *Wireless Fidelity*-internet (WiFi) samt *Third generation of mobile telecommunications technology* (3G-netværk) og ifølge den såkaldte *law of telecosm*, fordobles den globale båndbredde hver sjette måned (Kitchin, 2014:82).

Internettet og dets seneste tendenser er en altafgørende drivkraft for BD, da det for eksempel har gjort det muligt at indsamle data fra et anseligt antal sensorer og digitale enheder løbende, håndtere enorme mængder data gennem oplagring på blandt andet forskellige computere eller i "skyen" samt dele data hurtigt og let mellem flere aktører (Ibid.).

Udbredelsen af allestedsnærværende computerteknologi

Datagenereringens særdeles høje tempo og enestående diversitet kan spores til det moderne samfunds allestedsnærværende computerteknologi, der er blevet en realitet på grund af kreeringen af nye digitale apparater og enheder, som er tilsluttet internettet (Kitchin, 2014:83).

Udover nye digitale apparater og enheder er tidligere "uintelligente" objekter gjort "intelligente", eller "smarte", ved at blive digitale og sat i stand til at forbinde sig til internettet. Bilen er et eksempel på denne transformation:

"Although the car still appears the same and drives in the same fashion, its systems are increasingly aware of their performance through onboard diagnostics, its location through GPS-tracking, and the actions of the driver. The result is that cars are now capta rich generators (...)" (Dodge & Kitchin, 2005:871)

Det bedste eksempel på den allestedsnærværende computerteknologi er *smartphonen*, som de fleste mennesker render rundt med i deres lomme, og den bærbare computer. Med *smartphonen* er det muligt at tilgå IKT-netværk og –systemers services overalt og den bærbare computer tillader én at have computerkraft med sig, lige meget hvor.
Spredningen af digitale apparater og enheder har skabt fænomenet *everyware*:

"In everyware, all the information (…) becomes accessible from just about anywhere, at any time, and is delivered in a manner appropriate to our location and context (…) In everyware, the garment, the room and the street become sites of processing and mediation. Household objects from shower stalls to coffee pots are reimagined as places where facts about the world can be gathered, considered, and acted upon." (Greenfield, 2006:1)

Dominansen af indekserbar og maskinlæsbart data
Et andet element der bidrager til den voldsomme datagenerering, er den stigende brug af maskinlæsbare identifikationskoder og unikke identifikatorer både i relation til mennesker og objekter hvis udbredelse:

"(…) is derived from their ability to discriminate, to be rapidly and automatically monitored, updated, and processed, and to provide authentication and credentials that dictate various forms of access (…) Identification codes provide information systems with their technicity – that is, the ability to be able to process the capta they hold." (Dodge & Kitchin, 2005:855)

Med hensyn til mennesker er det for eksempel udviklingen af digitale og biometriske identifikatorer, som brugernavne og passwords, kreditkort og fingeraftryk, samt spatiale lokalitetsidentifikatorer, i form af eksempelvis *Global Positioning Systems*-placeringer (GPS), der danner indekserbar data. For objekters vedkommende er det særligt de unikke stregkoder, RFID-chips og den stigende opkobling til *Internet of Things* (IoT) der producerer data, som er indekserbar. (Dodge & Kitchin, 2005:852-869)

At disse identifikatorer er blevet maskinlæsbare skyldes den teknologiske udvikling på området, hvor stregkoder for eksempel indeholder en særegen kode der frigiver

information når en laserscanner føres over, sensorer eller transpondere kan automatisk registrere RFID-chips over afstande, nummerplader kan maskinelt identificeres og algoritmer kan genkende specifikke ansigtstræk (Kitchin, 2014:85).

I takt med yderligere teknologisk udvikling på området vil tendensen med forøget anvendelse af maskinlæsbare identifikationskoder og unikke identifikatorer fortsætte og skabe grundlag for endnu mere datagenerering (Ibid.).

Datalagringens hastigt stigende kapacitet og styrtdykkende omkostninger
Det sidste element der har skabt fundamentet for fremkomsten af BD, er det forløb på datalagringsområdet som har ført til den konstante kapacitetsstigning samt reduktionen af datalagringens omkostninger. Evnen til at oplagre og håndtere den voluminøse og evigt stigende mængde data er central for at kunne ibrugtage og udnytte BD.

Siden magnetbåndets erstatning af hulkortet i 1950'erne har datalagringskapaciteten været støt stigende i takt med opfindelsen og introduktionen af harddisken, disketten, CD'en, DVD'en, USB-stikket, den eksterne harddisk og senest *cloud computing*. For eksempel havde den første harddisk, præsenteret af IBM i 1956, en lagringskapacitet på 4,4 MB, mens Hitachis Deskstar 7K500 fra 2005 kunne rumme 500 GB. Det seneste skud på stammen, *cloud computing*, tilbyder dog næsten ubegrænset kapacitet. (Trikha, 2010:1449-1451)

Samtidig med væksten i datalagringskapacitet er udgifterne til at gemme og opbevare data faldet signifikant. Nogle vurderinger viser, at prisen pr. opbevaret GB er faldet: *"(...) an estimated factor of 10 million to one from 1956 to 2000 (...)"* (Kitchin, 2014:85).

3.2.3. Big data-kilder

Et sidste spørgsmål der kan stilles i forbindelse med defineringen af BD er, hvor alt det genererede data stammer fra? For at kunne besvare dette spørgsmål er det nyttigt at skelne mellem tre overordnede kategorier af datakilder: *Automatiserede datakilder*, *frivillige datakilder* og *styrede datakilder* (Kitchin, 2014: 87).

Automatiserede datakilder

De automatiserede datakilder producerer data autonomt uden, eller med begrænset, menneskelig indblanding med hjælp af digital teknologi. I de fleste tilfælde bliver dette data desuden behandlet automatisk af avancerede algoritmer, som herefter foretager en bestemt handling på baggrund af databehandlingen. (Ibid.)

Denne kategori af datakilder udgøres af: (1) Automatiseret overvågning, i form af for eksempel fjernaflæste målere, automatisk nummerpladegenkendelse, intelligente transportkort og offentlige skraldespande med RFID-chips; (2) digitale apparater som *smartphones*, GPS-enheder og kameraer; (3) sensorer der kan måle forskellige former for output, såsom gas- og kemikalieniveauer, og kan kommunikere med transpondere, så betalingen af motorvejsafgifter sker automatisk, når en lastbil kører forbi; (4) scanningsdata baseret på maskinlæsbare identifikationskoder der blandt andet består af stregkoder på forbrugsvarer og postpakker, magnetstriber eller elektroniske chips i kreditkort og pinkoder; og (5) interaktionsdata på IKT-netværk bestående af blandt andet *cookies* på *websites* og detaljeret registrering af for eksempel finansielle transaktioner, e-mails og telefonopkald. (Kitchin, 2014:89-93)

Frivillige datakilder

En stor del af datagenereringen er baseret på frivillige datakilder, hvormed der menes, at personer aktivt og frivilligt deltager i produktionen af data (Kitchin, 2014:93). Fænomenet hvor personer deltager i produktionen af et produkt eller en service, som de selv er med til at forbruge, kaldes for *prosumption* og er i de senere år blevet langt mere fremtrædende på grund af internettet og de sociale medier (Ritzer & Jurgenson, 2010:13).

Særligt fire *prosumption*-elementer er essentielle med hensyn til de frivillige datakilder: (1) Transaktioner i form af blandt andet online køb, hvor en række personoplysninger skal opgives, udfyldelse af onlineformularer, brug af kundekort ved forbrugskøb og kunde-*feedback* på ydet service og det købte produkt; (2) sociale medier bestående af sociale netværk, for eksempel Facebook og Twitter, billed- og videodelings *websites*, såsom Instagram og YouTube, blogs og *mashups*, der består af små applikationer bygget på kombineret data fra forskellige kilder, hvor brugerens data ”høstes” af de sociale mediers ejere; (3) *sousveillance* der er den udprægede tendens i det moderne samfund til

individuel selvovervågning og styring af eget liv og helbred gennem digital teknologi som eksempelvis Endomondo og Jawbone, der indsamler og journaliserer data om individet; (4) *crowdsourcing* som omhandler at en stor gruppe personer, fælles og frivilligt, mobiliseres og sammen bidrager til at løse en specifik opgave. Udviklingen af *Wikipedia* er et glimrende eksempel på *crowdsourcing*, men virksomheder kan tillige udnytte *crowdsourcing* til, for det første, at evaluere og vurdere forskellige løsninger eller produkter, som Amazon og Tripadvisor, og, for det andet, udvikle nye produkter eller løsninger gennem blandt andet afholdelsen af *hackathons*; og (5) folkevidenskab hvor individer gratis udfører og behandler empiriske observationer og målinger af fænomener. En hobbyvejrstation i en persons baghave symboliserer for eksempel dette. (Kitchin, 2014:93-98)

Styrede datakilder

De styrede datakilder har behov for menneskelig indblanding både i indsamlings- og behandlingsfasen for at kunne levere brugbart data. Sådanne manuelt afhængige datakilder er overvågningskameraer, undersøgelser såsom folketællinger og statistiske analyser, kontrolinspektioner, selvangivelser og luftfotografering til brug for 3D-kortlægninger af områder. (Kitchin, 2014:87-88).

4. Teorifelt

Dette kapitel identificerer først, på baggrund af en teoretisk diskussion, hvilke udfordringer BD giver for databeskyttelsen. Herpå rettes opmærksomheden imod teoretiske anskuelser på hvordan disse udfordringer kan afbødes. Anskuelserne opsummeres afslutningsvis i den føromtalte fempunktsmodel.

4.1. Udfordringer for databeskyttelsen

Fra 1950 til dets opløsning i 1990 beskæftigede den østtyske sikkerheds- og efterretningstjeneste *Ministerium für Staatssicherheit* (MfS), også kendt som Stasi, på sit højeste 91.015 fastansatte medarbejdere og 173.000 informanter. Disse medarbejdere holdt øje med bankkontoer, aflyttede boliger og telefoner, overvågede mere end 2.800 postadresser og gennemlæste 90.000 breve om dagen, hvilket resulterede i mere end 39 mio. kartotekskort og over 100 km dokumenter. Med Stasi er Østtyskland indskrevet i historien som et af de mest omfattende overvågningssamfund nogensinde. (Sloan & Warner, 2016:374; Mayer-Schönberger & Cukier 2014:150)

I dag genereres der dog langt mere data om den enkelte borger end Stasi nogensinde kunne drømme om:

> *"And clandestine three-letter government agencies are not the only ones spying on us. Amazon monitors our shopping preferences and Google our browsing habits, while Twitter knows what's on our minds. Facebook seems to catch all that information too, along with our social relationships. Mobile operators know not only whom we talk to, but who is nearby."* (Mayer-Schönberger & Cukier, 2014:150-151).

Det er klart at ikke alt BD angår personoplysninger. Rolls-Royce benytter sig eksempelvis ikke af personoplysninger i deres dataindsamling til brug i design- og produktionsfaserne af flymotorer og i deres efterfølgende serviceordning (Marr, 2016:25-30).

En stor del af BD er dog baseret på personoplysninger, som en rapport fra det hollandske datatilsyn blandt andet vidner om. I rapporten fremgår det, at den gennemsnitlige hollænder er registreret i 250-500 forskellige databaser, mens de mere aktive på sociale medier er inkluderet i op til 1.000 databaser (Koops, 2011:235).

Herudover har fremkomsten af BD betydet, at offentlige og private aktører har stærke grunde til at indsamle så meget data som muligt om borgeren og herefter lagre det i længere tid for at kunne genbruge det utallige gange i deres voksende datasæt (Mayer-Schönberger & Cukier, 2014:152)

Det 20. århundrede er fyldt med skrækeksempler på, hvor galt det kan gå, når personoplysninger misbruges: Det veludviklede hollandske folkeregister blev under den tyske besættelse af Holland i 2. Verdenskrig, benyttet af nazisterne til at identificere, pågribe og deportere hollandske jøder og sigøjnere; United States Census Bureau (USBC) udleverede, på baggrund af oplysninger fra dets folketælling i 1940, adresserne på de boligblokke, hvor der boede japansk-amerikanere til de amerikanske myndigheder under deres internering af amerikanere med japansk oprindelse; og folkeregistret i Rwanda, etableret under det belgiske kolonistyre, blev anvendt til at udføre etnisk udrensning af hutuer og tutsier ved folkedrabet i 1994. (Seltzer & Anderson, 2001:486-493)

BD vil formentlig ikke afstedkomme nye situationer med deportation, internering eller folkedrab. På trods af dette er dets udfordringer for databeskyttelsen dog fortsat alarmerende og relaterer sig særligt til den omfattende overvågning af borgeren, som BD foranlediger. Denne nye overvågning manifesterer sig ikke gennem Stasis metoder og er ej heller baseret hovedsagligt på videoovervågning. Overvågningen opstår på baggrund af alt det data, der genereres af og om borgeren.

Det *data trail* som borgeren efterlader, ihukom Thelma Arnold, giver BD-aktører mulighed for at stykke en meget dybdegående og omfattende profil sammen, der kan give en stor indsigt i borgerens dagligdag. Hermed kan BD være med til at bane vejen for manifestationen af det altseende *Panoptikon*, som den engelske filosof Jeremy Bentham designede, og dermed fremkalde de selvsamme negative konsekvenser for borgeren, kreeret af panoptikonnets evner inden for modificering af adfærd, oplæring og korrektion af borgere samt magtkoncentration (Foucault, 1995:200-205).

Konkret er udfordringerne fra BD todelt. På den ene side aktiverer BD en række fænomener, der truer borgerens ret til databeskyttelse og på den anden side, underminerer BD traditionelle tilgange til databeskyttelsen.

4.1.1. Den første del af big data-skyggesiden

Dataveillance og secondary uses

Der dannes en anseelig mængde data i samfundet om den enkelte borger. Dette data skabes både af borgeren selv og af andre, hvor førstnævnte går under navnet *data footprints*, mens sidstnævnte er kendt som *data shadows*. Problemet med disse er at:

> *"(...) in contrast to physical footprints and shadows, their digital counterparts are far from fleeting and ephemeral."* (Koops, 2011:230-231)

Data footprints og *data shadows* er vedblivende informationselementer om borgeren og udfolder herfor et *data trail*. Dette ville ikke være et problem, hvis det ikke var fordi, at BD-analyser gør det muligt at stykke borgerens ellers fragmenterede og spredte *data footprints* og *data shadows* sammen (Dodge & Kitchin, 2007:441). Sammenkædningen af en persons *data footprint* og *data shadow* giver en aktør mulighed for at opnå et oligoptikonsk, eller i værste fald et panoptisk, indblik i den pågældende persons liv (Kitchin, 2014:167).

I bund og grund skal der ikke meget sammenkædet data til, før der kan siges noget om en person, hvilket det følgende eksempel bevidner. Intelligente elmålere kan gennem det data de producerer, over en husstands elforbrug, afsløre nogle ret personlige oplysninger om dens beboere. Alene ville de forskellige data fra husstandens ageren på elforsyningsnettet ikke sige noget væsentligt, men kombineres denne data kan der konstrueres et indsigtsfuldt billede af beboernes dagligdag. Data om strøm, spænding, modstand og effekt kan for eksempel sige noget om, hvor tit beboerne spiser mikrobølgemad, hvor meget TV der ses og fra hvilket slags TV samt hvornår der normalt bades. (Quinn, 2009:9)

Anvendelsen af IKT til at indsamle, sammenkæde og analysere data for at kunne overvåge og opnå information om en person eller en gruppe af personer blev allerede i slutningen af 1980'erne begrebsliggjort med navnet *dataveillance* (Clarke, 1988:499-505).

Dataveillance har en række negative konsekvenser for borgere og samfundet. Borgerne påvirkes negativt gennem blandt andet manglende samtykke og indsigt i egne datastrømme, brug af data uden for kontekst eller med lav datakvalitet kan føre til forkerte

og uretfærdige beslutninger, sortlistning, heksejagt, *ex ante* diskriminering, nægtelse af behørig rettergang og eksponering for målrettede individuelle reklamer. De negative konsekvenser for samfundet knytter sig til udviklingen af mistillid, svækkelse af moral og sammenhængskraft samt risikoen for et skred henimod en repressiv stat. (Clarke, 1988:505)

Der er dog ligeledes positive aspekter forbundet med *dataveillance* såsom en øget sikkerhed i samfundet og finansielle gevinster på baggrund af bedre evner inden for identifikation og forebyggelse af fejl, misbrug og bedrageri (Ibid.).

For at vende tilbage til eksemplet med de intelligente elmålere vil der i første omgang være fordele at spore for både elforsyningsselskabet og forbrugeren. Elforsyningsselskabet kan med forbrugsdata i realtid blandt andet etablere en mere effektiv elproduktion ved at skalere produktionen efter de tidspunkter hvor der er størst efterspørgsel på elforsyningsnettet, mens forbrugerne, eksempelvis gennem en *app* på deres *smartphones*, kan følge med i deres elforbrug og forsøge at nedsætte dette.

I det ovenstående optræder *dataveillance* i sin positive form, men dette kan hurtigt ændre sig, hvis elforsyningsselskabet vælger at videresælge disse forbrugsdata til tredjeparter. Hermed bliver de intelligente elmåleres data brugt til et andet formål, for eksempel kan en marketingsvirksomhed sende skræddersyede reklamer eller rabatkuponer på mikrobølgeovne eller mikrobølgemad til husstanden med stor forkærlighed for mad lavet i mikrobølgeovne, end hvad der i første omgang var hensigten med dataindsamlingen.

