Datatilsynet

Kunnskapsdepartementet
Postboks 8119 Dep
0032 OSLO

Deres referanse	Vår referanse	Dato
14/3274	14/00736-2/EOL	14. oktober 2014

Datatilsynets høringsuttalelse - NOU 2014:5 MOOC til Norge

Vi viser til deres brev av 24. juni 2014 hvor dere sender utredningen NOU 2014: 5 MOOC til Norge - Nye digitale læringsformer i høyere utdanning, ut på høring.

Våre kommentarer er knyttet til de sidene av utredningen som berører problemstillinger som gjelder personvern. Slik vi ser det gjelder dette særlig tema som verifisering av identitet, læringsanalyse, kvalitetsutvikling/forskning og finansiering.

Innledningsvis ønsker vi å påpeke at utvalget ser ut til å ha gjort et grundig utredningsarbeid på de temaene de har valgt å vektlegge. Det er for eksempel positivt at det er innhentet betenkninger fra to kompetente fagpersoner innen opphavsrett.

På den annen side ser vi at utvalget ikke har sett nødvendigheten av å foreta en konsekvensutredning av personvernkonsekvenser i forbindelse med bruk av MOOC (Massive Open Online Courses). Sider ved MOOC som utfordrer personvernet er nevnt enkelte steder i utredningen, men er konsekvent utelatt når utvalget skal oppsummere sine anbefalinger og råd etter hvert kapittel.

Vi mener det er en svakhet ved utredningen at den ikke går nærmere inn på personvernproblematikk, og heller ikke gir klare anbefalinger til ansvarlige myndigheter og/eller utdanningsinstitusjonene om å avklare hvilke plikter som påløper etter personvernregelverket ved å ta i bruk MOOC.

Vi vil i det følgende komme med kommentarer til de kapitlene vi mener særlig reiser problemstillinger knyttet til personvern.

Kapittel 8 Dokumentasjon av oppnådd kompetanse.
I dette kapittelet er det særlig pkt 8.4 Verifisering av identitet som reiser problemstillinger som vedrører personvern.

Som utvalget selv fastslår handler verifisering av identitet i MOOC om å kreditere riktig person for oppnådd kompetanse. Dette har betydning for forvaltningen av identitetsopplysninger ved bruk av MOOC.

Postadresse:	Kontoradresse:	Telefon:	Telefaks:	Org.nr:	Hjemmeside:
Postboks 8177 Dep 0034 OSLO	Tollbugt 3	22 39 69 00	22 42 23 50	974 761 467	www.datatilsynet.no

Vi mener at de foreslåtte løsningene for bruk av biometri ikke er gode nok, og det vil ikke tilfredsstille noen trygg og sikker identifisering til dette formålet. Det vil være for enkelt å komme seg rundt slike løsninger. Vi foreslår heller at man avholder en fysisk eksamen på campus eller testsenter. Dersom man velger å la studenten avlegge eksamen over nett, bør identifisering gjøres ved at autorisert testpersonell foretar en manuell ID-kontroll i kombinasjon med bruk av e-ID ved levering. Det kan for eksempel være at man signerer elektronisk ved innlevering hvor man erklærer at det er riktig vedkommende som har utført eksamen.

Kapittel 10 Kvalitet og læringsutbytte
I dette kapittelet er det særlig pkt 10.3 Læringsanalyse og pkt 10.5.3 Virkemidler og premisser for kvalitetsutvikling som reiser problemstillinger som vedrører personvern.

Læringsanalyse
Utvalget trekker selv frem læringsanalyse som en av de mest sentrale endringene som heldigitale læringstilbud vil føre med seg. Sammenstilling av opplysningene lagret i digitale læringsressurser muliggjør analyser som tidligere ikke har vært mulig. Utvalget trekker frem at analyser kan gjøres på både makronivå (internasjonalt, regionalt, nasjonalt), mesonivå (institusjon) og mikronivå (deltakere og deltakergrupper).

Særlig analyse av data på mikronivå reiser problemstillinger som berører personvern. Viktige byggesteiner i personvernet er det enkelte menneskets rett og mulighet til selvbestemmelse, og medbestemmelse og kontroll med hvordan opplysninger om vedkommende blir brukt. I tillegg er formålsbestemthet, relevans, minimalitet, fullstendighet og kvalitet ved bruk av personopplysninger viktige prinsipper for ivaretakelse av personopplysningsvern[1].

Forskningsetikken er et eksempel på hvordan disse prinsippene er innbakt som spilleregler for all forskning som inkluderer behandling av personopplysninger. Prinsippet om selvbestemmelse gjenspeiles i hovedregelen om at bruk av personopplysninger i forskning forutsetter et frivillig, informert og uttrykkelig samtykke fra den opplysningene gjelder. Krav om formålsbestemthet er fremtredende i forskning og ivaretar personvernet ved at det for forskningsdeltaker er forutsigbart hva opplysningene skal brukes til.

I likhet med andre analysemetoder som baseres på stordataanalyse er det ved læringsanalyse utfordrende å ivareta enkeltpersoners rett til kontroll med personopplysningene sine. Dette fordi litt av poenget med å lagre store mengder personopplysninger er muligheten til bruk av disse personopplysningene til nye formål og bruksmuligheter som ikke nødvendigvis er forutsigbare på innsamlingstidspunktet.

Av de fem ulike formene for læringsanalyse som utvalget beskriver er det særlig prediktiv analyse og analyse av sosiale nettverk som er problematiske med hensyn til å ivareta studentenes personvern.

[1] Prinsippene er nærmere beskrevet i Datatilsynets rapport «Big data – personvernprinsipper under press»

Prediktiv analyse har som mål å forutsi hver enkelt students prestasjoner. Analysen gjøres ved hjelp av å kombinere demografiske data og tidligere studieresultater med opplysninger om for eksempel innloggingsmønstre på læringsplattformen, hvilke dokumenter studenten arbeider med og i hvilket omfang studenten deltar i nettdiskusjoner. Det er vanskelig å se for seg hvilke rammer som skal gjelde for denne type læringsanalyse, og hva som er yttergrensen for hvilke opplysninger som kan være nyttige. Det vil for eksempel kunne argumenteres med at god prediktiv analyse forutsetter at det meste av det studenten foretar seg i studietiden registreres og lagres, i tillegg til at studentens sosiale profil og forhistorie er tilgjengelig for studiestedet. En slik tilnærming utfordrer klart prinsippene om minimalitet og relevans fordi studiestedet med så bred datafangst vil sitte med mye overskuddsinformasjon. I tillegg vil det være svært utfordrende å forespeile studenten hva summen av de opplysningene som blir samlet inn vil vise, og hvilket bilde av vedkommende opplysningene samlet vil kunne gi.

Den type læringsanalyse som går ut på å analysere sosiale nettverk har som mål å identifisere deltakere som ikke er sosialt og faglig integrert. Utvalget viser til forskning som viser at det er en klar sammenheng mellom samarbeid med andre og testresultater. De som samarbeider oppnår bedre resultater. Vi mener det er et betimelig spørsmål om en så omfattende overvåking av studentenes atferd er et forholdsmessig virkemiddel for å oppnå mer samarbeid mellom studenter.

Datatilsynet publiserte for et år tilbake en rapport som tar for seg personvernutfordringene knyttet til Big Data. Flere av synspunktene som fremkommer i denne rapporten har overføringsverdi til temaet læringsanalyse. For eksempel gjelder dette bekymringen omkring det vi kaller ***datadeterminisme***. I vår rapport «Big data – personvernprinsipper under press» er dette begrepet beskrevet slik:

«(...)utstrakt bruk av (...) prediksjonsanalyse kan befeste eksisterende fordommer og forsterke sosial ekskludering og lagdeling. En utvikling der stadig flere beslutninger blir tatt basert på algoritmer, kan lede til et "dataenes diktatur"; vi blir ikke vurdert ut fra hva vi faktisk foretar oss, men på basis av hva alle dataene om oss sier at vi sannsynligvis kan komme til å gjøre[2]*. «*

Vi vedlegger rapporten til inspirasjon i det videre arbeidet.

Vår anbefaling er at det må gjennomføres en utredning av personvernkonsekvenser av særskilt læringsanalyse før MOOC tas i bruk.

Virkemidler og premisser for kvalitetsutvikling
Utvalget drøfter under dette punktet hvilke strategiske grep som må tas for at utdanningsinstitusjonene skal kunne ta i bruk de pedagogiske mulighetene ved MOOC.

Datatilsynet er helt enig i at det bør tas noen strategiske grep, og vi er spesielt opptatt av de nasjonale rammebetingelsene. Med rammebetingelser i sammenheng med personvern og digitale læringsressurser mener vi avklaring av yttergrensen for hva et studiested kan

[2] Big data – personvernprinsipper under press, s. 6

«pålegge» en student å oppgi av personopplysninger – særlig når det er tale om å oppgi opplysningene til en tredjepart utenfor studiestedets kontrollsfære.

Personvern i skolen har vært et prioritert område for Datatilsynet i 2013-2014. Vi har i den sammenheng vært på kontroll hos en rekke grunn- og videregående skoler og sett særskilt på bruk av digitale læringsressurser. Funnene som vi gjorde her mener vi er relevante også for høyere utdanning.

Et fremtredende funn er at med mindre bruk av digitale læringsressurser har en økonomisk side (koster penger) har skoleeiere ikke kontroll eller oversikt over hvilke digitale læringsressurser som blir tatt i bruk i skolen. Mange digitale læringsressurser er gratis i den betydning at de ikke koster penger. «Gratis» applikasjoner har imidlertid en pris. Valutaen det betales i er studentenes personopplysninger. Identitet, kjønn, alder, interesser, tilstedeværelse, prestasjoner, kontakter, kalender etc. er opplysninger som er verdt milliarder i en industri i kraftig vekst.

For noen er denne prisen akseptabel – for andre ikke. Det problematiske med bruk av personopplysninger som betalingsmiddel i utdanningssektoren er at det ikke er studenten selv som tar avgjørelsen. Det er studiestedet som aksepterer betaling med personopplysninger på vegne av studenten.

I 2012 hadde Datatilsynet en sak til behandling som illustrerer hvorfor dette er problematisk. En gruppe studenter ved Høyskolen i Vestfold reagerte på at studiestedet hadde valgt en arbeidsplattform hvor studentene måtte legge ut sine individuelle skriftlige oppgaver åpent for hele studentkullet. Det var riktignok gitt en mulighet til å lukke siden for andre, men et slikt valg må begrunnes spesielt. Valg av publiseringsløsning var begrunnet med at studentene skulle tilegne seg digitale ferdigheter. Det kom imidlertid også frem at plattformen som var valgt var gratis, og at plattformer med bedre muligheter for ivaretagelse av personvern koster penger.

Datatilsynet anerkjenner selvsagt at studiestedene har en viss instruksjonsmyndighet over sine studenter som følger av avtalen mellom student og lærested, samt universitetsloven. Opplysninger om hva man presterer underveis i studietiden er imidlertid et eksempel på opplysninger som for mange er et anliggende mellom student og lærer/sensor. Datatilsynet mener at å legge opp til et system der alle studenter skal levere sine arbeider i åpne rom på digital plattform er å gå for langt i instruksjonsmyndigheten.

Datatilsynet ønsker ikke å være en bremsekloss for utprøving og satsing på digitale verktøy og innlæring av digitale ferdigheter hos studentene. Tvert imot mener Datatilsynet at rett bruk av digitale verktøy i undervisningssektoren kan medføre godt personvern for både studenter og ansatte. Det er imidlertid avgjørende for en god balansegang at det defineres noen yttergrenser for hva studiestedene kan pålegge studentene å oppgi av opplysninger og til hvem.

Vi merker oss at utvalget i sine vurderinger knyttet til kvalitet og læringsutbytte er innom spørsmålet om hvorvidt det kan tenkes at de store datamengdene som genereres og lagres for

bruk i særskilt læringsanalyse kan være problematisk med hensyn til personvern. Utvalgets konklusjon om at prinsipielle og uavklarte faktorer knyttet til personvern burde adresseres, gjenspeiles dessverre ikke i de anbefalingene de gir i slutten av kapittelet.

I sin drøfting og vurdering av digital kompetanse fremhever utvalget at bruk av teknologi i læring skaper behov for sammensatt kompetanse av pedagogisk, teknologisk og administrativ karakter. Datatilsynet mener at det til denne listen må tilføyes behov for kompetanse på personvern og informasjonssikkerhet.

Kapittel 18 Økonomiske og administrative konsekvenser av utvalgets anbefalinger
Utvalget har innhentet en uttalelse fra universitetslektor Gisle Hannemyr som blant annet tar for seg mulige finansieringsordninger for MOOC.

Hannemyr slår fast at MOOC ikke er gratis. Han trekker så frem aktuelle måter å finansiere MOOC på. Blant de aktuelle finansieringsordningene er det han kaller «persondata-graving». Dette innebærer at studentene for å delta må samtykke til at personopplysninger om dem blir samlet inn, og at disse kan selges til virksomheter som er villige til å betale for dem.

Hannemyr sier avslutningsvis at han av personverngrunner ikke ser for seg at persondata-graving skal bli en aktuell finansieringsform i Norge. Vi håper han har rett i dette, men med bakgrunn i kontrollene vi har gjennomført på grunn- og videregående skoler det siste året er vi imidlertid ikke overbevisst om at bevisstheten omkring personvern er høy nok til at dette er en reell terskel.

Vi anbefaler derfor at Kunnskapsdepartementet finner en finansieringsordning som hindrer at det utvikler seg en praksis hvor personopplysninger er betalingsmidlet for bruk av MOOC.

Innebygd personvern og PIA
Innebygd personvern (Privacy by Design) innebærer at det tas hensyn til personvern i alle utviklingsfaser av et system, i rutiner og i forretningspraksisen. Standardinnstillinger bør settes mest mulig personvernvennlige, og man bør bygge personvernet inn i designet. Det er viktig å ivareta informasjonssikkerheten fra start til slutt, og særlig viktig med tanke på læringsanalyse er det at det vises åpenhet. Mulige bruksområder for de dataene som samles inn bør kartlegges på forhånd, slik at et samtykke er presist og dekkende. Alt i alt handler det om å respektere brukerens rett til selvbestemmelse.

Utvalget anbefaler at norske MOOC-tilbud samles og profileres gjennom en egen nasjonal portal[3]. I den grad det blir aktuelt å bygge en norsk portal anbefaler Datatilsynet at Kunnskapsdepartementet gjør en vurdering av personvernkonsekvenser (Privacy Impact assessment (PIA)). En vurdering av personvernkonsekvenser bør blant annet inneholde en gjennomgang av mulige rettslige grunnlag for utlevering og gjenbruk av personopplysninger, prinsippene for formålsbegrensning, proporsjonalitet og dataminimalisering, samt teknisk

[3] NOU 2014:5, s. 67

tilgang og sikkerhet. Når man gjennomfører en slik vurdering bør også potensielle konsekvenser for de registrerte gjennomgås nøye[4].

I EU er det laget et PIA-rammeverk for RFID-applikasjoner for å hjelpe til med å avdekke personvernkonsekvenser som kan følge bruk av RFID. Rammeverket er laget av RFID-miljøet på oppdrag fra Artikkel 29-gruppen og kan ha overføringsverdi for forberedelse til bruk av MOOC[5].

En annen viktig del av innebygd personvern er å sørge for åpenhet. Virksomheter som benytter MOOC må være åpne om hvordan de behandler personopplysningene de samler inn. Dette innebærer blant annet å gi den enkelte innsyn i hvilke beslutningskriterier (algoritmer) som ligger til grunn for utvikling av profiler, og fra hvilke kilder opplysninger er hentet. En nasjonal portal bør tilrettelegge for slik åpenhet.

Videre kan det å sørge for dataportabilitet være innebygd personvern. En nasjonal portal bør tilrettelegge for at den enkelte kan få utlevert alle data som lagres i tjenestene i portalen i et brukervennlig format. Dataportabilitet kan hindre at kunder låses til enkelttjenester.

[4] Big Data – personvernprinsipper under press, s. 53
[5] Opinion 09/2011 on the revised Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications

Oppsummering

- Det er nødvendig med en utredning av personvernkonsekvenser før Kunnskapsdepartementet legger til rette for bruk av MOOC.

- En utredning av personvernkonsekvenser bør særskilt ta for seg temaet læringsanalyse.

- Det er nødvendig med en avklaring av hvor langt den enkelte utdanningsinstitusjon kan gå i å gjøre utlevering av personopplysninger obligatorisk.

- Dersom man velger å la studenten avlegge eksamen over nett, bør identifisering gjøres ved at autorisert testpersonell foretar en manuell ID-kontroll i kombinasjon med bruk av e-ID ved levering.

- Kunnskapsdepartementet må finne en finansieringsordning som hindrer at det utvikler seg en praksis hvor personopplysninger blir betalingsmidlet for bruk av MOOC.

- En nasjonal satsing på MOOC må sørge for innebygd personvern i en MOOC-portal. Stikkord for innebygd personvern er: personvernvennlige standardinnstillinger, bygge personvern inn i designet, informasjonssikkerhet fra start til slutt, åpenhet, forhåndskartlegging av mulige bruksområder for de dataene som samles inn, dataportabilitet.

- Det bør utarbeides nasjonale rammebetingelser for bruk av MOOC og som inkluderer innebygd personvern.

Med vennlig hilsen

Bjørn Erik Thon
direktør

Eirin Oda Lauvset
seniorrådgiver

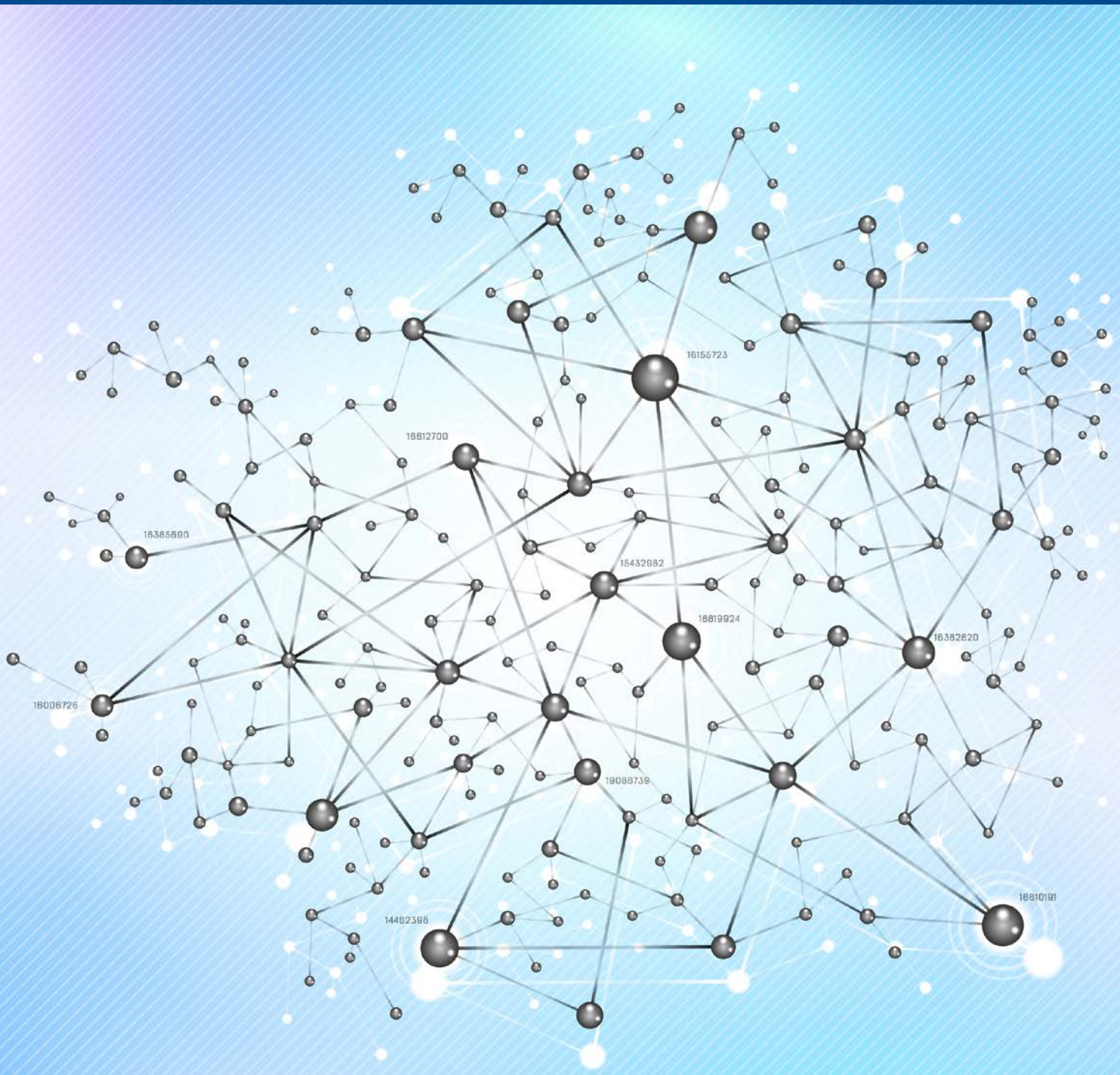
Kopi: Kommunal- og moderniseringsdepartementet
v/Statsforvaltningsavdelingen
Postboks 8112 Dep, 0032 OSLO

Datatilsynet

Big Data

– personvernprinsipper under press

September 2013



Innhold

Sammendrag 4
1 Innledning 6
1.1 Problemstilling 6
1.2 Definisjoner 7
1.2.1 Big Data 7
1.2.2 Personopplysningsbegrepet 8
1.2.3 Anonyme data og reidentifisering 8
1.3 Sentrale personvernprinsipper 8
1.4 Big Data i startgropen i Europa og Norge 9
2 Verdikjeden for Big Data: aktører, prosesser og teknologi 11
2.1 Innsamling av data 12
2.2 Lagring og aggregering 14
2.3 Analyse 15
2.4 Aktører 16
3 Personvernutfordringer knyttet til bruk av Big Data 18
3.1 Big Data i bruk blant internettbaserte selskaper 18
3.2 Big Data innen forsikring og kredittvurdering 19
3.3 Big Data på helseområdet 20
3.3.1 Helseforskning 20
3.3.2 Sensorer og selvlogging 21
3.4 Big Data innen politi, sikkerhet og etterretning 22
3.4.1 Smart politi 22
3.4.2 Ingen nål uten høystakk 23
3.5 Personvernutfordringer 24
3.5.1 Bruk av data til nye formål 24
3.5.2 Datamaksimalisering 25
3.5.3 Mangel på åpenhet – tap av kontroll 25
3.5.4 Ubalanse virksomhet – individ 26
3.5.5 Sammenstilling kan frembringe sensitiv informasjon 26
3.5.6 Farvel anonymitet 27
3.5.7 Feil faktagrunnlag 28
3.5.8 Datadeterminisme 28
3.5.9 Nedkjølingseffekt 29

3.5.10 Ekkokamre 29
4 Juridiske spørsmål 31
4.1 Big Data og loven 31
4.2 Personopplysningsbegrepet 32
4.2.1 Enhver form for informasjon 32
4.2.2 Tilknytningselementet 32
4.2.3 Identifiserbar enkeltperson 33
4.2.4 Oppsummering – personopplysningsbegrepet og Big Data 34
4.3 Rettslige krav til Big Data-behandling 34
4.3.1 Rettslig grunnlag for behandling av personopplysninger 35
4.3.2 Formålsbegrensningsprinsippet 39
4.3.3 Relevansprinsippet og dataminimalisering 40
4.3.4 Plikten til å sørge for korrekte data 40
4.4 Individets rettigheter 41
4.4.1 Gjennomsiktighet – informasjon og innsyn 41
4.4.2 Personprofiler og automatiserte avgjørelser 41
4.4.3 Retting og sletting 42
4.5 Noen internasjonale spørsmål 42
4.5.1 Lovvalg 42
4.5.2 Eksport av data til tredjeland 43
4.6 Nye regler om personvern 44
4.6.1 Samtykke og interesseavveining 44
4.6.2 Formålsbegrensningen 44
4.6.3 Innebygd personvern og standardinnstillinger 45
4.6.4 Utvidelse av databeskyttelsessonen 46
5 Oppsummering og anbefalinger 46
5.1 Samtykke fortsatt utgangspunktet 47
5.2 Rutiner for anonymisering og avidentifisering 48
5.3 Innsyn i profil og algoritme 49
5.4 ”Eiendomsrett” til egne personopplysninger 50
5.5 Innebygd personvern 51
5.5.1 Vurdering av personvernkonsekvenser (PIA) 51
5.5.2 Vurdering av personvernkonsekvenser i forbindelse med lovarbeid 52
5.6 Kunnskapsheving og bevissthet 52
Litteraturliste 54

Sammendrag

Big Data er et begrep som refererer til den enorme økningen i tilgang til, og automatisert bruk av, opplysninger. Det refererer til gigantiske mengder digitale data som er kontrollert av selskap, myndigheter og andre store organisasjoner, og som gjøres til gjenstand for omfattende analyse ved bruk av algoritmer.

Bruk av Big Data utfordrer sentrale personvernprinsipper. Enkelte hevder derfor at dagens personvernlovgivning må tilpasses en ny virkelighet. Datatilsynet deler ikke denne oppfatningen. I en tid der stadig flere opplysninger om oss samles inn er det viktigere enn noen gang å verne om grunnleggende personvernprinsipper. Prinsippene er vår garanti mot å bli gjort til gjenstand for profilering i stadig nye og flere sammenhenger.

Big Data kan brukes til mange gode og samfunnsnyttige formål. Analyseteknikkene benyttes til å analysere anonymiserte data for å identifisere og forutsi trender og sammenhenger. Bruk av anonymiserte data utfordrer i utgangspunktet ikke personvernet. Men Big Data kan også brukes slik at det berører enkeltindivider direkte. I rapporten trekker vi frem ti sentrale personvernutfordringer knyttet til Big Data:

1. *Bruk av data til nye formål:* Big Data handler i stor grad om gjenbruk av data. Dette utfordrer personvernprinsippet om formålsbegrensning. I henhold til dette prinsippet må virksomheter som benytter innsamlede personopplysninger som grunnlag for prediktiv analyse, forsikre seg om at prediksjonsanalysen ikke er uforenlig med det opprinnelige innsamlingsformålet. Dette kan innebære en betydelig utfordring for kommersiell Big Data-analyse.
2. *Datamaksimalisering:* Big data innebærer et nytt syn på data, der data får en verdi i seg selv. Verdien ligger i dataenes *fremtidige* bruksmuligheter. Et slikt syn på data påvirker virksomhetenes ønske om og motivasjon til å slette data. Verken private eller offentlige virksomheter vil ønske å slette data som på et senere tidspunkt, og sammenstilt med andre datasett, kan vise seg å bringe ny innsikt og penger.
3. *Mangel på åpenhet:* Mangel på åpenhet og informasjon om hvordan data benyttes og sammenstilles kan føre til at vi blir offer for beslutninger vi ikke forstår og ikke har kontroll over. Den alminnelige internettbruker har for eksempel liten innsikt i hvordan personopplysninger samles inn og utnyttes av kommersielle interesser. Flere av aktørene som opererer i dette markedet er i tillegg ukjente for folk flest.
4. *Sammenstilling kan frembringe sensitiv informasjon:* En utfordring ved Big Data-analyse er at innsamlede opplysninger som hver for seg ikke er sensitive, gjennom sammenstilling kan gi et sensitivt resultat. Det er viktig at virksomheter som benytter Big Data er kjent med denne problemstillingen, og ikke utvikler algoritmer som kan komme til å avsløre vitale personverninteresser.
5. *Risiko for reidentifisering:* En av de virkelig store utfordringene ved Big Data-analyse er risikoen for reidentifisering. Gjennom sammenstilling av data fra flere kilder kan det oppstå risiko for at enkeltindivider kan identifiseres fra i utgangspunktet anonyme datasett. Dette gjør anonymisering som metode for å hindre personvernulemper ved profilering og annen dataanalyse mindre virkningsfull.

6. *Ubalanse virksomhet – individ:* Big Data øker ubalansen mellom de store virksomhetene på den ene side og enkeltindividet på den andre. Det er virksomhetene som *samler inn* personopplysninger som henter ut den stadig voksende merverdien som ligger i analyse og bearbeiding av disse opplysningene, og ikke vi som *avgir* opplysningene. Snarere kan denne transaksjonen være til vår ulempe i den forstand at den kan utsette oss for fremtidig sårbarhet.

7. *Feil faktagrunnlag:* Det er et viktig personvernprinsipp at beslutninger som får konsekvenser for den enkelte skal være basert på korrekte opplysninger. En svakhet ved Big Data-analyse er at den ofte ikke tar hensyn til kontekst. Å basere beslutninger på opplysninger som er tiltenkt andre formål kan gi resultater som ikke samsvarer med den faktiske situasjonen.

8. *Datadeterminisme:* Utstrakt bruk av automatiserte avgjørelser og prediksjonsanalyse kan befeste eksisterende fordommer og forsterke sosial ekskludering og lagdeling. En utvikling der stadig flere beslutninger i samfunnet blir tatt basert på algoritmer, kan lede til et "dataenes diktatur"; vi blir ikke vurdert ut fra hva vi faktisk foretar oss, men på basis av hva alle dataene om oss sier at vi sannsynligvis kan komme til å gjøre.

9. *Nedkjølingseffekt:* Hvis alle sporene vi etterlater oss, på Internett og andre steder, blir brukt til stadig nye og for oss ukjente formål, kan dette legge bånd på hvordan vi deltar i samfunnet. Uklarhet og usikkerhet knyttet til myndighetenes bruk av Big Data kan true tillitten vi har til myndighetene. Utstrakt bruk av Big Data, kan i verste fall ha en nedkjølende effekt på ytringsfriheten hvis premissene rundt bruken er holdt skjult og ikke er etterprøvbar.

10. *"Ekkokamre":* Med økt personalisering av nettet vil den enkelte i stadig mindre grad bli eksponert for meninger som avviker fra deres egne. Dette vil kunne påvirke rammebetingelsene for offentlig debatt og meningsbryting. Dette er ikke i første rekke en personvernutfordring, men en samfunnsutfordring.

Selv om Big Data reiser flere personvernutfordringer, er det mulig å benytte denne analyseformen og samtidig respektere personvernet til den enkelte. I rapporten gir vi blant annet følgende anbefalinger:

- Behandling av personopplysninger skal som hovedregel baseres på *samtykke*. Hvis det ikke er mulig eller ønskelig å benytte samtykke, bør opplysningene *anonymiseres*.

- Gode rutiner for anonymisering og avidentifisering av opplysningene er av stor betydning. Dette vil bidra til å redusere faren for reidentifisering.

- Big Data bør brukes etter prinsippene for innebygd personvern.

- Virksomheter som benytter Big Data må være åpne om hvordan de behandler personopplysningene de samler inn. Dette innebærer blant annet å gi den enkelte innsyn i hvilke *beslutningskriterier* (algoritmer) som ligger til grunn for utvikling av profiler, og fra hvilke *kilder* opplysningene er hentet.

- Den enkelte bør gis mulighet til å få utlevert alle data virksomheten besitter om en i et brukervennlig format. Dataportabilitet vil hindre at kunder låses til tjenester som har uakseptable vilkår.

1 Innledning

Data er overalt. Mesteparten av disse dataene er generert av oss forbrukere i form av opplastede videoer på YouTube, twitter-meldinger, treningsapplikasjoner, e-poster, lokasjonsdata fra mobilen, Facebook-oppdateringer, musikkstreaming, nettsøk, kjøp av bøker på Amazon og så videre. Med utbredelse av Tingenes Internett[1] vil nye datastrømmer komme til. Talløse sensorer vil laste opp opplysninger i nettskyen om hvordan vi mennesker samhandler med tingene rundt oss. Det er estimert at innen 2020 vil det finnes mer enn 50 milliarder sensorer som kan kommunisere med hverandre (IBM 2013).

Stadig flere kommersielle virksomheter og myndigheter oppdager at disse enorme datastrømmene kan utnyttes strategisk. Dette kalles Big Data og er spådd å ha en omveltende effekt på samfunnet. Hensikten med Big Data er å utnytte de gigantiske datamengdene til å lete etter mønstre og sammenhenger det ikke var mulig å få øye på tidligere. Slik kunnskap er verdifull ikke bare innenfor markedsføring og salg, men også for myndighetene, med henblikk på blant annet sykdoms- og kriminalitetsbekjempelse.

Big Data kan brukes til mange gode formål, men fenomenet utfordrer også personvernet. Big data innebærer et nytt syn på data, der data får en verdi i seg selv. Verdien ligger i dataenes *fremtidige* bruksmuligheter. Et slikt syn på data utfordrer sentrale personvernprinsipper om formålsbegrensning og dataminimalisering. En annen sentral utfordring ved Big Data er at denne analyseformen medfører risiko for reidentifisering, noe som gjør anonymisering som metode for å hindre personvernulemper ved profilering og annen dataanalyse mindre virkningsfull.

Bruk av Big Data er foreløpig i startgropen. Hvordan vi håndterer utviklingen mot stadig mer omfattende utnyttelse av de enorme datastrømmene som genereres, er kritisk for personvernet. Sterke krefter – både kommersielle aktører og offentlige myndigheter – omfavner Big Data. Utnyttelsen av datastrømmene – omtalt som den nye oljen – vil være viktig for å fremme konkurransekraft og innovasjon i samfunnet. Men dette må gjøres på en måte som ikke truer personvernet til den enkelte.

1.1 Problemstilling

Rapportens overordnede problemstillinger er: Hvilke potensielle personvernutfordringer medfører bruk av Big Data og hvordan ser det juridiske handlingsrommet ut?

Big Data er et forholdsvis uklart begrep, med kontaktflate mot en rekke ulike problemstillinger, aktiviteter og aktører. En målsetting med rapporten er derfor å få bedre oversikt over aktørbildet og hvilke konkrete aktiviteter som utgjør kjernen i Big Data.

For å få en dypere forståelse av bruken av Big Data har vi sett på hvordan denne formen for analyseteknikk benyttes innenfor ulike sektorer. Vi har valgt ut fire sektorer: internettbaserte selskap, kredittvurderings- og forsikringsbransjen, helsesektoren og justissektoren. Sektorene er valgt

[1] Tingenes Internett refererer til at gjenstander blir utstyrt med enheter som kommuniserer trådløst med hverandre i nettverk.

for å vise bredden i bruk av Big Data innenfor både privat og offentlig sektor. Valget av sektorene er videre truffet på bakgrunn av at dette er områder hvor Big Data allerede er, eller er spådd å bli, utnyttet i høy grad.

Avgrensninger

Big Data berører som sagt mange og ulike problemstillinger av betydning for personvernet. Flere av problemstillingene kunne enkeltvis vært gjenstand for omfattende rapporter, slik som for eksempel utfordringer knyttet til sporingsteknologi, profilering og faren for reidentifisering av anonyme data.

Rammene for denne rapporten har ikke vært å behandle enkeltstående problemstillinger i detalj. Vår målsetting har vært å kartlegge på overordnet nivå hvilke utfordringer Big Data medfører for personvernet og hvilke begrensninger dagens personvernlovgivning legger på storskala dataanalyse.

Oppbygning av rapporten

I kapittel to omtaler vi hva vi kaller verdikjeden for Big Data. Vi viser hvilke prosesser, teknologi og aktører som er knyttet til Big Data. I kapittel tre ser vi på hvordan Big Data blir brukt innenfor ulike sektorer, og hvilke potensielle personvernutfordringer bruken medfører. I kapittel fire diskuterer vi det juridiske handlingsrommet for bruk av Big Data. Avslutningsvis komme vi med anbefalinger og tiltak vi mener er viktige for å sikre at bruk av Big Data respekterer den enkeltes personvern.

1.2 Definisjoner

Vi vil i det følgende definere sentrale begreper i rapporten, slik som Big Data, personopplysning, anonyme data og reidentifisering.

1.2.1 Big Data

Det finnes ikke én omforent definisjon av begrepet Big Data. Big Data benyttes om mye, og betydningen av begrepet er vag. Betegnelsen refererer både til dataene i seg selv og aktiviteten knyttet til å samle inn, lagre og analysere dem.

EU-kommisjonens rådgivende organ i personvernspørsmål, Artikkel 29-gruppen, definerer Big Data slik:[2]

Big Data refererer til den enorme økningen i tilgang til, og automatiserte bruk av, opplysninger: det refererer til gigantiske mengder digitale data som er kontrollert av selskap, myndigheter og andre store organisasjoner, og som gjøres til gjenstand for omfattende analyse ved bruk av algoritmer. Big Data kan bli brukt til å identifisere generelle trender og sammenhenger, men kan også bli benyttet slik at det berører enkeltindivider direkte.

Vi vil ta utgangspunkt i denne definisjonen, men legge til hva som etter vår oppfatning er det mest sentrale aspektet ved Big Data, nemlig at det handler om sammenstilling av data fra flere ulike kilder. Det er altså ikke kun volumet i seg selv som er interessant, men det at man henter ut sekundærverdi fra data ved å gjenbruke og analysere dem. Dette aspektet ved Big Data, og hvilke konsekvenser det har, er etter vår mening det mest sentrale og utfordrende sett fra et personvernperspektiv.

[2] Opinion 03/2013 on purpose limitation

1.2.2 Personopplysningsbegrepet

Personopplysningsbegrepet er sentralt i en analyse av personvernutfordringer knyttet til bruk av Big Data. For at en opplysning skal kunne defineres som en personopplysning, må opplysningen direkte eller indirekte kunne knyttes til et enkeltindivid.[3] Big Data dreier seg i stor utstrekning om analyse av anonyme eller anonymiserte data. I henhold til personopplysningsloven er slike data ikke å regne som personopplysninger. Big Data-analyse som omfatter bruk av anonyme eller anonymiserte opplysninger, vil derfor normalt sett falle utenfor personopplysningslovens virkeområde.

1.2.3 Anonyme data og reidentifisering

Anonyme data er av Artikkel 29-gruppen definert som data som det ikke er mulig, ved hjelp av alle rimelige tekniske hjelpemidler, å knytte tilbake til en enkeltperson. Anonymiserte data er definert som anonyme data som *tidligere* var knyttet til en enkeltperson, men der identifisering ikke lenger er mulig.[4]

En av de store utfordringene ved Big Data-analyse sett fra et personvernperspektiv, er faren for reidentifisering. Reidentifisering innebærer at data som i utgangspunktet fremstår som anonyme, ved hjelp av ulike teknikker, tilbakestilles til identifiserende opplysninger. Gjennom sammenstilling av flere datasett, slik man gjør i Big Data-analyse, kan det oppstå risiko for at enkeltindivider kan identifiseres fra i utgangspunktet anonymiserte data.

1.3 Sentrale personvernprinsipper

Det europeiske personverndirektivet fastsetter enkelte kjerneprinsipper for hvordan personopplysninger skal behandles på en rimelig, lovlig og legitim måte. [5] Følgende prinsipper har særlig relevans i forhold til Big Data:

Formålsbestemthet. Personopplysninger skal bare samles inn for bestemte formål, og disse må være uttrykkelig angitt og legitime. Dessuten må formålene for den videre behandlingen av opplysningene ikke være uforenlige med de formål opplysningene opprinnelig ble samlet inn for. Personopplysninger skal heller ikke utleveres til andre uten at det foreligger samtykke eller et annet rettslig grunnlag for utlevering.

Relevans og minimalitet. Personopplysninger skal bare innhentes, lagres og behandles i den grad det er nødvendig for å oppnå formålet. Innsamlede data som ikke lenger er nødvendige for det angitte formålet må slettes eller anonymiseres.

Fullstendighet og kvalitet. Personopplysninger må være relevante, korrekte og fullstendige ut fra formålene de skal benyttes til. Opplysninger lagret i et register brukes ofte som grunnlag til å fatte beslutninger om de registrerte. Dette prinsippet sikrer at beslutningene ikke blir fattet på et ufullstendig eller feilaktig grunnlag.

---

[3] Personopplysningsbegrepet er nærmere omtalt i avsnitt 0.
[4] Opinion 04/2007 on the concept of personal data
[5] DIRECTIVE 95/46/EC, Europaparlamentet og Rådets direktiv 95/46/EF av 24. oktober 1995 om beskyttelse av fysiske personer i forbindelse med behandling av personopplysninger og om fri utveksling av slike opplysninger.

Informasjon og innsyn. Som registrert har man rett til å bli informert om innsamling og bruk av sine personopplysninger. Man har også rett til å få innsyn i de opplysninger som er registrert om seg. Man har videre rett til å få en manuell vurdering av avgjørelser som fullt ut er basert på automatisert behandling av personopplysninger, dersom den avgjørelsen som tas er av vesentlig betydning for den som er registrert.

1.4 Big Data i startgropen i Europa og Norge

En rapport fra McKinsey Global Institute (2011) viser til hvordan bruk av Big Data vil ha en transformerende effekt på hele sektorer fra markedsføring til helsesektoren og i politisk kampanjevirksomhet. Det blir fremhevet at Big Data vil øke konkurransekraften og innovasjonsgraden i næringslivet. I Europa vil offentlige myndigheter kunne hente ut over 100 milliarder euro i effektiviseringsgevinster ved å benytte Big Data, hevder McKinsey. I dette tallet har man ikke tatt med bruk av Big Data til å bekjempe skatte- og trygdesvindel.

Tidligere har kun svært store virksomheter hatt mulighet til å drive med avansert dataanalyse ettersom det har forutsatt tilgang til enorm datalagringskapasitet og datakraft. Med mulighet til å lagre data i nettskyen, samt tilgang til rimelig analyseprogramvare, er Big Data nå i praksis mulig å ta i bruk for både små og store selskap.

Foreløpig er bruk av Big Data ikke spesielt utbredt i Europa sammenlignet med i USA (NESSI White Paper 2012). Kunnskapen om Big Data er til en viss grad til stede blant større europeiske virksomheter, men er fortsatt lav blant små og mellomstore bedrifter. Kun et fåtall virksomheter spesialisert på å levere Big Data-relaterte tjenester og produkter er etablert i Europa (NESSI White Paper 2012). De fleste slike selskap er etablert i USA.

Av hindre som trekkes frem for å forklare hvorfor Europa er en Big Data-sinke, er for eksempel mangel på kompetanse (det utdannes ikke tilstrekkelig mange med såkalt data science-kompetanse), et fragmentert marked og fravær av tilstrekkelig mange store markedsaktører som kan drive utviklingen fremover. I tillegg trekkes dagens europeiske personvernlovgivning frem som et vesentlig hinder for at Europa ikke har kommet like langt som USA i satsningen på Big data (NESSI White Paper 2012).

I EUs Digital Agenda-program fremheves viktigheten av å legge til rette for bruk av Big Data (Whelan 2013). [6] Satsning på Big Data knyttes sammen med EUs strategi for åpne data og fokus på nettskytjenester. Modernisering av personvernregelverket blir sett på som avgjørende for at EU skal ha mulighet til å hente ut gevinster ved bruk av Big Data i fremtiden.

I forbindelse med rapporten har Datatilsynet vært i kontakt med ulike aktører i det norske markedet. Vårt inntrykk på bakgrunn av disse samtalene er at bruk av Big Data heller ikke har kommet langt i Norge. Et fåtall aktører har tatt teknologien i bruk, blant annet virksomheter innen telekommunikasjon, medier og detaljhandel. Det har også vokst frem enkelte analyseselskap som er

[6] The Digital Agenda for Europe (DAE) har til målsetning å hjelpe europeiske borgere og bedrifter å hente ut størst mulig gevinst av digital teknologi.

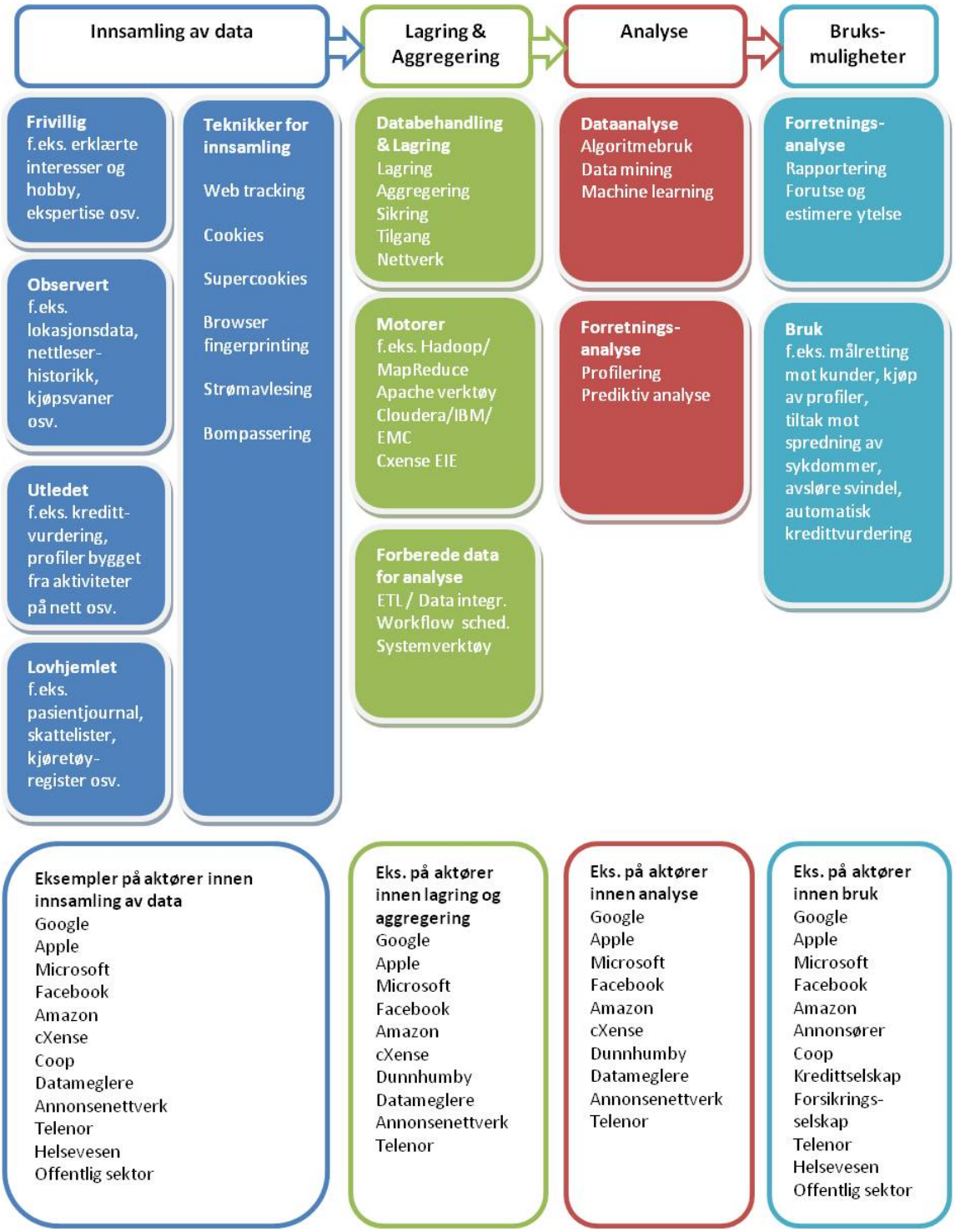
spesialisert på Big Data, og som tilbyr sin ekspertise og teknologi til norske og utenlandske selskap.[7] Sterke forskningsmiljøer på Big Data finnes ved Universitetene i Oslo, Trondheim og Tromsø.

[7] For eksempel cXense som leverer blant annet ekspertise og teknologi til store norske og utenlandske medieselskap: http://www.cxense.com/

2 Verdikjeden for Big Data: aktører, prosesser og teknologi

I det følgende vil vi gjennomgå verdikjeden for Big Data, det vil si prosessen fra innsamling av data til lagring, aggregering og analyse. Formålet med fremstillingen er å gi et overblikk over de ulike trinnene i analyseprosessen og hvordan ulike aktører er involvert på ulike punkter i verdikjeden.

Figur 1, verdikjeden:



2.1 Innsamling av data

Det første trinnet i verdikjeden for Big Data er innsamling av opplysninger som skal danne grunnlag for videre analyse. Et av karaktertrekkene ved Big Data er at det benyttes et mangfold av ulike datakilder, og både strukturerte og ustrukturerte data[8]. Det kan være datakilder som inneholder personopplysninger, eller kilder som ikke inneholder slike opplysninger, slik som for eksempel værdata og opplysninger generert fra sensorer i ulike typer produksjonsutstyr i en fabrikkhall.

Det er datakilder som inneholder personopplysninger som er interessante fra vårt perspektiv, og slike kilder er det mange av; vi legger igjen elektroniske spor i forbindelse med gjennomføringen av de fleste av våre aktiviteter fra vi står opp til vi legger oss. Dagen starter ved å sjekke Facebook. Bomringen registrerer når du kjører til og fra jobb. Kundekortet i butikken og kredittkortet registrerer dine innkjøp. Adgangskortet på jobben noterer når du starter og slutter dagen. Mobiltelefonen i lomma og bruk av lokasjonsbaserte applikasjoner registrerer ditt bevegelsesmønster gjennom hele dagen. Alt dette kan være relevante og attraktive kilder i Big Data-sammenheng.