Anvendelsen af data til et andet formål end dataindsamlingens oprindelige hensigt går under betegnelsen *secondary uses* (Solove, 2007:767). Det er her, i kombinationen af *dataveillance* og *secondary uses*, at problemet for databeskyttelsen særligt opstår, når data benyttes til at udnytte personer (Kitchin, 2014:168).

Dataveillance har særdeles gode vilkår i en BD-verden, hvilket selvsagt skyldes den omfattende generering af data om borgerne:

"(...) what programmes one watched on television used to be known only to those present in the room, whereas now cable companies can monitor what is being viewed (...) What one bought was only known by the customer and the person serving them, whereas now it is routinely collected through digital checkout tills. Where one travelled was largely unknown to everyone but the traveler (...) but now GPS embedded into cars or mobile phones, or signal triangulation with phone mast, or the use of RFID cards on public transport systems or at toll booths, reveals the location and paths traversed." (Kitchin, 2014:168-169)

Personlige oplysninger som tidligere var forbeholdt en håndfuld personer er nu tillige blevet globalt tilgængelige: CV'et ligger frit tilgængeligt på LinkedIn, feriebilleder uploades til Instagram, interesser og ting, som den enkelte synes godt om, tilkendegives på Facebook og personlige meninger og holdninger deles på Twitter (Minelli et al., 2012:11-12).

Samtidig sender *apps*, digitale enheder og intelligente apparater en lind strøm af data tilbage til deres udviklere. For eksempel kan *apps*:

"(...) trace your Web habits, look into your contact list, make phone calls without your knowledge, track your location, examine your files and more. They can also automatically send information such as location data to mobile ad networks." (Gralla et al., 2011).

Profilering og weblining

Profilering og *weblining* er to fænomener der udspringer fra *dataveillance* og som derfor ligeledes er blevet forstærket af BD. Disse udgør tillige en trussel imod databeskyttelsen.

Tidligere var profilering hovedsagligt et redskab for statslige myndigheder af sikkerhedshensyn og for at bekæmpe kriminalitet og bedrageri. Nu om stunder sker størstedelen af profileringen i den private sektor. Her benyttes BD til at lave, ikke bare profilering baseret på den gennemsnitlige person i en gruppe, men profilering helt ned på enkeltpersonniveau. Denne profilering af enkeltpersoner er blevet muliggjort ved at kombinere den pågældendes *data footprints* og *data shadows*. (Kitchin, 2014:176)

Profilering i den private sektor appliceres for det meste til at udføre mere målrettet og effektiv markedsføring samt at lave skræddersyede priser, tilbud og rabatter. Alt sammen for at påvirke forbrugeradfærden. (Ibid.)

Et eksempel på dette kan hentes fra den amerikanske detailhandelskæde Target. Ved hjælp af en mængde data om dens kvindelige kunder kan Target forudsige, hvem der er gravide, hvilket trimester de befinder sig i og hvornår de har termin. Hermed kan Target sende skræddersyede reklamer og rabatkuponer til de identificerede gravide kunder. (Duhigg, 2012)

Eksemplet med Target understreger tydeligt, hvorfor BD udgør en udfordring for databeskyttelsen. I forlængelse af profilering er der dog endnu et fænomen, som for alvor sætter databeskyttelsen under pres; *weblining* (Danna & Gandy, Jr., 2002:382). *Weblining* dækker over at IKT-redskaber, i denne sammenhæng profilering, bruges til at differentiere mellem en gruppe personer, hvor nogle diskrimineres og enten ekskluderes eller ikke nyder det samme niveau inden for et gode, service eller en ydelse (Graham, 2005:566)

På baggrund af profilering kan virksomheder blandt andet sortere deres kunder, efter hvem der er mest og mindst profitable samt hvem der er høj- eller lavrisikokunder. Hermed kan virksomhederne fokusere på at give de mest profitable kunder en privilegeret behandling, mens de mindre profitable og mere risikofyldte kundegrupper enten oplever forringede vilkår eller ekskluderes fra ydelser og services. (Danna & Gandy, Jr., 2002:375-382)

Weblining er i særdeleshed udbredt inden for feltet af kreditvurdering, hvor risikostyring er helt central. Det seneste skud på stammen er udførelsen af kreditvurdering baseret på data fra sociale netværk (Wei et al., 2015:234). FICO, en amerikansk dataanalysevirksomhed med fokus på kreditvurderinger, eksperimenterer med at inkorporere data fra Facebook i deres kreditvurderingsmodeller. Dette indebærer for eksempel at overvåge statusopdateringer og kommentarer, som virksomhedens administrerende direktør Will Lansing forklarer:

"If you look at how many times a person says 'wasted' in their profile, it has some value in predicting whether they're going to repay their debt." (McLannahan, 2015)

I tilknytning til dette har Facebook selv taget patent på en ny kreditvurderingsalgoritme, der kreditvurderer efter sammensætningen af individers sociale netværk. Tankegangen er grundlæggende, at en person med et netværk bestående af investeringsrådgivere, ph.d.er og dataloger vil have større sandsynlighed for at tilbagebetale et lån, end en person hvor det sociale netværk udgøres af rengøringsassistenter og fængselsindsatte. (O'Neil, 2016:156)

Big datas sorte boks

BD blev i det indledende kapitel betegnet som et *Weapon of Math Destruction*, hvilket skyldes problemet med dens sorte boks.

BD-analyser bygger på avancerede og komplekse matematiske algoritmer, som oftest kun forstås af dens udviklere. Dette er et alvorligt problem, idet BD-algoritmer spiller en voksende rolle i samfundet hvor deres forudsigelser og afgørelser anvendes på mange områder med konsekvenser for borgerne.

En person der oplever at blive udsat for *weblining*, fordi en algoritme har vendt tommelfingeren nedad for vedkommende, kan have ualmindeligt svært ved, at få indsigt i hvorfor algoritmen reagerede som den gjorde. Hermed kan det være næsten umuligt at forsvare sig over for algoritmen og måske få omgjort dens beslutning ved at modbevise algoritmens vurderingskriterier, som kan være fejlbehæftet. (O'Neil, 2016:3-11)

BD's sorte boks udgør dermed en trussel imod borgernes retssikkerhed og bringer mindelser om forfatteren Franz Kafkas roman *Processen*, hvor hovedpersonen arresteres og kommer for en domstol på grund af en kriminel handling. Hovedpersonen får dog aldrig at vide, hvem der har anklaget ham og hvad anklagen består i.

Datasikkerhed

Det sidste element relaterer sig til datasikkerhed. Alt det data der genereres om borgerne og som oplagres er i risikozonen for at blive udsat for kriminelle handlinger fra

eksempelvis *hackere*, der tiltvinger sig adgang til data for at misbruge det til egen vinding. (Kitchin, 2014:174)

Datasikkerhed er et område der er af stor vigtighed og kræver kritisk opmærksomhed. Dette skyldes den nuværende situation på området, hvor den mængde af data der har behov for at blive beskyttet vokser hurtigere end evnen til at opstille de nødvendige sikkerhedsforanstaltninger. Opgørelser estimerer at blot halvdelen af den samlede mængde data, som det er påkrævet at beskytte, er omgivet af sikkerhedsværn. (Gantz & Reinsel, 2011:1-8)

Det har længe været kendt, at computere, *smartphones* og andre digitale enheder kan udsættes for *hacking*, men den store bekymring på datasikkerhedsområdet i disse år er spredningen af intelligente apparater og objekter koblet op på IoT, som vil kræve helt nye procedurer for datasikkerhed (Rezendes & Stephenson, 2013).

For at apparater og objekter, såsom termostater, køleskabe, fjernsyne og biler, kan forbindes trådløst til IoT, kræver det inkorporeringen af minicomputere. Kapaciteten på disse er dog så lille, at der kun er installeret basale sikkerhedsforanstaltninger, som er sårbare over for forsøg på *hacking*.

Det forventes at antallet af apparater og objekter opkoblet til IoT vil stige fra omkring 11 mia. i 2013 til 50 mia. inden 2020 og hermed er det tydeligt, hvorfor datasikkerhed er et betydningsfuldt område med hensyn til databeskyttelse. (The Economist, 2014c)

4.1.2. Den anden del af big data-skyggesiden

Mens den første del af BD-skyggesiden aktiverede en række fænomener, der udgør en trussel imod databeskyttelsen, drejer den anden del sig om, at BD underminerer de traditionelle tilgange til sikring af databeskyttelse. Særligt tre gængse databeskyttelsesstrategier har BD obstrueret: *Notice and consent*, *opt-out* og anonymisering.

Notice and consent

En af grundstenene for databeskyttelse er begrebet *notice and consent*, hvorved borgeren får oplyst, før en dataindsamling påbegyndes, hvilke informationer og oplysninger der indhentes og til hvilket formål. Borgeren har herefter muligheden for enten at afvise eller acceptere denne dataindsamling. (Mayer-Schönberger & Cukier, 2014:153)

Et velkendt problem med *notice and consent* er dets indbyggede dilemma, der drejer sig om at:

> *"(...) people do not read privacy policies (...) if people read them, they do not understand them (...) if people read and understand them, they often lack enough background knowledge to make an informed choice (...) if people read them, understand them, and can make an informed choice, their choice might be skewed by various decisionmaking difficulties."* (Solove, 2013:1888)

BD har medført et nyt problem for *notice and consent*-konceptet gennem tre tendenser. For det første opstår den store værdi i BD, når BD-aktører finder en innovativ *secondary uses*-løsning. *Secondary uses* er dog ofte ikke blevet identificeret på dataindsamlingstidspunktet (Rubinstein, 2013:78). Denne kendsgerning medfører problemstillingen om, hvordan BD-aktører kan underrette borgere om formålet for en dataindsamling, som endnu ikke er opfundet (Mayer-Schönberger & Cukier, 2014:153)?

For det andet, og i forlængelse af den første tendens, opstår endnu en problemstilling: Hvordan kan disse borgere give et informeret samtykke baseret på noget ubekendt? (Ibid.) Det virker urealistisk at BD-aktører vil bruge enorme ressourcer på at informere borgere, hver gang tidligere indsamlet data skal bruges i en ny kontekst. For eksempel er det svært at forestille sig, at Google, i forbindelse med udviklingen af *Google Flu Trends*, ville tage kontakt til alle de personer, hvis søgeforespørgsler blev anvendt til influenza-forudsigelserne og bede om deres samtykke. Samtidig ville det reelt sætte databeskyttelsesreglerne ud af kraft, hvis personer gennem deres samtykke ved en dataindsamling accepterede, at deres data måtte bruges til alle mulige formål fremover (Mayer-Schönberger & Cukier, 2014:154).

For det tredje dækker databeskyttelsesreglerne kun personoplysninger men hvad så med det faktum, at BD-analyser kan sammenkæde data, der i første omgang ikke kan kategoriseres som personoplysninger, på en sådan måde, at der alligevel afsløres ganske personlige oplysninger om en person (Rubinstein, 2013:78)?

Opt-out

Personer har muligheden for at kræve sig selv fjernet fra databaser og datasæt, hvis de kan anvise en legitim grund. Men at gøre brug af muligheden for *opt-out* kan i visse tilfælde medvirke til at skabe ny data. Et eksempel på dette kan hentes fra *Google Street View*, der i Tyskland skabte stor furore:

"People feared that pictures of their homes and gardens could aid gangs of burglars in selecting lucrative targets. Under regulatory pressure, Google agreed to let homeowners opt out by blurring their houses in the image. But the opt-out is visible on Street View – you notice obfuscated houses – and burglars may interpret this as a signal that they are especially good targets." (Ibid.)

Anonymisering

Den sidste gængse databeskyttelsestilgang som er blevet gjort forældet af BD er anonymiseringsteknikker. Anonymiseringen omhandler at data der kan identificere en person, for eksempel navn, adresse, personnummer, telefonnummer og kreditkortnummer, fjernes fra datasæt. Efterfølgende kan datasættet behandles og videregives til en tredjepart uden at personens identitet og personoplysninger kompromitteres (Ibid.).

BD-analyse er dog imidlertid blevet så avanceret, at det på trods af anonymiseringsbestræbelserne er muligt at re-identificere anonymiserede personer ved hjælp af komplekse algoritmer (Ibid.; Minelli et al., 2013:31). I denne sammenhæng er det igen relevant at huske tilbage på eksemplet med Thelma Arnold.

4.2. Teoretiske anskuelser på hvordan big data-udfordringerne afbødes

I det foregående afsnit blev det påvist at BD udspænder en række udfordringer for databeskyttelsen. Gennem den teoretiske litteratur på området kan der anvises fem

overordnede strategier for at afbøde udfordringerne fra BD, som på samme tid giver plads til at udnytte BD's fordele og gevinster. Disse diskuteres i det følgende.

4.2.1. Privacy through accountability

En grundpille inden for databeskyttelse har været, at udstyre borgerne med rettigheder til at bevare kontrollen med deres personoplysninger gennem retten til, at bestemme om personoplysningerne må anvendes samt hvordan og af hvem (Mayer-Schönberger & Cukier, 2014:173).

Dette er blandt andet evident i de såkaldte *Fair Information Practices* (FIP*)* hvor der i den originale version lægges stor vægt på, at borgerne skal have mulighed for at vide hvilke personoplysninger der er registreret og hvordan de anvendes, forhindre anvendelsen af personoplysningerne samt ændre og korrigere i disse (Gellman, 2017:3).

Disse FIP's er gradvist blevet transformeret til *notice and consent*-konceptet, der hovedsagligt er blevet gjort utidssvarende af BD's *secondary uses* og dermed ikke længere er egnet til at sikre databeskyttelsen.

Frihed under ansvar

På baggrund af dette foreslår Mayer-Schönberger og Cukier et paradigmeskifte på databeskyttelsesområdet, hvor fokus flyttes fra: *"(...) 'privacy by consent' to 'privacy through accountability' (...)"* (Mayer-Schönberger & Cukier, 2014:175). Det vil sige at der skal placeres et udvidet ansvar hos BD-aktørerne i forbindelse med databeskyttelse, mens borgernes ansvar skal indtage en mindre fremtrædende rolle (Mayer-Schönberger & Cukier, 2014:173).

Den grundlæggende tankegang bag *privacy through accountability* (PTA) er, at BD-aktørerne ikke behøves at opnå et nyt samtykke fra en registreret, når dennes personoplysninger skal anvendes til et andet formål end det formål, der blev givet for dataindsamlingen i første omgang, hvis BD-aktøren vurderer, at det nye anvendelsesformål ikke udgør en risiko for databeskyttelsen. (Mayer-Schönberger & Cukier, 2014:173-174).

I forlængelse af dette skal der i databeskyttelseslovgivning skelnes mellem to grupper af anvendelsesmuligheder. Den ene gruppe indeholder anvendelsesmuligheder som anses for værende tilforladelige og derfor er tilladte med få, eller ingen, standardiserede databeskyttelsesforanstaltninger fra BD-aktørens side. Omvendt rummer den anden gruppe de risikobetonede anvendelsesmuligheder hvor der etableres regler for, hvordan BD-aktørerne skal vurdere risikoen ved en anvendelsesmulighed og hvordan denne risiko afmonteres. (Mayer-Schönberger & Cukier, 2014:173)

PTA skal i bund og grund forstås som frihed under ansvar for BD-aktørerne. Lever BD-aktørerne op til deres ansvar på området for databeskyttelse, betyder det, at de har friheden til at anvende indsamlede personoplysninger til *secondary uses* uden at skulle bruge ressourcer på at anskaffe en ny runde af samtykker fra berørte registrerede. (Ibid.)

Frihed under ansvar kræver dog, at der forefindes sanktionsmuligheder over for eventuelle BD-aktører der lader hånt om PTA. Derfor skal BD-aktørernes ansvar være retligt bindende og der skal i databeskyttelseslovgivningen formuleres sanktionsmuligheder for kontrolinstanserne i form af påbud, bøder og retsforfølgelse. (Ibid.)

Centre for Information Policy Leadership (CIPL) har identificeret fire elementer som er nødvendige for PTA: (1) ny og forøget gennemskuelighed; (2) bedre risikoanalyse; (3) fair databehandling; og (4) udviklingen af dataetik. (CIPL, 2015:3-5)

Ny og forøget gennemsigtighed

Gennemsigtighed i databehandlingsprocessen har altid været et vigtigt element og optræder eksempelvis i *direktivets* kapitel to, sektion fire og fem i artiklerne 10-12, der omtaler oplysningspligten ved indsamling af oplysninger og den registreredes ret til indsigt (European Parliament & the Council of the European Union, 1995:41-42).

Kompleksiteten af BD og dens *secondary uses* betyder dog, at BD konstant vil:

"(...) outstrip the ability of individuals to understand fully how and by whom their information is being used. This reality requires a new application of transparency that

extends beyond its traditional function of providing legal notice of specific uses." (CIPL, 2015:3)

I stedet skal gennemsigtighedselementet styrkes ved, at BD-aktørerne gør mere kontekstuel information om databehandlingen tilgængelig for de borgere, databehandlingen vedrører. Denne information skal omfatte:

"(...) the general value of the intended uses of personal information for the individual, including any unexpected, out-of-context and non-obvious future uses (...) how individual and society may benefit from such uses and address any associated concerns and how the organization will mitigate them." (Ibid.)