Utviklingen mot Tingenes Internett vil bidra til å generere nye strømmer av data. Tingenes Internett betyr at stadig flere gjenstander og personer blir utstyrt med enheter som kommuniserer trådløst med hverandre i nettverk. Enhetene kan være sensorer som samler data eller RFID-brikker som brukes for å identifisere gjenstander, dyr eller personer. Slike enheter kan brukes til å overvåke og styre gjenstander, personer eller prosesser, og vi kan fjernstyre disse gjennom apper eller nettsider. For eksempel kan du sjekke om døren er låst hjemme og låse den dersom den ikke er det. Dine treningssko kan kommunisere med smarttelefonen slik at du får alle detaljer om din treningsrunde, og alarmklokken kan snakke med kaffetrakteren og lysbryteren, og tilpasse seg dine behov. Men det er ennå ikke slik at alle ting styres over Internett. Mange enheter kan kun styres innenfor et lokalt nettverk. Teknologien finnes i dag, og med IPv6[9] ligger det til rette for at hver enkelt ting kan få sin unike identifikator (URI) og dermed bli tilgjengelig over Internett. Tingenes Internett er en disruptiv[10] teknologi og når dette tar av, vil Big Data bli virkelig stort.

Personvernutfordringer ved Tingenes Internett oppstår når det blir samlet inn og aggregert deler av data som relaterer seg til ulike gjenstander eller tjenester. En samling av mange fragmenter av data kan plutselig bli personopplysninger når hendelser blir vurdert i kontekst av sted, tid og gjentakelser. Bruk av sensorteknologi i dagligvarebutikker kan for eksempel gi indikasjoner om religion, eller gi grunnlag for antakelser om en persons livsstil og helse ved at det avslører rutinemessig innkjøp av

---

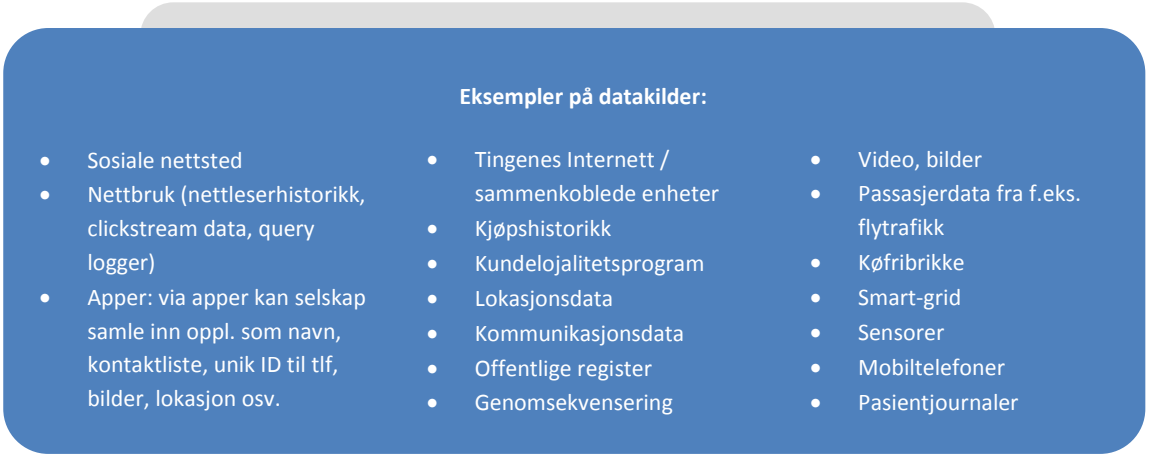
[8] *Strukturerte data* kan beskrives som data som følger en formell struktur av datamodeller, slik som rader og kolonner i en relasjonsdatabase, for eksempel et kunderegister. *Ustrukturerte data* refererer til data som mangler en gjenkjennelig struktur. For eksempel regnes bilder, video, e-post, word-dokumenter og ren tekst for å være ustrukturerte data innen et datasett eller fil. *Semistrukturerte data* er en form for strukturerte data som ikke følger den formelle strukturen av datamodeller. Deler har fast format og deler består av fritekst. Det er ofte kjent som skjemaløs eller selvdokumenterende struktur. Semistrukturerte data er gjerne XML-filer, bøker, websider, e-post og EDI – ofte data innsamlet fra sensorer og maskiner.

[9] Internet Protocol versjon 6 (IPv6) er et sett med protokoller som datamaskiner bruker til å utveksle informasjon over Internett og over hjemme- og firmanettverk. Med IPv6 kan det tilordnes mange flere IP-adresser enn med IPv4.

[10] En *disruptiv innovasjon* er en nyskapning som forstyrrer et eksisterende marked. Begrepet brukes i forretnings- og teknisk litteratur for å beskrive innovasjoner som forbedrer et produkt eller en tjeneste på en måte som markedet ikke forventer, vanligvis ved å senke prisen eller designe for et annet sett med forbrukere.

bestemte matvarer. Sensordata hentet ut fra kjøleskapet kan for eksempel si noe om eierens hverdagsrutiner; når man er hjemme, når på døgnet man spiser, hvor ofte osv. Med ny teknologi og nye ting som kobles til Internett vil det også dukke opp nye sårbarheter som gjør at applikasjoner og systemer kan angripes. TV-en kan bli hacket, og appen som styrer sikkerhetsalarmen hjemme kan bli hacket eller lekke opplysninger til en tredjepart (ars technica 2012).

Figur 2, datakilder:



Personopplysninger kan hentes inn på ulike måter:

- Det kan skje frivillig ved at personer eksplisitt stiller personopplysninger om seg selv til rådighet. Dette kan for eksempel skje ved at man oppretter en profil på et sosialt nettsted, oppgir opplysninger for å bli medlem av et lojalitetsprogram til en butikkjede, laster ned en applikasjon på mobiltelefonen eller registrerer opplysninger om seg selv for å få tilgang til en tjeneste.
- Personopplysninger kan bli automatisk registrert av virksomheter i forbindelse med at man benytter en bestemt tjeneste – dette i motsetning til opplysninger man oppgir på forespørsel. Eksempler på slike data er lokasjonsdata, nettleserhistorikk, handlevaner, besøkshistorikk på treningssenteret og passeringsdata i bomringen.
- Personopplysninger kan bli *utledet* gjennom bearbeiding og analyse av data innsamlet for tidligere og andre formål. Personopplysninger kan også bli utledet fra ulike sett med tilsynelatende anonyme opplysninger.
- Innsamling av personopplysninger til de statlige og kommunale myndighetene kan være nedfelt i egen lov og forskrift. Dette gjelder for eksempel personopplysninger i pasientjournal, i skattelistene og i kjøretøyregisteret.

Svært mange av opplysningene som benyttes i Big Data-sammenheng er generert på nett. Disse opplysningene kan samles inn eksplisitt (når man registrere en sosial profil på nett), eller mer i det skjulte, slik det gjøres ved bruk av sporingsteknologi. Bruk av cookies krever nå samtykke slik at innsamlingen skal bli mer synlig for brukerne og gi dem mer kontroll. Vi beskriver her ulike teknikker som benyttes for å samle inn personopplysninger på nett:

- *Web tracking* (sporing på nett) kan defineres som innsamling, analyse og bruk av data bestående av brukeraktivitet fra en datamaskin eller annen enhet mens man bruker ulike nettjenester. Hensikten for de som samler inn data er å kombinere og analysere opplysningene til ulike formål. Det som skaper størst utfordringer ved sporing på nett er når det gjøres av en tredjepart. Et eksempel på dette er når man som registrert bruker på et nettsted klikker på en bannerannonse på dette nettstedet, og e-postadressen dermed blir sendt videre til et dusin andre selskaper (Mayer 2011).
- *Cookies* (informasjonskapsler) er små tekstfiler som plasseres på en brukers datamaskin når man laster ned en nettside. En cookie kan inneholde informasjon som brukeren har registrert på siden, for eksempel brukernavn og passord i kryptert form, og som oversendes nettsiden ved påfølgende besøk. På den måten er det ikke nødvendig å registrere informasjonen mer enn én gang.
- *Supercookies* er en slags cookie som blir permanent lagret på en brukers datamaskin. Supercookies er generelt vanskeligere for brukerne å finne og fjerne fra deres enheter siden disse ikke kan slettes på samme måte som vanlige cookies.
- *Browser fingerprinting* kan brukes som sporingsteknikk mot folk som begrenser bruken av cookies. Metoden går ut på å samle inn data om hva slags nettleser man har, installerte plugins, systemfonter, type nettleser, operativsystem, tidssone, skjermoppløsning og fargedybde, og om cookies er blokkert. En måte å bruke browser fingerprinting på er å kombinere metoden med en IP-adresse for å finne ut hvilke enheter som skjuler seg bak den enkelte IP-adressen. Browser fingerprinting er en kraftig teknikk og et slikt fingeravtrykk kan likestilles med cookies, IP-adresser og supercookies når det snakkes om sporing på nett (Eckersley 2010).

2.2 Lagring og aggregering

Etter at data er samlet inn, kan de bli lagret og aggregert. Individuelle dataelementer organiseres og lagres i datasett som kan bli brukt til videre prosessering og analyse. Noen aggregerer og anonymiserer data før de lagres, mens andre lagrer data med personopplysninger. I denne sammenhengen betyr aggregering at data blir slått sammen til større mengder, slik at data ikke kan relateres til noen, eller identifisere noen.[11] Utfordringen med dette er at det er mulig at det aggregerte datasettet ved en senere anledning kan kobles sammen med andre datasett og på den måten reidentifisere enkeltpersoner.

For at Big Data-aktørene skal være konkurransedyktige, må teknologien de bruker kunne håndtere et stort volum av data, kunne behandle ulike typer data (strukturerte, ustrukturerte og semistrukturerte) fra ulike kilder, og få dataene indeksert og søkbare innen svært kort tid. En sentral forutsetning for Big Data er fremveksten av skytjenester som kan tilby nær ubegrenset lagringskapasitet til en stadig rimeligere pris. Virksomheter kan samle inn og behandle langt større datamengder enn tidligere. Begrenset lagringskapasitet representerer ikke lenger et hinder.

[11] Eksempel på aggregering er avrunding av tall. Dersom en oppføring i en tabell har en verdi på 10 000, som er det antallet mennesker som utfører en bestemt aktivitet en måned, endres til 10 001 den neste måneden, vil en uvedkommende kunne sammenlikne tabellene og finne den ene forekomsten som er forskjellig. Avrunding av dette tallet vil kunne forebygge det.

Som følge av kravene om å håndtere mengde, hastighet og variasjon av data [12], er det utviklet mange nye verktøy, infrastruktur og rammeverk. Big Data-teknologi bryter med den tradisjonelle tankegangen om lagring og prosessering av data ved bruk av stormaskiner. Ny teknologi gjør det mulig å behandle og få ut verdier, av nye og ustrukturerte datakilder. Med dette har det også dukket opp utfordringer innen informasjonssikkerhet – utfordringer som kan få konsekvenser for personvernet. Sikkerhetsutfordringene dukker eksempelvis opp ved bruk av flere infrastrukturlag for å prosessere Big Data, ny type infrastruktur for å takle den enorme gjennomstrømmingen av data og ved bruk av ikke-skalerbar kryptering av store datasett.

Enkelte mener imidlertid at lagring av data i forbindelse med Big Data tilhører fortiden, og at det er analyse av sanntidsdata som er fremtiden. Et selskap som går imot tanken om datalagring i storskala er Numenta og dets gründer Jeff Hawkins. Hawkins mener at den eneste grunnen til at man skal se på historiske data er hvis man mener at verden ikke kommer til å forandre seg (*The New York Times*, 2012a). Numenta har et produkt som kalles Grok, og som automatisk analyserer datastrømmer i sanntid. Det er en skybasert tjeneste som henter jevne datastrømmer fra for eksempel termostater, klikk på Internett eller maskiner. Til å begynne med observerer Grok flyten av data, og etter hvert begynner den å gjøre gjetninger på hva som kommer til å skje. Jo mer data, jo mer nøyaktig blir spådommene (mer om "sensemaking" i neste kapittel) (*The New York Times* 2012a).

2.3 Analyse

I det tredje trinnet i verdikjeden tar man de innsamlede og lagrede dataene og *kombinerer dem med annen informasjon*. En sentral del av verdiskapingen i dette trinnet er sammenslåing av data fra ulike kilder for å lage profiler, og å bruke analyseverktøy til å utlede informasjon som ikke ellers er tilgjengelig. Virksomheten kan enten velge å sammenstille kun egne virksomhetsinterne data, eller kjøpe data fra andre aktører (evt. innhente data fra åpne kilder) og så sammenstille disse med egne data.

I det følgende presenterer vi en oversikt over noen av analyseteknikkene som blir brukt i Big Data:

- *Data mining* (datautvinning) går ut på å lete etter struktur og mening i store mengder med ustrukturerte data.
- *Machine Learning* (maskinlæring) er en type kunstig intelligens, innen fagområdet informatikk. Det er en vitenskapelig disiplin som er opptatt av design og utvikling av algoritmer, og som gjør datamaskiner i stand til å utvikle atferd basert på empiriske data. Algoritmer brukes for å gjenkjenne og forutse korrelasjoner og mønstre i dataene. En algoritme (i matematikk og informatikk) er en presis beskrivelse av en endelig serie operasjoner som skal utføres for å løse et problem eller et sett med flere problem. Algoritmer brukes altså for å fortelle et program hva som skal utføres og hvordan det skal utføres. Machine learning-prosessen er ganske lik data mining-prosessen. Begge systemene søker gjennom data for å lete etter mønstre. Der data mining pakker ut data for å få menneskelig forståelse, bruker machine learning data for å forbedre programmets egen

[12] En hyppig brukt definisjon av Big Data knytter begrepet opp mot de tre V-ene: volum, variasjon og velositet (hastighet): http://www.forbes.com/sites/oreillymedia/2012/01/19/volume-velocity-variety-what-you-need-to-know-about-big-data/

forståelse. Machine learning gjenkjenner mønstre i data og justerer programmets handlinger etter det. Facebooks nyhetsstrøm er et eksempel på dette. Nyhetsstrømmen endres etter brukerens samhandlinger med andre brukere. Dersom en bruker ofte "tagger" en venn på bilder, skriver på hans vegg eller "liker" hans linker, vil nyhetsstrømmen vise mer av denne vennens aktivitet i brukerens nyhetsstrøm. Dette er basert på antatt nærhet både i slektskap og tiltrekning, samt tid og mengde. Algoritmen som brukes kalles EdgeRank[13].

- *Social network analysis* er en analyse av sosiale nettverk. Social network analysis ser på sosiale relasjoner med utgangspunkt i nettverksteori. Nettverkene består av noder (som representerer individuelle aktører i nettverket) og bånd (som representerer relasjoner mellom individer, slik som vennskap, slektskap, stilling i organisasjoner). Disse nettverkene er ofte avbildet i et "sosialt nettverk-diagram", der nodene er representert som punkter og bånd er representert som linjer.
- *Prediksjonsanalyse* dreier seg om å anslå fremtidige sannsynligheter og trender. Det sentrale elementet i prediksjonsanalyse er "the predictor" – en variabel som kan måles på en enkeltperson eller annen enhet for å forutsi fremtidig atferd. For eksempel er det sannsynlig at et forsikringsselskap vil ta hensyn til potensielle "predictors" innen sikker kjøring slik som alder, kjønn, og kjørehistorikk, når de skal utstede bilforsikring. En "predictive model" består av flere sammenkoblede "predictors", og brukes i analyse for å forutsi fremtidige sannsynligheter med en akseptabel grad av pålitelighet.
- *Sensemaking* går ut på å integrere nye transaksjoner (observasjoner) med tidligere transaksjoner, på samme måte som man tar en brikke i et puslespill og finner de etterfølgende brikkene på bordet. I motsetning til andre analytiske metoder, som krever at brukerne stiller spørsmål til systemet, kan de nye systemene operere med et annet prinsipp: data finner data, og relevans finner brukeren.

Andre analyseteknikker er A/B testing, Classification, Cluster analysis (clustering), Natural language processing (NLP), (Social) Neural networks, Optimization, Spatial analysis, Simulation, Time series analysis og Visualization.

2.4 Aktører

En lang rekke ulike aktører er involvert i verdikjeden for Big Data. Vi kan grovt skille mellom dataeiere, datameklere, Big Data-selskap og Big Data analyse- og konsulentselskap. Enkelte virksomheter er involvert i samtlige etapper fra innsamling av data til bruk av det ferdige analyseresultatet (se figur 1, verdikjeden). Andre aktører opptrer kun på bestemte punkter i verdikjeden.

Dataeiere

Dette er den aktuelle virksomheten som tar Big Data i bruk. Det kan være små og store virksomheter, offentlige så vel som private. Det er ofte data som virksomheten selv besitter som er gjenstand for analyse, selv om også andre datakilder i økende grad ønskes benyttet for å vinne ny innsikt. Dataene har virksomheten samlet inn som et ledd i utøvelsen av virksomhetens ulike aktiviteter. Det kan for

[13] En enkel forklaring på hvordan EdgeRank virker: http://blog.hubspot.com/understanding-facebook-edgerank-algorithm-infographic

eksempel være en butikkjede som samler inn data via kundekort, lagrer og aggregerer disse, samt analyserer dem for å forbedre sin egen forretningsmodell.

Datameklere

Datameklere, også kalt databrokers, samler inn data med henblikk på analyse og videresalg av opplysningene. Dette er et marked som vokser seg større med bruk av Big Data, ettersom stadig flere virksomheter, innen stadig flere sektorer, er interessert i å kjøpe data som de kan sammenstille med egne virksomhetsinterne data. Acxiom er en av de virkelig store datameklerne. Det er et amerikansk selskap som samler, analyserer og tolker kunde- og forretningsinformasjon for klienter, og hjelper dem med målrettede annonsekampanjer og lignende. Klientmassen i USA består for det meste av virksomheter innen finans, forsikring, direkte markedsføring, media, varehandel, teknologi, helse og telekommunikasjon, samt myndighetene. Selskapet er en av verdens største behandlere av forbrukeropplysninger. De sies å ha opplysninger om 96 prosent av USAs husholdninger (*The New York Times* 2012b).

Et annet selskap som få har hørt om, men som har vokst seg stort i det skjulte, er Flurry. Det er en gratis analysetjeneste som leverer brukerstatistikk til app-eieren mot at Flurry får tilgang til opplysninger om brukerne. Flurry besitter på denne måten enorme mengder informasjon om mobilbrukere, samlet inn via ulike applikasjoner. Dette er opplysninger som Flurry aggregerer og selger videre som profiler til markedsførere og andre som er ute etter å treffe bestemte målgrupper.

I Norge finnes det ikke rene datameklingsselskap. Heller ikke på europeisk plan finnes det mange slike selskap.

Big Data-selskap

Dette er selskap som bygger nye forretningsmodeller og lager nye tjenester basert på tilgjengelige data generert på internett eller åpne offentlige data. Eksempel på et slikt selskap er kredittvurderingsselskapet Kreditech, som vil bli omtalt i punkt 3.2.

(Big) Dataspesialister

Noen selskap hjelper andre selskap å hente ut verdi av dataene de besitter. Store virksomheter som Google og Facebook, har ressurser til å analysere dataene de samler inn i eget hus. For mindre selskaper blir dette vanskelig og man kjøper kompetanse fra spesialiserte (Big Data) analyseselskaper. Dette er selskaper med spisskompetanse og spesialutviklet programvare til å drive med Big Data-analyser. Av programvareselskaper med analyse- og visualiseringsverktøy kan vi nevne SAS (Visual Analytics), QlikView, Tableau, Tibco Spotfire og Panopticon. Datasift er eksempel på en aktør som selger sanntidsanalyse fra sosiale nettverk. Palantir leverer analyseprogramvare til blant annet politi- og etterretningsmyndigheter i USA. En norsk aktør innen analyse og Big Data er cXense.

3 Personvernutfordringer knyttet til bruk av Big Data

Det overordnede formålet med bruk av Big Data er grovt sett det samme innen samtlige sektorer: å vinne ny innsikt om sammenhenger som blant annet kan benyttes til å predikere handling eller hendelser. Hvilke data som samles inn, hvordan de behandles og hvorvidt analyseresultatet er ment benyttet overfor enkeltindivider, eller på et mer overordnet nivå, varierer. I det følgende vil vi se på ulike bruksområder for Big Data og hvilke mulige personvernutfordringer bruk av Big Data reiser.

3.1 Big Data i bruk blant internettbaserte selskaper

Internettbaserte selskap er pionerer innen utnyttelse av Big Data-teknologi (Bollier 2010). Alle de store internettselskapene – Google, Facebook, Amazon, eBay, Microsoft, Apple og Yahoo! – benytter Big Data i en eller annen form til å hente ut sekundærverdi av de gigantiske datamengdene de besitter. Google er et godt eksempel på dette. Ikke bare benytter de dataene de samler inn til å drive målrettet markedsføring. Dataene benyttes også til å forbedre søkealgoritmene og til å utvikle nye dataintensive tjenester. Et illustrerende eksempel er Googles nylig lanserte tjeneste *Google Now*, omtalt som en typisk Big Data-applikasjon. Hensikten med mobilapplikasjonen er å gi folk hjelp *før* de selv innser at de behøver det. For eksempel kan dette gjøres ved å gi beskjed om at bussen er forsinket før de går fra jobb for dagen. Google Now's algoritme benytter data fra brukernes e-post, kalender og søkehistorikk for å lære folks vaner å kjenne (*MIT Technology Review* 2013a).

Facebook, verdens største sosiale nettsamfunn, besitter enorme mengder personopplysninger. Ikke bare om sine over 900 millioner medlemmer, men også om ikke-medlemmer hvis disse besøker nettsider og apper hvor Facebooks "liker-knapp" er installert. Via denne funksjonen kan Facebook spore nettaktivitet utenfor nettstedets egne sider (Datatilsynet 2011a). Facebook har et eget *Data Science Team*. Deres jobb er å analysere disse enorme datamengdene for å avdekke mønstre og trender i folks samhandling og aktivitet. Dette er verdifull kunnskap i utviklingen av nye tjenester og produkter, og noe som Facebook kan tjene penger på. Det er også kunnskap som er svært verdifull å selge videre til en lang rekke andre aktører som ønsker å nå spesielle målgrupper. Facebook hevder at de ikke selger informasjon om sine medlemmer videre til tredjeparter.

Internetts grunnleggende forretningsmodell har blitt gratistjenester som tjener penger på personopplysninger generert fra nettbrukerne. Personopplysninger samles inn enten ved at brukerne selv oppgir dem, eller ved å spore brukernes nettaktivitet ved hjelp av ulike sporingsverktøy. Sporingsverktøyene benytter seg av unike identifikatorer som kan sammenstille brukeratferd om den enkelte på tvers av mange ulike nettjenester over tid. Opplysningene som samles inn blir bruk til å bygge opp brukerprofiler. Dette gjør det mulig å skreddersy reklame, tilbud og tjenester til bestemte kunder. Denne aktiviteten omtales som atferdsrettet reklame. Artikkel 29-gruppen har skrevet en uttalelse om personvernutfordringer knyttet til denne formen for markedsføring på nett.[14]

Bruk av Big Data innen atferdsrettet markedsføring representerer ikke noe grunnleggende nytt. Teknologien gjør det imidlertid mulig å prosessere og sammenstille et *enda større volum* og å

[14] Opinion 02/2010 on online social networking

innhente data fra et *bredere spekter av datakilder* enn tidligere. Bruk av Big Data har derfor blitt kalt *"data mining on steroids"* (Rubinstein 2012).

Med utbredelsen av Tingenes Internet vil markedet for omsetning av personopplysninger øke i omfang. Vi kan få en utvikling der Internetts grunnleggende forretningsmodell blir overført til andre markeder. De smarte joggeskoene med sensorer kan tilbys gratis mot at brukeren samtykker til at data om joggeturene samles inn og analyseres til ulike formål. Den smarte tannbørsten gis bort gratis mot at brukeren deler opplysningene tannbørsten samler inn med forskningsinstitusjoner, forsikringsselskap, matvarekjeder etc. Nye virksomheter og forretningsmodeller vil vokse frem for å hente ut merverdien av de gigantiske mengdene med personopplysninger som genereres i stadig flere sammenhenger. PRISM-saken har vist at også andre enn kommersielle aktører er interessert i å utnytte disse dataene. Denne saken omtales nærmere i kapittel 3.4.2.

3.2 Big Data innen forsikring og kredittvurdering

I forsikrings- og kredittvurderingsbransjen er bruk av korrelasjonsanalyse og profilering heller ikke nytt. Korrelasjonsanalyse blir benyttet til å vurdere risikoprofil og kredittverdighet. Norske forsikrings- og kredittvurderingsselskap kan imidlertid ikke hente inn og benytte personopplysninger etter eget forgodtbefinnende. Selskapene har konsesjon fra Datatilsynet til å behandle personopplysninger på bestemte vilkår. I konsesjonen er det nedfelt hvilke datakilder som kan benyttes. Når et forsikringsselskap skal gjøre en risikoanalyse av en kunde, er det for eksempel forbud mot å bruke en del typer data, slik som kjønn, etnisitet, religion eller betalingsanmerkninger, i prissettingen.

Bruk av ny og kraftfull datamining-teknologi er spådd å bli stort innen forsikrings- og kredittvurderingsbransjen internasjonalt. Dette er en trend som trolig også vil påvirke norske aktører. Selv små forbedringer av treffsikkerheten i analysene kan gi stor gevinst (Dagens IT 2013). Big data legger til rette for at et langt bredere spekter av datakilder kan inngå i utarbeidingen av kredittscore og risikoprofiler. Nye kredittvurderingsselskap som spesialiserer seg på bruk av Big Data har etter hvert dukket opp, slik som for eksempel tyske Kreditech[15], som hevder å være ledende på dette i Europa.

Selskapet sier følgende om hvilke datakilder de benytter seg av i sine kredittvurderingsanalyser:

> *"Kreditech works liks a Big mosaic - Any online data that can be found about an individual will be used for fraud detection, identification, scoring: Location data (GPS, micro-geographical), social graph (likes, friends, locations, posts), behavioral analytics (movement and duration on the webpage), people's e-commerce shopping behavior and device data (apps installed, operating systems) are just some examples of up to 8,000 data points that are processed in real-time for any single scoring unit."*

[15] http://www.kreditech.com/#where-we-are

Figur 3:

(kilde: www.kreditech.com)

Ikke bare vil flere datakilder bli benyttet, nye datakilder vil også kunne erstatte dem som benyttes i dag. Pasientjournaler har for eksempel vist seg å være kostbare å analysere. Eksperter tror derfor at andre datakilder vil bli benyttet av forsikringsbransjen for å avgjøre folks helsetilstand og risikoprofil, fordi dette vil være billigere og mer effektivt (*The Economist* 2012). Opplysninger samlet inn fra sosiale medier kan gi informasjon om folks aktivitetsnivå og dermed antatte helsetilstand: Er vedkommende sosial og har mange interesser, eller sitter vedkommende mye inne? Sammenstilling av store datasett vil også kunne avdekke sammenhenger om folks risikoprofil og helsetilstand, noe som kan være av interesse for selskapene. Tidsskriftet *The Economist* (2012) viser til et eksempel der bruk av Big Data-analyse har identifisert at mennesker som foretar hyppige uttak fra minibanker lever lengre enn de som benytter kredittkort og sjekker. Det er etter hvert også mulig for forsikrings- og kredittopplysningsselskapene å kjøpe opplysninger om folks forbruksmønster fra datameklere og andre selskap som besitter store databaser med slike opplysninger.

Også i Norge finnes det kredittvurderingsselskaper som sier de vil satse på Big Data (*Dagens IT* 2013). Dagens konsesjon setter imidlertid begrensninger for hvordan norske selskap kan ta teknologien i bruk. Bruk av opplysninger hentet fra sosiale medier, og innhenting av forbrukerdata fra eksterne selskap, slik som skissert over, vil for eksempel ikke være tillatt.

3.3 Big Data på helseområdet

McKinsey Globale Institute (2011) hevder at helsevesenet vil kunne hente store effektiviseringsgevinster ved å benytte Big Data, for eksempel ved å benytte teknologien til å få ned antallet feilbehandlinger ved sykehusene. Big Data er videre spådd å bli viktig både i forbindelse med individuell pasientbehandling og i forebyggende helsearbeid på populasjonsnivå.

3.3.1 Helseforskning

Foreløpig er det først og fremst innen forskning og i forebyggende helsearbeid at vi finner eksempler på Big Data. Big Data har blant annet vist seg å kunne predikere utbrudd og spredning av epidemier

med stort presisjonsnivå. Et forskningsprosjekt i regi av *Harvard school of public health* har undersøkt spredningsmønstret til malaria ved å innhente lokasjonsdata fra mobiltelefonene til 15 millioner kenyanere og sammenstille disse dataene med kenyanske myndigheters database over opplysninger om malariautbrudd (HSPH News 2012). Sammenstillingen av datasettene gjorde det mulig å predikere hvordan malaria sprer seg mellom ulike deler av landet. Tradisjonell statistikkinnsamling forteller først i etterkant at et sykdomsutbrudd har funnet sted. Da kan det være for sent å agere. En datamaskin som i stedet ser etter mønstre i for eksempel kommunikasjon på sosiale medier, eller ved å analysere lokasjons- og kommunikasjonslogger, kan gi tidlige indikasjoner på at en negativ utvikling har startet og hvor, slik at tiltak kan settes inn tidsnok.

Ved å benytte Big Data-analyse har forskere videre lykkes i å oppdage farlige bieffekter ved ulike medikamenter. Forskere ved Stanford oppdaget at to ulike medikamenter (en antidepressiva og en hodepinetablett) kunne få fatale konsekvenser for brukeren hvis de ble tatt i kombinasjon (Tatonetti et al. 2011). Dette fant de ut ved å sammenstille aggregerte data fra helsejournaler med nasjonale registre over innrapporterte bieffekter. Disse opplysningene ble så sammenstilt med 82 millioner søk foretatt på Microsofts søketjeneste Bing. Resultatet fra analysen viste at personer som tok *begge* preparatene i større grad enn de som kun tok *ett* av preparatene, foretok søk på ord relatert til bieffekter som ”hodepine” og ”trøtthet”. Forskerne identifiserte altså et mønster som indikerte at inntak av de to preparatene samtidig, kunne utløse alvorlige bivirkninger.

Bruken av Big Data i pasientbehandling er foreløpig lite utbredt (Bollier 2010 og HealtWorks Collective 2013). Dette skyldes flere forhold. For det første er kunnskapen om Big Data lav innen helsesektoren. For det andre er muligheten til å behandle og sammenstille data fra pasientjournaler strengt regulert i de fleste land. Og for det tredje er kobling og sammenstilling av data i helsetjenesten utfordrende fordi det ikke finnes en felles teknisk infrastruktur som legger til rette for slik analyse.

3.3.2 Sensorer og selvlogging

En utvikling som trolig vil stimulere bruken av Big Data på helseområdet, er den økende bruken av mobilt selvmonitoreringsutstyr. Folk tar mer og mer styring over egen helse. De søker etter helserelatert informasjon på nett, utveksler informasjon om sykdomssymptomer i nettsamfunn som *patienslikeme.com* og benytter mobilapplikasjoner til å måle egen helsetilstand (Bollier 2010). Smarttelefonen kan fungere som både stetoskop og blodtrykksmåler. En undersøkelse gjennomført av Datatilsynet og Teknologirådet i 2013, viste at 33 prosent av respondentene hadde brukt minst én helse- eller treningsapp. De mest ekstreme brukerne av slikt utstyr kalles ”life loggers” (selvloggere), og er en bevegelse som oppfordrer til logging av egen helse og alle andre tenkelige gjøremål. Formålet med innsamlingen er å dele, sammenstille og analysere dataene slik at de kan frembringe ny innsikt for enkeltindividet, og for samfunnet som helhet (Morozov 2013).[16]

[16] Gary Wolf, teknologijournalist, skrev manifestet som lanserte ”Quantified Self”-bevegelsen. Han trekker frem fire faktorer som forklarer fremveksten av selvlogging: 1) elektroniske sensorer har blitt mindre og kraftigere 2) sensorene har blitt allestedsnærværende når de nå ligger inne i mobiltelefonene 3) sosiale medier har normalisert og lagt til rette for en delingskultur 4) fremveksten av skytjenester har gjort det mulig å lagre og sammenstille helsedata (og andre personopplysninger) på nye måter (Morozov, 2013).

En konsekvens av denne trenden er at flyten av helseopplysninger vil øke. Helseopplysninger vil være tilgjengelige for analyse på nye måter og for flere – og andre typer – aktører enn i dag.[17] Det offentlige helsevesenet vil ha problemer med å ta i mot helseopplysningene som genereres fra bruk av mobilt selvmonitoreringsutstyr. Brukere av slik teknologi vil derfor trolig være avhengige av private aktører som kan analysere opplysningene. Kommersielle virksomheter vil slik kunne bygge opp store databanker med helseopplysninger. Dette er opplysninger de kan benytte selv, eller selge tilgang til. Helseopplysninger forteller svært mye om oss og vil derfor kunne være attraktive for mange aktører, slik som forskningsinstitusjoner, forsikringsselskaper, arbeidsgivere og banker.

3.4 Big Data innen politi, sikkerhet og etterretning

I det følgende vil vi se på hvordan storskala dataanalyse benyttes innen politiet og etterretningstjenestene.

3.4.1 Smart politi

Big Data kan gjøre politiet smartere. Ved å benytte avanserte analyseteknikker kan politiet avdekke tidligere ukjente sammenhenger i kriminalitetsdata og andre tilgjengelige datakilder. Trender og mønstre kan brukes til å sannsynliggjøre en fremtidig utvikling. Dette kan hjelpe politiet til å forutsi hendelser, fordele ressurser og kanskje til og med avverge at hendelser inntreffer (Datatilsynet og Teknologirådet 2013). Bruk av Big Data i politiet kalles *predictive policing*, og flere mener dette vil revolusjonere måten politiarbeid drives på (Morozov 2013).

Politiet i Norge har foreløpig ikke tatt i bruk slik avansert analyseteknologi[18]. I USA er det derimot flere eksempler på bruk av Big Data i politiet. Politiet i Los Angeles (LAPD) har for eksempel tatt i bruk et analyseverktøy kalt *PredPol*, opprinnelig utviklet for å forutsi jordskjelv og etterskjelv. *PredPol* fores med lokal kriminalitetsstatistikk over biltyverier, innbrudd og annen relevant informasjon med henblikk på å bekjempe kriminalitet. LAPD kan nå forutsi hvor og når det er sannsynlig at en gitt kriminell handling finner sted – og det innenfor områder ned til 150 m². Ved hjelp av mobile digitale kart kan politipatruljene bruke denne informasjonen til å ligge i forkant av kriminelle hendelser. At politiet er på stedet allerede før forbrytelsen har skjedd, har naturlig nok ført til en kraftig reduksjon i kriminaliteten (Datatilsynet og Teknologirådet 2013).

I Europa er det politiet i Storbritannia som har kommet lengst i å ta i bruk avansert analyseteknologi. I forbindelse med avviklingen av de olympiske leker i 2010, benyttet politimyndighetene i London Big Data-teknologi til å drive sanntids sentimentanalyse på ord og uttrykk sammenstilt fra sosiale media (Hewlett-Packard 2013).

Prediksjonsanalyse blir benyttet også innenfor andre deler av justissektoren. Flere delstater i USA prøver ut et system designet for å forutsi hvor sannsynlig det er at en fengselsinnsatt vil drepe eller

[17] Innenfor den såkalte selvmonitoreringsbevegelsen (life-loggers movement) vektlegges viktigheten av eierskap til egne helsedata; alle skal ha rett til å få en kopi av sine helseopplysninger utlevert og mulighet til å benytte disse slik man selv vil (Bollier 2010).

[18] 22. juli-kommisjonen var knusende i sin dom over politiets bruk av teknologi. De påpekte i sin rapport at norsk politi må bli flinkere til å utnytte potensialet som ligger i informasjons- og kommunikasjonsteknologi: http://www.22julikommisjonen.no/Rapport

bli drept mens vedkommende er i permisjon. På grunnlag av dette systemet avgjør myndighetene tilslag eller avslag på permisjonssøknader (Mayer-Schönberger og Cukier 2013).

3.4.2 Ingen nål uten høystakk

Etterretningstjenestene i USA ligger i front når det gjelder å ta i bruk nye og kraftige analyseteknologier. En artikkel i *The New York Times* (2013), viser til at NSA og CIA har prøvd ut IBMs Big Data-teknologi i to år. NSA har nylig bygget et enormt datasenter i Utah, fem ganger større enn Capitol Hill, viet Big Data-analyse.

Big Data har endret måten å drive etterretningsarbeid på. Tidligere tok man utgangspunkt i mistenkelige enkeltpersoner og forsøkte å kartlegge dem så detaljert som mulig ved hjelp av telefonavlytting og andre hjelpemidler. Ved hjelp av Big Data kan man gå den andre veien. Man starter med å samle inn enorme datamengder, og benytter disse til å lete etter mønstre og korrelasjoner som kan avsløre mistenkelige hendelser og personer. Det er lite interessant for etterretningsmyndighetene å benytte anonymiserte data i denne sammenhengen. Å kartlegge enkeltindivider og relasjoner mellom enkeltindivider er et sentralt formål.

I kjølvannet av 11. september 2001, har etterretningsmyndighetene i mange land fått utvidede fullmakter til å samle inn og analysere personopplysninger fra en lang rekke kilder. I juni 2013 lekket den britiske avisen *The Guardian*, via en intern varsler i NSA, topphemmelige dokumenter om det såkalte PRISM-programmet (*The Guardian* 2013). Dokumentene avslørte at NSA skal ha blitt innvilget direkte tilgang til eposter, chat-korrespondanse, voice calls og dokumenter fra Microsoft, Yahoo, Google, Facebook, Paltalk, AOL, Skype, YouTube og Apple. Representanter fra de berørte selskapene sier imidlertid at de verken kjenner til programmet eller at myndighetene skal ha hatt en slik direkte tilgang til deres servere. Selskapene hevder de kun utleverer informasjon på grunnlag av spesifikke henvendelser fra myndighetene og at disse henvendelsene håndteres etter fastsatte regler. EU har kritisert amerikanske myndigheter i forbindelse med avsløringen rundt PRISM, og EU kommisjonær Viviane Reding har blant annet uttalt at:

> *"The concept of national security does not mean that "anything goes": States do not enjoy an unlimited right of secret surveillance. In Europe, even in cases involving national security, every individual – irrespective of their nationality – can go to a Court, national or European, if they believe that their right to privacy or to data protection has been infringed. I have made my point clearly: this is what I want for European citizens also in the US". (Reding 2013).*

Alle fakta i denne saken er foreløpig ikke på bordet. Det er uansett en kjensgjerning at de enorme datamengdene med personopplysninger som de store internettselskapene besitter, er av svært stor interesse for mange aktører. Etterretningstjenestene er i denne sammenheng ikke noe unntak. Til hvilke aktører data fra de store nettselskapene flyter, er av stor betydning for den enkelte borger. Bruk av Big Data har potensiale i seg til å legge til rette for et massivt overvåkingssamfunn.

I hvilken grad norske etterretningstjenester har tatt i bruk Big Data-teknologi er ikke kjent for Datatilsynet. Etterretningstjenestene i Norge er ikke underlagt personopplysningsloven, men er regulert i egen særlovgivning. Det er EOS-utvalget, et kontrollorgan oppnevnt av Stortinget, som fører tilsyn med etterretnings-, overvåkings- og sikkerhetstjenestene.

3.5 Personvernutfordringer

Slik vi har sett i de forutgående eksemplene, kan Big Data bli benyttet til mange samfunnsnyttige formål. Big Data blir i stor grad brukt til å identifisere overordnede trender og sammenhenger som ikke nødvendigvis har en slagside mot personvernet. Men Big Data kan også bli benyttet på en slik måte at det berører enkeltindivider direkte.

I noen sektorer er bruk av Big Data mer rettet mot enkeltindividet enn i andre. Innen helseforskning er formålet med Big Data i hovedsak å identifisere mønstre og sammenhenger på populasjonsnivå – ikke å få dypere innsikt om enkeltindivider. For internettbaserte selskap innen sosiale medier og netthandel, samt for forsikrings- og kredittopplysningsvirksomheter, er det imidlertid svært interessant å få dypere innsikt i handlingsmønsteret til enkeltindivider. For politiet og etterretningstjenestene er det interessant å benytte Big Data både til å avdekke mønstre i kriminalitetsbildet, og til å avsløre mistenkelig atferd hos enkeltindivider.

Enkelte former for bruk av Big Data kan komme i direkte konflikt med dagens personvernlovgivning. Andre former vil ikke være direkte lovstridige, men kan føre til et press på sentrale personvernprinsipper, noe som igjen kan ha uheldige konsekvenser for samfunnet som helhet. I det følgende vil vi trekke frem sentrale personvernutfordringer knyttet til bruk av Big Data, hjemmehørende i begge de to kategoriene nevnt overfor. Hvordan utfordringene må håndteres innenfor gjeldende lovverk diskuteres i kapittel 4.

3.5.1 Bruk av data til nye formål

Big Data handler i stor grad om å hente ut sekundærverdi av innsamlede data:

> "*Unlike material things – the food we eat, a candle that burns – data's value does not diminish when it is used; it can be processed again and again. (…) Just as data can be used many times for the same purpose, more importantly, it can be harnessed for mulitiple purposes as well*" (Mayer-Schönberger og Cukier 2013:101)

Dette utfordrer personvernprinsippet om at data kun skal samles inn og brukes videre til klart angitte formål.[19]

Googles innsamling og bruk av opplysninger fra sine brukere i forbindelse med utviklingen av tjenester som Google Now, er et eksempel på dette. Europeiske personvernmyndigheter krever nå at Google må klargjøre hva som er formålet med innsamlingen av opplysninger fra brukerne og hvordan opplysningene kobles mellom selskapets ulike tjenester. [20]

---

[19] Jf. personopplysningsloven § 11 første ledd bokstav c)

[20] I oktober 2012 sendte de europeiske personvernmyndighetene et brev til direktør Page i Google med et krav om at selskapet skulle implementere en rekke krav for å tilfredsstille det europeiske personverndirektivet 95/46/EC: http://www.cnil.fr/fileadmin/documents/en/20121016-letter_google-article_29-FINAL.pdf. Overordnet forklart var det tre krav som ble fremsatt, men som enda ikke er gjennomført:

- At Google skulle gi tilstrekkelig informasjon til sine brukere om hva som er formålet med behandlingen av personopplysninger og hvilke typer data som behandles.
- At Google skulle klargjøre hva som var formålet og virkemidler i forbindelse med kobling av data mellom selskapets tjenester.
- At Google skulle gi informasjon om hvor lenge de lagret personopplysninger for de ulike tjenestene.

Offentlige virksomheter har også interesse av å sammenstille data på nye måter for å skape bedre og mer effektive tjenester. Et smart politi for eksempel, vil være sultent på data. Data som opprinnelig var samlet inn i forvaltningsøyemed kan være nyttig i en etterforskningssak når de for eksempel sammenstilles med informasjon fra et annet register, og på denne måten brukes til å avdekke sammenhenger og mønstre (Datatilsynet og Teknologirådet 2013). Deling og gjenbruk av data kan gi et mer slagkraftig politi, men dette innebærer at opplysninger blir brukt til andre formål og i andre sammenhenger enn opprinnelig tenkt.

Muligheten som ligger i Big Data til å sammenstille stadig større datasett, og – den i mange tilfeller svært verdifulle – kunnskapen man kan trekke ut av slik analyse, kan sette prinsippet om formålsbestemthet under press. Denne utfordringen må sees i klar sammenheng med utfordringen under, om datamaksimalisering.

3.5.2 Datamaksimalisering

Big data innebærer et nytt syn på data, der data får en verdi i seg selv. Verdien ligger i dataenes fremtidige bruksmuligheter. Et slikt syn på data påvirker virksomhetenes ønske om og motivasjon til å slette data. Verken private eller offentlige virksomheter vil ønske å slette data som på et senere tidspunkt, og sammenstilt med andre datasett, kan vise seg å bringe ny og verdifull innsikt. Hvem kunne for eksempel forutse at søkehistorikken til Bing kunne bidra til å avsløre alvorlige bieffekter ved samtidig inntak av to bestemte medisinske preparater (ref. punkt 3.3.1)?

Big Datas forretningsmodell er antitesen til dataminimalisering, som er et sentralt personvernprinsipp[21]: Det skal ikke samles inn og lagres mer data om personer enn nødvendig for å oppfylle nærmere angitte formål. Dataene skal slettes når de ikke lenger er nødvendige for formålet. Mer utstrakt bruk av Big Data kan føre til at det fremover vil bli enda mer utfordrende for Datatilsynet å følge med på at sletteplikten overholdes i henhold til personopplysningsloven.[22]

3.5.3 Mangel på åpenhet – tap av kontroll

Innsynsretten og retten til informasjon om behandlingen av egne personopplysninger er viktige personvernprinsipper. Mangel på åpenhet og informasjon om hvordan data benyttes og sammenstilles kan føre til at vi blir offer for beslutninger vi ikke forstår og ikke har kontroll over: Hvilke profiler finnes om meg der ute? På basis av hvilke personopplysninger blir de utformet? Er jeg vurdert som en attraktiv eller verdiløs kunde?

Den alminnelige internettbruker har for eksempel liten innsikt i hvordan annonsemarkedet på Internett og andre digitale plattformer fungerer og hvordan vedkommendes personopplysninger samles inn og benyttes av kommersielle interesser (Turow 2011). I en undersøkelse foretatt av Datatilsynet kom det frem at svært få leverandører av mobilapplikasjoner opplyste om hvilke personopplysninger som ble samlet inn og hvordan disse ble behandlet (Datatilsynet 2010). Flere av aktørene som opererer i dette markedet, særlig tredjepartsaktører som datameklere og analyseselskaper, er dessuten ukjente for folk flest. Rettigheten den enkelte har til å be om innsyn i hvilke personopplysninger som er samlet inn, blir da vanskelig å praktisere.

---

[21] Jf. artikkel 6 (1) c i direktiv 95/46/EFC

[22] Jf. lovens § 11 første ledd bokstav e, jf. § 28.

Det finnes også store datameklingsselskaper som henter inn personopplysninger via andre plattformer enn Internett, som tidligere nevnte Acxiom. Selskapet har ikke gitt publikum rett til innsyn i personopplysningene selskapet besitter om den enkelte. Etter kritisk mediedekning og press fra amerikanske reguleringsmyndigheter, har imidlertid Acxiom signalisert at de vil åpne opp sine databaser for innsyn i løpet av 2013 (*The New York Times* 2012b, Federal Trade Commission 2012 og Forbes 2013).

3.5.4 Ubalanse virksomhet – individ

Big Data øker ubalansen mellom de store virksomhetene på den ene siden og enkeltindividet på den andre. Det er de virksomhetene som *samler inn* personopplysninger som henter ut den stadig voksende merverdien som ligger i analyse og bearbeiding av disse opplysningene, og ikke vi som *avgir* opplysningene. Snarere kan denne transaksjonen være til forbrukerens ulempe i den forstand at den kan utsette oss for potensiell fremtidig sårbarhet.

OECD (2013) har viet denne problemstillingen oppmerksomhet og forsøkt å utvikle en metode for å bestemme den monetære verdien på personopplysninger. I følge OECDs rapport vil en metode for å fastsette verdien på personopplysninger bidra til å gi større åpenhet og innsikt i hvordan markedet for omsetning av personopplysninger fungerer. En forhøyet bevissthet hos den enkelte om hvilken verdi våre personopplysninger har, vil kunne bidra til å jevne ut forholdet mellom virksomheter på den ene siden og enkeltindividet på den andre. Blant annet kan det bidra til at vi stiller høyere krav og forventninger til hvordan våre personopplysninger håndteres.

OECD-rapporten peker på en trend som kanskje kan bidra til å gjøre det lettere å fastsette verdien på personopplysninger. Det har dukket opp selskaper som tilbyr såkalte ”data lockers”. Tjenester med *data lockers* gir kundene mer kontroll over bruken av deres personopplysninger, blant annet ved at de selv kan bestemme om deres opplysninger skal selges videre til tredjeparter. Velger brukeren å selge opplysningene vil de motta en provisjon av salget.

Selskapet *reputation.com* vil gjøre det lettere for nettbrukere å tjene penger på sine personopplysninger. De vil lansere en tjeneste som skal gjøre det mulig for nettbrukere å dele personopplysninger med ulike virksomheter, i bytte mot rabatter eller andre goder (*MIT Technology Review* 2013b).

3.5.5 Sammenstilling kan frembringe sensitiv informasjon

En annen personvernutfordring ved Big Data-analyse er at data som hver for seg ikke er sensitive, gjennom sammenstilling kan gi et sensitivt resultat. Big Data-verktøy kan identifisere mønstre som kan si noe om folks helse, politiske overbevisning eller seksuelle legning. Slike opplysninger har krav på særlig vern. Virksomheter som benytter Big Data må derfor være dette bevisst i utviklingen av algoritmer slik at de unngår at data sammenstilles på en slik måte at vitale personverninteresser avsløres. Et mye brukt eksempel for å illustrere denne utfordringen ved bruk av Big Data er den amerikanske næringsmiddelkjeden Targets såkalte ”graviditetsalgoritme” (*The New York Times* 2012c). Target utviklet en algoritme som kan predikere hvilke kunder som er gravide basert på hvilke varer som handles inn. Target sendte tilbudskuponger på ”graviditetsprodukter” til disse kundene. I et tilfelle førte utsendelse av en slik kupong til at faren i huset fikk kjennskap til datterens graviditet før hun selv fikk anledning til å fortelle det. Slik bruk av innsamlede opplysninger bryter med enkeltindividets forventninger til hvilke formål opplysningene kan bli benyttet.