En sådan ny og forstærket gennemsigtighed vil gøre det muligt for offentligheden at udvikle tillid til, at BD-aktørerne vil handle korrekt med hensyn til anvendelse af personoplysninger. Tillid fra offentligheden til BD-aktørerne er afgørende for, at PTA skal kunne fungere. (CIPL, 2015:3-4)

Udover kommunikation af kontekstuel information, fremfor blot retligt påkrævede bekendtgørelser, skal informationen ligeledes foregå på nye måder, som øger overskueligheden for offentligheden. I stedet for at informere igennem det traditionelle lange dokument, enten i fysisk eller onlineform, skal nye moderne metoder, som udnytter teknologiens muligheder, finde anvendelse. Dette kan eksempelvis ske gennem inkorporeringen af den kontekstuelle information i selve BD-aktørens teknologi. (Ibid.)

Privacy Impact Assessment

En central opgave for BD-aktøren i PTA er at vurdere, hvorvidt en ny anvendelse af indsamlede personoplysninger eller en ny teknologi udgør en trussel imod databeskyttelsen. Herfor er en kompetent risikoanalyse, også kaldet *Privacy Impact Assesment* (PIA), altafgørende. (CIPL, 2015:4)

PIA har, ligesom gennemsigtighedselementet, i lang tid været en grundbestanddel på databeskyttelsesområdet (Article 29 Data Protection Working Party, 2014:2) og kan for

eksempel genfindes i *direktivets* artikel 17 og 20 (European Parliament & the Council of the European Union, 1995:43-44).

Selvom PIA har en mangeårig tilstedeværelse på databeskyttelsesområdet, er en egentlig metodologi i dette felt underudviklet i forhold til andre områder, hvor risikoanalyser er væsentlige (CIPL, 2014:13). Denne problemstilling blev blandt andet noteret af The US National Institute of Standards and Technology (NIST) i et diskussionsoplæg:

> *"In the security field, risk management models, along with technical standards and best practices, are key components of improving security. Similarly, the safety risk management field also has well-developed models, technical standards and best practices. To date, the privacy field has lagged behind in the development of analogous components."* (NIST, 2014:1)

Den underudviklede metodologi skyldes den manglende konsensus blandt databeskyttelsesområdets aktører om en fælles standardiseret tilgang til PIA. På området er der eksempelvis mangel på overensstemmelse i relation til opfattelsen af, hvilke skadelige effekter og negative konsekvenser som en PIA skal søge at identificere og afbøde. Dette har blandt andet medført at der i databeskyttelseslovgivning, der påbyder PIA's, ofte ikke specificeres, hvilke skadelige effekter og negative påvirkninger der skal vurderes. (CIPL, 2014:13-16)

For at en PIA skal være et effektivt remedie i PTA er der derfor behov for udviklingen af en fælles metodologi, der sikrer konvergens mellem de påvirkninger, som skal vurderes og stræbes efter at blive formindske (CIPL, 2014:19).

En anden udfordring er at BD skaber en række nye udfordringer for databeskyttelsen, som den sædvanlige PIA ikke kan håndtere. De nuværende PIA's er beregnet til håndgribelige databeskyttelsesudfordringer som økonomiske tab, trusler for den fysiske velvære og trivsel, sikkerhedsbrud og overskridelse af tavshedspligt. (Polonetsky et al., 2014:3)

Udover at forstærke disse håndgribelige udfordringer skaber BD ligeledes databeskyttelsesudfordringer, der er uhåndgribelige og abstrakte. Førstnævnte udgøres for

eksempel af diskriminering, eksklusion og isolering, mens sidstnævnte blandt andet består af panoptisk overvågning, paranoia, manglende tillid og trusler mod demokratiet. Disse ligger uden for håndteringsevnerne af den traditionelle PIA og skal derfor søges inkorporeret i en ny PIA-metodologi. (Ibid.)

En ny kompetent PIA kan være med til at:

> *"(…) help prioritize the investment of scarce resources in protecting privacy and enforcing privacy obligations. It can identify serious risks to privacy and measures for mitigating them. It can expand our collective thinking about the range of risks that the processing of personal data can present to individuals, organizations, and society (…) And it can help bring rigor and discipline to our thinking about data processing and how to maximize its benefits while reducing its costs."* (Kuner et al., 2015:96)

For at opnå en sådan kompetent og moderne PIA er der behov for seks tiltag: (1) PIA's skal, i tråd med risikoanalyser på andre områder, professionaliseres gennem forskning, ensartede standarder og fælles fastsatte principper og processer, hvilket kræver et tæt samarbejde mellem kontrolinstanserne, BD-aktørerne og den akademiske verden; (2) der skal udvikles en fælles struktur for, hvilke skadelige effekter og negative konsekvenser der skal tages højde for og modvirkes i en PIA. I særdeleshed skal de uhåndgribelige og abstrakte udfordringer indlemmes på lige fod med de mere håndgribelige; (3) ligesom for de skadelige effekter og negative konsekvenser skal der etableres konsensus om en ny model for hvordan gevinsterne ved en databehandling vurderes; (4) PIA's skal baseres på forståelsen af, at ikke alle identificerede risici kan elimineres helt og derfor skal målsætningen være at fokusere på at minimere de risici, der udgør den største udfordring for databeskyttelsen. Herefter skal det kommunikeres klart og tydeligt til offentligheden om de tilbageværende risici og hvilke tiltag der tages i tilknytning til disse; (5) PIA's skal være effektive, fleksible og skalerbare, så de kan anvendes af både store og små BD-aktører; og (6) PIA's må ikke stå alene i BD-aktørernes bestræbelser på at imødegå udfordringerne for databeskyttelsen. (CIPL, 2016:7-13)

Fair databehandling

Princippet om fair databehandling har været fast inventar i europæisk databeskyttelse længe og er blandt andet repræsenteret i *direktivets* artikel 6 (European Parliament & the Council of the European Union, 1995:40). Et problem med princippet om fair databehandling er imidlertid, at det ofte reduceres til blot at indebære kommunikationen af databeskyttelsesbemærkninger til personer. (CIPL, 2015:4)

Fair databehandling vedrører i PTA mere end blot bemærkninger om eksempelvis databeskyttelsesrettigheder (Ibid.). UK Information Commissioner's Office (ICO) har peget på tre faktorer, som BD-aktører skal tage hensyn til for at sikre en fair databehandling: Effekter, forventninger og gennemsigtighed. (ICO, 2014:19-20)

For det første tager en fair databehandling hensyn til den effekt, behandlingen af data har på de registrerede. BD-aktører skal derfor, som led i en fair databehandling, være bevidste om og tage højde for potentielle indvirkninger på individet, samfundsgrupper og samfundet. (ICO, 2014:20-22)

For det andet tager en fair databehandling hensyn til, hvorvidt databehandlingen pågår inden for de registreredes forventninger til, hvordan en fair databehandling burde foregå eller om de forventer, at én bestemt aktivitet medfører kreeringen af data som efterfølgende behandles. En registreret må for eksempel forvente, at der skabes data gennem brugen af kundekort, som herefter bliver databehandlet, mens den registrerede ikke kan forvente, at det data der kreeres på et socialt medie bliver videresolgt til en markedsanalysevirksomhed. (ICO, 2014:22-27)

For det tredje tager en fair databehandling hensyn til, om behandlingen er gennemskuelig. Er den ikke dette søsættes der tiltag, som kan medvirke til at øge databehandlingens transparens for de registrerede. (ICO, 2014:27-28)

Dataetik

CIPL's sidste element er behovet for udviklingen af en decideret dataetik, der kan bidrage til at sikre BD-aktørernes efterlevelse af PTA (CIPL, 2015:5). EDPS udtrykker ligeledes et ønske om fastlæggelse af en dataetik, der kan være med til at sikre, at fremtidig

databeskyttelseslovgivning ikke kun bliver: *"(...) respected on paper while in effect being neutralised in cyberspace."* (EDPS, 2015:13)

O'Neil foreslår eksempelvis en dataetik, der tager form efter lægeløftet og som henter inspiration fra virksomhedskonsulenterne Emanuel Derman og Paul Wilmotts ed i kølvandet på finanskrisen i 2008:

> *"I will remember that I didn't make the world, and it doesn't satisfy my equations. Though I will use models boldly to estimate value, I will not be overly impressed by mathematics. I will never sacrifice reality for elegance without explaining why I have done so. Nor will I give the people who use my model false comfort about its accuracy. Instead, I will make explicit its assumptions and oversights. I understand that my work may have enormous effects on society and the economy, many of them beyond my comprehension."* (O'Neil, 2016:205-206)

Udløbsdato for personoplysninger og samtykker

PTA betyder at BD-aktørerne ikke længere er juridisk forpligtiget til at slette personoplysninger, når indsamlingsformålet er opfyldt eller opnå et nyt samtykke. Personoplysningerne kan anvendes til *secondary uses*, så længe disse ikke udgør en trussel imod databeskyttelsen. (Mayer-Schönberger & Cukier, 2014:174)

I forlængelse af dette opstår risikoen dog for at personoplysninger gemmes for altid og dermed skabes begrebet *permanent memory*, hvor en person aldrig helt kan slippe ud af sin fortids skygge, idet digitale oplysninger om personen altid kan bringes frem i lyset. (Ibid.)

For at modarbejde dette skal der fastsættes tidsrammer for, hvor længe forskellige typer af oplysninger må oplagres. De typer med den største risiko for databeskyttelsen skal have en hurtigere udløbsdato end andre kategorier af oplysninger. (Mayer-Schönberger & Cukier, 2014:174-175)

En anden løsning på problemstillingen om *permanent memory* er at indføre udløbsdatoer på afgivne samtykker. Det vil sige at hvis et samtykke ikke forlænges inden for en bestemt

tidsperiode, udløber samtykket og personoplysningerne skal efterfølgende slettes. (Custers, 2016:4)

Innovative tekniske løsninger

PTA skal afslutningsvis funderes på nye tekniske databeskyttelsesløsninger for at sikre databeskyttelsen i de tilfælde, der ikke kan afhjælpes gennem de ovenstående elementer (Mayer-Schönberger & Cukier, 2014:175)

Differential privacy og *privacy by design* er to nye tekniske databeskyttelsesløsninger, som er hensigtsmæssige at implementere i denne sammenhæng. Den første innovative tekniske løsning, *differential privacy*, er grundlæggende baseret på at der i datasæt med personoplysninger indsættes tilfældig "støj", således at det bliver umuligt at identificere enkeltpersoner (Dwork, 2011:91-92).

Et eksempel på *differential privacy* er Facebooks *Advertising Reach Data* som gøres tilgængelig for potentielle annoncører, så de kan se hvor mange personer fra en bestemt målgruppe de kan nå. Annoncørerne får dog ikke oplyst et eksakt antal personer i en målgruppe. I stedet oplyses et omtrent antal:

> *"(...) the target audience is always reported as a multiple of twenty people, with estimates of less than forty reported as 'fewer than 20 people' (...) Facebook's respective estimates for a targeted group and for the same group broken into disjoint groups often reveal small but substantial deviations from additivity."* (Chin & Klinefelter, 2012:1433)

Privacy by design handler om at der allerede i designfasen af et nyt produkt eller teknologi indopereres foranstaltninger, der sikrer databeskyttelse gennem blandt andet minimering af data som bliver indsamlet og behandlet (Schaar, 2010:267-268).

4.2.2. Foranstaltninger imod big data-forudsigelser

BD har sat dens aktører i stand til, med stigende nøjagtighed, at profilere og komme med forudsigelser på, hvordan personer vil agere i givne situationer, hvilket blandt andet kan munde ud i *weblining*.

Åbenhed, certificering og bevislighed

Profileringer, forudsigelser og *weblining* foretages af avancerede og komplekse algoritmer, hvor den menneskelig indblanding, for det meste, er henvist til et minimum. Når beslutninger, der kan påvirke personer, træffes på baggrund af algoritmer, er der behov for foranstaltninger til at sikre, at disse personer ikke diskrimineres, behandles uretfærdigt eller oplever brud på deres retssikkerhed (Mayer-Schönberger & Cukier, 2014:176).

Herfor anbefales tre foranstaltninger: (1) der skal være åbenhed omkring algoritmen og den data der anvendes; (2) den pågældende algoritme skal certificeres og godkendes af en udenforstående ekspertgruppe til at blive brugt på følsomme områder; og (3) det skal være muligt for registrerede at modbevise beslutninger truffet på baggrund af algoritmen, gennem specificeringen af konkrete måder hvorpå dette kan lade sig gøre. (Ibid.)

Right to be forgotten

Alt det genererede data om en person danner grundlag for algoritmers profilering, forudsigelser og potentielle *weblining*. Det vil sige at en persons tidligere handlinger er med til at afgøre deres fremtidige muligheder, hvilket kan have en række vidtrækkende konsekvenser for den enkelte. (Conley, 2010:53)

Der er derfor behov for et redskab til at modvirke dette pres på personers personoplysninger og privatliv (Conley, 2010:53-54). I den teoretiske litteratur er der stor enighed om, at defineringen af en ret til at få slettet oplysninger om ens fortid er et hensigtsmæssigt værktøj i denne sammenhæng. Et sådant redskab benævnes *right to be forgotten* og handler dybest set om:

> *"(...) that someone has a significant interest (possibly to be protected in the form of a legal right) in not being confronted by others with elements of her past, more in particular with data from the (more remote) past that are not relevant for present-day decisions or views about her."* (Koops, 2011:232)

De basale principper for en *right to be forgotten* findes allerede i *direktivets* artikel 12(b) og 12(c) hvor det udtrykkes, at en person har ret til, at få den dataansvarlige til at slette

eller blokere oplysninger, der ikke er behandlet i overensstemmelse med *direktivets* bestemmelser (European Parliament & the Council of the European Union, 1995:42).

Det er dog tydeligt at *direktivets* artikel 12 er for vag i relation til den teoretiske litteraturs definition på en *right to be forgotten* og der er derfor behov for en eksplicit formulering af denne rettighed, der afspejler den ovenstående definition (Koops, 2011:232-233).

4.2.3. Åbningen af big datas sorte boks

BD er baseret på algoritmer der er så avancerede og komplekse, at de er for indviklede til at forstå for de fleste individer. Hermed er der en risiko for at algoritmerne og deres forudsigelser:

"(...) will become black boxes that offer us no accountability, traceability, or confidence."
(Mayer-Schönberger & Cukier, 2014:179)

For at vende denne udvikling og åbne op for BD's sorte boks er det nødvendigt med udviklingen af ny ekspertise, som skal komplementere de nuværende nationale datatilsyn og EDPS. Disse skal monitorere, granske og sikre gennemsigtighed i relation til algoritmerne samt bistå personer, der føler sig uretfærdigt behandlet ovenpå en beslutning der grunder i en algoritme, med at få oprejsning eller erstatning (Ibid.).

Algoritmisten

Én måde at skabe denne fornødne ekspertise på er at oprette en ny jobfunktion og institution, der skal monitorere BD-aktørers anvendelse af algoritmer både internt og eksternt à la virksomheders brug af interne bogholdere og eksterne revisorer:

"These new professionals would be experts in the areas of computer science, mathematics, and statistics; they would act as reviewers of big-data analyses and predictions. Algorithmists would take a vow of impartiality and confidentiality (...) They would evaluate the selection of data sources, the choice of analytical and predictive tools, including algorithms and models, and the interpretation of results. In the event of a dispute, they would have access to the algorithms, statistical approaches, and datasets that produced a given decision." (Mayer-Schönberger & Cukier, 2014:180)

Den eksterne algoritmist skal påtage sig en række funktioner såsom at fungere som en uvildig auditor, der vurderer nøjagtigheden og validiteten af BD-forudsigelser, når dette kræves af eksempelvis en offentlig myndighed, udføre revisioner for BD-aktører der ønsker eksperthjælp, certificere algoritmer, rådgive offentlige myndigheder om hvordan BD kan implementeres i den offentlige sektor og bistå personer der forsøger at få omgjort en BD-beslutning. (Mayer-Schönberger & Cukier, 2014:181)

Som intern algoritmist skal den pågældende BD-aktørs aktiviteter på området overvåges med det sigte at varetage interesserne for både BD-aktøren men også registrerede. Den interne algoritmist skal ligeledes godkende nye BD-aktiviteter, før disse initieres, og er kontaktperson for personer, som vil klage over BD-aktøren. (Mayer-Schönberger & Cukier, 2014:182)

For at få inspiration til rollen som intern algoritmist kan der skeles til den tyske databeskyttelseslovgivning, der påkræver, at private og offentlige organisationer udpeger en uafhængig intern databeskyttelsesrepræsentant udstyret med opgaven om at sikre organisationens efterlevelse af databeskyttelseslovgivningen (Mayer-Schönberger, 2010:1881).

Adgang for forskere

En alternativ måde hvorpå BD's sorte boks kan åbnes op for offentligheden på er, at lade den akademiske verden få adgang til algoritmerne. Denne adgang medfører flere fordele. Eksempelvis kan bias i algoritmer afsløres, flere vil opnå ekspertise med algoritmerne og videnskabelige udgivelser kan være med til at øge indsigten og forståelsen for BD og dens metoder:

"If you consider mathematical model as the engines of the digital economy (…) these auditors are opening the hoods, showing us how they work. This is a vital step, so that we can equip these powerful engines with steering wheels – and brakes." (O'Neil, 2016:211).

Der er allerede forsøg på at undersøge BD-algoritmerne på universiteter som Princeton og Massachussetts Institute of Technology (MIT), men de møder modvillighed fra BD-aktører, der ønsker at holde granskninger af deres algoritmer interne og ikke vil afsløre

disses funktionsmåder. Denne modvillighed kan løses gennem fastsættelse af lovgivning, der pålægger BD-aktører at åbne mere op for blandt andet forskning. (O'Neil, 2016:210-212)

4.2.4. Featurization af big data

PTA tildeler BD-aktørerne frihed under ansvar i relation til deres behandling af personoplysninger. Forbundet med denne frihed er dog risikoen for at BD-aktører, for egen vindings skyld, ikke vælger at honorere dette ansvar. Elementer som sanktioneringsmuligheder og forøget transparens er svar på denne risiko.