3.5.6 Farvel anonymitet

En av de virkelig store utfordringene ved Big Data er risikoen for reidentifisering. Gjennom sammenstilling av data fra flere kilder kan det oppstå risiko for at enkeltindivider kan identifiseres fra i utgangspunktet anonyme datasett.

Big Data kan bestå av en kombinasjon av identifiserbare og ikke-identifiserbare opplysninger. Selv om datasettene som benyttes er anonymiserte, er det en risiko for at sammenstillingen av datasettene kan medføre reidentifisering av enkeltpersoner (eller til og med reidentifisering av feilaktig *antatte* enkeltpersoner, såkalte falske positiver).

Reidentifisering kan skje ved at noen tar personlige data de allerede har om andre og søker etter treff i et anonymisert datasett, eller ved at man tar et treff fra et anonymt datasett og søker etter treff på offentlig tilgjengelig informasjon.

Risiko for reidentifisering kan reduseres ved å sikre at kun anonymiserte data inngår i analysen. Det er imidlertid ikke alltid enkelt å få overblikk over om et datasett er fullstendig anonymisert, eller om det fortsatt inneholder personopplysninger. Dette kan være vanskelig av to grunner:

Begrepet *å identifisere* – og derav *å anonymisere* – er komplisert fordi man kan identifisere enkeltpersoner på en rekke ulike måter. Det inkluderer direkte identifikasjon, hvor noen er eksplisitt identifiserbare fra en enkelt datakilde (for eksempel en liste med fullt navn), og indirekte identifikasjon, hvor to eller flere datakilder må kombineres for at identifisering skal skje.

Organisasjoner som har planer om å slippe et anonymisert datasett kan være tilfreds med at dataene i seg selv ikke skal identifisere noen. Men det de ikke vet er at det i noen tilfeller kan være mulig at det finnes andre data tilgjengelig som gjør det mulig for en tredjepart å reidentifisere personer i det anonymiserte datasettet.

Selv etter at identifiserende opplysninger har blitt slettet, er det fortsatt mulig å koble spesifikk informasjon til et individ på bakgrunn av koblinger som finnes i ulike Big Data-samlinger. Et eksempel på dette er *"How to break anonymity of the Netflix Prize Dataset"* (Narayanan og Shmatikov 2008). Netflix annonserte en konkurranse for utviklere med en premie på én million amerikanske dollar. Målet var at noen skulle utvikle en løsning som ga en forbedring på 10 prosent på deres anbefalingsmodul. I den forbindelse slapp Netflix et "treningsdatasett" til de konkurrerende utviklerne som de kunne bruke for å trene sine system. Med datasettet fulgte en "disclaimer" (ansvarsfraskrivelse) hvor det stod *"for å beskytte kundenes personvern, har all personlig informasjon som identifiserer den enkelte kunde blitt fjernet og alle kundens ID-er har blitt erstattet med tilfeldig tildelte ID-er."* Det finnes flere filmvurderingsportaler på Internett, blant annet IMDb. På IMDb kan enkeltpersoner registrere seg og rangere filmer, og stå frem med fullt navn. Forskerne Narayanan og Shmatikov koblet Netflix sin avidentifiserte treningsdatabase med IMDbs database (basert på datoen for vurdering av en bruker) og klarte på den måten delvis å reidentifisere brukerne i Netflix treningsdatabase. Det finnes også en lang rekke andre slike eksempler. [23]

[23] Et annet eksempel er et case der amerikanske forskere klarte å reidentifisere DNA-materiale som lå lagret i av-identifisert form (Gymrek et al. 2013).

Akkurat som et menneskes fingeravtrykk kan identifisere en enkeltperson på et sted hvor en kriminell handling har funnet sted, kan også "data fingerprints" gjøre det. Det er en kombinasjon av dataverdier som ikke er lik noen annen kombinasjon av dataverdier i en tabell. Forskere har funnet slike data fingerprints i anonymiserte datasett mye lettere enn de fleste kan tro er mulig. Straks noen finner et unikt data fingerprint, kan vedkommende linke de aktuelle dataene med tilleggsopplysninger. Mange anonymiseringsteknikker hadde vært perfekte om ikke uvedkommende visste noe som helst annet om hele verdens befolkning. I virkeligheten er det motsatte tilfellet og det lages nye databaser hver dag med informasjon om mennesker. Netflix-studien viser at det er svært enkelt å identifisere enkeltmennesker i anonymiserte data.

3.5.7 Feil faktagrunnlag

Det er et viktig personvernprinsipp at beslutninger som får konsekvenser for den enkelte skal være basert på korrekte opplysninger. Å basere avgjørelser på opplysninger hentet inn og sammenstilt fra for eksempel sosiale medier, medfører imidlertid en risiko for at beslutningene treffes på feil faktagrunnlag. Avgjørelser basert på sammenstilling av slike opplysninger vil ikke være transparente og etterprøvbare i samme grad som avgjørelser basert på opplysninger hentet fra offisielle registre. Det er videre viktig å være bevisst at data samlet inn fra for eksempel sosiale medier ikke nødvendigvis gir et korrekt bilde av en person. En svakhet ved Big Data-analyse er at den ofte ikke tar hensyn til kontekst[24]. Å basere beslutninger på opplysninger tiltenkt andre formål og oppstått innenfor en annen kontekst, kan gi resultater som ikke samsvarer med den faktiske situasjonen.

3.5.8 Datadeterminisme

Big Data-tankegangen hviler på en antagelse om at jo mer data man samler inn og har tilgang til, jo bedre, mer begrunnede og treffsikre beslutninger kan man ta. Men mer datainnhøsting betyr ikke nødvendigvis mer kunnskap. Mer data kan også bety mer forvirring og flere falske positiver: "*Big Data is driven more by storage capabilities than by superior ways to ascertain useful knowledge*" (Bollier 2010).

Mer utstrakt bruk av Big Data, og dertil hørende bruk av automatiserte avgjørelser og prediksjonsanalyse, kan få uheldige konsekvenser for enkeltindividet. Big Data kan befeste eksisterende fordommer og stereotypier, og forsterke sosial ekskludering og lagdeling. Vi kan få et A- og et B-lag i samfunnet, der bare de med "riktig" profil blir prioritert. Profiler på aggregert nivå kan også slå helt feil ut for enkeltpersoner. En utvikling der stadig flere beslutninger i samfunnet blir tatt basert på algoritmer, kan lede til et "dataenes diktatur" (Mayer-Schönberger og Cukier 2013); at vi ikke blir vurdert på basis av hva vi faktisk foretar oss, men på basis av hva alle dataene om oss sier at vi sannsynligvis kan komme til å gjøre. Enkeltindividet er mer enn summen av de digitale sporene vi legger igjen.

Det er også viktig å være bevisst at algoritmer ikke er objektive: *"The prejudices of a society are reflected in the algorithms that are searched"*, har forfatteren av boka The Numerati, Stephen Baker, uttalt (Bollier 2010). Algoritmer er sannsynlighetsmodeller basert på historiske data og er født i en samfunnsmessig kontekst (Morozov 2013). Mer bruk av profileringsteknikker og prediktiv analyse i

[24] danah boyd og Kate Crawford er to forskere som har poengtert viktigheten av å ta hensyn til kontekst i Big Data-analyse (2012).

politiet for eksempel, kan føre til at folk vil føle seg feilaktig kriminalisert på bakgrunn av etnisitet, bosted og lignende.

Vil man i fremtiden kunne bli arrestert på bakgrunn av en algoritme? Kan slike sannsynlighetsprognoser være tilstrekkelig til å gi politiet "skjellig grunn til mistanke" og anledning til å ransake personer som ellers ikke ville blitt stoppet, eller banke på dører som ellers ikke ville blitt banket på (Datatilsynet og Teknologirådet 2013)? Å mistenkeliggjøre mennesker, ikke på bakgrunn av hva de har gjort, men for hva de kan komme til å gjøre, utfordrer selve prinsippet som rettsstaten er bygget på; at man er uskyldig til det motsatte er bevist. En slik bruk av algoritmer vil også undergrave synet på at mennesket har en fri vilje og selv kan foreta moralske valg.

I Norge og mange andre land er det de siste årene vedtatt lover som gjør at det skal lite til å bli arrestert på bakgrunn av mistenkelig atferd. Bruk av Big Data sett i lys av slike lover kan gjøre enkeltindividet ekstra sårbart i forhold til uforvarende å havne i politiets søkelys.

3.5.9 Nedkjølingseffekt

Hvis vi får en utvikling der kredittscore og forsikringspremier baseres på nær sagt alle opplysninger vi legger igjen i ulike sammenhenger på nett og ellers i dagliglivet, kan dette få konsekvenser for personvernet, og hvordan vi oppfører oss. Om ti år er det kanskje slik at barna dine ikke får forsikring fordi du har delt på sosiale nettverk at du er disponert for en arvelig sykdom. Dette kan føre til at vi legger bånd på hvordan vi deltar i samfunnet, eller aktivt tilpasser våre handlinger – både online og ellers. Vi frykter konsekvenser med hensyn til om vi vil innvilges lån, få bilforsikring, bli leietager og så videre.

Når det gjelder etterretningsmyndighetenes bruk av Big Data, truer hemmeligholdet rundt hvilke datakilder de samler inn opplysninger fra og hvordan de brukes, ikke bare den enkeltes personverninteresser. Det utfordrer også tilliten til myndighetene og i siste instans selve fundamentet for et åpent og velfungerende demokrati.

Hvilke konsekvenser har det når vi blir usikre på hvordan opplysninger vi avgir i ulike sammenhenger på Internett og via mobiltelefonen blir analysert og potensielt brukt til nye, og for oss ukjente, formål? Vil vi da våge å ytre oss like fritt? Mennesker som vet at de blir iakttatt endrer oppførsel fordi konteksten er en annen – tillitten til omgivelsene endres. Et dårlig ivaretatt personvern kan svekke demokratiet ved at borgerne begrenser sin deltakelse i åpen meningsutveksling. Utstrakt bruk av Big Data i etterretningssammenheng og av politiet generelt, kan i verste fall ha en nedkjølende effekt på ytringsfriheten hvis premissene rundt bruken er holdt skjult og ikke er etterprøvbar. Det finnes også andre offentlige virksomheter som kan ha interesse av å hente inn opplysninger om borgerne for kontrollformål. Toll- og skattemyndighetene og NAV kan ved bruk av Big Data stå bedre rustet i kampen mot for eksempel smugling, skatteunndragelser og trygdemisbruk. Det er usikkerheten knyttet til hvilke opplysninger som betraktes som relevante og interessante for myndighetene, ikke nødvendigvis hva som er realiteten i dette, som vil kunne påvirke borgernes frimodighet.

3.5.10 Ekkokamre

Personaliseringen av nettet, med skreddersydde tilbud basert på den enkeltes atferd på nett, påvirker også rammebetingelsene for meningsbryting – et viktig fundament for et velfungerende demokrati. Dette er ikke i første rekke en personvernutfordring, men en samfunnsutfordring. Eli Pariser (2011) og Joseph Turow (2011) trekker frem faren ved "ekkokamre" eller "filter bubbles" –

det vil si at personalisering av nettet resulterer i at man kun eksponeres for innhold som svarer til ens profil. Nettet, og dermed samfunnet, blir inndelt i ulike ”bokser” som ikke har kontakt med hverandre. Dette vil legge en demper på meningsbrytingen, fordi man i mindre grad enn tidligere blir utsatt for meninger som avviker fra ens egne.

4. Juridiske spørsmål

Analyser av Big Data kan føre til verdifulle resultater innenfor flere sektorer. Big Data reiser imidlertid også noen juridiske spørsmål. Svarene på disse spørsmålene ligger ikke i dagen. Det finnes lite norsk og europeisk rettskildemateriale som direkte berører tematikken.[25] Dermed må problemstillingene vurderes på bakgrunn av de generelle, teknologinøytrale og skjønnsmessige lovreglene som finnes på området. For vår del vil det rettslige utgangspunktet først og fremst være *personopplysningsloven*.[26]

Nedenfor ser vi på noen av de juridiske spørsmålene som oppstår ved bruk eller *behandling* av Big Data, i lys av personopplysningsloven. Vi skal også undersøke hva som ligger i lovens personopplysningsbegrep, ettersom uttrykket har en sentral betydning.

4.1 Big Data og loven

Det ligger ikke noen bestemt form for databehandling innbakt i begrepet Big Data. Uttrykket mangler som nevnt et presist innhold, og kan omfatte både innsamling og analyse av informasjon. De ulike trinnene i analyseprosessen kan dessuten variere, slik vi har sett i kapittel 2. At databehandlingen gjelder opplysninger om fysiske personer er heller ikke opplagt – formålene med analysene kan ofte oppnås ved hjelp av anonym eller anonymisert informasjon. I andre tilfeller er det nettopp opplysninger om enkeltindivider som er det interessante.

Behandling av Big Data kan utløse plikter og rettigheter etter den norske personopplysningsloven. Forutsetningen er at behandlingen angår *personopplysninger*. Loven gjelder nemlig for "*behandling av personopplysninger som helt eller delvis skjer med elektroniske hjelpemidler*", ifølge lovens § 3. At behandling av Big Data skjer med elektroniske hjelpemidler, ligger i sakens natur.

Dermed blir spørsmålet om hva som omfattes av personopplysningsbegrepet avgjørende for lovens anvendelse. Dette spørsmålet er imidlertid ikke alltid like enkelt å besvare, som vi skal se nedenfor.

---

[25] Vi kjenner ikke til at det er fattet noen avgjørelser i norske domstoler eller i norsk statsforvaltning som omtaler *Big Data* som fenomen. Søk på uttrykket i Lovdata gir null treff den 30. mai 2013. Søket omfatter samtlige rettsinstanser, lover, forskrifter og avgjørelser fra Personvernnemnda, i tillegg til juridisk litteratur.
Søk på "Big Data", i kombinasjon med begreper som *privacy* og *data protection* i internasjonale juridiske artikkeldatabaser, gir kun en håndfull treff, og artiklene tar som oftest utgangspunkt i amerikanske regler. Søk i *LexisNexis* i juni 2013 ga 13 treff, hvorav et fåtall var relevante.

[26] Personopplysningsloven, lov av 14. april 2000 nr. 31 om behandling av personopplysninger, bygger på det europeiske personverndirektivet (Europaparlamentets og Rådets direktiv 95/46/EF av 24. oktober 1995 om beskyttelse av fysiske personer i forbindelse med behandling av personopplysninger og om fri utveksling av slike opplysninger).

4.2 Personopplysningsbegrepet

Ifølge lovens definisjon av personopplysningsbegrepet, dreier det seg om "*opplysninger og vurderinger som kan knyttes til en enkeltperson*".[27] Definisjonen[28] inneholder tre hovedbestanddeler:[29]

- enhver form for informasjon
- som kan knyttes til
- identifiserbar eller identifisert enkeltperson

Selv om definisjonens bestanddeler henger nært sammen, og de påvirker hverandre gjensidig, går vi gjennom hver og en av dem for seg nedenfor.

4.2.1 Enhver form for informasjon

Det første elementet kan tolkes svært vidt. Kort sagt dekker det all tenkelig informasjon, uavhengig av art, innhold eller format.

Formuleringen omfatter for det første *objektive* opplysninger, slik som informasjon om en persons alder, bosted eller inntekt. Det samme gjelder *subjektive* opplysninger, for eksempel én persons vurderinger eller karakteristikker av en annen. Hvis Peder twitrer at Hans ikke er til å stole på, er dette et eksempel på en slik subjektiv opplysning.

I denne sammenhengen spiller innholdet i informasjonen ingen rolle. Om opplysningene fremstår som betydningsfulle for den de angår eller andre, er uten betydning. At opplysningene ikke er troverdige eller umulige å bevise, er også uvesentlig. Det har heller ikke noe å si om opplysningene direkte angår forhold som tradisjonelt forbindes med privatlivets fred. Opplysninger som skriver seg fra det offentlige rom, eller fra arbeidsplassen, er også omfattet av informasjonsbegrepet i loven. Det samme gjelder både sensitive og alminnelige opplysninger.

Begrepet omfatter all informasjon, uavhengig av format. Opplysningene kan komme til uttrykk gjennom skrift, tall, tegninger, fotografier, lyd eller biometriske kjennetegn. Når det gjelder informasjonskildene, kan det dreie seg om alt fra e-poster til post-it-lapper, offentlige saksdokumenter, meldinger på sosiale medier, sms-er, digitale og analoge fotografier eller lydopptak, og så videre.

4.2.2 Tilknytningselementet

Informasjonen må imidlertid kunne knyttes til en fysisk person. Dette betyr at opplysningen kan si noe om – eller at informasjonen angår – et enkeltindivid. Noen ganger er denne tilknytningen ganske åpenbar, slik som i en pasientjournal, i personalmappen på jobben eller i registrene til kredittopplysningsselskapet.

Andre ganger er tilknytningen mindre opplagt. Opplysninger om tilstanden til et hus eller en bil vil rimeligvis først og fremst assosieres med tingen i seg selv. Den samme informasjonen kan imidlertid

---

[27] Lovens § 2 nr. 1.
[28] Opinion 04/2007 on the concept of personal data
[29] I enkelte fremstillinger er tilknytningselementet og identifikasjonskravet behandlet sammen (Schartum og Bygrave 2011)

også avsløre andre forhold, for eksempel om personer som har hatt befatning med tingen. Opplysninger om verdien av et hus eller en bils verkstedshistorikk kan også si noe om eieren av huset eller den som har brukt bilen. I visse tilfeller kan dessuten opplysninger om én person samtidig være opplysninger om en annen, for eksempel innenfor medisin og genetikk.

Det er med andre ord tilstrekkelig at tilknytningen mellom informasjon og person er indirekte. Indirekte vil tilknytningen også være når opplysningene i første rekke angår en gruppe, men samtidig kan si noe om individene som gruppen består av. Dette forutsetter at individene er *identifiserbare*. Hva som ligger i dette uttrykket, ser vi på nedenfor.

4.2.3 Identifiserbar enkeltperson

Informasjonen må kunne knyttes til en *identifisert eller en identifiserbar enkeltperson*.[30] Dette innebærer for det første at opplysninger om juridiske personer faller utenfor personopplysnings-lovens saklige virkeområde.[31] For det andre må enkeltpersonen være *identifiserbar*.

Enkeltperson

Hvor man skal trekke grensen mellom opplysninger angående juridiske personer og opplysninger om enkeltpersoner, kan være vanskelig å avgjøre i praksis. Opplysninger om enkeltpersonforetak vil for eksempel også kunne si noe om innehaveren, ikke minst dersom foretaket ikke har noen ansatte. Andre opplysninger som primært fremstår som foretaksopplysninger, kan også avsløre forhold som kan knyttes til et enkeltindivid. Behandlingen av dem vil i så fall være underlagt bestemmelsene i personopplysningsloven.

Identifiserbarhetskriteriet

At enkeltpersonen må kunne la seg identifisere, er ikke direkte uttrykt i legaldefinisjonen i personopplysningsloven.[32] Dette fremgår imidlertid av den tilsvarende definisjonen i direktivet, som henviser til at opplysningene må kunne knyttes til en *identifisert* eller *identifiserbar* person.[33]

At personen P er *identifisert* vil si at P kan skilles ut fra en gruppe av personer. P er således *identifiserbar* når det er *mulig* å identifisere ham, selv om identifiseringen ennå ikke har skjedd. At P er identifisert, vil normalt være enkelt å konstatere. Derfor er det identifiserbarhetskriteriet som definerer personopplysningsbegrepets ytre grense.

Dette betyr at personopplysningsloven kommer til anvendelse hvis det er mulig å skille ett individ fra et annet. At identifikasjonen kan tenkes å finne sted på et eller annet tidspunkt i fremtiden, er altså tilstrekkelig. Det er heller ikke nødvendig at det er den behandlingsansvarlige selv som har muligheten til å sammenkoble informasjonen som gjør identifiseringen mulig.[34]

---

[30] Jf. ordlyden i direktiv 95/46/EF artikkel 2.
[31] Personopplysningsloven gjelder imidlertid også for behandling av kredittopplysninger om andre enn enkeltpersoner, jf. personopplysningsforskriften § 4-1.
[32] Lovens § 2 nr. 1.
[33] I forarbeidene til personopplysningsloven er det da også gitt uttrykk for at identifiserbarhetskriteriet ligger implisitt i begrepet "enkeltperson", se Ot.prp. nr. 92 (1998-1999) merknadene til § 2 i kapittel 16.
[34] Jf. avsnitt 26 i direktivets fortale: "*for at afgøre, om en person er identificerbar, tages alle de hjælpemidler i betragtning, der med rimelighed kan tænkes bragt i anvendelse for at identificere den pågældende enten af den registeransvarlige eller af enhver anden person*".

Opplysninger som fremstår som anonyme, kan vise seg å være personopplysninger i lovens forstand. Forklaringen er at det er mulig å identifisere en eller flere personer *indirekte*. For eksempel regnes IP-adresser som personopplysninger i visse sammenhenger.[35] Internettleverandøren besitter lister over abonnenter og tildelte IP-adresser. Disse kan sammenstilles, slik at identiteten til abonnenten avdekkes.[36] Øvrige opplysninger som forekommer sammen med slike adresser, vil dermed også regnes som personopplysninger.

4.2.4 Oppsummering – personopplysningsbegrepet og Big Data

Summen av de ovennevnte elementene blir et vidtfavnende personopplysningsbegrep. Konklusjonen er at personopplysningsloven, som definerer sitt eget virkeområde ved hjelp av dette begrepet, har et tilsvarende bredt nedslagsfelt. Lovverket gjelder i utgangspunktet behandling av alle former for personlig informasjon, så lenge det er mulig å knytte den til et identifiserbart individ i fremtiden. Den som behandler slike opplysninger i forbindelse med en eller annen form for Big Data, har dermed et sett med lovfestede begrensninger å forholde seg til.

Hvis opplysningene derimot er *anonyme* eller *anonymisert*, vil reglene ikke sette noen begrensninger på behandlingen av dem. Big Data-analytikere har med andre ord et friere spillerom i omgangen med slike opplysninger. Anonyme opplysninger kan defineres som opplysninger som det ikke er mulig å knytte til et identifiserbart individ, når man tar i betraktning alle de hjelpemidlene som med rimelighet kan tenkes brukt for å identifisere vedkommende, enten av den behandlingsansvarlige eller av en hvilken som helst annen person. Uttrykket "anonymiserte opplysninger" refererer slik til opplysninger som det tidligere har vært mulig å knytte til en identifiserbar person, men hvor tilknytningen til identifiserbare personer er gjort umulig.[37]

De rettslige kravene som vi omtaler nedenfor, gjelder altså bare dersom opplysningene kan knyttes til identifiserbare personer, og ikke behandling av anonyme opplysninger.

4.3 Rettslige krav til Big Data-behandling

Den som har ansvaret for at det behandles Big Data – den behandlingsansvarlige[38] – må altså forholde seg til bestemmelsene i personopplysningsloven. Grunnkravene i lovens § 11 står helt sentralt i den forbindelse. Hvis databehandlingen ikke oppfyller alle vilkårene som oppstilles i denne paragrafen, er behandlingen ulovlig. Den som planlegger å behandle personopplysninger, må med

---

[35] Jf. praksis hos Datatilsynet og Personvernnemnda (se for eksempel nemndas sak 2011-10), den svenske Datainspektionen (Dnr. 1402-2007) og den franske Commission Nationale de l'Information et des Libertés (CNIL), se http://www.cnil.fr/linstitution/actualite/article/article/ladresse-ip-est-une-donnee-a-caractere-personnel-pour-lensemble-des-cnil-europeennes/

[36] Abonnenten kan naturligvis være en juridisk person, for eksempel innehaveren av en kafé. I så fall vil kafeens IP-adresser ikke uten videre knyttes til en identifiserbar enkeltperson. Hvis kafeen fører lister over internettbrukerne, eller lignende, kan det gjøre identifisering mulig, og dermed vil IP-adressene regnes som personopplysninger.

[37] Opinion 04/2007 on the concept of personal data.