Begrebet *featurization* af BD er et andet værktøj til at sikre håndhævelse af databeskyttelsesansvaret. *Featurization* af BD indebærer *empowerment* af personer i tilknytning til personoplysninger og databeskyttelse (Tene & Polonetsky, 2013:263).

Denne *empowerment* af personer er baseret på en *quid pro quo*, hvor BD-aktørerne, for at kunne nyde mindre restriktiv regulering på området for dataindsamling og –behandling, både skal inddrage de personer hvis personoplysninger anvendes i databehandlingen og efterfølgende dele de resulterende gevinster med disse (Tene & Polonetsky, 2013:264).

Dette involverer at BD-aktørerne skal give personer adgang til data omhandlende dem selv i let og overskuelig form, hvorefter den enkelte person for eksempel kan anvende applikationer fra tredjeparter til at opnå en række fordele og gevinster (Ibid.).

Featurization af BD vil medføre ny innovation og skabe fundamentet for et nyt marked inden for applikationer og programmer, der baserer sig på personoplysninger og andet data, som den enkelte person selv stiller til rådighed. Derudover vil *featurization* af BD føre til øget bevågenhed omkring BD-aktørernes adfærd med hensyn til dataindsamling og –behandling og dermed afsløre misbrug eller uetisk brug af personoplysninger. (Tene & Polonetsky, 2013:264-268)

Øget inddragelse og *empowerment* af personer vil dermed styrke databeskyttelsen idet:

"Privacy suffers not only when individuals are unaware of data practices, but also when they are uninterested or disengaged. Such an environment, regardless of the regulatory mechanisms in place, provides insufficient checks on data collection and use. Where individuals can access data in a manner that is engaging, useful, or valuable, they will give rise to natural checks on behavior, thus serving as a useful compliance mechanism for privacy law." (Tene & Polonetsky, 2013:269-270).

4.2.5. Redefinering af personoplysningsbegrebet

Det nuværende personoplysningsbegreb består af en binær objektiv definition på personoplysninger og særligt sensitive personoplysninger (Hon et al., 2011:213). Personoplysninger defineres i *direktivets* artikel 2(a) som:

"(...) any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity." (European Parliament & the Council of the European Union, 1995:38)

Mens de såkaldte særligt sensitive personoplysninger, også benævnt specielle kategorier af personoplysninger, i *direktivets* artikel 8 angår:

"(...) personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership (...) and data concerning health or sexual life." (European Parliament & the Council of the European Union, 1995:40)

Det fremgår af *direktivets* artikel 2 og 8 at personoplysninger er omfattet af *direktivets* anvendelsesområde, mens de særligt sensitive personoplysninger er dækket af mere restriktive bestemmelser (European Parliament & the Council of the European Union, 1995:38-40).

Punkt 26 i *direktivets* præambel opstiller dog en fortolkning om, at personoplysninger der anonymiseres, hvorledes det ikke længere bliver muligt at identificere enkeltpersoner,

fritages for *direktivets* bestemmelser og regler (European Parliament & the Council of the European Union; 1995:33).

BD udfordrer dog dette personoplysningsbegreb på to måder. Det er blevet teknisk muligt at sætte anonymiseringsforløbet i bakgear og kombineringen samt sammenkoblingen af tilsyneladende uskyldige oplysninger kan føre til en panoptisk indsigt, der kan afsløre private og intime ting om en person. (Hon et al., 2011:214)

Herfor er det nødvendigt med en redefinering af personoplysningsbegrebet for at imødegå BD-udfordringerne for databeskyttelsen. W. Kuan Hon, Christopher Millard og Ian Walden (2011) foreslår at personoplysningsbegrebet opbygges efter en to-trins metode hvor:

> *"First, the 'personal data' definition should be based on a realistic likelihood of identification. Secondly, rather than applying all the DPD's* (*direktivet*, red.) *requirements to information determined to be 'personal data', it should be considered, in context, which requirements should apply, and to what extent, based on a realistic risk of harm and likely severity."* (Hon et al., 2011:224)

Direktivets "alt eller intet"-tilgang, hvor oplysninger om en person enten kan eller ikke kan kategoriseres som personoplysninger og som er afgørende for om *direktivets* mekanismer er gældende, skal dermed erstattes af en model, hvor risikoen for identifikation og skade på den registrerede, i forbindelse med databehandling, er afgørende for hvilke foranstaltninger, der er behørige at iværksætte:

> *"In considering whether particular information should trigger data protection obligations, the key factor ought to be, what is the realistic risk of identification? Only where risk of identification is sufficiently realistic (…) should information be considered 'personal data' (…) Criteria for triggering the application of the DPD's requirements should be more nuanced, not 'all or nothing' (…) A better starting point might be to require explicit consideration of risk of harm (…) from intended processing (…) More sensitive situations, with greater risk of resulting harm and/or greater severity of likely harm, would require greater precautions."* (Hon et al., 2011:226)

4.3. Fempunktsmodellen

Den ovenstående teoretiske diskussion, om hvilke strategier der kan være med til at afbøde BD-skyggesidens udfordringer for databeskyttelsen, kan opsummeres i den nedenstående fempunktsmodel.

Tabel 3: Fempunktsmodellen

Strategi	Instrument(er)
Privacy through accountability	- *Frihed under ansvar*: BD-aktører kan anvende personoplysninger til *secondary uses* uden nyt samtykke, hvis det vurderes, at dette *secondary use* ikke udgør en trussel for retten til databeskyttelse. - *Gennemsigtighed*: BD-aktører skal gøre mere kontekstuel information om deres databehandling tilgængelig for registrerede på mere overskuelige måder. - *PIA*: Udvikling af en fælles og kompetent metodologi for risikoanalyser der kan håndtere BD's abstrakte og uhåndgribelige udfordringer. - *Fair databehandling*: BD-aktører skal tage hensyn til de effekter, behandlingen af personoplysninger vil have for personen, samfundsgrupper og samfundet. Databehandlingen skal foregå inden for de realistiske forventninger, som personer har til dataindsamling og –behandling samt være gennemskuelig. - *Dataetik*: Der skal etableres en kollektivt anerkendt etik på databeskyttelsesområdet. - *Udløbsdatoer på personoplysninger og samtykker*: Fastsættelse af tidsrammer for hvor længe BD-aktører må opbevare personoplysninger og hvornår afgivne samtykker, i relation til dataindsamling og –behandling af personoplysninger, skal fornyes. - *Innovative tekniske løsninger*: Nye tekniske foranstaltninger, såsom *differential privacy* og *privacy by design*, skal implementeres for at styrke databeskyttelsen.
Foranstaltninger imod *big data*-forudsigelser	- *Åbenhed, certificering og bevislighed*: BD-aktører, der træffer beslutninger vedrørende personer på baggrund af BD-forudsigelser, skal implementere foranstaltninger, der sikrer åbenhed omkring algoritmerne og deres anvendte data, sørge for at algoritmerne lever op til certificeringer og standarder samt gøre det muligt for registrerede at gøre indsigelse imod sådanne beslutninger. - *Right to be forgotten*: Specifik formulering af en ret til at få slettet personoplysninger.
Åbningen af *big datas* sorte boks	- *Algoritmisten*: Etablering af en ny ekstern og intern jobfunktion til at monitorere BD-aktørernes efterlevelse af databeskyttelsesforpligtigelser og deres anvendelse af algoritmer. - *Adgang for forskere*: Vedtagelse af initiativer der sikrer bedre og lettere adgang til BD-algoritmerne for den akademiske verden.
Featurization* af *big data	- *Empowerment*: Større inddragelse, fra BD-aktørernes side, af personer i dataindsamlingen og –behandlingen af personoplysninger samt deling af de opnåede gevinster med de registrerede. - *Adgang til egen personoplysninger*: BD-aktører skal give personer adgang til egne personoplysninger i et let og overskueligt format.

Strategi	Instrument(er)
Redefinering af personoplysningsbegrebet	- *To-trinsmetoden*: Personoplysningsbegrebet skal baseres på en to-trinsmodel. Først afgøres det om bestemte oplysninger falder inden for databeskyttelsesområdet på basis af risikoen for en specifik identificering af en registreret. Dernæst afhænger graden af databeskyttelsesrestriktioner af, hvor stor risikoen er for at den registrerede skades.

5. Analyse

Dette kapitel rummer analysen, der kronologisk vil gennemgå de identificerede strategier fra det foregående kapitel. Herefter vurderes det om *forordningen* kan afbøde BD-udfordringerne.

5.1. Privacy through accountability i forordningen?

PTA består af instrumenterne frihed under ansvar, gennemsigtighed, PIA, fair databehandling, dataetik, udløbsdatoer på personoplysninger og samtykker samt innovative tekniske løsninger, der alle skal være tilstede, for at strategien bliver effektiv.

5.1.1. Frihed under ansvar

Det centrale i frihed under ansvar er, om BD-aktøren har mulighed for at behandle personoplysninger til et andet formål, *secondary uses*, end det formål, som blev givet i første omgang, da dataindsamlingen hos en registreret startede.

Artikel 5 omtaler principperne for databehandling af personoplysninger, som den dataansvarlige, jævnfør artikel 5(2), har ansvaret for at efterleve og påvise. Et af disse principper, der fremgår af artikel 5(1.b), handler om formålsbegrænsning, der betyder at personoplysninger skal være:

"(...) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes (...)."

Hermed fremgår det at den dataansvarlige har mulighed for at gennemføre en ny databehandling med et andet formål, hvis dette formål ikke er uforeneligt med det oprindelige formål og er lovligt baseret på bestemmelserne om lovlig behandling i artikel 6(1). Den dataansvarlige skal, jævnfør artikel 6(4), i sin vurdering af, hvorvidt det nye formål er foreneligt med det oprindelige formål tage hensyn til:

"any link between the purposes for which the personal data have been collected and the purposes of the intended further processing; the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the

controller; the nature of the personal data, in particular whether special categories of personal data are processed, pursuant to Article 9, or whether personal data related to criminal convictions and offences are processed, pursuant to Article 10; the possible consequences of the intended further processing for data subjects; the existence of appropriate safeguards, which may include encryption or pseudonymisation."

I henhold til den dataansvarliges oplysningspligt, jævnfør artikel 13(3) og 14(4), skal den dataansvarlige dog orientere den registrerede før personoplysninger, indsamlet til et formål, behandles til et andet formål.

Frihed under ansvar opererer ligeledes med en differentiering af anvendelsesmuligheder, alt efter hvor risikobetonede de er. *Forordningen* foretager ikke eksplicit en sådan opdeling af anvendelsesmuligheder. Artikel 24(1) pålægger dog den dataansvarlige ansvaret for at implementere tekniske og organisatoriske foranstaltninger, der sikrer, at databehandlingen lever op til *forordningen*. Omfanget af disse tekniske og organisatoriske foranstaltninger afhænger af databehandlingens karakter, omfang, kontekst og formål samt mulige risici og konsekvenser for den registreredes rettigheder og friheder.

Artikel 24(1) ligger dermed op til en differentiering, hvor omfanget af de tekniske og organisatoriske foranstaltninger stiger i takt med, hvor betænkelig og risikobetonet databehandlingen er.

Frihed under ansvar kræver sanktioneringsmuligheder over for de BD-aktører, som ikke honorerer deres forpligtigelser i forbindelse med den større frihed de nyder på databeskyttelsesområdet. En række repressalier kan lokaliseres i *forordningen*, som derved opfylder kravet om straffeforanstaltninger i frihed under ansvar.

Artikel 58(2) tildeler tilsynsmyndighederne korrigerende beføjelser i form af advarsler, kritik, påbud, midlertidige udstedelser af begrænsninger og forbud imod databehandling, fjernelse af certificeringer, administrativ bødepålæggelse samt forbud imod overførsel af personoplysninger til tredjelande.

De administrative bøder er et interessant element og specificeres yderligere i artikel 83. Her fremgår det for eksempel af artikel 83(5), at tilsynsmyndigheden kan udstede administrative bøder på op til 20 mio. € eller 4% af den samlede globale omsætning, hvis der er tale om en virksomhed og dette beløb er større end de 20 mio. €, hvis der sker overtrædelser af *forordningens* artikler 5-9, 12-22 og 44-49, forpligtigelser pålagt gennem national lovgivning i medlemsstaterne funderet i *forordningens* kapitel IX omhandlende bestemmelser for særlige databehandlingssituationer eller tilsynsmyndighedens undersøgende eller korrigerende beføjelser udfoldet i artikel 58(1) og 58(2).

Samtidig grundfæster artikel 82(1), at en person der har været udsat for materiel eller immateriel skade, på grund af overtrædelser af *forordningen*, har ret til erstatning fra den dataansvarlige eller databehandleren. Herudover bevirker artikel 84(1), at de enkelte medlemsstater kan vedtage regler om andre sanktionsmuligheder i relation til brud på *forordningen* og særligt de overtrædelser der ikke er dækket af de administrative bøder.

5.1.2. Gennemsigtighed

Allerede i artikel 5(1.a) nævnes gennemsigtighed hvor det fremgår, at behandlingen af personoplysninger blandt andet skal foregå på en fair og transparent måde i forhold til den registrerede.

Gennemsigtighedselementet skal dog, som før omtalt, gøres bredere end blot at berøre for eksempel oplysningspligten ved indsamling af oplysninger og den registreredes ret til indsigt. Den bredere gennemsigtighed skal komme til udtryk gennem øget tilgængelighed af mere kontekstuel information om databehandlingen såsom den værdi den tilsigtede og eventuelle fremtidige anvendelser af personoplysningerne skaber, hvilke fordele og ulemper der er forbundet med anvendelsen ud fra både et samfunds- og individperspektiv samt hvordan dataansvarlige eller databehandlere vil begrænse ulemperne.

Artikel 12(1) pålægger den dataansvarlige at gøre en række informationer tilgængelige for den registrerede i et: *"(...) concise, transparent, intelligible and easily accessible form, using clear and plain language"*. Informationerne skal dertil, så længe det er hensigtsmæssigt, gives elektronisk blandt andet ved brug af, jævnfør artikel 12(7), standardiserede ikoner, der sikrer et brugervenligt overblik over databehandlingen.

Der sker dog ikke en fastlæggelse af, at denne informationskommunikation skal foregå på innovative måder gennem ny teknologi på basis af eksempelvis en inkorporering i selve den teknologi, som den dataansvarlige benytter til at indsamle oplysninger om en registreret. Hermed leves der ikke op til gennemsigtighedselementets påpegning af, at kommunikationen til den registrerede skal funderes på innovative måder, der benytter teknologiens nye muligheder for at blive så overskuelig og meningsfuld som muligt.

Informationen der skal videregives af den dataansvarlige bestemmes i artikel 13 og 14, der henholdsvis omhandler oplysningspligten når der indsamles oplysninger hos den registrerede og oplysningspligten ved oplysninger der ikke er indsamlet hos den registrerede.

Artikel 13(1) udstikker at den dataansvarlige, på tidspunktet for dataindsamlingen, skal oplyse den registrerede om: (1) kontaktoplysninger på den dataansvarlige og eventuelt dennes *Data Protection Officer* (DPO); (2) formålet med databehandlingen; (3) den dataansvarliges, eller en tredjeparts, legitime interesse hvis databehandlingen er baseret på artikel 6(1.f); (4) øvrige potentielle modtagere af personoplysningerne; og (5) hvorvidt den dataansvarlige påtænker at overføre personoplysningerne til et tredjeland.

Herudover bestemmer artikel 13(2), at den dataansvarlige skal oplyse den registrerede om: (1) hvor længe personoplysningerne vil blive oplagret eller kriterierne for fastlæggelse af tidsperioden for oplagringen; (2) den registreredes rettigheder i form af retten til indsigt, berigtigelse eller sletning af personoplysninger, retten til at begrænse behandlingen af personoplysninger eller gøre indsigelser imod databehandlingen, retten til dataportabilitet, retten til at tilbagekalde samtykke og retten til at indbringe en klage for en tilsynsmyndighed; (3) hvorvidt den registrerede er forpligtiget til at oplyse personoplysninger samt konsekvenser forbundet ved ikke at oplyse disse; og (4) om der foretages automatiske beslutninger og profilering samt hvordan disse fungerer.

Med få undtagelser har artikel 14 stort set samme ordlyd. Undtagelserne udgøres af artikel 14(1.d) og 14(2.f) som henholdsvis omhandler, hvilke kategorier af personoplysninger der er tale om og fra hvilke kilder disse personoplysninger stammer fra.

Dermed er det tydeligt at indholdet af gennemsigtighedselementets kontekstuelle information ikke efterkommes i *forordningen*.

5.1.3. PIA

Et tredje instrument iboende i PTA er kravet om, at BD-aktørerne benytter sig af PIA's, der er forankret i en fælles metodologi og som kan håndtere BD's mere abstrakte og uhåndgribelige databeskyttelsesudfordringer.

PIA kan identificeres i artikel 35, der omtaler databeskyttelseskonsekvensanalyser. Ifølge artikel 35(1) skal den dataansvarlige udføre en konsekvensanalyse før påbegyndelsen af en databehandling, hvis denne gør brug af ny teknologi samt at dens karakter, omfang, kontekst og formål sandsynligvis kan udgøre en trussel imod en fysisk persons rettigheder og friheder. Artikel 35(7) fastlægger at en konsekvensanalyse minimum skal indeholde:

"a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller; an assessment of the necessity and proportionality of the processing operations in relation to the purposes; an assessment of the risks to the rights and freedoms of data subjects (…) and the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation (…)."