[38] Ifølge legaldefinisjonen i personopplysningsloven § 2 nr. 4 er den behandlingsansvarlige den som bestemmer formålet med behandlingen av personopplysninger og hvilke hjelpemidler som skal brukes. Den engelske versjonen av direktivet benytter begrepet "data controller".

andre ord på forhånd forsikre seg om at databehandlingen ikke er i strid med personopplysningsloven § 11.[39]

Nedenfor skal vi se nærmere på hva som ligger i disse kravene. Hvilken betydning de vil ha i forbindelse med ulike former for behandling av Big Data, er av særlig interesse. I den sammenheng viser vi til eksemplene som er presentert i avsnittene ovenfor.

4.3.1 Rettslig grunnlag for behandling av personopplysninger

Personopplysningsloven § 11 innleder med å vise til at ingen kan behandle personopplysninger uten at det foreligger et rettslig grunnlag for databehandlingen. Utgangspunktet er med andre ord et alminnelig forbud mot slik databehandling som loven omfatter, men loven er altså utstyrt med konkrete rettslige grunnlag som likevel gjør databehandlingen lovlig. Loven angir flere forskjellige grunnlag, som samtykke, lovhjemmel og ulike nødvendighetsgrunner.

Lovhjemmel

Lovhjemlet behandling av personopplysninger vil si at det er gitt regler i, eller i medhold av, formell lov som pålegger eller gir adgang til en bestemt behandling av personopplysninger. Som oftest dreier det seg om lover som pålegger myndighetene oppgaver og plikter. Eksempler på dette er utlendingsloven[40] og politiregisterloven.[41] Begge lovene fastsetter konkrete regler om utlendingsmyndighetenes og politiets adgang til å behandle visse kategorier personopplysninger. Gjennomføringen av datalagringsdirektivet i norsk rett vil for øvrig føre til en hjemmelsbestemt lagringsplikt i privat sektor, nemlig tilbydere av tele- og internettjenester.

Hvorvidt en bestemmelse i lov gir et rettslig grunnlag for en bestemt databehandling, må avgjøres ved fortolkning av den enkelte bestemmelsen. Hjemmelskravet er relativt, det vil si at det kreves klarere hjemmel jo større personvernmessige konsekvenser behandlingen kan få.[42] Behandling regulert i annen lov må også fylle de kravene som oppstilles for lovlig behandling i EU-direktivet artikkel 7, så fremt den aktuelle loven sorterer under områdene som er dekket av EØS-avtalen.

Vi skal ikke gå nærmere inn på dette hjemmelsgrunnlaget i denne rapporten; nedenfor går vi gjennom de alternativene som er mest relevante i Big Data-sammenheng.[43]

Samtykke

Et av spørsmålene som behandlingen av Big Data reiser, er om den som samler inn og analyserer informasjonen må innhente samtykke fra de personene som opplysningene angår. Et gyldig samtykke i lovens forstand skal være *frivillig*, *informert* og *uttrykkelig*.[44]

---

[39] Med mindre de generelle unntakene i lovens § 3 andre ledd (private formål) eller § 7 (ytringsfriheten) kommer til anvendelse, eller de generelle betingelsene i lovens innledning ikke er oppfylt, slik som bestemmelsene om geografisk virkeområde i § 4, etc.
[40] Lov 15. mai 2008 nr. 35 om utlendingers adgang til riket og deres opphold her.
[41] Lov 28. mai 2010 nr. 16 om behandling av opplysninger i politiet og påtalemyndigheten; loven trer i kraft når Kongen bestemmer.
[42] Dette følger av *legalitetsprinsippet*, og er understreket av i Ot.prp. nr. 92 (1998-1999), se merknadene til lovens § 8 i proposisjonens kapittel 16.
[43] Vi antar at lovens øvrige nødvendighetsgrunner – ivaretakelse av rettslige forpliktelser, beskyttelse av vitale interesser og utøvelse av offentlig myndighet – har begrenset interesse i denne sammenhengen.
[44] Ifølge personopplysningsloven § 2 nr. 7.

Av og til fremstilles samtykke som lovens hovedregel – det vil si at behandling av personopplysninger i hovedsak skal baseres på samtykke fra den som opplysningene angår.[45] Tar man utgangspunkt i den private autonomi, og tanken om at den enkelte skal kunne ha bestemmelsesretten over egne personopplysninger – slik som i teorien om det *integritetsfokuserte* personvernet – fremstår dette som både fornuftig og rettferdig. [46]

Imidlertid er det ikke rettslig dekning for et slikt utgangspunkt. Alternativene i loven er likeverdige – dette fremgår av lovens ordlyd, hvor de ulike grunnlagene er adskilt av konjunksjonen "eller".[47] Samtykke må imidlertid innhentes dersom ingen av de andre rettslige grunnlagene i loven er anvendelige.[48] Samtykke vil dermed – i alle fall i noen sammenhenger – kunne fremstå som det mest anvendelige rettsgrunnlaget i praksis. Begrunnelsen er at de andre behandlingsgrunnlagene oppstiller terskler som det kan være vanskelig å overstige.[49]

Alternativ til samtykke – kontrakt

Innsamling av personopplysninger over Internett kan i visse tilfeller baseres på alternativet i personopplysningsloven § 8 bokstav a. Denne bestemmelsen sier at personopplysninger kan behandles når det er *nødvendig for å oppfylle en avtale med den registrerte*.[50]

Utveksling av kontakt- og betalingsinformasjon ved kjøp av varer og tjenester er et eksempel på behandling av personopplysninger som kan forankres i dette rettsgrunnlaget. Et selskap som selger varer over Internett kan derfor uten videre kreve de opplysningene om en kjøper som er nødvendige for å gjennomføre salget. Alternativet kan også være anvendelig for visse opplysningskategorier i andre tilfeller. Et eksempel på dette er lokasjonsopplysninger i forbindelse med nedlasting av en kartapplikasjon for smarttelefon.[51]

Det er imidlertid viktig å være oppmerksom på at loven her oppstiller et *nødvendighetskrav.* Opplysningene kan med andre ord ikke behandles hvis formålet med behandlingen kan nås på annen måte. Dette alternativet har dermed et nokså snevert anvendelsesområde, og bør benyttes med varsomhet.

Alternativ til samtykke – interesseavveining

Det forekommer at personopplysninger samles inn via Internett uten at brukerne er klare over det, for eksempel til profileringsformål. Databehandlingen må likefullt baseres på minst ett av de rettslige grunnlagene i loven. I disse tilfellene vil avtalealternativet åpenbart ha sine begrensninger. Samtykke vil kunne være et alternativ, men det kan være mange grunner til at datainnsamlerne ikke ønsker å gi brukerne en reell valgfrihet – en *opt in*-løsning vil for eksempel kunne redusere datagrunnlaget

---

[45] På side 108 i Ot.prp. nr. 92 (1998-1999) er det blant annet uttalt at "*(b)ehandling av personopplysninger bør i størst mulig utstrekning baseres på samtykke fra den registrerte, selv om den også kan hjemles i de grunnlagene som oppstilles i bokstavene a-f. For det første vil dette styrke den registrertes muligheter til å råde over opplysninger om seg selv. For det annet vil man ved å basere behandlingen på samtykke unngå mulig tvil om de mer skjønnsmessige vilkårene i bokstavene a til f er oppfylt*".

[46] Ref NOU 1997:19 (Justis- og politidepartementet 1997)

[47] I samme retning går Personvernnemnda i avgjørelsen PVN-2012-1.

[48] Jf. personopplysningsloven § 8, og § 9 som kommer inn dersom det behandles sensitive personopplysninger.

[49] Jf. PVN-2012-1.

[50] Det er en forutsetning at de sivilrettslige krav som stilles til gyldig avtaleinngåelse er oppfylt (se Opinion 15/2011 on the definition of consent s. 6)

[51] Se for eksempel Datatilsynets sak 12/00276 (nedlasting av applikasjon for smarttelefon).

betraktelig. Det eneste relevante rettsgrunnlaget som da står igjen, er lovens *interesseavveiningsalternativ*.[52]

Ifølge dette alternativet kan personopplysninger bare behandles hvis behandlingen er nødvendig for at den behandlingsansvarlige, eller tredjepersoner som opplysningene utleveres til, kan ivareta en berettiget interesse, *og* hensynet til den registrertes personvern ikke overstiger denne interessen.[53]

Man står altså overfor en avveining mellom to motstående interesser. Utfallet avhenger av hvilke interesser som befinner seg i vektskålene, og disse må veies mot hverandre i det enkelte tilfellet.

Det kan dreie seg om svært skjønnsmessige størrelser på begge sider av vektstangen. De *berettigede interessene* som bestemmelsen viser til har vært tolket nokså vidt i norsk forvaltningspraksis. Økonomisk vinning er for eksempel utvilsomt en slik interesse i denne sammenhengen. Det reelle hinderet på denne siden av vektstangen er *nødvendighetskriteriet*. For å oppfylle dette kriteriet må den behandlingsansvarlige kunne godtgjøre at de anførte og legitime behandlingsformålene ikke kan oppnås på annen måte enn gjennom nettopp den spesifikke behandlingen som avveiningen gjøres i lys av.

Samtidig vil utfallet av avveiningen avhenge av hvilke personverninteresser som står på spill. Det er altså ikke tilstrekkelig at behandlingen som sådan er nødvendig – interessen som behandlingen tjener må i tillegg veie tyngre enn de motstående personverninteressene. Det er vanskelig å si noe generelt om hvordan de hvilke ulike personverninteressene skal vektlegges i seg selv, men man kan ta utgangspunkt i en risikobasert tilnærming. Jo større behovet er for å beskytte informasjonen ut fra konfidensialitetshensyn, jo større vekt må man legge på personverninteressen. I forarbeidene til den norske loven er det for øvrig uttrykt at *"(g)enerelt må hensynet til privatlivets fred tillegges betydelig vekt i avveiningen mot kommersielle interesser."*[54] Personverninteressene må også vurderes i lys av at privatlivets fred er beskyttet av menneskerettighetene.[55]

Kompenserende tiltak i form av informasjonssikkerhetsmessige grep eller *pseudonymisering* av personopplysningene kan også spille inn. De kan veie opp for personvernulempene behandlingen ellers ville ha medført.[56]

Datatilsynets anbefaling

Selv om vi ikke kan utelukke at interesseavveiningen *kan* legitimere persondatabehandling i enkelte tilfeller, må dette alltid vurderes konkret. Alternativet kan derfor vanskelig benyttes som rettslig grunnlag for innsamling og analyse av Big Data generelt. Under enhver omstendighet er det den behandlingsansvarlige som har bevisbyrden for at vilkårene i interesseavveiningsalternativet er til

---

[52] Personopplysningsloven § 8 bokstav f.

[53] Bestemmelsen har sitt motstykke i direktivets artikkel 7 – alternativet omtales gjerne som *"the balancing of interests test"*, eller simpelthen *"legitimate interests"*.

[54] Se også *"Who Owns The Future?"* (Lanier 2013) der forfatteren argumenterer for at det er umoralsk at et fåtall personer med herredømme over de største serverne skal berike seg på innsamling og analyse av våre data, uten at vi får noe igjen for det.

[55] Jf. for eksempel personverndirektivet artikkel 1, Charter of Fundamental Rights of the EU artikkel 8, og Den europeiske menneskerettskonvensjonen (EMK) artikkel 8. EMK er inkorporert i norsk rett ved lov av 21. mai 1999 nr. 30 om styrking av menneskerettighetenes stilling i norsk rett.

[56] Opinion 04/2007 on the concept of personal data.

stede. Med bestemmelsens mange skjønnsmessige bestanddeler og usikkerhetsmomenter, kan dette by på utfordringer.[57]

Samtidig vil det i mange tilfeller være unødvendig å analysere data om identifiserbare personer for å nå de samfunnsnyttige målene som det ofte henvises til. Storskalaanalyse av folks bevegelsesmønstre i en storby, for å effektivisere kollektivtrafikktilbudet, kan medføre både miljømessige og økonomiske gevinster. Disse fordelene kan imidlertid nås uten at man kjenner identiteten til de som forflytter seg rundt i byen.[58]

Når det gjelder lovens rettslige grunnlag, er Datatilsynets anbefaling derfor at den behandlingsansvarlige primært baserer seg på å innhente gyldig samtykke fra de registrerte på forhånd – dette er trolig den beste forsikringen mot å begå lovbrudd. I de tilfellene hvor det eksisterer en direkte relasjon mellom den behandlingsansvarlige og den registrerte, for eksempel et kundeforhold, vil en slik tilnærming fremstå som praktisk og overkommelig. For eksempel kan et nettsted som selger bøker gi brukerne mulighet til å *klikkakseptere* at opplysninger om hva de titter på og kjøper, blir analysert og brukt til profileringsformål.

Konstant etterspørsel av samtykke på Internett vil føre til en form for samtykkeapati hos internettbrukerne, hevder noen (Hordern 2013). Dette kan paradoksalt nok føre til redusert beskyttelse for individene, er argumentet (Hordern 2013, Hildebrandt 2009, Tene og Polonetsky 2012). Den behandlingsansvarlige vil komme til å innhente så vide samtykker som mulig, og belage seg på at samtykkeerklæringene stort sett ikke vil bli lest, for deretter å behandle personopplysningene etter eget forgodtbefinnende. Samtykke er derfor ikke egnet som rettslig grunnlag for behandling av Big Data, hevdes det.

Kravet om at samtykket skal være *informert* innebærer ikke noe mer enn at den behandlingsansvarlige gir de registrerte en oversikt over hva databehandlingen innebærer.[59] Det trenger ikke være seksti sider lange vilkår, slik det av og til harseleres over.[60] Kortfattet og presis informasjon, som i tillegg utelukkende tar for seg relevante aspekter ved databehandlingen, vil med andre ord kunne redusere faren for slik apati.

For det andre er det ikke alltid nødvendig å innhente samtykke, for eksempel i de tilfellene hvor opplysningene må oppgis for å oppfylle en kjøpsavtale. Er personverntrusselen lav, samtidig som de motstående interessene er tungtveiende, kan dette også danne grunnlag for behandling av Big Data. Og for det tredje må den behandlingsansvarlige uansett oppfylle sin plikt til å informere de registrerte. Samtykkeskeptikerne ser av og til ut til å overse disse faktorene.

---

[57] Artikkel 29-gruppen har annonsert at det i løpet av 2013 vil komme en uttalelse om "*legitimate interests*", som skal sikre homogenitet i fortolkningen av kriteriene på tvers av Europa.
[58] Sml. Datatilsynets sak 09/00163 ("Ruter AS").
[59] Se for eksempel personopplysningsloven § 19.
[60] Disse vilkårene er da heller ikke ment å skulle opplyse den registrerte i henhold til kravene til et gyldig samtykke, men snarere å gi tjenesteleverandørene kontraktrettslig ryggdekning, ofte med utgangspunkt i angloamerikanske rettsprinsipper, jf. det såkalte "*four corners*"-prinsippet, se for eksempel http://en.wikipedia.org/wiki/Four_corners_(law)

4.3.2 Formålsbegrensningsprinsippet

Gjenbruk av informasjon er et særtrekk ved Big Data.[61] Grunnkravene omtalt over må være oppfylt også for slik gjenbruk.[62] Dette gjelder uavhengig av om personopplysningene behandles i medhold av samtykke, eller om behandlingen baseres på et annet rettsgrunnlag.

Formålsbegrensningsprinsippet står særlig sentralt i den forbindelse.[63] Prinsippet innebærer at innsamlede personopplysninger ikke kan brukes til formål som er *uforenlige* med innsamlingsformålet, med mindre den registrerte samtykker til slik bruk. Prinsippet setter de registrerte i stand til å ta et informert valg om å overlate sine data til den behandlingsansvarlige. Samtidig er de registrerte sikret en viss forutsigbarhet, ettersom prinsippet gir forsikringer om at opplysningene ikke vil bli behandlet til helt andre formål i neste omgang, uten de registrertes kunnskap.

Hva som skal regnes som uforenlig, kan være vanskelig å avgjøre.[64] Sikkert er det imidlertid at referansepunktet for vurderingen er det formålet som den behandlingsansvarlige er forpliktet til å oppgi ved innsamlingen.[65] Det er uansett en forutsetning for diskusjonen at det andre formålet skiller seg fra det første. Den behandlingsansvarlige kan imidlertid ikke påberope seg et hvilket som helst formål for å omgå denne begrensningen. Formålene skal være uttrykkelig angitt, og de må være legitime, i den forstand at de finnes en saklig tilknytning mellom behandlingsformål og den behandlingsansvarliges virksomhet.[66] Den behandlingsansvarlige bør dermed tenke nøye gjennom hvordan behandlingsformålene formuleres før datainnsamlingen settes i gang.

Hva som nærmere ligger i uforenlighetsbegrepet, må avgjøres på bakgrunn av en *kompatibilitetsvurdering*.[67] Dette betyr ikke at enhver annen bruk er utelukket. Sentrale momenter i vurderingen er blant annet om den videre bruken av opplysningene innebærer ulemper for den registrerte, og om bruken bryter med den registrertes berettigede forventninger, eller skiller seg sterkt fra den som lå til grunn for innsamlingen.[68]

Data som er samlet inn av offentlige myndigheter, for eksempel i medhold av lovbestemmelser under trussel om straffansvar, vil slik vanskelig kunne behandles senere til prediksjon av individuell atferd i forsikringsøyemed.[69] Ved prediktiv analyse av innsamlede data, må altså den behandlingsansvarlige forsikre seg om at prediksjonsanalysen ikke er uforenlig med det opprinnelige innsamlingsformålet.

Prinsippet oppstiller altså noen rammer for gjenbruk av personopplysninger som allerede er samlet inn. I dette kan det ligge en betydelig utfordring for kommersiell Big Data-analyse. Innhenting av forhåndssamtykke som omfatter analyseformålene, kan være en løsning. Ofte vil imidlertid

---

[61] Jf. sitatet innledningsvis i avsnitt 3.5.1.
[62] Personopplysningsloven § 11, som tilsvarer artikkel 6 i direktivet.
[63] Lovens § 11 første ledd bokstav c; prinsippet omtales også som *formålsbestemthetsprinsippet* eller *finalitetsprinsippet*.
[64] Hva som skal regnes som én behandling, er igjen avhengig av behandlingsformålet, se Ot.prp. nr. 92 (1998-1999), kapittel 16 merknadene til § 2.
[65] Formålet må som nevnt også være legitimt.
[66] Personopplysningsloven § 11 første ledd bokstavene b og c.
[67] Se Opinion 03/2013 on purpose limitation, jf. uttrykket "incompatible" over.
[68] Ot.prp. nr. 92 (1998-1999) kapittel 16, merknadene til lovforslagets § 11.
[69] Se Ot.prp. nr. 92 (1998-1999), på samme sted, se også Høyesteretts avgjørelse i Rt. 2013 s. 143.

analysene foregå hos andre enn de som samlet inn opplysningene – i Big Data dreier det seg, som nevnt i avsnitt 3.5.2, ofte om videre bruk av innsamlede data til nye formål, for å utvinne dataenes iboende sekundærverdi.

Den sikreste foranstaltningen mot å krenke dette prinsippet vil derfor være å sørge for at opplysningene anonymiseres. I så fall vil den videre behandlingen ikke være regulert av lovgivningen, uavhengig av om det er den opprinnelige innsamleren eller en ny aktør som står ansvarlig for behandlingen. Anonymiseringen må imidlertid være *reell* – se blant annet mer om dette i avsnitt 3.5.6.

4.3.3 Relevansprinsippet og dataminimalisering

Et annet av lovens grunnkrav slår fast at personopplysninger bare kan behandles hvis de er relevante for behandlingsformålet. Relevanskravet må først og fremst forstås som et krav om at den behandlingsansvarlige skal begrense seg til bare å behandle personopplysninger som er *nødvendige* for oppnåelse av behandlingsformålet.[70] Relevanskravet omtales av og til som *data minimisation*, eller dataminimalisering, og kan med fordel sees i sammenheng med formålsbegrensningsprinsippet.

Vi ser umiddelbart at datamaksimaliseringstankegangen som vi beskrev i avsnitt 3.5.2, kommer i direkte konflikt med relevanskravet. Kravet om dataminimalisering vil med andre ord kunne sette en stopper for storstilt innsamling og akkumulasjon av persondata som er motivert av datasettenes potensielle verdi i fremtiden.

Nok en gang ser vi at anonymisering av personlig informasjon kan gi datainnsamlerne større handlingsrom, så fremt de sørger for reell anonymisering av opplysningene.

4.3.4 Plikten til å sørge for korrekte data

Et annet av personopplysningslovens grunnkrav fastsetter at den behandlingsansvarlige skal sørge for at opplysningene som behandles er *korrekte*.[71] Kravet bør i denne sammenhengen forstås på bakgrunn av at det gjelder alle former for behandling av personopplysninger, og ikke bare på innsamlingsstadiet.[72] Den behandlingsansvarlige har dermed en plikt til å sørge for at slutninger om enkeltindivider på bakgrunn av analyser av Big Data er korrekte. Hvis analysene viser at personer som liker X med åtti prosent sannsynlighet vil komme til å utsettes for Y, er det ikke mulig å konkludere med at årsakssammenhengen vil inntreffe i 100 prosent av tilfellene. Diskriminering på bakgrunn av statistiske analyser kan dermed også bli et spørsmål om personvern. Spørsmålet er kanskje særlig aktuelt i forbindelse med ulike former for profilering eller prediktiv analyse på individnivå.

Hvis personopplysninger som ligger åpent tilgjengelig, for eksempel på Internett, senere benyttes til nye formål av en annen behandlingsansvarlig, følger dette ansvaret med. Hvis Peder twitrer at Kari er

[70] Schartum og Bygrave (2004) viser til at relevanskravet kan gi uttrykk for flere faktorer, som *logisk*, *rettslig* og *kognitiv* relevans. Vi går ikke nærmere inn på hva som skiller disse faktorene fra hverandre her.
[71] Personopplysningsloven § 11 første ledd bokstav e, jf. direktivet artikkel 6 (d), som viser til at dataene skal være "*accurate*".
[72] Dette poenget illustreres ved en henvisning til direktivteksten, som tilføyer at informasjonen "*must be (…), where necessary, kept up to date*".

en svindler, kan selskapet S, som yter forbrukslån med høy rente, bare behandle opplysninger om at denne Kari *er* en svindler, så fremt selskapet kan godtgjøre at påstanden er korrekt.[73]

Gjennomsiktighet, for eksempel i form av retten til å gjøre seg kjent med innholdet i opplysninger som behandles om en selv, er for øvrig en forutsetning for at den registrerte skal kunne ivareta sine interesser. Opplysninger, vurderinger og påstander som viser seg å ikke være korrekte, kan i medhold av regelverket kreves korrigert eller slettet.[74]

4.4 Individets rettigheter

4.4.1 Gjennomsiktighet – informasjon og innsyn

Personopplysningsloven pålegger den behandlingsansvarlige å gi informasjon om databehandlingen til de som berøres av den.[75] Informasjonsplikten oppstår ved informasjonsinnsamlingstidspunktet, og gjelder uavhengig av om personopplysningene samles inn på grunnlag av samtykke eller et annet rettslig fundament. Hvis opplysningene samles inn fra den registrerte selv, skal informasjonen gis på forhånd. Samles de inn fra andre enn den registrerte, skal informasjonen gis så snart opplysningene er innhentet. Informasjonen skal dessuten gis uoppfordret. I tillegg har den registrerte *innsynsrett* i opplysningene, det vil si en rett til å kreve ytterligere informasjon om databehandlingen.[76]

Bestemmelsene er ment å skulle gi den registrerte muligheten til å skaffe seg en oversikt over hvilke opplysninger som behandles om ham. Hvordan opplysningene behandles, hva formålet er med behandlingen, om opplysningene skal utleveres og hvor de er hentet fra, er også forhold det må informeres om. Denne informasjonen er en forutsetning for den registrertes rett til å kontrollere at det ikke behandles feilaktig informasjon, jf. avsnitt 4.3.4, og for retten til å kreve mangelfull personinformasjon korrigert eller slettet.

4.4.2 Personprofiler og automatiserte avgjørelser

Det oppstår en utvidet informasjonsplikt i to tilfeller. Det første tilfellet er hvis den behandlingsansvarlige treffer avgjørelser som fullt ut er basert på *automatisk* behandling av personopplysninger, og avgjørelsen har vesentlig betydning for den registrerte. I så fall har den registrerte en rett til å kreve at den behandlingsansvarlige gjør rede for regelinnholdet – det vil si *logikken* eller *algoritmen* – i programvaren som "fatter" avgjørelsen.

Den andre situasjonen er når den behandlingsansvarlige henvender seg til den registrerte eller treffer avgjørelser som retter seg mot ham, på bakgrunn av personprofiler.[77] Innholdet i denne informasjonen skiller seg ikke vesentlig fra den som skal gis etter de alminnelige reglene i §§ 19 og 20, men her er det ikke lenger *datainnsamlingen* som utløser plikten. Det fremgår av bestemmelsen at plikten til å informere inntreffer ved avgjørelsestidspunktet.

Så fremt den konkrete behandlingen av Big Data innebærer slike henvendelser eller avgjørelser som nevnt over, får altså den ansvarlige en utvidet plikt til å informere personene som informasjonen

---

[73] Ettersom kravene i personopplysningsloven § 11 er kumulative, er forutsetningen for denne diskusjonen at de øvrige grunnkravene er til stede.
[74] Se personopplysningsloven § 27.
[75] §§ 19 og 20.
[76] § 18
[77] §§ 21 og 22.

angår. Det finnes få eksempler på at disse bestemmelsene er påberopt i praksis, men de kan fort tenkes å få fornyet aktualitet i forbindelse med behandling av Big Data. Personalisert markedsføring på bakgrunn av profiler som er laget ved hjelp av ulike sporingsteknikker på nettet er ett av eksemplene beskrevet i kapittel 2. Retten til å kreve redegjørelser for innholdet i algoritmene som ligger til grunn for avgjørelser av vesentlig betydning, vil dessuten kunne vise seg som en viktig garanti for ivaretakelsen av den enkeltes rettssikkerhet.

4.4.3 Retting og sletting

Innsynsretten og informasjonsplikten er forutsetninger for den registrertes rett til å kreve korrigering av feilaktig eller mangelfull personinformasjon. Som vi var inne på i avsnitt 3.5.2, vil ulike variasjoner av fenomenet Big Data kunne føre til en aversjon mot sletting av data, ettersom dataene innehar en verdi i seg selv.

Reglene om sletting kan deles inn i to grupper. For det første er den behandlingsansvarlige forpliktet til å slette personopplysninger som det ikke lenger er nødvendige å behandle for å oppfylle det opprinnelige behandlingsformålet.[78] I denne situasjonen ser vi at opplysningene har vært relevante, men at de etter en stund har "utspilt sin rolle". Et eksempel på dette er trafikkinformasjon som teleselskapene oppbevarer for faktureringsformål. Når faktureringen er et faktum, skal opplysningene slettes.[79] Ønsker teleselskapene å gjøre disse dataene til gjenstand for Big Data-analyse, må teleselskapene i så fall enten innhente samtykke fra kundene, eller sørge for at opplysningene anonymiseres.[80]

For det andre må opplysninger som det ikke er adgang til å behandle også slettes.[81] Vi kan tenke oss at ulovlig innsamlede personopplysninger ønskes anonymisert, slik at det kan gjennomføres analyser av de anonymiserte opplysningene. Det er vanskelig å si hvorvidt loven er til hinder for en slik praksis, ofte likestilles anonymisering med den i regelverket omtalte slettingen. I Big Data-sammenheng har da også anonymisering blitt en mer uforutsigbar affære enn det var for bare noen år siden. Under en hver omstendighet vil det kunne virke urimelig om datainnsamleren kan oppnå en økonomisk vinning på bakgrunn av en handling eller praksis som i utgangspunktet var ulovlig.

4.5 Noen internasjonale spørsmål

4.5.1 Lovvalg

Hvilke lands lover og regler regulerer behandling av Big Data? Svaret er de rettsregler som gjelder i den behandlingsansvarliges etableringsland.[82] Det er altså i utgangspunktet ikke mulig å utlede forpliktelser på bakgrunn av den norske personopplysningsloven, med mindre den behandlingsansvarlige er etablert i landet.

---

[78] Personopplysningsloven § 28, jf. § 11 første ledd bokstav e.
[79] Når endringsloven om datalagringsdirektivet trer i kraft, vil de samme opplysningene kunne lagres lenger, av hensyn til kriminalitetsbekjempelsesformål, se lov 15. april 2011 nr. 11 om endringer i ekomloven og straffeprosessloven mv. (gjennomføring av EUs datalagringsdirektiv i norsk rett).
[80] Med mindre analysene er kompatible med faktureringsformålet.
[81] Personopplysningsloven § 27 fastsetter at opplysninger som er "*uriktige, ufullstendige eller som det ikke er adgang til å behandle*" skal rettes hvis den registrerte krever det. Når det ikke er anledning til å behandle dataene i det hele tatt, vil slik retting kunne skje ved hjelp av sletting, fremgår det av Ot.prp. nr. 92 (1998-1999), se merknadene til § 27 i kapittel 16 i proposisjonen.
[82] Personopplysningsloven § 4 første ledd.

Dette kriteriet byr i praksis på en rekke utfordringer som vi ikke skal gå nærmere inn på her. Det er tilstrekkelig å nevne at det i utgangspunktet dreier seg om faktisk utøvelse av aktiviteter innenfor en forholdsvis fast struktur, og at denne strukturens rettslige status ikke er av avgjørende betydning. Det er imidlertid nødvendig også å nevne *hjelpemiddelkriteriet*. Dersom den behandlingsansvarlige ikke er etablert i EØS-området, kan loven likevel komme til anvendelse hvis den behandlingsansvarlige benytter hjelpemidler i Norge.[83] Den norske personopplysningsloven kan dermed få ekstraterritoriell virkning, for eksempel overfor en virksomhet i USA.

Det er imidlertid høyst uklart hva som ligger i hjelpemiddelbegrepet. For eksempel har enkelte antydet at dataprogrammer og informasjonskapsler er en form for hjelpemidler i regelverkets forstand. Spørsmålet er interessant, men foreløpig uavklart.[84] Hvis det ikke skal mer til for å utløse virkninger etter norske regler enn at det benyttes informasjonskapsler, vil det i alle fall kunne medføre en dramatisk utvidelse i den norske personopplysningslovens geografiske virkeområde. Bestemmelsen om regelverkets geografiske virkeområde vil imidlertid kunne komme til å bli drastisk endret, se nedenfor.

4.5.2 Eksport av data til tredjeland

Hvis databehandlingen innebærer overføring av personopplysninger til utlandet, må den behandlingsansvarlige overkomme visse rettslige hindre.[85] En (fiktiv) norsk butikkjede har for eksempel samlet inn og lagret opplysninger om alle transaksjoner med kjedens kunder de siste ti årene. Nå ønsker kjeden å analysere informasjonen. Ettersom kjeden ikke selv har den nødvendige kompetansen for å gjennomføre disse analysene, skal opplysningene sendes til tre ulike tjenesteleverandører i henholdsvis USA, Tyskland og Israel.[86]

Opplysningene kan uten videre overføres til Tyskland, siden det er fri flyt av personopplysninger i EØS-området.[87] Opplysningene kan også overføres til Israel, ettersom landet er godkjent som trygg dataimportør ved beslutning av EU-kommisjonen.[88] Når det gjelder USA, må det undersøkes om leverandøren er Safe Harbor-sertifisert.[89] I så fall kan opplysningene også sendes dit uten videre.[90] Hvis databehandleren i USA ikke er oppført på den offisielle listen over Safe Harbor-sertifiserte selskaper, må butikkjeden stille *garantier*, som skal godkjennes av Datatilsynet før dataene kan overføres.[91] Den sikreste måten å få slik godkjennelse er å benytte EUs standardkontrakter for overføring av personopplysninger.

Når selskapet Amazon samler inn data om hvilke bøker norske kunder titter på, skjer det ikke noen overføring i rettslig forstand. Personopplysningene må først være samlet inn av en

---

[83] Bestemmelsens andre ledd.
[84] Et lignende spørsmål er reist for EF-domstolen – den kommende uttalelsen vil kunne få avgjørende betydning i så måte. Case C-131/12: Reference for a preliminary ruling from the Audiencia Nacional (Spain) lodged on 9 March 2012 — Google Spain, S.L., Google Inc. v Agencia Española de Protección de Datos, Mario Costeja González
[85] Personopplysningsloven §§ 29 og 30.
[86] De tre samarbeidspartnerne i eksemplet er *databehandlere*, jf. lovens § 15.
[87] Jf. lovens hovedregel i § 29 første ledd.
[88] Personopplysningsforskriften § 6-1, jf. Commission Decision 2011/61/EU.
[89] Commission Decision 2000/520/EC.
[90] https://safeharbor.export.gov/list.aspx
[91] Personopplysningsloven § 30 andre ledd.

behandlingsansvarlig som er etablert i Norge for at personopplysningsloven skal få anvendelse.[92] I dette eksemplet går informasjonen direkte fra den registrerte til Amazon, og datainnsamlingen er ikke vernet av den norske personvernlovgivningen.

4.6 Nye regler om personvern

Personverndirektivet, som den norske loven bygger på, er fra 1995. Selv om direktivets målsetting og de generelle prinsippene fremdeles er gyldige, har globaliseringen og den teknologiske utviklingen endret verden. EU-kommisjonen uttalte derfor i 2010 at det var på tide å revidere personvernlovgivningen i Europa.[93]

Kommisjonen har nå utarbeidet et utkast til generell forordning om behandling av personopplysninger.[94] EU-parlamentet har i sin tur kommet med kommentarer og endringsforslag til utkastet.[95] Forordningen vil bli gjeldende rett for Norge, ettersom lovgivningen på databeskyttelses-området er EØS-relevant.[96] Om det blir en forordning, vil reglene bli direkte inntatt i norsk lovgivning.[97] Reglene vil også ha forrang fremfor andre regler i norsk rett som ikke er EØS-relevante.[98]

Det er særlig fire elementer i den foreslåtte reguleringen som er av interesse i forbindelse med Big Data. Vi ser nærmere på disse nedenfor.

4.6.1 Samtykke og interesseavveining

EU-kommisjonen uttalte allerede i 2010 at det var en målsetting å styrke og tydeliggjøre reglene om samtykke.[99] Styrking av samtykke har også vært en prioritet for Parlamentet, noe som blant annet tydelig fremgår av forslaget til endringer i ordlyden i forordningsforslagets artikkel 7.

Får Parlamentet gjennomslag for sine synspunkter, vil det bli vanskeligere å samle inn og bruke data uten den registrertes samtykke. I praksis vil det i så fall innebære at rekkevidden av interesseavveiningsalternativet innsnevres tilsvarende. Parlamentet går da også imot EU-kommisjonens forslag om at Kommisjonen selv, gjennom såkalte "*delegated acts*",[100] skal kunne legge avgjørende føringer for interesseavveiningen. Parlamentet foreslår i samme åndedrag å innføre en uttømmende liste over hvilke interesser som kan regnes som legitime i forordningsteksten. Hvorvidt en eller flere Big Data-relaterte interesser dukker opp en slik liste, gjenstår å se.

4.6.2 Formålsbegrensningen

Kommisjonen har foreslått å *snevre inn* formålsbegrensningen. Dermed vil de behandlingsansvarliges handlingsrom *utvides* tilsvarende. Kommisjonen har foreslått at det ikke lenger skal være nødvendig

---

[92] Se avsnittet over om jurisdiksjon.
[93] COM(2010) 609 final.
[94] Gjeldende utkast er datert 25.1.2012.
[95] Den såkalte Albrecht-rapporten: COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)
[96] Jf. EØS-avtalen vedlegg XI.
[97] Flere medlemsland mener at regelverket bør innføres som direktiv. Forskjellene mellom forordninger og direktiver er forklart i Arnesen, Finn (2009).
[98] Jf. EØS-loven § 2.
[99] Jf. COM (2010) 609 final
[100] En form for rettslig instrument innført ved Lisboa-traktaten; Kommisjonen har fått delegert myndighet til å gi regler vedrørende "*non essential elements*" i allerede eksisterende lovgivning, se traktatens artikkel 290.

å innhente samtykke fra de registrerte for å behandle allerede innsamlede personopplysninger til nye og uforenlige formål. I stedet mener Kommisjonen at det bør være tilstrekkelig at opplysningene kan behandles i medhold av minst ett av de øvrige rettsgrunnlagene.[101] Det vil si at det i praksis vil bli *enklere* for de som behandler Big Data å overvinne denne terskelen.

Parlamentet ønsker på sin side å bevare prinsippet slik det er i dag.[102] Det samme gjør Artikkel 29-gruppen, som viser til at forslaget vil medføre så vidtrekkende unntak fra formålsbegrensningen at prinsippet i praksis vil miste sin betydning.[103]

Dersom en justering av formålsbegrensningsprinsippet som foreslått av kommisjonen blir en realitet, kan det forenkle hverdagen for Big data-brukerne og -analytikerne. På den annen side, får Parlamentet og Artikkel 29-gruppen gjennomslag for sine innvendinger, vil rettstilstanden forbli uendret.

4.6.3 Innebygd personvern og standardinnstillinger

I artikkel 23 har Kommisjonen foreslått å kodifisere prinsippene om innebygd personvern og personvernvennlige standardinnstillinger, under overskriften "*Data protection by design and by default*". Hovedregelen i den foreslåtte artikkelens første ledd er gitt følgende utforming:

> *"Having regard to the state of the art and the cost of implementation, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures and procedures in such a way that the processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject."*

Bestemmelsen fastslår altså at den behandlingsansvarlige ("the controller") har en plikt til å implementere egnede tekniske tiltak for å ivareta de kravene som forordningen stiller. Dette er for så vidt ikke noe nytt, i alle fall ikke sett med norske øyne. Tilsvarende forpliktelser kan utledes av dagens regelverk, av blant annet personopplysningsloven § 14 og de utfyllende bestemmelsene i personopplysningsforskriftens tredje kapittel.

Imidlertid kan det muligens regnes som en nyvinning at denne plikten skal gis virkning allerede fra tidspunktet før selve databehandlingen har begynt ("*at the time of the determination of the means for processing")*. Det siste innebærer altså en rettslig forpliktelse til å følge *privacy* eller *data protection by design*-tankegangen allerede på forberedelsesstadiet, før databehandlingen er iverksatt.[104] Artikkelen kan med fordel leses i lys av de grunnleggende kravene til behandling av personopplysninger og reglene om risikovurdering (*Data Protection/Privacy Impact Assessment*) i forslagets artikkel 33.

I artikkelens andre ledd er det foreslått å ta inn en bestemmelse om *data protection by default,* eller personvernvennlige standardinnstillinger, som skal sikre at det ikke behandles opplysninger utover

---

[101] Det dreier seg her om de samme rettsgrunnlagene som er omtalt i avsnitt 4.3.1.
[102] COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)
[103] Opinion 03/2013 on purpose limitation.
[104] Albrecht påpeker for øvrig at skjæringspunktet er tidspunktet for "determination of *the purposes* and means for processing".

det som er strengt nødvendig for de spesifiserte og legitime behandlingsformålene som den behandlingsansvarlige kan vise til:

> *"The controller shall implement mechanisms for ensuring that, by default, only those personal data are processed which are necessary for each specific purpose of the processing and are especially not collected or retained beyond the minimum necessary for those purposes, both in terms of the amount of the data and the time of their storage. In particular, those mechanisms shall ensure that by default personal data are not made accessible to an indefinite number of individuals."*

Innsamling av og øvrig behandling av persondata skal altså begrenses til et minimum gjennom standardinnstillinger, og målestokken er hva som kan anses som nødvendig i relasjon til de legitime formålene som ligger til grunn for den aktuelle databehandlingen. Siktemålet med denne bestemmelsen er å lette overholdelsen av de kravene som regelverket stiller, for slik å sikre beskyttelsen av individets rettigheter.[105]

4.6.4 Utvidelse av databeskyttelsessonen

Et endringsforslag fra EU-kommisjonen som vekker stor interesse, er utkastets artikkel 3. Samtidig som artikkelen befester etableringslandsprinsippet som hovedregelen når det gjelder forordningens territoriale omfang,[106] foreslår kommisjonen i artikkelens andre ledd å utvide forordningens geografiske virkeområde. Kommisjonen ønsker å gi forordningen ekstraterritoriell effekt i forbindelse med visse aktiviteter.

Dersom selskap utenfor EU tilbyr varer og tjenester til individer i EU,[107] eller samler inn personopplysninger for å overvåke de samme individenes atferd, må selskapene respektere de europeiske reglene på området. Amerikanske selskap som ikke har noen tilknytning til Europa i selskapsrettslig forstand, vil dermed kunne komme til å måtte respektere fremtidens europeiske standarder for databeskyttelse og personvern. Sosiale nettsamfunn og andre tjenester som tilbys uten vederlag via nettet, og som norske borgere i dag bruker på eget ansvar, vil altså i fremtiden kunne underlegges de felleseuropeiske personvernreglene i fremtiden.

5. Oppsummering og anbefalinger

Ikke alle former for bruk av Big data innebærer personvernutfordringer. Sammenstilling av anonymiserte data for å predikere mønstre og utviklingstrekk på aggregert nivå faller utenfor personopplysningslovens virkeområde.

Personvernutfordringene knyttet til Big Data er først og fremst relatert til at dette representerer en samfunns- og forretningstrend som kan sette etablerte personvernprinsipper under press. Vi lever i en ekstremt datarik tid der nær sagt alt vi foretar oss kan måles, spores og analyseres. Med fremveksten av Tingenes Internett vil disse datamengdene øke dramatisk. De tekniske hindrene for å

---

[105] Se Albrecht-rapportens *Amendment 178*. COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)
[106] Jf. artikkelens første ledd.
[107] Ettersom forordningen er EØS-relevant, vil borgere i Norge og de andre EØS-statene nødvendigvis også utløse slike rettsvirkninger som nevnt i artikkelen.

lagre og utnytte alle dataene er borte. Mange og sterke krefter ønsker å ta de gigantiske datamengdene som daglig genereres i bruk. Å hindre utvanning av den rettslige beskyttelsen som dagens regler gir mot fri bruk av personopplysninger, kan derfor bli en utfordring fremover.

Big Data utfordrer særlig prinsippene om formålsbestemthet og dataminimalisering. Enkelte hevder at disse prinsippene ikke vil være mulig å håndheve i en tidsalder preget av Big Data (Tene og Polonetsky 2012 og World Economic Forum 2013), og at personvernet først og fremst må ivaretas gjennom god informasjon fra virksomhetene om hvordan personopplysninger blir behandlet. Datatilsynet mener imidlertid at et vern om disse prinsippene er viktigere enn noen gang i en tid der stadig flere opplysninger om oss samles inn. Prinsippene er vår garanti mot å bli til gjort gjenstand for profilering i stadig nye og flere sammenhenger. En utvanning av sentrale personvernprinsipper, i kombinasjon med mer utstrakt bruk av Big Data kan få uheldige konsekvenser for personvernet og for andre viktige samfunnsverdier som ytringsfrihet og vilkårene for meningsbryting.

En annen sentral utfordring er at Big Data gjør skillet mellom anonyme og ikke-anonyme opplysninger uklart og uforutsigbart. Big Data slår benet under tidligere teknikker for å anonymisere data. Ved bruk av Big Data blir anonymisering som metode for å hindre personvernulemper ved dataanalyse og profilering mindre virkningsfull.

Til tross for at bruk av Big Data reiser flere personvernutfordringer, innebærer ikke det at bruk av denne analyseformen er uaktuell innenfor dagens personvernlovgivning. Under vil vi kort trekke opp hvilke sentrale forutsetninger som bør være tilstede, samt hvilke tiltak som bør iverksettes, for at bruk av Big Data skal foregå innenfor gjeldende rett og respektere personvernet til den enkelte.

5.1 Samtykke fortsatt utgangspunktet

Å innhente gyldig samtykke fra de registrerte i forbindelse med bruk av personopplysninger til analyse- og profileringsformål, er den beste forsikringen mot å bryte personvernlovgivningen. I den nye europeiske personvernforordningen er det dessuten foreslått å innskjerpe mulighetene for å behandle personopplysninger på annet rettslig grunnlag enn samtykke.

Det har blitt hevdet at samtykke som rettslig fundament vil fungere dårlig i tidsalderen for Big Data. Begrepet samtykkeapati blir benyttet for å beskrive hvordan konstant etterspørsel etter samtykke på Internett paradoksalt nok kan føre til redusert beskyttelse for individene. Vi kan få en utvikling der virksomheter ber om svært brede samtykker fra sine kunder, og spekulere i at samtykkeerklæringene ikke vil bli lest i detalj, slik at de har "albuerom" til å benytte opplysningene til senere og andre formål.

Slik bruk av samtykke vil imidlertid ikke være lovlig. Det er ikke anledning til å samle inn all slags opplysninger og behandle disse uten hemninger, selv om man baserer seg på samtykke. De øvrige ufravikelige grunnkravene som regelverket stiller – for eksempel relevansprinsippet og formålsbegrensningsprisnippet må respekteres.[108]

[108] Prinsippene er uttrykt i henholdsvis lovens § 11 første ledd bokstav d og direktivet artikkel 6 (c), og § 11 første led bokstav c og artikkel 6 (b).

Virksomheter som ønsker å benytte innsamlede data til formål som er uforenelige med det opprinnelige formålet, må be om samtykke fra den registrerte. Hvis det ikke er mulig eller ønskelig å be om slikt samtykke, er anonymisering av dataene som ønskes sammenstilt og analysert et praktisk alternativ. Da vil ikke opplysningene lenger være å regne som personopplysninger i rettslig forstand, og behandlingen vil falle utenfor lovens virkeområde.

I enkelte tilfeller vil ikke anonymisering av dataene være praktisk gjennomførbart eller meningsfullt sett i lys av formålet med analysen. Da kan eventuelt ulike teknikker for avidentifisering av opplysningene benyttes for å begrense personvernulempene ved (gjen)bruk av opplysningene. Ved bruk av teknikker for avidentifisering vil imidlertid lovens hovedkrav for behandling av opplysningene fortsatt gjelde. Avidentifisering kan likevel virke som et kompenserende tiltak som kan få innflytelse på interesseavveiningsvurderingen som ligger innbakt i prinsippet om formålsbegrensning. Gjennom avidentifisering er det dermed en mulighet for at man kan gjøre det unødvendig å be om nytt samtykke ved gjenbruk av innsamlede opplysninger til nye formål.

5.2 Rutiner for anonymisering og avidentifisering

Den behandlingsansvarlige må på et tidlig tidspunkt beslutte om personopplysningene som skal inngå i Big Data-analysen skal anonymiseres, avidentifiseres eller om de skal være identifiserbare. Dette valget påvirker hvordan virksomheten må forholde seg til personopplysningsloven i den videre behandlingen av opplysningene. Velges anonymisering vil som nevnt den videre prosessen falle utenfor personopplysningslovens virkeområde.

Anonymisering av opplysninger blir imidlertid stadig mer utfordrende. I forbindelse med Big Data-analyse, kan en virksomhet ofte ikke med sikkerhet vite hvorvidt de anonymiserte datasettene kan la seg reidentifisere av en ukjent tredjepart. Det er derfor viktig å teste anonymiserte data i henhold til akseptabelt risikonivå. Dette bør dokumenteres, for eksempel som del av en vurdering av personvernkonsekvenser (beskrevet i kapittel 5.4.1).

Anonymisering av opplysningene som inngår i analysen bør skje så tidlig som mulig. Dersom den behandlingsansvarlige ikke sitter på god kompetanse kan det benyttes en tiltrodd tredjepart, som handler på vegne av den behandlingsansvarlige.[109]

Teknikker for å anonymisere data kan for eksempel være:

- *Aggregering.* Når data blir fremvist som totalsummer, slik at ingen data som relateres til – eller identifiserer – enkeltpersoner, blir vist frem. Lave summer blir ofte skjult ved å gjøre de "uklare", eller ved å slette de. Eksempel på aggregering er å vise frem gjennomsnittsverdier.

- *Avledede dataelementer.* Når man bruker et sett av verdier som gjenspeiler noe av kildedataene, men som skjuler de eksakte opprinnelige verdiene. Eksempler på dette er når man erstatter fødselsdato med alder eller år, og erstatter adresse med kommune.

Virksomheter som velger å avidentifisere opplysningene fremfor å anonymisere dem, må være oppmerksomme på at den videre behandlingen av opplysningene da vil falle inn under

[109] Opinion 06/2013 on Open data and public sector information (PSI) reuse.

personopplysningsloven. Teknikker for å avidentifisere data er for eksempel (Information Commissioner's Office 2012): [110]

- *Data masking*. Innebærer å sladde/slette opplagte personopplysninger (avidentifisere). Et eksempel er å slette navn og adresse fra et datasett, men beholde andre opplysninger som fødselsdato.

- *Pseudonymisering*[111]. En type avidentifisering av data hvor personidentifikasjon er endret etter en bestemt nøkkel. Data er assosiert til en bestemt person uten at personen er identifiserbar i det aktuelle datasettet.

Før man tilgjengeliggjør pseudonymiserte[112] eller avidentifiserte[113] datasett er det viktig at man viser høy grad av forsiktighet. Dersom data er detaljerte, mulig å koble til et annet datasett [114], og inneholder personopplysninger, bør tilgangen begrenses og kontrolleres nøye. Dersom data er aggregerte og det er mindre fare for at de kan kobles til et annet datasett, er det høyere sannsynlighet for at de kan tilgjengeliggjøres uten vesentlige risikoer.