PIA har dermed en central placering i *forordningen* men problemstillingen omkring etableringen af en fælles metodologi, på baggrund af de seks nævnte tiltag i kapitel fire, adresseres ikke. På trods af at de seks tiltag ikke benævnes i *forordningen*, eksisterer der dog fortsat en chance for, at der kan igangsættes en bevægelse henimod en fælles metodologi.

Denne chance grunder i artikel 35(4), som foreskriver, at tilsynsmyndigheden skal udarbejde en offentlig fortegnelse over hvilke databehandlingsaktiviteter, der er underlagt kravet om en konsekvensanalyse. Efterfølgende skal denne fortegnelse meddeles til *European Data Protection Board* (EDPB), der blandt andet har til opgave, jævnfør artikel 70(1), at sikre en konsistent anvendelse af *forordningen*.

Bestræbelsen på en konsistent anvendelse af *forordningen* skal, i henhold til artikel 70(1.v) og 70(1.w), blandt andet ske ved, at EDPB understøtter fælles uddannelsesprogrammer og deling af viden om lovgivning og praksis på databeskyttelsesområdet samt faciliterer udveksling af personale mellem tilsynsmyndighederne i medlemsstaterne og i tredjelande.

Ydermere fastlægger artikel 35(6), at tilsynsmyndigheden skal aktivere den såkaldte sammenhængsmekanisme, se artikel 63, hvis den førnævnte fortegnelse over databehandlingsaktiviteter vedrører databehandling i forbindelse med udbud af varer og services, overvågning af registrerede på tværs af medlemsstater eller hvis fortegnelsen har indflydelse på den frie udveksling af personoplysninger i EU.

Endeligt skal den dataansvarlige, på baggrund af artikel 36(1), konsultere tilsynsmyndigheden inden igangsættelsen af en databehandling, hvis konsekvensanalysen har vist, at der er en høj risiko forbundet med denne behandling. Er tilsynsmyndigheden af den opfattelse, at den dataansvarlige ikke i tilfredsstillende grad har identificeret og undersøgt behandlingens mulige risici samt opstillet tilstrækkelige foranstaltninger imod disse, skal tilsynsmyndigheden skriftligt rådgive den dataansvarlige.

De ovenstående elementer fra *forordningen* må tilnærmelsesvis siges at kunne relateres til de tre første tiltag grundet det store fokus på ensretning og standardisering på tværs af medlemsstaterne og dermed er der gode vilkår for fremvæksten af en fælles metodologi. Dog står det hen i det uvisse, om *forordningens* konsekvensanalyse kan kapere de mere abstrakte og uhåndgribelige BD-udfordringer, da dette ikke fremgår af *forordningens* bestemmelser.

5.1.4. Fair databehandling

I PTA skal princippet om fair databehandling, der ligeledes omtales i artikel 5(1.a), indeholde faktorerne effekter, forventninger og gennemsigtighed. Den sidstnævnte faktor, gennemsigtighed, er allerede nævnt i den ovenstående diskussion men kan dertil identificeres i artikel 15(1), der angiver den registreredes ret til indsigt. Dette bevirker at den registrerede har ret til, at blive bekendt med om en dataansvarlig behandler personoplysninger. I givet fald har den registrerede ret til at få adgang til disse personoplysninger og en række øvrige informationer påkrævet i artikel 13 og 14.

Herudover foreskriver artikel 15(2) og artikel 15(3), at den registrerede har ret til at blive gjort opmærksom på behørige foranstaltninger, der legitimerer overførslen af personoplysninger til tredjelande samt at den dataansvarlige skal stille en kopi af de behandlede personoplysninger til rådighed.

Gennemsigtigheden styrkes endvidere ved at den dataansvarlige, i artikel 19, pålægges en underretningspligt, der betyder at den dataansvarlige, i en situation hvor den registrerede har fået berigtiget eller slettet personoplysninger eller fået indskrænket databehandlingen, skal sørge for at dette viderekommunikeres til tredjeparter, der har modtaget personoplysningerne. Hermed bliver det mere overskueligt for den registrerede, at kontrollere sine personoplysninger.

Samtidig er det indlysende, at retten til dataportabilitet i artikel 20 øger gennemsigtigheden ved en databehandling. Retten til dataportabilitet indebærer, jævnfør artikel 20(1), at den registrerede:

"(...) have the right to receive the personal data concerning him or her, which he or she has provided to a controller (dataansvarlig, red.) *in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without the hindrance from the controller to which the personal data have been provided (...)."*

Gennemsigtigheden gavnes tillige af kravet i artikel 27(1) om, at en dataansvarlig eller databehandler etableret i et tredjeland, der behandler personoplysninger fra en registreret i EU, skal udpege en repræsentant placeret i EU. Denne repræsentant skal, på baggrund af artikel 27(4), tage sig af henvendelser fra tilsynsmyndigheder og registrerede i forbindelse med databehandlingen på vegne af den dataansvarlige eller databehandleren. Dermed gøres det lettere for registrerede, at kontrollere deres personoplysninger, selvom den dataansvarlige eller databehandleren er placeret uden for EU's grænser.

Endnu et punkt der kan knyttes til gennemsigtigheden, er kravene i artikel 30(1) og 30(2) om, at en dataansvarlig eller databehandler, eller deres repræsentant, skal føre detaljerede protokoller over de databehandlingsaktiviteter, der finder sted. Disse protokoller skal

ifølge artikel 30(4) være tilgængelig for tilsynsmyndigheden, hvis dette kræves. Herved har tilsynsmyndigheden gode vilkår for at monitorere den dataansvarlige eller databehandlerens gøren og laden.

Forordningen sikrer ligeledes et højt gennemsigtighedsniveau i forbindelse med brud på persondatasikkerheden gennem kravet i artikel 33(1) om, at den dataansvarlige, inden for 72 timer efter sikkerhedsbruddet, skal informere tilsynsmyndigheden, hvis der er tale om, at sikkerhedsbruddet udgør en trussel for registreredes rettigheder og friheder.

Informationen til tilsynsmyndigheden skal, jævnfør artikel 33(3), minimum indeholde: (1) en beskrivelse af sikkerhedsbruddet; (2) hvilke kategorier af personoplysninger og hvilke registrerede der er ramt; (3) en redegørelse for sikkerhedsbruddets mulige konsekvenser; og (4) hvordan den dataansvarlige vil håndtere sikkerhedsbruddet og reducere de mulige konsekvenser.

Indebærer sikkerhedsbruddet en stor trussel imod registreredes rettigheder og friheder, foreskriver artikel 34(1) og 34(2), at den dataansvarlige ligeledes skal notificere de berørte registrerede så hurtigt som muligt og stille de selvsamme oplysninger fra artikel 33(3) til rådighed.

Sidst men ikke mindst udstyrer artikel 57(1.b) og 57(1.e) tilsynsmyndigheden med opgaverne om at fremme offentlighedens bevidsthed om problemstillinger relateret til databehandling og gøre registrerede opmærksomme på deres rettigheder. Dette er ligeledes med til at højne gennemsigtigheden.

Den tredje faktor inden for fair databehandling må dermed siges at stå stærkt i *forordningen*, men er det ligeledes gældende for de to øvrige faktorer; effekterne og forventningerne?

At BD-aktøren skal tage hensyn til databehandlingens effekt på den registrerede kommer først til syne i artikel 6(4.d). Denne angår at den dataansvarlige, i forbindelse med databehandling til et andet formål end det som i første omgang var givet og som ikke er baseret på fornyet samtykke fra den registrerede eller anden lovgivning, skal tage den

intenderede databehandlings konsekvenser for den registrerede i betragtning, når det skal afgøres, om den nye databehandlings formål er forenelig med det oprindelige formål.

Derudover fremgår effekt-faktoren blandt andet af artiklerne 13(2.f) og 14(2.g), der begge drejer sig om, at den dataansvarlige, i forbindelse med databehandling hvor der forekommer automatiske afgørelser og profilering, skal oplyse, hvilke konsekvenser en sådan databehandling kan have for den registrerede. Denne ordlyd kan ydermere findes i forlængelse af indsigtsretten, hvor den registrerede, jævnfør artikel 15(1.h), har ret til at få adgang til information angående den dataansvarliges vurdering af mulige konsekvenser for den registrerede ved anvendelse af automatiske afgørelser og profilering.

I tæt tilknytning til artiklerne 13(2.f), 14(2.g) og 15(1.h) omhandler artikel 22 bestemmelser for anvendelsen af automatiske afgørelser og profilering. Artikel 22(1) bekendtgør at den registrerede har ret til, at afgørelser truffet på baggrund af automatik og profilering ikke må stå alene, hvis dette har retsvirkning eller på anden vis øver indflydelse på den registrerede.

Artikel 24 beskæftiger sig med den dataansvarliges ansvar, som ligeledes understreger, at hensynet til databehandlingens effekter på den registrerede er tilstede i *forordningen*, idet artikel 24(1) slår fast, at den dataansvarlige skal implementere passende tekniske og organisatoriske foranstaltninger, der bunder i en hensyntagen til:

"(...) the nature, scope, context and purposes as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons (...)."

På to andre områder fremgår opmærksomheden på databehandlingens effekt.
Artikel 32(1) omhandler persondatasikkerheden i forbindelse med databehandling og fastslår, at den dataansvarlige og databehandleren skal implementere egnede tekniske og organisatoriske foranstaltninger, der sikrer et højt sikkerhedsniveau og flugter med blandt andet de risici og skader, et sikkerhedsbrud potentielt ville medføre for den registrerede.
Artikel 32(2) uddyber dette:

"In assessing the appropriate level of security account shall be taken in particular of the risks that are presented by processing, in particular from accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to personal data transmitted, stored or otherwise processed."

Databeskyttelseskonsekvensanalysen optræder i artikel 35 og bevirker at den dataansvarlige, ved en databehandling der baserer sig på ny teknologi, skal udføre en PIA, der blandt andet tager hensyn til mulige risici og konsekvenser for den registrerede. Artikel 35(7.c) og 35(7.d) fastlægger at denne PIA skal indeholde en vurdering af de mulige risici og konsekvenser, som databehandlingen vil have for den registreredes rettigheder og friheder samt hvilke modvirkende foranstaltninger den dataansvarlige vil søsætte.

Det fremgår i kapitel fire at databehandlingens effekt skal ses fra både den enkeltes, forskellige befolkningsgruppers og samfundets synsvinkel. Som det ovenstående viser tages der stort hensyn til databehandlingens effekt på den enkelte, mens samfunds- og befolkningsgruppesynsvinklen slet ikke kommer til syne i *forordningen*. Hermed må det siges at denne del af fair databehandling kun delvist opfyldes.

I forlængelse af det ovenstående kan der dog argumenteres for, at *forordningen* udviser hensyn til databehandlingens effekt på børn, idet artikel 8(1) slår fast, at en lovlig behandling af børns personoplysninger, som grunder i artikel 6(1.a), på baggrund af udbudte informationstjenester kun er lovlig, hvis barnet er mindst 16 år. Er det pågældende barn under 16 år, er databehandlingen kun lovlig, hvis der er indhentet samtykke fra indehaverne af forældremyndigheden. Dette rokker dog ikke ved den overordnede opfattelse af at den fair databehandlings effekt kun efterkommes delvist.

Faktoren der omhandler at databehandlingen skal foregå inden for de berørte registreredes forventninger figurerer ikke i selve *forordningens* bestemmelser og dermed fremstår denne faktor som den svageste del af fair databehandling. Forventnings- og effektfaktorernes henholdsvis manglende og delvise tilstedeværelse i *forordningen* medvirker til, at fair databehandling, samlet set, har en svag fremtoning.

5.1.5. Dataetik

Udviklingen af en form for dataetik er et vigtigt element i PTA for at sikre, at BD-aktørerne lever op til det ansvar, der følger med den større frihed og at selve databeskyttelseslovgivningen ikke blot bliver en papirtiger.

Forordningen opererer specifikt med ønsket om at etablere et såkaldt adfærdskodeks i artikel 40:

> *"The Member States, the supervisory authority, the Board* (EDPB, red.) *and the Commission shall encourage the drawing up of codes of conduct intended to contribute to the proper application of this Regulation, taking account of the specific features of the various processes sectors and the specific needs of micro, small and medium-sized enterprises."* (Artikel 40(1))

Adfærdskodekset kan, jævnfør artikel 40(2), udarbejdes af foreninger eller organisationer, som repræsenterer forskellige grupper af dataansvarlige eller databehandlere og skal tage hensyn til: (1) principperne for rimelig og gennemsigtig databehandling; (2) den dataansvarliges legitime interesser; (3) indsamlingen af personoplysninger; (4) pseudonymiseringen af personoplysninger; (5) videregivelsen af information til offentligheden og registrerede; (6) den registreredes rettigheder; (7) hvilken information der skal kommunikeres til børn, hvordan børn skal beskyttes og hvordan der skal indhentes samtykke fra personer med forældremyndigheden; (8) foranstaltningerne udbredt i artiklerne 24, 25 og 32; (9) underretningen om brud på persondatasikkerheden til tilsynsmyndighederne og registrerede; (10) overførslen af personoplysninger til tredjelande; og (11) udenretslige sager og andre konfliktløsningsprocedurer mellem dataansvarlige og registrerede.

Artikel 40(4) fastsætter at hvis der er tale om et adfærdskodeks, udfærdiget af en forening eller organisation som repræsenterer en gruppe af dataansvarlige og databehandlere, skal en mekanisme inkorporeres i adfærdskodekset, der sætter et organ, jævnfør artikel 41(1), i stand til at kontrollere om adfærdskodekset overholdes af de dataansvarlige og databehandlere, der har påtaget sig dets regelsæt.

En række bestemmelser er med til at sikre, at disse adfærdskodekser så vidt muligt bliver gjort homogene i og mellem medlemsstaterne. Et adfærdskodeks skal, efter artikel 40(5), indberettes til tilsynsmyndigheden, som herefter godkender adfærdskodekset, hvis det overholder *forordningen* og rummer tilstrækkelige foranstaltninger til at sikre databeskyttelsen. Godkendes adfærdskodekset foreskriver artikel 40(6), at tilsynsmyndigheden skal registrere og offentliggøre dette.

Finder adfærdskodekset anvendelse i flere medlemsstater skal tilsynsmyndigheden, i henhold til artikel 40(7), videresende udkastet til EDPB som på vegne af tilsynsmyndigheden godkender adfærdskodekset, hvis det overholder de førnævnte godkendelseskrav. Godkendes adfærdskodekset rapporterer EDPB, se artikel 40(8) og 40(9), adfærdskodekset til Kommissionen som herefter kan tage stilling til, om adfærdskodekset skal gøres generelt gyldigt i EU. I tillæg hertil udstyrer artikel 40(11) EDPB med opgaven om at indsamle alle godkendte adfærdskodekser og gøre dem offentligt tilgængelige.

Efterlevelsen af et adfærdskodeks kan, som tidligere nævnt i forbindelse med artikel 40(4), foretages af et organ, der besidder en høj ekspertise på databeskyttelsesområdet. Organet skal dog have denne kompetence delegeret fra tilsynsmyndigheden, hvilket slås fast i artikel 41(1) og som er baseret på en række betingelser knæsat i artikel 41(2).

Vurderes det at en dataansvarlig eller databehandler har overtrådt adfærdskodekset, kan organet med hjemmel i artikel 41(4) suspendere eller udelukke den dataansvarlige eller databehandleren fra adfærdskodekset og underrette tilsynsmyndigheden herom.

Svigter organet sit ansvar, kan tilsynsmyndigheden, på baggrund af artikel 83(4), udstede en administrativ bøde på op til 10 mio. € eller, hvis der er tale om en virksomhed, 2% af den samlede årlige globale omsætning, hvis dette beløb ikke er mindre end de førnævnte 10 mio. €.

Forordningens adfærdskodeks minder ikke om O'Neils ønske om en dataetik med lægeløftet som forbillede, men bestemmelserne i artiklerne 40 og 41 tegner dog et billede af, at adfærdskodekset vil komme til at indtage en markant rolle, når *forordningen* træder

i kraft. Risikoen for administrative bøder vil virke som et stærkt incitament for et kontrolorgan til at håndhæve adfærdskodekset blandt de deltagende dataansvarlige og databehandlere, hvor *naming and shaming* kan stimulere til overholdelsen af det pågældende adfærdskodeks.

5.1.6. Udløbsdatoer på personoplysninger og samtykker

En ulempe ved PTA er, som sagt, at risikoen for *permanent memory* vokser, idet BD-aktørerne ikke er forpligtiget til at slette personoplysninger, når indsamlingsformålet er opnået eller indhente et nyt samtykke, hvis personoplysningerne kan anvendes til et *secondary use*, der ikke udgør en trussel imod databeskyttelsen. For at undgå en sådan situation er det nødvendigt at indføre udløbsdatoer på personoplysninger og samtykker, så disse ikke kan opbevares for altid og dermed afmontere risikoen for en digital *permanent memory*.

Forordningen fastsætter dog ikke udløbsdatoer for hverken personoplysninger eller samtykker. Det tætteste *forordningen* kommer på dette, er princippet om opbevaringsbegrænsning i artikel 5(1.e) med ordlyden om, at personoplysninger skal være:

> *"kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed (...)."*

Kort sagt fastsættes det at personoplysninger må opbevares, indtil de ikke længere er nødvendige for opfyldelsen af de givne formål for dataindsamlingen. *Forordningens* artikel 6(3) åbner dog op for, at medlemsstaterne kan vedtage national lovgivning, der blandt andet opstiller betingelser for opbevaringsperioder i forbindelse med lovligheden af den dataansvarliges databehandling. Dette ændrer dog ikke på det faktum, at der ikke opstilles udløbsdatoer i *forordningen* hverken med hensyn til personoplysninger eller samtykker.