Datatilsynet anbefaler å etablere et nettverk eller et organ hvor de som har behov for å anonymisere eller avidentifisere data kan diskutere utfordringer relatert til anonymisering og utveksle erfaringer. I England finnes det et slikt nettverk (The UK Anonymisation Network (UKAN)) som koordineres av universitetene i Manchester og Southampton, the Open Data Institute, og the Office for National Statistics. [115]

5.3 Innsyn i profil og algoritme

Retten til informasjon om, og innsyn i, behandlingen av egne personopplysninger er viktige personvernprinsipper. Virksomheter som benytter Big Data på en åpen og gjennomsiktig måte, vil tjene på dette i form av økt tillitt blant kunder, brukere og i samfunnet forøvrig. Virksomheter som benytter innsamlede data til profilering og prediksjonsanalyse på en lite transparent måte, risikerer å støte folk fra seg. Dette gjelder særlig dersom de bryter med kundenes forventninger til hvordan dataene blir benyttet.

For at et samtykke skal være gyldig må det være informert. Det innebærer at den enkelte skal få informasjon om hvilke data som samles inn, hvordan de behandles, til hvilke formål de vil bli brukt og om dataene eventuelt videreformidles. I Big Data-sammenheng innebærer dette også en rett til å få innsyn i sin profil. For å sikre størst mulig åpenhet knyttet til bruken av Big Data, bør det også gis

[110] Det er også et pågående arbeid i Technology Subgroup i Artikkel 29-gruppen om anonymiseringsteknikker.
[111] I denne rapporten begrenser vi oss ikke til pseudonymisering slik det er å forstå etter legaldefinisjonen i helseregisterloven, men enhver bruk av en alternativ identifikator.
[112] Etter legaldefinisjonen i helseregisterloven: helseopplysninger der identitet er kryptert eller skjult på annet vis, men likevel individualisert slik at det lar seg gjøre å følge hver person gjennom helsesystemet uten at identiteten røpes.
[113] Etter legaldefinisjonen i helseregisterloven: helseopplysninger der navn, fødselsnummer og andre personentydige kjennetegn er fjernet, slik at opplysningene ikke lenger kan knyttes til en enkeltperson, og hvor identitet bare kan tilbakeføres ved sammenstilling med de samme opplysninger som tidligere ble fjernet.
[114] Det er mulig å koble de avidentifiserte opplysningene i et datasett med opplysninger i et annet datasett ved at det er brukt for eksempel samme unike ID til den enkelte.
[115] http://www.ukanon.net/

innsyn i hvilke beslutningskriterier (algoritmer) som ligger til grunn for utvikling av profilen. Den enkelte bør også få informasjon om fra hvilke kilder de ulike personopplysningene er hentet. I henhold til den norske personvernlovgivningen har den enkelte i dag rett til å få vite regelinnholdet i algoritmer som ligger til grunn for automatiserte avgjørelser som har vesentlig betydning for den enkelte[116]. Dette for blant annet å hindre urettmessig diskriminering og at avgjørelser av betydning for vedkommende blir tatt på feil grunnlag.

Forfatteren Evgeny Morozov (2013), tar i sin bok *"To save everything, click here",* til orde for at blant annet politiets algoritmer bør gjøres til gjenstand for offentlig revisjon. Eksterne aktører bør jevnlig revidere politiets algoritmer slik at offentligheten er kjent med hvilke variabler som inngår i Big Data-analysene. Dette vil bidra til å gjøre beslutninger truffet ved hjelp av Big Data-teknologi mer transparente og dermed mer demokratiske. En slik rettighet eksisterer i den norske personvernlovgivningen i dag, men saker som avgjøres eller etterforskes i medhold av de såkalte rettspleielovene, er ikke direkte omfattet av personopplysningsloven.[117]

Åpenhet og innsyn i hvordan politiet tar i bruk Big Data, er ekstra viktig sett i lys av at terskelen for å mistenke noen for å planlegge terror senkes. Hvis premissene rundt bruken er holdt skjult og ikke er etterprøvbar, kan det i verste fall ha en nedkjølende effekt på ytringsfriheten.

Når det gjelder etterretningstjenestenes bruk av Big Data, ligger dette utenfor personopplysningslovens virkeområde. For å lempe på negative personvernkonsekvenser ved bruk av Big Data i disse tjenestene, er det imidlertid viktig at beslutningskriteriene som ligger til grunn for deres dataanalyser, samt hvilke datakilder som inngår i dem, er gjenstand for demokratisk kontroll. Vi vil særlig trekke frem viktigheten av at kontrollorganet for disse tjenestene, EOS-utvalget, besitter den nødvendige kompetansen og innsikten i hvilke konsekvenser bruk av storskala dataanalyse kan ha for personvernet og andre viktige samfunnsverdier.

5.4 "Eiendomsrett" til egne personopplysninger

Big Data forsterker den økonomiske ubalansen mellom enkeltindividet på den ene siden og de store virksomhetene på den andre. Personopplysninger har blitt en svært verdifull vare og bestanddel i utviklingen av nye tjenester. Det er industrien alene som henter ut verdien av våre personopplysninger, ikke vi som har avgitt dem.

Det finnes ulike måter å bøte på dette misforholdes på. En løsning er at virksomheter pålegges å gi den registrerte/kunden tilgang til alle dataene som virksomheten besitter i et brukervennlig, portabelt og maskinlesbart format. Dette er omtalt som dataportabilitet og vil bidra til å styrke kontrollen den enkelte har over sine personopplysninger. Det vil gjøre det lettere å skifte fra en tjenesteleverandør til en annen slik at man kan velge den tjenesten som tilbyr best vilkår, også hva personvern angår. Dataportabilitet vil hindre at kunder låses til tjenester som har uakseptable vilkår. På sikt kan et slikt pålegg bidra til å presse frem mer personvernvennlige tjenester. Forslaget til ny europeisk personvernforordning har inkludert dataportabilitet som en rettighet.

[116] Ref. personopplysningslovens § 22
[117] Dette er slått fast i personopplysningsforskriften § 1-3.

Selskap som tilbyr såkalte "data lockers", som omtalt tidligere i rapporten, er en annen måte å gi brukeren større eierskap til egne personopplysninger på. "Data lockers" gir kunden mulighet til selv å tjene på videresalg og sekundærbruk av egne personopplysninger. Fremveksten av slike løsninger viser at hensyn til brukernes personvern kan utnyttes som forretningsmodell og konkurransefortrinn.

5.5 Innebygd personvern

Innebygd personvern (Privacy by Design) innebærer at det tas hensyn til personvern i alle utviklingsfaser av et system, i rutiner og i forretningspraksisen. Standardinnstillinger bør settes mest mulig personvernvennlige, og man bør bygge personvernet inn i designet. Det er viktig å ivareta informasjonssikkerheten fra start til slutt, og særlig viktig med tanke på Big Data er det at det vises åpenhet. Mulige bruksområder for de dataene som samles inn bør kartlegges på forhånd, slik at samtykket er presist og dekkende. Alt i alt handler det om å respektere brukerens personvern (Datatilsynet, 2013).

5.5.1 Vurdering av personvernkonsekvenser (PIA)

For å ivareta tillitten til de som får sine personopplysninger samlet inn, behandlet og analysert i Big Data-sammenheng, er det viktig å være i forkant og vurdere utfordringene for personvernet så tidlig som mulig. Dette kan gjøres ved å gjennomføre en vurdering av personvernkonsekvenser (Privacy Impact Assessment – PIA). Som nevnt tidligere er det viktig å gjøre en vurdering når man skal bruke eller frigjøre anonymiserte data. Dette for å vurdere risikoen for reidentifisering. Noen faktorer som kan være til hjelp i arbeidet:

- Kartlegg hvilke andre data som er tilgjengelige, enten offentlig eller kun tilgjengelig for enkeltvirksomheter/enkeltindivider. Undersøk om de dataene som skal bli gjort tilgjengelige kan kobles til andre tilgjengelige datasett.
- Kartlegg graden av sannsynlighet for at noen vil forsøke å reidentifisere opplysningene (enkelte data er mer attraktive for interessenter enn andre).
- Kartlegg hvor sannsynlig det er at et forsøk på reidentifisering vil lykkes (dersom noen prøver), med vurdering av de foreslåtte anonymiseringsteknikkene.[118]

En vurdering av personvernkonsekvenser bør blant annet inneholde en gjennomgang av mulige rettslige grunnlag for utlevering og gjenbruk av personopplysninger, prinsippene for formålsbegrensning, proporsjonalitet og dataminimalisering, samt teknisk tilgang og sikkerhet. Når man gjennomfører en slik vurdering bør også potensielle konsekvenser for de registrerte gjennomgås nøye.[119],[120]

Datatilsynet anbefaler at aktører som skal ta i bruk Big Data-teknologi følger de syv stegene for innebygd personvern og gjør en vurdering av personvernkonsekvenser (PIA). I EU er det laget et PIA-rammeverk for RFID-applikasjoner for å hjelpe til med å avdekke personvernkonsekvenser som kan

[118] Opinion 06/2013 on Open data and public sector information (PSI) reuse.
[119] Opinion 06/2013 on Open data and public sector information (PSI) reuse.
[120] Et eksempel på en virksomhet som har konsekvensvurdert sin bruk av Big Data, er FN-prosjektet *UN Global Pulse*. Global Pulse-prosjektet benytter Big Data for å forbedre og effektivisere organisasjonens bistands- og utviklingsarbeid. Prosjektet har utviklet en personvernerklæring som på en god måte informerer om hvordan organisasjonen benytter Big Data. http://www.unglobalpulse.org/privacy-and-data-protection

følge bruk av RFID. [121] Dette rammeverket sett i lys av Tingenes Internett er også aktuelt for aktører innen Big Data. Rammeverket er laget av RFID-miljøet på oppdrag fra Artikkel 29-gruppen. Datatilsynet oppfordrer Big Data-miljøet å etablere et liknende rammeverk for bruk av Big Data-teknologi.

I Stortingsmeldingen *"Personvern – utsikter og utfordringar" (*Fornyings-, administrasjon og kyrkjedepartementet 2013) poengteres det at offentlige styresmakter bør være en pådriver for bruk av innebygd personvern. Det presiseres i meldingen at innebygd personvern bør spille en viktig rolle allerede på planleggings- og lovgivningsstadiet.

5.5.2 Vurdering av personvernkonsekvenser i forbindelse med lovarbeid

Gjenbruk av personopplysninger kan være ønskelig ut i fra et samfunnsperspektiv (Fornyings-, administrasjon og kyrkjedepartementet 2013). Potensialet som ligger i Big Data til å hente ut effektiviseringsgevinster i offentlig sektor, kan føre til at ønsket om å gjenbruke innsamlede data vil øke. Gjenbruk krever som tidligere nevnt et nytt behandlingsgrunnlag. For det offentlige vil lov ofte være best egnet som behandlingsgrunnlag. Innhenting av samtykke til ny behandling kan være krevende og i mange tilfeller også lite hensiktsmessig. Ved lovfesting av gjenbruk av personopplysninger er det viktig at lovgiveren gjør en grundig vurdering av personvernkonsekvenser i lovgivingsprosessen. Dette er stadfestet ved at personvern er tatt inn som et eget punkt i veiledningen til utredningsinstruksen.

Det er også utarbeidet en egen veileder til utredningsinstruksen om vurdering av personvernkonsekvenser (Fornyings-, administrasjons- og kirkedepartementet 2008). Datatilsynet har gjentatte ganger påpekt manglende personvernutredning før nye lov- og forskriftsforslag sendes ut på høring. I Big Data-alderen, med mulighet for å sammenstille og analysere stadig større datasett med personopplysninger, er det enda viktigere enn før at denne veilederen blir fulgt.

5.6 Kunnskapsheving og bevissthet

Kunnskap og bevissthet om personvernutfordringer knyttet til Big Data er viktig blant virksomheter som tar teknologien i bruk. Datatilsynet oppfordrer bransjeorganisasjoner til å sette utfordringene på dagsorden, og gi opplæring i hvordan de kan håndteres blant annet gjennom bruk av innebygd personvern. Kunnskap om personvern og personvernutfordringer ved Big Data-bruk bør inn på pensum ved universiteter og høgskoler som underviser i dataanalyse/data science.

Det er videre avgjørende at også kontrollmyndighetene innehar den nødvendige kunnskap og bevissthet om potensialet som ligger i Big Data. Dette er viktig for at de skal kunne fungere som effektive og slagkraftige håndhevere av regelverket som er oppstilt for å beskytte sentrale samfunnsverdier.

Forskning på samfunns- og personvernmessige konsekvenser av Big Data er dessuten av stor betydning. Big Data er foreløpig et forholdsvis nytt fenomen. Det vil være viktig å forske på hvordan tilgangen til stadig mer, og flere typer data, vil påvirke hvordan vi treffer beslutninger og organiserer samfunnet fremover.

---

[121] Opinion 09/2011 on the revised Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications

Litteraturliste

Article 29 Data Protection Working Party – uttalelser og arbeidsdokumenter:

- Working document on a common interpretation of Article 26(1) of Directive 95/46/EC (WP 114)
- Opinion 4/2007 on the concept of personal data (WP 136)
- Opinion 5/2009 on online social networking (WP 163)
- Opinion 2/2010 on online behavioural advertising (WP 171)
- Opinion 8/2010 on applicable law (WP 179)
- Opinion 9/2011 on the revised Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications (WP 180)
- Opinion 15/2011 on the definition of consent (WP 187)
- Opinion 04/2012 on Cookie Consent Exemption (WP 194)
- Opinion 05/2012 on Cloud Computing (WP 196)
- Opinion 03/2013 on purpose limitation (WP 203)
- Opinion 06/2013 on Open data and public sector information ('PSI') reuse (WP 207)

Blixrud, K. B. & Ottesen, C. A. (2010), "Personvern i finanssektoren", Gyldendal Akademisk, Oslo

Bollier, D. (2010), "The promise and perils of Big Data", The Aspen Institute, Washington DC, http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf

boyd, d. & Crawford, K. (2012), "Critical Questions for Big Data", *Information, Communication & Society 15:5, 662-679*, http://dx.doi.org/10.1080/1369118X.2012.678878

COM (2010) 609 final, "Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee og the Regions. A comprehensive approach on personal data protection in the European Union", *European Commission*

COM (2012) 11 final 2012/0011 (COD), "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 and on the free movement of such data (General Data Protection Regulation)", *European Commission*

COM (2012)0011 – C7-0025/2012 – 2012/0011(COD), "Draft report on the proposal for a regulation of the European Parliament and of the Council on the protection of individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)", *Committee on Civil Liberties, Justice and Home Affairs*, *European Parliament*

Commission Nationale de l'Information et des Libertés (2012), "Vie privée à l'horizon 2020", *Cahiers IP Innovation & Prospective N°01*, Paris

Dagens IT, (15.06.2013), "Facebook-data kan gi deg lån – eller avslag", http://www.dagensit.no/article2629493.ece, [nedlastet: 10.09.2013]

Datatilsynet (2011a), "Social Network Services and Privacy – A case study of Facebook", www.datatilsynet.no/Global/english/11_00643_5_PartI_Rapport_Facebook_2011.pdf

Datatilsynet (2011b), "Hva vet appen om deg", http://www.datatilsynet.no/Global/04_veiledere/app_rapport_DT2011.pdf

Datatilsynet (2013), ”7 steg til innebygd personvern”, http://www.datatilsynet.no/Teknologi/Innebygd-personvern/

Datatilsynet & Teknologirådet (2013), “Personvern. Tilstand og trender 2013”, http://www.datatilsynet.no/Global/04_veiledere/personvernrapport_tilstand_trender2013.pdf

Eckersley, P. (2010), “How Unique Is Your Web Browser?”, *Electronic Frontier Foundation*, https://panopticlick.eff.org/browser-uniqueness.pdf, [nedlastet: 10.09.2013]

The Economist, (02.06.2012), “Very personal finance, Marketing information offers insurers another way to analyze risk”, http://www.economist.com/node/21556263, [nedlastet: 15.08.2013]

Federal Trade Commission, (18.12.2012), “FTC to Study Data Broker Industry’s Collection and Use of Consumer Data”, http://www.ftc.gov/opa/2012/12/databrokers.shtm, [nedlastet: 10.09.2013]

Fornyings-, administrasjons- og kirkedepartementet (2008), ”Vurdering av personvernkonsekvenser. Veileder til utredningsinstruksen”, Oslo

Fornyings-, administrasjon og kyrkjedepartementet (2013), ”Personvern – utsikter og utfordringar”, St.meld. nr. 11 (2012-2013), Oslo, Fornyings-, administrasjon og kyrkjedepartementet

Forbes, (25.06.2013), ”Finally You'll Get To See The Secret Consumer Dossier They Have On You”, http://www.forbes.com/sites/adamtanner/2013/06/25/finally-youll-get-to-see-the-secret-consumer-dossier-they-have-on-you/, [nedlastet: 10.09.2013]

*The Gua*rdian, (07.06.2013), “PRISM scandal: tech giants flatly deny allowing NSA direct access to servers”, http://www.guardian.co.uk/world/2013/jun/07/prism-tech-giants-shock-nsa-data-mining, [nedlastet: 15.08.2013]

Gymrek, M. , McGuire, A. L., Golan, D., Halperin, E. og Erlich, Y. (2013), ”Identifying Personal Genomes by Surname Inference”, *Science 18 January 2013: 339 (6117), 321-324*, [DOI:10.1126/science.1229566]

Hewlett-Packard (2013), “Metropolitan Police leverage social media to engage in local community”, http://www8.hp.com/h20195/v2/GetDocument.aspx?docname=4AA4-5393EEW,[nedlastet: 10.09.2013]

HealthWorks Collective, (26.02.2013), “Big Data in Healthcare – Hype or Reality?”, http://healthworkscollective.com/shahidshah/85441/guest-article-try-not-fall-Big-data-healthcare-hype-focus-actionable-data, [nedlastet: 10.09.2013]

Hildebrandt, M. (2009), “Who is profiling who? Invisible Visibility”, i *Reinventing Data Protection?*, red: Gutwirth, S., Poullet, Y., De Hert, P., de Terwangne, C. og Nouwt, S., Springer

HSPS News, (11.10.2012), “Using cell phone data to curb the spread of malaria”, *Harvard school of public health*, http://www.hsph.harvard.edu/news/press-releases/cell-phone-data-malaria/, [nedlastet: 10.09.2013]

Hordern, V. (2013), “Consent – the silver bullet?”*, Data Protection Ireland* (DPI 6 1 (13)

IBM (2013), “Connect to millions of devices and sensors with event-driven, near-real-time communications”, *IBM MessageSight*, http://public.dhe.ibm.com/common/ssi/ecm/en/wsd14115usen/WSD14115USEN.PDF, [11.09.2013]

Information Commissioner’s Office (2012), “Anonymisation: managing data protection risk code of practice”, http://www.ico.org.uk/Global/~/media/documents/library/Data_Protection/Practical_application/anonymisation_code.ashx

Justis- og politidepartementet (1997), “Et bedre personvern – forslag til lov om behandling av personopplysninger”, NOU 1997:19, Oslo, Statens forvaltningstjeneste

Lanier, J. (2013), ”Who owns the future?”, Simon & Schuster, New York

Mayer, J. (11.10.2011), ”Tracking the Trackers: Where everybody knows your username”, *The Center for Internet and Society, Stanford Law School*, http://cyberlaw.stanford.edu/node/6740, [nedlastet: 10.09.2013]

Mayer-Schönberger, V. (2009), ”Delete: The Virtue of Forgetting in the Digital Age”, Princeton University Press

Mayer-Schönberger, V. & Cukier, K. (2013), ”Big Data. A Revolution That Will Transform How We Live, Work and Think”, John Murray, London

McKinsey Globale Institute (2011),”Big Data: The next frontier for innovation, competition, and productivity”, http://www.mckinsey.com/insights/business_technology/Big_data_the_next_frontier_for_innovation

MIT Technology Review (2013a),”Big Data Gets Personal”, *Business Report*

MIT Technology Review, (30.07.2013b),”If Facebook Can Profit from Your Data, Why Can’t You?”, <http://www.technologyreview.com/news/517356/if-facebook-can-profit-from-your-data-why-cant-you/>, [nedlastet: 10.09.2013].

Morozov, E. (2013), “To save everything click here. Technology, solutionism and the urge to fix problems that don’t exist”, Penguine Books, London

Narayanan, A. og Shmatikov, V. (2008), ”Robust De-anonymization of Large Datasets. (How to Break Anonymity of the Netflix Prize Dataset)”, http://arxiv.org/pdf/cs/0610105v2.pdf, [nedlastet: 10.09.2013]

The New York Times, (28.11.2012a), “Jeff Hawkins Develops a Brainy Big Data Company”, http://bits.blogs.nytimes.com/2012/11/28/jeff-hawkins-develops-a-brainy-Big-data-company/, [nedlastet: 10.09.2013]

The New York Times, (16.12.2012b), “You for Sale. Mapping, and Sharing, the Consumer Genome”, http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html?pagewanted=all, [10.09.2013]

The New York Times (16.12.2012c), "How Companies Learn Your Secret", http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all [nedlastet: 10.09.2013]

The New York Times (08.06.2013a), "How the U.S. Uses Technology to Mine More Data More Quickly", http://www.nytimes.com/2013/06/09/us/revelations-give-look-at-spy-agencys-wider-reach.html?_r=0 [nedlastet: 10.09.2013]

Nikulainen, T. (2013), "Big Data Revolution – What is it?", *ETLA Brief No 10.* http://pub.etla.fi/ETLA-Muistio-Brief-10.pdf

Nordbeck, P. & Lundqvist, D. S. (2012), "Data som skapas i molnet – hur långt sträcker sig personuppgiftsansvaret?", *Lov&Data nr. 110*

OECD (2013), "Exploring the Economics of Personal Data: A survey of methodologies for measuring monetary value", *OECD Digital Economy Papers*, No. 220, OECD Publishing. http://dx.doi.org/10.1787/5k486qtxldmq-en

Ot.prp. nr. 92 (1998-1999) om lov om behandling av personopplysninger

Pariser, E. (2011), "The Filter Bubble. What the Internet is Hiding from You", Penguin Books, London.

Personvernnemnda, avgjørelser:
- PVN-2004-1
- PVN-2011-10
- PVN-2012-1

Reding, V. (14.06.2013), "PRISM scandal: Vice-President Reding makes it clear the data protection rights of EU citizens are non-negotiable", *European Commission*, http://ec.europa.eu/commission_2010-2014/reding/multimedia/news/2013/06/20130612_en.htm [nedlastet: 10.09.2013].

Rubinstein, I. S. (2012), "Big Data: The End of Privacy or a New Beginning?", *Public law & legal theory research paper series, Working paper NO. 12-56*, New York University School of Law

Schartum, D. W. & Bygrave, L. A. (2006), "Utredning av behov for endringer i personopplysningsloven", Rapport 2006, Justisdepartementet og Moderniseringsdepartementet.

Tatonetti N. P.,Denny J. C., Murphy, S. N.,Fernald G. H., Krishnan, G., Castro, V., Yue, P., Tsau P. S., Kohane, I., Roden, D.M. & Altman, R. B. (2011), "Detecting Drug Interactions From Adverse-Event Reports: Interaction Between Paroxetine and Pravastatin Increases Blood Glucose Levels", *Clinical Pharmacology & Therapeutics*, 90:1, 133–142, doi:10.1038/clpt.2011.83, http://www.nature.com/clpt/journal/v90/n1/abs/clpt201183a.html

Tene, O. & Polonetsky, J. (2012), "Big Data for All: Privacy and User Control in the Age of Analytics", *Northwestern Journal of Technology and Intellectual Property*, Forthcoming, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2149364

Turow, J. (2011), "The Daily You. How the New Advertising Industry Is Defining Your Identity and Your Worth", Yale University Press, New Haven & London

VINT research report (2013), ”Creating clarity with Big Data”, http://www.sogeti.se/upload/SV/Kalendarium/Dokument/Big-data1.pdf

Whelan, A. (3.11.2012), “Big Data – The Digital Agenda for Europe and Challenges for 2012”, *The Institute of International European Affairs*, http://www.iiea.com/events/Big-data--the-digital-agenda-for-europe-and-challenges-for-2012 [nedlastet: 10.09.2013]

World Economic Forum (2013), “Unlocking the value of personal data: From collection to usage”, Prepared in collaboration with The Boston Consulting Group

Datatilsynet

Gateadresse: Tollbugata 3, Oslo

Postadresse: Pb 8177 Dep, 0034 Oslo

E-post: postkasse@datatilsynet.no

Telefon: 22 39 69 00

Faks: 22 42 23 50

www.datatilsynet.no
www.personvernbloggen.no