5.1.7. Innovative tekniske løsninger

Det sidste instrument inden for PTA er de innovative tekniske løsninger, der skal fungere som et sidste værn for databeskyttelsen, hvis de øvrige PTA-instrumenter ikke har

elimineret truslen. *Differential privacy* og *privacy by design* er eksempler på sådanne innovative tekniske løsninger.

Forordningens artikel 25 omhandler innovative tekniske løsninger i form af, det føromtalte, *privacy by design* samt det som kan benævnes *privacy by default*. Artikel 25(1) vedrører *privacy by design* og foreskriver, at den dataansvarlige, med det aktuelle tekniske niveau, implementeringsomkostninger, den respektive databehandlings karakter, omfang, kontekst og formål samt databehandlingens risici for fysiske personers rettigheder og friheder in mente, skal:

"(...) both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organizational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimization, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation (...)."

Privacy by default omtales i artikel 25(2) og drejer sig om at den dataansvarlige, gennem tekniske og organisatoriske foranstaltninger, skal sikre at teknologien, som standardindstilling, kun indsamler og behandler de personoplysninger, som er nødvendige for at opnå et givent formål.

Forordningen medfører dermed at BD-aktører skal inkorporere innovative tekniske løsninger i form af *privacy by design* og *privacy by default* i deres dataindsamling og –behandling. Desværre indgår *differential privacy*, som er en særlig effektiv databeskyttelsesmetode, ikke i *forordningens* bestemmelser vedrørende innovative tekniske løsninger. På trods af dette er det troværdigt at opfatte instrumentet innovative tekniske løsninger som opfyldt i *forordningen*.

5.2. Foranstaltninger imod big data-forudsigelser i forordningen?

Strategien omhandlende foranstaltninger imod BD-forudsigelser baserer sig på to instrumenter: Åbenhed, certificering og bevislighed samt *right to be forgotten*. For at kunne sige at denne strategi er aktiveret, skal disse to instrumenter derfor kunne identificeres i *forordningen*.

5.2.1. Åbenhed, certificering og bevislighed

Automatiske afgørelser baseret på komplekse algoritmer, og disses profilering, forudsigelser og *weblining*, udgør en trussel imod fysiske personer i form af diskriminering, uretfærdig behandling og brud på retssikkerheden. Afhjælpningen af disse trusler skal foregå gennem foranstaltninger, der sikrer åbenhed, certificering og bevislighed.

Åbenhed i relation til algoritmer og deres automatiske afgørelser sikres gennem *forordningens* oplysningspligt ved indsamling af personoplysninger hos den registrerede eller fra en anden kilde i artiklerne 13 og 14. Artikel 13(2.f) og 14(2.g) pålægger at den dataansvarlige har pligt til, ved en dataindsamling, at oplyse den registrerede om eksistensen af automatiske afgørelser i henhold til artikel 22(1) og 22(4), hvilken logik den automatiske afgørelse bygger på og hvilken betydning samt konsekvenser dette kan have for den registrerede.

Samtidig indeholder den registreredes indsigtsret i artikel 15 bestemmelsen om, jævnfør artikel 15(1.h), at den registrerede har ret til at få at vide, hvorvidt den dataansvarlige anvender den registreredes personoplysninger til at foretage automatiske afgørelser og, i tråd med artikel 13(2.f) og 14(2.), hvilken logik der ligger bag og hvilken betydning samt konsekvenser denne automatiske afgørelse vil have.

Konsekvensanalysen i artikel 35 bidrager ligeledes til åbenheden omkring automatiske afgørelser og de bagvedliggende algoritmer, idet en konsekvensanalyse blandt andet påkræves, i henhold til artikel 35(3.a.), når der foretages en: *"(...) systematic and extensive evaluation of personal aspects relating to natural persons (...)"* baseret på automatisk databehandling og profilering.

Åbenheden opstår i denne situation, idet den dataansvarlige, for det første, skal konsultere tilsynsmyndigheden, når konsekvensanalysen påviser, at en påtænkt databehandling er særligt risikobetonet og for det andet åbner artikel 35(9) op for, at den dataansvarlige, når det er relevant, skal orientere sig om registreredes, eller deres repræsentanters, synspunkter angående den planlagte databehandling.

Forordningen indeholder dermed en del elementer, som gør det muligt at opnå indsigt i de algoritmer, der benyttes til at træffe automatiske afgørelser og dermed må åbenhed siges at være tilstede i *forordningen*.

Certificering af algoritmer, som foretager automatiske afgørelser, kan både betyde større åbenhed omkring selve algoritmen, da den derved skal leve op til nogle fastsatte krav, som for eksempel en registreret kan orientere sig om, og bedre kontrol med algoritmerne, da de fastsatte certificeringskrav kan forhindre uansvarlig og skadelig anvendelse af algoritmerne.

Artikel 42 udbreder dette certificeringselement hvor medlemsstaterne, tilsynsmyndighederne, EDPB og Kommissionen, i artikel 42(1), skal arbejde for udviklingen og etableringen af certificeringsmekanismer og –mærkninger på databeskyttelsesområdet. Certificeringsmekanismer og –mærkninger skal demonstrere, at dataansvarliges eller databehandleres behandlingsaktiviteter overholder *forordningen*.

Selve certificeringen af en dataansvarlig eller databehandler, jævnfør artikel 42(5), udstedes af et decideret certificeringsorgan, omtalt i artikel 43, eller af den kompetente tilsynsmyndighed på baggrund af kriterier, som er blevet fastlagt af enten tilsynsmyndigheden eller EDPB med udgangspunkt i henholdsvis artikel 58(3) eller artikel 63. Sker certificeringen på baggrund af kriterier godkendt af EDPB, kan det resultere i en fælles EU-certificering kaldet: *"(…) the European Data Protection Seal"* (artikel 42(5)).

Certificering er dermed, ligesom åbenhed, at finde i *forordningen*, men certificerings-elementet adskiller sig dog, idet det fundamentalt står ganske svagt, hvilket skyldes bestemmelsen i artikel 42(3) om, at certificeringen skal være frivillig. Hermed er det ikke en selvfølge, at certificerings-elementet kommer til at spille en større rolle på databeskyttelsesområdet.

Bevislighed er udtrykt i artikel 22(3) og angiver at en dataansvarlig, såfremt dennes automatiske databehandling og afgørelse er lovlig, skal:

"(...) implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision."

Forordningen sikrer dermed at fysiske personer har mulighed for at give deres besyv med i relation til en automatisk afgørelse og bestride denne. Samtidig danner den dataansvarliges oplysningspligt, i artiklerne 13 og 14, og den registreredes indsigtsret, fra artikel 15, et stærkt fundament for bevisligheden, idet denne mulighed for indsigt i en algoritmes logik samt dens betydning og konsekvenser gør det lettere for en fysisk person at bestride den automatiske afgørelse.

Dertil indeholder *forordningen* en række bestemmelser, der kan støtte en fysisk person i bestridelsen af automatiske afgørelser. Artikel 57(1.e) bekendtgør at tilsynsmyndigheden, efter anmodning, skal rådgive registrerede om udøvelsen af deres rettigheder udlagt i *forordningen* og, jævnfør artikel 57(2), facilitere indgivelsen af klager fra registrerede på en let og overskuelig måde. Endvidere fastlægger artikel 79(1) at den registrerede har ret til effektive juridiske retsmidler, hvis den registrerede vurderer, at han eller hendes rettigheder i *forordningen* er blevet overtrådt og artikel 80(1) slår fast, at den registrerede har ret til at lade sig repræsentere. Det er alt andet lige lettere at håndhæve sine rettigheder, eller bestride afgørelser, med hjælp fra aktører med ekspertise inden for området.

Bevisligheden må derfor vurderes at være ganske stærk, om end *forordningen* ikke udformer udtrykkelige måder hvorpå, en fysisk person kan modbevise automatiske afgørelser. Alt i alt fremstår åbenheds, certificerings og bevisligheds-instrumentet dermed tydeligt i *forordningen*.

5.2.2. Right to be forgotten

Et effektivt middel til at begrænse BD-forudsigelser og automatiske afgørelser er ved at gøre det lettere for registrerede, at få slettet deres indsamlede personoplysninger. Hermed undermineres BD-algoritmernes datagrundlag som anvendes til at producere forudsigelser og afgørelser.

Registreredes ret til at få slettet personoplysninger figurerer, som tidligere nævnt, vagt i *direktivets* artikel 12, der bemærker, at den registrerede har ret til at få slettet personoplysninger, hvis dataindsamlingen eller –behandlingen ikke overholder *direktivets* forskrifter (European Parliament & the Council of the European Union, 1995:42).

Det afgørende for at dette instrument kan siges at være succesfuldt inkorporeret i *forordningen* er, hvorvidt der i *forordningen* kan identificeres en mere kraftfuld formulering af *right to be forgotten*.

Formuleringen af denne rettighed sker i artikel 17, der omtaler retten til sletning. Artikel 17(1) fastlægger at den registrerede har ret til at få den dataansvarlige til at slette personoplysninger, hvis et af seks forhold gør sig gældende: (1) personoplysningerne er ikke længere nødvendige for at opfylde det formål, som de blev indsamlet til eller behandlet for; (2) den registrerede trækker samtykket til databehandlingen tilbage og behandlingen ikke er funderet på et andet retsgrundlag; (3) den registrerede gør indvending imod databehandlingen på baggrund af artikel 21(1) og 21(2) og der ikke forefindes andre legitime grunde for databehandlingen; (4) personoplysningerne er blevet behandlet ulovligt; (5) en retlig forpligtelse, enten i EU-retten eller i national lovgivning, medfører at den dataansvarlige skal slette personoplysningerne; og (6) hvis personoplysningerne er blevet indsamlet i forlængelse af en udbudt informationssamfundsservice i henhold til artikel 8(1).

Herudover er den dataansvarlige, jævnfør artikel 17(2), forpligtiget til at:

”(…) take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data.”

Sammenlignes *forordningens right to be forgotten* med *direktivets* ditto er det tydeligt, at rettigheden i *forordningen* er mere udspecificeret. Selvom *forordningens right to be forgotten* kunne have været mere omfattende, den registrerede kunne for eksempel være blevet udstyret med retten til at kræve sine personoplysninger slettet uden aktiveringen af

et af de førnævnte seks forhold, er den nye version mere omfattende end tidligere og må derfor siges at være succesfuldt implementeret.

To andre artikler understøtter retten til sletning af personoplysninger i bestræbelsen på at mindske udbredelsen af BD-forudsigelser og automatiske afgørelser. Artikel 21(1) giver den registrerede indsigelsesret i relation til behandling af personoplysninger, hvilket betyder, at den dataansvarlige ikke må behandle disse personoplysninger, såfremt der ikke kan demonstreres legitime grunde til databehandlingen, som overstiger hensynet til den registreredes rettigheder og friheder.

Artikel 22 (1) slår fast at den registrerede har ret til ikke at være underlagt beslutninger udelukkende baseret på automatisk behandling og afgørelser. Dette er dog, jævnfør artikel 22(2), ikke gældende, hvis den automatiske behandling og afgørelse er nødvendig for indgåelsen eller opfyldelsen af et kontraktligt forhold mellem den dataansvarlige og den registrerede, hvis den automatiske behandling og afgørelse har hjemmel i enten EU-retten eller i national lovgivning samt hvis den registrerede har givet sit utvetydige samtykke.

5.3. Åbningen af big datas sorte boks i forordningen?

Åbningen af den sorte boks, der udgør BD's maskinrum, er en strategi, der baserer sig på instrumenterne algoritmisten og større adgang for den akademiske verden.

5.3.1. Algoritmisten

Strategiens omtale af interne og eksterne algoritmister, til at kaste lys over BD-algoritmerne, genfindes i det store hele i *forordningens* kapitel IV, afdeling 4 omhandlende den såkaldte DPO.

Artikel 37(1) kræver at en dataansvarlig eller databehandler udnævner en DPO når:

"(a) the processing is carried out by a public authority or body, except for courts acting in their judicial capacity; (b) the core activities of the controller or the processor consist of processing operations which, by virtue of their nature, their scope and/or their purposes, require regular and systematic monitoring of data subjects on a large scale; or (c) the

core activities of the controller or the processor consist of processing on a large scale of special categories of data pursuant to Article 9 and personal data relating to criminal convictions and offences referred to in Article 10."

Den dataansvarlige og databehandleren skal, ifølge artikel 38(3), sikre, at den udnævnte DPO i sin opgavevaretagelse skal kunne fungere uafhængigt og ikke må blive afskediget eller straffet for sit virke.

DPO'ens opgavevaretagelse, se artikel 39(1), udgøres af minimum fem opgaver: (1) informere og rådgive den dataansvarlige eller databehandleren, og deres relevante ansatte, om forpligtigelser udstukket i *forordningen* samt i anden EU-ret eller i national lovgivning omhandlende databeskyttelse; (2) monitorere overholdelsen af *forordningen* samt anden EU-ret eller national lovgivning på databeskyttelsesområdet og overvåge den dataansvarliges eller databehandlerens databeskyttelsespolitikker, hvilket blandt andet indebærer ansvarsfordeling, bevidstgørende kampagner, efteruddannelse af ansatte involveret i databehandling og databeskyttelsesrevisioner; (3) rådgive om konsekvensanalysen og tilse dens opfyldelse; (4) samarbejde med tilsynsmyndigheden; og (5) agere kontaktpunkt for tilsynsmyndigheden i databeskyttelsessager.

Tillige giver artikel 38(4) en registreret mulighed for at tage kontakt til en DPO for at få svar på spørgsmål angående brugen af personoplysninger samt muligheden for udøvelse af rettigheder udfoldet i *forordningen*.

Udpegelsen af en DPO, jævnfør artikel 37(5), afhænger af dennes faglige kvalifikationer inden for særligt databeskyttelseslovgivningsområdet og kapabiliteten til at varetage de nævnte opgaver i artikel 39(1).

DPO'en minder dermed i udgangspunktet om algoritmisten. Et par elementer gør dog, at *forordningens* DPO ikke er identisk med algoritmisten. De teoretiske overvejelser om algoritmisten opererer med både en intern og ekstern algoritmist, hvilket *forordningen* ikke berører. Derudover lægges der vægt på, at de personer der skal bestride posten som algoritmist er eksperter inden for datalogi, matematik og statistik, mens *forordningen* kun opstiller kravet om stærke kompetencer inden for databeskyttelseslovgivning og til at

håndtere de føromtalte arbejdsopgaver. Samtidig er en hovedopgave for algoritmisten, at denne blandt andet skal udøve dybdegående kontrol med algoritmerne samt deres forudsigelser og automatiske afgørelser. En rolle som ikke kommer til syne i *forordningen*.

Overordnet set minder algoritmisten og DPO'en om hinanden men graves der lidt i overfladen, kommer der en række forskelligheder til syne. Disse vidner om at DPO'en ikke kommer til at indtage en rolle, som til fulde lever op til de teoretiske anvisninger omhandlende algoritmisten.

5.3.2. Adgang for forskere

Belysningen af BD's sorte boks kan ligeledes fremmes gennem en større inddragelse af og adgang for den akademiske verden i relation til algoritmerne. Det står dog klart, at dette instrument ikke får nogen gang på jorden, da der ikke på noget tidspunkt i *forordningen* etableres initiativer eller bestemmelser som fremmer inddragelsen af de akademiske verden.

5.4. Featurization af big data i forordningen?

Featurization af BD er en måde, udover sanktioneringsmulighederne og den forøgede gennemsigtighed fra PTA, hvorpå det sikres, gennem *empowerment* af fysiske personer, at BD-aktørerne overholder deres databeskyttelsesforpligtigelser.

5.4.1. Empowerment af fysiske personer

Featurization-strategiens noget idealistiske kald for en deling af BD-gevinsterne mellem BD-aktørerne og de registrerede eksisterer ikke direkte i *forordningen* men kan opstå som en afledt effekt af de *empowerment*-tiltag, der kan peges på i *forordningen*.

Den stærkeste *empowerment* af fysiske personer forekommer gennem retten til dataportabilitet, som artikel 20 udbreder. Retten til dataportabilitet betyder, i henhold til artikel 20(1), at:

"The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and

machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided (...).”

Ydermere bevirker artikel 20(2), at den registrerede har ret til, hvis det er teknisk muligt, at få overført personoplysninger direkte fra en dataansvarlig til en anden, når den registrerede udøver sin ret til dataportabilitet. Retten til dataportabilitet er gældende, jævnfør artikel 20(1), når databehandlingen er baseret på samtykke eller et kontraktligt forhold og gennemføres automatisk.

Med retten til dataportabilitet ved hånden har en registreret mulighed for at videregive sine personoplysninger til en anden dataansvarlig eller tredjepart og kan derved for eksempel få adgang til en tjeneste eller service, som den registrerede kan kapitalisere på. Den registrerede kunne eksempelvis give en *app* adgang til den datastrøm, som en intelligent elmåler producerer, hvorefter *appen* kommer med forslag til, hvordan elforbruget kan nedsættes, hvilket i sidste ende kommer den registreredes pengepung til gode. Dermed tilfalder gevinsterne ikke blot BD-aktører men også fysiske personer.

Øget inddragelse og *empowerment* af fysiske personer på området skal ligeledes skabe en større interesse og bevågenhed for databeskyttelsen, hvilket vil øge presset på BD-aktører til at efterleve *forordningen*. Dette skyldes at negative databeskyttelsessager kan resultere i dårlig omtale for en BD-aktør. I forlængelse af dette kan det tænkes, at registrerede i en sådan situation gør brug af deres ret til dataportabilitet og flytter personoplysninger over til en dataansvarlig, der har et stærkt ry inden for databeskyttelse og herefter benytter sig af retten til at begrænse den oprindelige dataansvarliges databehandling fastlagt i artikel 18.

Retten til dataportabilitet er det stærkeste eksempel på *empowerment* i *forordningen*, men et par andre elementer bidrager tillige til *empowerment*-indholdet. Her kan der peges på: (1) oplysningspligten i artikel 13 og 14 hvor den dataansvarlige skal stille en række oplysninger til rådighed for den registrerede; (2) indsigtsretten i artikel 15 hvor den registrerede har ret til at få adgang til et væld af oplysninger; (3) kravet om *privacy by default* i artikel 25(2) der grundlæggende skal sørge for, at personoplysninger ikke kan

videregives til et ubegrænset antal fysiske personer uden den registreredes indblanding; og (4) bestemmelsen i artikel 35(9) om at den dataansvarlige, i forlængelse af en konsekvensanalyse, blandt andet skal tage hensyn til de registreredes synspunkter.

Featurization af BDkan derfor vurderes at være til stede i *forordningen*.

5.5. Redefinering af personoplysningsbegrebet i forordningen?

Behovet for en redefinering af personoplysningsbegrebet er den femte og sidste strategi til at modvirke BD-udfordringerne, hvor den såkaldte to-trins metode anbefales af Hon, Millard og Walden.

I *forordningen* optræder personoplysningsbegrebet dog ikke i en revideret udgave. *Forordningens* anvendelsesområde fastlagt i artikel 2(1), som bekendtgør, at *forordningen* er gældende for behandling af personoplysninger, der foregår helt eller delvist automatisk samt ikke-automatisk databehandling, der vil blive indeholdt i et register, er enslydende med artikel 3 i *direktivet*, som udfolder dets anvendelsesområde (European Parliament & the Council of the European Union; 1995:39).

Ligeledes er selve definitionen på personoplysninger, jævnfør artikel 4(1), stort set den samme som i *direktivet*, med enkelte tilføjelser der tager hensyn til den teknologiske udvikling på området siden 1995. Herfor lyder *forordningens* definition på personoplysninger således:

"'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable person natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".

Dette gør sig tillige gældende for de særlige kategorier af personoplysninger, hvor *forordningens* artikel 9(1), der forbyder behandlingen af personoplysninger der afslører race eller etnisk oprindelse, politiske holdninger, religiøse eller filosofiske overbevisninger,

fagforeningsmedlemskab, helbredstilstand, seksuelle forhold, seksuelle orientering samt behandlingen af genetiske eller biometriske data med det formål at identificere en fysisk person, er identisk med *direktivets* artikel 8, hvor de enkelte ændringer der forekommer, igen skyldes hensynet til den teknologiske udvikling (European Parliament & the Council of the European Union, 1995:40-41).

Det står dermed klart at *forordningen* ikke indeholder spor af to-trins metoden eller andre revideringer af personoplysningsbegrebet. Herfor aktiveres denne strategi ikke i *forordningen*.

5.6. Kan forordningen modvirke big data-udfordringerne?

Vurderingen af hvorvidt *forordningen* kan afbøde BD-udfordringerne afhænger af i hvor høj grad, de fem strategier og deres instrumenter er integreret i *forordningen*. Den nedenstående tabel giver et overblik over strategiernes vurderede tilstedeværelse i *forordningen*.

Tabel 4: Er strategierne at finde i forordningen?

Strategi	Ja	Delvist	Nej
Privacy through accountability		/	
Foranstaltninger imod *big data*-forudsigelser	✓		
Åbning af *big datas* sorte boks			✘
Featurization* af *big data	✓		
Redefinering af personoplysningsbegrebet			✘

Selvom instrumenterne frihed under ansvar, PIA, dataetik og innovative tekniske løsninger forefindes i *forordningen* vurderes PTA samlet set til blot at være delvist opfyldt i *forordningen*. Dette skyldes at vigtige elementer til at sikre databeskyttelsen, i en kontekst hvor BD-aktører har større frihedsgrader, ikke eksisterer eller fremstår mangelfulde i

forordningen. Førstnævnte gør sig gældende for instrumenterne gennemsigtighed og udløbsdatoer på personoplysninger og samtykker, mens sidstnævnte er aktuel i relation til instrumentet fair databehandling.

Strategierne foranstaltninger imod BD-forudsigelser og *featurization* af BD kan omvendt opfattes som værende indbygget i *forordningen*, idet disses instrumenter i form af åbenhed, certificering og bevislighed, *right to be forgotten* og *empowerment* ganske klart kan genfindes i *forordningen*.

Den manglende inkorporering af de to strategier omhandlende åbningen af BD's sorte boks og redefineringen af personoplysningsbegrebet er dog en malurt i bægeret. At disse to strategier ikke aktiveres gennem *forordningen* skyldes den utilfredsstillende fremtoning af algoritmist-instrumentet, den manglende implementering af adgang for den akademiske verden og den udeblevne opdatering af personoplysningsbegrebet.

Spørgsmålet der nu mangler at blive besvaret er om *forordningen* kan modvirke BD-udfordringerne for databeskyttelsen?

Først og fremmest er det tydeligt, at *forordningen* giver en klart bedre databeskyttelse sammenlignet med *direktivet*. *Forordningens* indhold er mere bredt og dybt funderet end *direktivet* og er ligeledes tilpasset det nuværende teknologiske udviklingsniveau. Men dette er jo i bund og grund det mindste der kunne forventes af en forordning, der er over 20 år yngre end sin forgænger på databeskyttelsesområdet.

Desværre er den samlede pallette af strategier ikke inkorporeret, hvilket betyder, at *forordningen* ikke fuldt ud kan modvirke BD-udfordringerne for databeskyttelsen. *Forordningen* udstyrer BD-aktører, i overensstemmelse med PTA, med en større frihed til at anvende personoplysninger til at udforske *secondary uses*, hvis det vurderes, at dette *secondary use* fortsat er foreneligt med det formål, som personoplysningerne i første omgang blev indsamlet til. Dette styrker BD-aktørers innovationsevne og mulighed for at opdage nye forretningsmuligheder og er derfor et vigtigt led i opnåelsen af *forordningens* delmål om både at understøtte personoplysningers frie bevægelighed og etableringen af et digitalt indre marked.

Denne større frihed skal balanceres med instrumenter, der styrker databeskyttelsen, *forordningens* andet delmål, og sørger for at BD-aktørerne er deres frihedsgrader værdige samt bevidste om deres ansvar.

Forordningen går et stykke af vejen for at skaffe denne balance gennem: (1) robuste sanktioneringsmuligheder; (2) differentierede krav til databehandling afhængig af hvor risikobetonet denne behandling kan kategoriseres som; (3) konsekvensanalyser før påbegyndelsen af visse former for databehandling; (4) udviklingen af dataetik; (5) indbyggede tekniske løsninger i teknologi der sikrer bedre databeskyttelse i form af *privacy by design* og *privacy by default*; (6) større åbenhed og bedre bevislighed vis-a-vis automatiske afgørelser baseret på BD; (7) certificering af BD-algoritmer; (8) retten til at begære personoplysninger slettet; og (9) *empowerment* af fysiske personer.

Forordningen efterlader dog fortsat en række databeskyttelseshuller. Disse optræder i skikkelse af: (1) en manglende tilgængelighed af bredere og kontekstuel information som ej heller påkræves at skulle viderekommunikeres til registrerede på mere overskuelige måder; (2) princippet om fair databehandling der under overfladen virker udhulet; (3) den ikke eksisterende beskæftigelse med specifikke udløbsdatoer for opbevaringen af personoplysninger og samtykker; (4) en etablering af en DPO-stilling der ikke er omfattende nok til at modsvare algoritmisten; (5) den fortsatte udelukkelse af den akademiske verden i relation til BD-algoritmerne; og (6) fastholdelsen af personoplysningsbegrebet fra *direktivet*.

Samtidig viser analysen at der kan stilles spørgsmålstegn ved, hvorvidt den påkrævede konsekvensanalyse er omfattende nok til at kunne håndtere BD's abstrakte og uhåndgribelige udfordringer, om certificeringen af algoritmer bliver udvandet, idet denne er frivillig, og hvorfor den såkaldte *right to be forgotten* ikke udfoldes helt.

Dette betyder at *forordningen*, til en vis grad, vil formå at modvirke BD-udfordringerne og dermed sikre en bedre databeskyttelse. De påpegede huller og spørgsmålstegn gør dog at to vigtige elementer på databeskyttelsesområdet, i en tid med stigende udbredelse af BD, ikke adresseres godt nok; nemlig funktionsmåden, mekanikken og opbygningen af det der

befinder sig under motorhjelmen på BD, dens sorte boks, samt det forældede personoplysningsbegreb.

Før der for alvor åbnes op for BD's sorte boks så registrerede, beslutningstagere og andre databeskyttelsesaktører kan få en fuldbyrdet indsigt i BD's indre virkemåde, kan der ikke etableres en databeskyttelse, der dækker alle dele af BD-skyggesiden og dens udfordringer.

Det forældede personoplysningsbegreb bevirker at truslen, selvom *forordningen* rummer krav om tekniske databeskyttelsesløsninger, om re-identificering af anonymiserede registrerede fortsat er omnipotent.

I lyset af disse overvejelser er det dermed tydeligt, at der i *forordningen* er plads til klare forbedringer for at sikre en fyldestgørende databeskyttelse i EU. Heldigvis giver to ting i *forordningen* grund til optimisme.

Den første grund er den førnævnte artikel 6(3.b), der indeholder bemærkningen om, at medlemsstaterne i deres nationale retsgrundlag kan fastsætte yderligere lovgivning, der tilpasser anvendelsen af *forordningens* bestemmelser. En medlemsstat kan dermed forsøge at lukke nogle af de huller og fjerne nogle af de problemstillinger der må siges at præge *forordningen*, så længe den nationale lovgivning holder sig inden for *forordningens* rammer.

Den anden grund er *forordningens* oprettelse af EDPB, se artikel 68(1), der blandt andet i artikel 70(1.b) og 70(1.e) udstyres med opgaverne om at rådgive Kommissionen i spørgsmål vedrørende databeskyttelse i EU herunder i relation til forslåede ændringsforlag til *forordningen* og udstedelse af retningslinjer, henstillinger og *best practice*-procedurer på databeskyttelsesområdet. EDPB skal fungere uafhængigt, jævnfør artikel 69, og dermed kan en aktivistisk EDPB-bestyrelse arbejde for at styrke *forordningen* på de punkter, der halter efter.

6. Konklusion

Denne afhandling har beskæftiget sig med BD og databeskyttelse i EU, hvor det blev undersøgt, hvilke udfordringer BD giver for databeskyttelse og hvorvidt *forordningen* kan modvirke disse.

Databeskyttelse på EU-niveau kom i starten af 1970'erne på den politiske agenda, men først i 1990 præsenterede Kommissionen en databeskyttelsespakke, som blandt andet mundede ud i vedtagelsen af *direktivet* i 1995. Teknologiens ofte hurtige fremskridt har dog tidligere gjort national databeskyttelseslovgivning forældet, hvilket var én af grundene til, at Kommissionen i 2009 påbegyndte et reformarbejde på databeskyttelsesområdet, der endte ud i *forordningens* vedtagelse i 2016.

Når *forordningen* i 2018 finder anvendelse, har den været ni år undervejs og i mellemtiden er den teknologiske udvikling fortsat med fremkomsten af blandt andet BD. BD rummer en anseelig mængde fordele, men der findes dog ligeledes en skyggeside som gør, at BD kan defineres som et potentielt *Weapon of Math Destruction*, der medfører en række udfordringer for databeskyttelsen.

I forlængelse af denne situation undersøgte afhandlingen det centrale spørgsmål om, hvorvidt *forordningen* er i stand til at modvirke disse udfordringer eller om *forordningen* allerede inden den finder anvendelse er gjort utidssvarende?

For at afklare dette spørgsmål kortlagde afhandlingen først BD-skyggesidens udfordringer gennem en teoretisk diskussion. Gennem denne diskussion blev det afdækket, at BD's udfordringer for databeskyttelsen er todelt. På den ene side skaber eller forstærker BD fænomener som *dataveillance*, *secondary uses*, profilering, *weblining*, BD's sorte boks og forøgede trusler imod datasikkerheden, der bevirker at databeskyttelse som rettighed sættes under pres og på den anden side underminerer BD traditionelle databeskyttelsesmetoder såsom *notice and consent*, *opt-out* og anonymisering.

Ved at inddrage et antal teoretiske anskuelser blev afhandlingen i stand til at udvikle en fempunktsmodel for, hvordan disse BD-udfordringer kan afmonteres.

Fempunktsmodellen indeholdte strategierne PTA, foranstaltninger imod BD-forudsigelser, åbningen af BD's sorte boks, *featurization* af BD og redefinering af personoplysningsbegrebet.

Herefter anvendte afhandlingen en dokumentanalytisk indholdsanalyse af selve *forordningen* for at iagttage, hvorvidt fempunktsmodellens strategier kunne identificeres efter tesen om, at jo mere af fempunktsmodellens indhold der kunne lokaliseres i *forordningen*, jo stærkere ville den være til at afbøde BD-udfordringerne for databeskyttelsen.

Analysen viste at blot to af strategierne, foranstaltninger imod BD-forudsigelser og *featurization* af BD, var indeholdt i *forordningen*, mens PTA kun delvist kunne siges at være tilstede. Omvendt stod det klart at åbningen af BD's sorte boks og redefinering af personoplysningsbegrebet ikke var inkorporeret i *forordningen*.

Selvom *forordningen* går et stykke af vejen for at sikre databeskyttelsen over for BD, må det konkluderes, at *forordningen* ikke er fuldt ud i stand til at modvirke BD-udfordringerne for databeskyttelsen idet en række centrale databeskyttelseshuller efterlades uopfyldte. Disse huller kredser i særdeleshed om at logikken bag BD ikke blotlægges i tilfredsstillende grad, hvilket betyder at etableringen af en hensigtsmæssig databeskyttelse umuliggøres og at personoplysningsbegrebet er altmodisch i en tid, hvor blandt andet re-identificering af anonymiserede fysiske personer er mulig i mange situationer.

Forordningens manglende integrering af essentielt indhold fra strategierne betyder, at den allerede nu ligner en databeskyttelseslovgivning der, endda inden den overhovedet finder anvendelse, er på vej til at blive overhalet indenom af den teknologiske udvikling. Der kan dog findes trøst i, at *forordningen* efterlader plads til, at proaktive medlemsstater og EDPB kan lave ændringer eller tilføjelser til *forordningen*, som måske kan bringe den tilbage i kapløbet med den teknologiske udvikling.

Litteraturliste

Article 29 Data Protection Working Party (2014): *Statement on the role of a risk-based approach in data protection legal frameworks*, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf (senest tilgået d. 17/5-17).

Barbaro, Michael & Zeller Jr., Tom (2006): *A Face Is Exposed for AOL Searcher No. 4417749*, Aug. 9, 2006, http://www.nytimes.com/2006/08/09/technology/09aol.html (senest tilgået d. 18/4-17).

Boyd, Danah & Crawford, Kate (2011): *Six Provocations for Big Data*, Oxford Internet Institute's "A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society" on September 21, 2011.

Bryman, Alan (2012): *Social Research Methods*, 4th edition, Oxford University Press.

Chin, Andrew & Klinefelter, Anne (2012): *Differential Privacy as a Response to the Reidentification Threat: The Facebook Advertiser Case Study*, North Carolina Law Review, vol. 90, no. 5, 2012, 1417-1456.

CIPL (2014): *The Role of Risk Management in Data Protection*, Paper of the Project on Privacy Risk Framework and Risk-based Approach to Privacy, Centre for Information Policy Leadership at Hunton & Williams LLP.

CIPL (2015): *The Role of Enhanced Accountability in Creating a Sustainable Data-driven Economy and Information Society*, Protecting Privacy in a World of Big Data, Paper 1, Centre for Information Policy Leadership at Hunton & Williams LLP.

CIPL (2016): *The Role of Risk Management*, Protecting Privacy in a World of Big Data, Paper 2, Centre for Policy Leadership at Hunton & Williams LLP.

Clarke, Roger A. (1988): *Information Technology and Dataveillance*, Communications of the ACM, vol. 31, no. 5, 1988, 498-512.

Commission of the European Communities (1972): *The European Community and Data Processing – Government Development Aids Permitted*, Information (Competition) 21/72, http://aei.pitt.edu/7734/1/31735055263010_1.pdf (senest tilgået d. 18/4-17).

Commission of the European Communities (1973): *Communication of the Commission to the Council: A Community Policy for Data Processing*, SEC(73) 4300 final, http://aei.pitt.edu/6337/1/6337.pdf (senest tilgået d. 18/4-17).

Commission of the European Communities (1981): *Commission Recommendation of 29 July 1981 relating to the Council of Europe convention for the protection of individuals with regard to automatic processing of personal data (81/679/EEC)*, Official Journal of the European Communities No L 246/31 29.8.81.

Commission of the European Communities (1990a): *Commission Communication on the protection of individuals in relation to the processing of personal data in the Community and information security*, COM(90) 314 final – SYN 287 and 288, http://aei.pitt.edu/3768/1/3768.pdf (senest tilgået d. 22/4-17).

Commission of the European Communities (1990b): *Proposal for Council Directive concerning the protection of individuals in relation to the processing of personal data*, COM(90) 314 final – SYN 287, Official Journal of the European Communities No C 277/3 5.11.90.

Commission of the European Communities (1992): *Amended proposal for a Council Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data*, COM(92) 422 final – SYN 287, Official Journal of the European Communities No C 311/30 27.11.92.

Conley, Chris (2010): *The Right to Delete*, AAAI Spring Symposium Series 2010, 53-58.

Constine, Josh (2012): *How Big Is Facebook's Data? 2.5 Billion Pieces Of Content And 500+ Terabytes Ingested Every Day*, TechCrunch, Aug 22, 2012, https://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/ (senest tilgået d. 1/5-17).

Council of Europe (1981): *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*, ETS No. 108, 28/01/1981.

Council of the European Communities (1977): *Council Decision of 27 September 1977 on adopting a series of studies in support of the use of informatics (77/616/EEC)*, Official Journal of the European Communities No L 255/25 6.10.77.

Council of the European Communities (1984): *Council Decision of 10 April 1984 amending Decision 79/783/EEC adopting a multiannual programme (1979 to 1983) in the field of data processing (84/254/EEC)*, Official Journal of the European Communities No L 126/27 12.5.84.

Council of the European Union (2008): *Council Framework Decision 2008/977/JHA of 27 November 2008 on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters*, Official Journal of the European Union L 350/60 30.12.2008.

Computing Research Association (2003): *Grand Research Challenges in Information Systems*, A Conference Series on Grand Research Challenges in Computer Science and Engineering, http://archive.cra.org/reports/gc.systems.pdf (senest tilgået d. 3/5-17).

Custers, Bart (2016): *Click here to consent forever: Expiry dates for informed consent*, Big Data & Society, vol. 3, no. 1, 2016, 1-6.

Danna, Anthony & Gandy, Jr., Oscar H. (2002): *All That Glitters is Not Gold: Digging Beneath the Surface of Data Mining*, Journal of Business Ethics, vol. 40, no. 4, 2002, 373-386.

Davenport, Thomas H.; Barth, Paul & Bean, Randy (2012): *How Big Data is Different*, MIT Sloan Management Review, vol. 54, no. 1, 2012, 43-46.

de Vaus, David A. (2001): *Research Design in Social Research*, SAGE Publications.

Dodge, Martin & Kitchin, Rob (2005): *Codes of life: identification codes and the machine-readable world*, Environment and Planning D: Society and Space, vol. 23, no. 6, 2005, 851-881.

Dodge, Martin & Kitchin, Rob (2007): *Outlines of a World Coming into Existence: Pervasive Computing and the Ethics of Forgetting*, Environment and Planning B: Urban Analytics and City Science, vol. 34, no. 3, 2007, 431-445.

Driscoll, Kevin (2012): *From Punched Cards to "Big Data": A Social History of Database Populism*, communication +1, vol. 1, no. 1, article 4.

Duhigg, Charles (2012): *How Companies Learn Your Secrets*, The New York Times, Feb. 16, 2012, http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html (senest tilgået d. 11/5-17).

Dwork, Cynthia (2011): *A Firm Foundation for Private Data Analysis*, Communications of the AMC, vol. 54, no. 1, 2011, 86-95.

EDPS (2015): *Towards a new digital ethics: Data, dignity and technology*, European Data Protection Supervisor, Opinion 4/2015, 11 September 2015.

European Commission (1999): *Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by the institutions and bodies of the Community and on the free movement of such data*, COM (1999) 337 final – 1999/0153(COD), Official Journal of the European Communities No CE 376/24 28.12.1999.

European Commission (2010): *European Commission sets out strategy to strengthen EU data protection rules*, http://europa.eu/rapid/press-release_IP-10-1462_en.htm (senest tilgået d. 10/6-17).

European Commission (2012a): *Commission proposes a comprehensive reform of data protection rules to increase users' control of their data and to cut costs for businesses*, http://europa.eu/rapid/press-release_IP-12-46_en.pdf (senest tilgået d. 17/4-17).

European Commission (2012b): *Communication from the Commission to The European Parliament, The Council, The European Economic and Social Committee and The Committee of the Regions: Safeguarding Privacy in a Connected World. A European Data Protection Framework for the 21st Century*, http://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:52012DC0009 &from=en (senest tilgået d. 17/4-17).

European Commission (2015): *Agreement on Commission's EU data protection reform will boost Digital Single Market*, http://europa.eu/rapid/press-release_IP-15-6321_en.pdf (senest tilgået d. 17/4-17).

European Parliament (1974): *Oral Question No 193/73, with debate, on protecting the privacy of the Community's citizens*, Debates of the European Parliament, Sitting of Wednesday, 13 March 1974, http://www.europarl.europa.eu/pdf/jadis/082012/001-en.pdf (senest tilgået d. 19/4-17).

European Parliament (1975a): *Interim report drawn up on the behalf of the Legal Affairs Committee on the protection of the rights of the individual in the face of developing technical progress in the field of automatic data processing*, Legal Affairs Committee of the European Parliament, European Parliament Working

Documents 1974-1975, Document 487/74, PE 39.608/fin, 19 February 1975, http://aei.pitt.edu/65083/1/WD3126.pdf (senest tilgået d. 19/4-17).

European Parliament (1975b): *Resolution on the protection of the rights of the individual in the face of developing technical progress in the field of automatic data processing*, Official Journal of the European Communities No C 60/48 13.3.75.

European Parliament (1976): *Resolution on the protection of the rights of the individual in the face of developing technical progress in the field of automatic data processing*, Official Journal of the European Communities No C 100/27 3.5.76.

European Parliament (1979a): *Resolution on the protection of the rights of the individual in the face of technical developments in data processing*, Official Journal of the European Communities No C 140/34 5.6.79.

European Parliament (1979b): *Data protection in the Community: Oral question with debate (Doc. 1-287/79)*, Debates of the European Parliament, Sitting of Monday, 24 September 1979, http://aei.pitt.edu/64681/1/B2813.pdf (senest tilgået d. 21/4-17).

European Parliament (1982): *Resolution on the protection of the rights of the individual in the face of technical developments in data processing*, Official Journal of the European Communities No C 87/39 5.4.82.

European Parliament & the Council of the European Union (1995): *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data*, Official Journal of the European Communities No L 281/31 23.11.95.

European Parliament & the Council of the European Union (1997): *Directive 97/66/EC of the European Parliament and of the Council of 15 December 1997 concerning the processing of personal data and the protection of privacy in the telecommunications sector*, Official Journal of the European Communities No L 24/1 30.1.98.

European Parliament & the Council of the European Union (2000): *Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000*

on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data, Official Journal of the European Communities No L 8/1 12.1.2001.

European Parliament & the Council of the European Union (2002): *Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)*, Official Journal of the European Communities L 201/37 31.7.2002.

European Parliament & the Council of the European Union (2006): *Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC*, Official Journal of the European Communities L 105/54 13.4.2006.

European Parliament & the Council of the European Union (2016): *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, Official Journal of the European Union No L 119/1 4.5.2016.

European Union (1997): *Treaty of Amsterdam amending the Treaty on European Union, the Treaties establishing the European Communities and certain related acts, signed at Amsterdam, 2 October 1997*, Official Journal of the European Communities No C 340/1 1.11.97.

European Union (2000): *Charter of Fundamental Rights of the European Union*, Official Journal of the European Communities No C 364/1 18.12.2000.

European Union (2004): *Treaty establishing a Constitution for Europe*, Official Journal of the European Union No C 310/3 16.12.2004.

European Union (2010): *Consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union*, Official Journal of the European Union No C 83/1 30.3.2010.

Foucault, Michel (1995): *Discipline and Punish: The Birth of the Prison*, translated from the French by Alan Sheridan, 2nd edition, Vintage Books.

Fuster, Gloria González (2014): *The Emergence of Personal Data Protection as a Fundamental Right of the EU*, Springer International Publishing.

Gantz, John & Reinsel, David (2011): *Extracting Value from Chaos*, IDC iView sponsored by EMC Corporation, IDC Go-to-Market Services.

Gellman, Robert (2017): *Fair Information Practices: A Basic History*, SSRN.

Gillies, James & Cailliau, Robert (2000): *How the Web was Born: The Story of the World Wide Web*, Oxford University Press.

Ginsberg, Jeremy; Mohebbi, Matthew H.; Patel, Rajan S.; Brammer, Lynnette, Smolinski, Mark S. & Brilliant, Larry (2009): *Detecting influenza epidemics using search engine query data*, nature, vol. 457, no. 19, February 2009, 1012-1015.

Gleeson, Justin; Kitchin, Rob; Bartley, Brendan & Treacy, Caroline (2009): *New Ways of Mapping Social Inclusion in Dublin City*, Dublin City Partnership/Dublin City Development Board, National University of Ireland, Maynooth.

Graham, Stephen D.N. (2005): *Software-sorted geographies*, Progress in Human Geography, vol. 29, no. 5, 2005, 562-580.

Gralla, Preston; Sacco, Al & Faas, Ryan (2011): *Smartphone apps: Is your privacy protected?*, Computerworld, jul 7, 2011, http://www.computerworld.com/article/2509878/data-privacy/smartphone-apps--is-your-privacy-protected-.html (senest tilgået d. 11/5-17).

Greenfield, Adam (2006): *Everyware: The Dawning Age of Ubiquitous Computing*, New Riders Publishing.

Hilbert, Martin & López, Priscila (2011): *The World's Technological Capacity to Store, Communicate, and Compute Information*, Science, vol. 332, no. 6025, 2011, 60-65.

Hon, W. Kuan; Millard, Christopher & Walden, Ian (2011): *The problem of "personal data" in cloud computing: What information is regulated? – the cloud of unknowing*, International Data Privacy Law, vol. 1, no. 4, 2011, 211-228.

Hornung, Gerrit (2012): *A General Data Protection Regulation for Europe? Light and Shade in the Commission's Draft of 25 January 2012*, SCRIPTed, vol. 9, no. 1, 2012, 64-81.

IBM (2017): *What is big data?*, https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html (senest tilgået d. 1/5-17)

ICO (2014): *Big data, artificial intelligence, machine learning and data protection*, UK Information Commissioner's Office.

Kitchin, Rob (2014): *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*, SAGE Publications.

Koops, Bert-Jaap (2011): *Forgetting Footprints, Shunning Data Shadows: A Critical Analysis of the "Right to be Forgotten" in Big Data Practice*, SCRIPTed, vol. 8, no. 3, 2011, 229-256.

Kuner, Christopher; Cate, Fred H.; Millard, Christopher; Svantesson, Dan Jerker B. & Lynskey, Orla (2015): *Risk management in data protection*, International Data Privacy Law, vol. 5, no. 2, 2015, 95-98.

Lynggaard, Kenneth (2010): *Dokumentanalyse*, kapitel 6. i Brinkmann, Svend & Tanggaard, Lene (red.) (2010): *Kvalitiative Metoder – En Grundbog*, Hans Reitzels Forlag.

Manyika, James; Chui, Michael; Brown, Brad; Bughin, Jacques; Dobbs, Richard; Roxburgh, Charles & Hung Byers, Angela (2011): *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute.

Marr, Bernard (2016): *Big Data in Practice: How 45 Succesful Companies used Big Data Analytics to Deliver Extraordinary Results*, John Wiley & Sons.

Marz, Nathan & Warren, James (2015): *Big Data: Principles and best practices of scalable realtime data systems*, Manning Publications.

Mayer-Schönberger, Viktor (1998): *Generational Development of Data Protection in Europe*, kap. 8 i Agre, Philip E. & Rotenberg, Marc (ed.) (1998): *Technology and Privacy: The New Landscape*, The MIT Press.

Mayer-Schönberger, Viktor (2010): *Beyond Privacy, Beyond Rights: Toward a "Systems" Theory of Information Governance*, California Law Review, vol. 98, no. 6, 2010, 1853-1885.

Mayer-Schönberger, Viktor & Cukier, Kenneth (2014): *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Mariner Books.

McLannahan, Ben (2015): *Being 'wasted' on Facebook may damage your credit score*, Financial Times, October 15, 2015, https://www.ft.com/content/d6daedee-706a-11e5-9b9e-690fdae72044 (senest tilgået d. 12/5-17).

Minelli, Michael; Chambers, Michele & Dhiraj, Ambiga (2012): *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*, John Wiley & Sons.

NIST (2014): *NIST Privacy Engineering Objectives and Risk Model Discussion Draft*, https://www.nist.gov/sites/default/files/documents/itl/csd/nist_privacy_engr_objectives_risk_model_discussion_draft.pdf (senest tilgået d. 17/5-17).

Nordhaus, William D. (2001): *The Progress of Computing*, Cowles Foundation for Research in Economics, Yale University, IDEAS Working Paper Series from RePEc.

O'Neil, Cathy (2016): *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Allen Lane.

O'Neill, Judy E. (1995): *The Role of ARPA in the Development of the ARPANET, 1961 -1972*, IEEE Annals of the History of Computing, vol. 17, no. 4, 1995, 76-81.

Open Data Center Alliance (2012): *Big Data Consumer Guide*, Open Data Center Alliance, Inc.

Polonetsky, Jules; Tene, Omer & Jerome, Joseph (2014): *Benefit-Risk Analysis for Big Data Projects*, Future of Privacy Forum.

Quinn, Elias Leake (2009): *Smart Metering and Privacy: Existing Laws and Competing Policies*, paper presented to the Colorado Public Utilities Commission in May 2009, SSRN.

Rezendes, Christopher J. & Stephenson, W. David (2013): *Cyber Security in the Internet of Things*, Harvard Business Review, June 21, 2013, https://hbr.org/2013/06/cyber-security-in-the-internet (senest tilgået d. 12/5-17).

Rial, Nerea (2013): *Free access to data can improve sectors like health care*, New Europe, https://www.neweurope.eu/article/power-big-data-europe/ (senest tilgået d. 1/5-17).

Ritzer, George & Jurgenson, Nathan (2010): *Production, Consumption, Prosumption: The nature of capitalism in the age of the digital 'prosumer'*, Journal of Consumer Culture, vol. 10, no. 1, 2010, 13-36).

Rubinstein, Ira S. (2013): *Big Data: The End of Privacy or a New Beginning?* International Data Privacy Law, vol. 3, no. 2, 2013, 74-87.

Schaar, Peter (2010): *Privacy by Design*, Identity in the Information Society, vol. 3, no. 2, 2010 267-274.

Scherer, Michael (2012): *Inside the Secret World of the Data Crunchers Who Helped Obama Win*, Time, Nov. 07, 2012, http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/2/ (senest tilgået d. 3/5-17).

Seltzer, William & Anderson, Margo (2001): *The Dark Side of Numbers: The Role of Population Data Systems in Human Rights Abuses*, Social Research, vol. 68, no. 2, 2001, 481-513.

Sloan, Robert H. & Warner, Richard (2016): *The Self, the Stasi, the NSA: Privacy, Knowledge, and Complicity in the Surveillance State*, Minnesota Journal of Law, Science & Technology, vol. 17, no. 1, 2016, 347-408.

Solove, Daniel J. (2007): *"I've Got Nothing to Hide" and Other Misunderstandings of Privacy*, San Diego Law Review, vol. 44, no. 4, 2007, 745-772.

Solove, Daniel J. (2013): *Introduction: Privacy Self-management and the Consent Dilemma*, Harvard Business Review, vol. 126, no. 7, 2013, 1880-1903.

Strohm, Chris & Homan, Timothy R. (2013): *NSA Spying Row in Congress Ushers in Debate over Big Data*, Bloomberg, 25. juli 2013, https://www.bloomberg.com/news/articles/2013-07-25/nsa-spying-row-in-congress-ushers-in-debate-over-big-data (senest tilgået d. 1/5-17).

TechAmerica (2012): *Demystifying Big Data: A Practical Guide To Transforming The Business of Government*, TechAmerica Foundation's Federal Big Data Commission.

Tene, Omer & Polonetsky, Jules (2013): *Big Data for All: Privacy and User Control in the Age of Analytics*, Northwestern Journal of Technology and Intellectual Property, vol. 11, no. 5, 2013, 239-273.

The Economist (2010a): *Data, data everywhere*, Special report: Managing information, The Economist, Feb 25th 2010, http://www.economist.com/node/15557443 (senest tilgået d. 1/5-17).

The Economist (2010b): *All too much*, Special report: Managing information, The Economist, Feb 25th 2010, http://www.economist.com/node/15557421 (senest tilgået d. 1/5-17).

The Economist (2014): *The internet of things (to be hacked)*, The Economist, Jul 12th 2014, http://www.economist.com/news/leaders/21606829-hooking-up-gadgets-web-promises-huge-benefits-security-must-not-be (senest tilgået d. 12/5-17).

Townsend, Anthony M. (2013): *Smart Cities: Big Data, Civic Hackers, and the Quest for a new Utopia*, W. W. Norton & Company.

Trikha, Bindu (2010): *A Journey from floppy disk to cloud storage*, International Journal on Computer Science and Engineering, vol. 2, no. 5, 2010, 1449-1452.

Wei, Yanhao; Yildirim, Pinar; Van den Bulte, Christophe & Dellarocas, Chrysanthos (2015): *Credit Scoring with Social Network Data*, Marketing Science, vol. 35, no. 2, 2015, 234-258.

Zikopoulos, Paul C.; Eaton, Chris; deRoos, Dirk; Deutsch, Thomas & Lapis, George (2012): *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill.