Definition og analyse af personoplysninger, forholdet til PSI-loven, og muligheden for anonymisering af personhenførbar information

- *Et juridisk dokument af Michael Agerholm Juhl, jurist hos ODAA. April 2015.*

Indledning

Offentliggørelse af data til videreanvendelse via ODAA, er ligesom alt andet data der offentliggøres med henblik på videreanvendelse, reguleret af PSI-loven (lov om videreanvendelse af den offentlige sektors informationer). PSI-loven er baseret på EU-direktivet 2003/98/EF af 17. november 2003 om videreanvendelse af den offentlige sektors informationer, der søger at fremme offentlige myndigheders frigørelse af data til kommerciel og ikke-kommerciel videreanvendelse.

Adgangen til at frigøre den data som de offentlige myndigheder har i deres besiddelse er ikke ubegrænset: Det følger af PSI-lovens § 4, stk. 1, at offentlige myndigheder kan stille dokumenter og datasamlinger til rådighed, medmindre anden lovgivning er til hinder herfor. Anden lovgivning skal derfor respekteres, og her kommer bl.a. persondataloven i spil. Det følger også af artikel 1 i direktiv 2003/98/EF som loven er baseret på, at der er visse undtagelser til den generelle adgang til at frigøre data når det kommer til data der er omfattet af databeskyttelseslovgivningen. Hvis nogle oplysninger udgør personoplysninger i persondatalovens forstand, så er det persondataloven der regulerer hvorledes disse oplysninger må behandles, og hvilke betingelser der skal være opfyldt førend de må behandles.

Der vil derfor i det følgende blive set nærmere på hvilken type data der konstituerer persondata, for at kunne vurdere om den pågældende data er omfattet af behandlingsreglerne i persondataloven, hvilke regler der gælder for behandling af personoplysninger, og hvordan disse regler hænger sammen med reglerne i PSI-loven. Derudover vil der blive set nærmere på muligheden for at anonymisere data såfremt dette er nødvendigt.

Personoplysninger generelt og begrebet personhenførbarhed

Personoplysninger er defineret i persondatalovens § 3 som enhver form for information om en identificeret eller identificerbar fysisk person. Ved udtrykket identificerbar person skal forstås en person, der direkte eller indirekte kan identificeres ved et eller flere elementer der er særlige for den pågældende persons fysiske, fysiologiske, psykiske, økonomiske, kulturelle eller sociale identitet.

Det ligger i begrebet personoplysninger, at der er tale om enhver oplysning der kan henføres (dvs. en personhenførbar oplysning) til én fysisk person, selv om dette forudsætter kendskab til

personnummer, registreringsnummer, eller lignende særlige identifikationer, som fx et løbenummer. Begrebet er således meget bredt defineret.
En oplysning kan derfor karakteriseres som en personoplysning, og være omfattet af loven, selvom det først er når den kombineres med andre oplysninger at den kan kan henføres til en fysisk person. Hvis en oplysning betragtes som en personoplysning er den derfor pr. definition personhenførbar.
Ved vurderingen af om en person er identificerbar ud fra en given oplysning, skal alle hjælpemidler der med rimelighed kan tænkes bragt i anvendelse for at identificere den pågældende person, tages i betragtning.

Et eksempel: Der indsamles spørgeskemaer fra en række personer. Spørgeskemaerne er alene påført et løbenummer. Løbenummeret henviser til en oversigt over navnene på de personer der har besvaret spørgeskemaerne. Spørgeskemaet vil i sig selv isoleret set ikke indeholde personoplysninger, men da oplysningerne i spørgeskemaet kan sammenkobles med oversigten over personer, vil det derfor ved behandlingen af oplysningerne være muligt at identificere de pågældende personer. Der er således tale om personoplysninger, og behandlingen vil derfor være omfattet af persondataloven.

Der findes forskellige typer af personoplysninger, og persondataloven er struktureret således, at alle oplysningstyper som ikke specifikt er opregnet i § 7, § 8, eller som personnummeret særreguleret i § 11, er omfattet af § 6, stk. 1 der altså gælder for de fleste personoplysninger.

Behandlingsbetingelser for de almindelige personoplysninger i § 6
Som nævnt gælder der, at de personoplysninger der ikke er specifikt opregnet i og dermed omfattet af persondatalovens § 7, § 8 og § 11, skal betragtes som almindelige personoplysninger, og dermed omfattet af behandlingsbetingelserne i § 6.
Behandling af almindelige personoplysninger må kun finde sted i medfør af § 6, hvis en af betingelserne i § 6, stk. 1-3 er opfyldt. Det er således kun nødvendigt at én enkelt betingelse er opfyldt førend behandling kan ske.

Der kan således behandles oplysninger efter persondatalovens § 6, hvis:

- den registrerede har givet udtrykkeligt samtykke hertil;

- behandlingen er nødvendig af hensyn til en aftale eller til gennemførelse af foranstaltninger der træffes på den registreredes anmodning
- behandlingen er nødvendig for at overholde en retlig forpligtelse for den dataansvarlige
- behandlingen er nødvendig for at beskytte den registreredes vitale interesser
- behandlingen er nødvendig af hensyn til udførelsen af en opgave i samfundets interesse
- behandlingen er nødvendig af hensyn til udførelsen af en opgave, der henhører under offentlig myndighedsudøvelse som den dataansvarlige eller tredjemand til hvem oplysningerne videregives, har fået pålagt,
- behandlingen er nødvendig for, at den dataansvarlige eller den tredjemand til hvem oplysningerne videregives kan forfølge en berettiget interesse, og hensynet til den registrerede person ikke overstiger denne interesse
- en virksomhed må videregive oplysninger om en forbruger til en anden virksomhed hvis forbrugeren har givet udtrykkeligt samtykke hertil,
    - videregivelse af oplysninger kan dog ske uden samtykke, hvis er tale om generelle kundeoplysninger

Af de ovennævnte behandlingsbetingelser, er betingelsen om samtykke klart den stærkeste behandlingsgrund. I forhold til alle datatyper og alle behandlingsformer der er omfattet § 6, men også af §§ 7-8, kan behandling af personoplysninger finde sted, såfremt den registrerede person udtrykkeligt samtykker hertil. Samtykket er grundlæggende en viljeserklæring fra den registrerede person og kan gives af denne, men kan ligeledes gives af en person der har fået fuldmagt hertil. Samtykket vil typisk gives til den dataansvarlige som vurderer om det er gyldigt, men det er i praksis accepteret at samtykket kan gives til tredjemand som herefter formidler det til den dataansvarlige.

Som nævnt ovenfor, er en oplysning først en personoplysning når den i sig selv eller i kombination med andre oplysninger kan henføres til en specifik fysisk person, og dette gælder uanset om der er tale om en almindelig personoplysning, en følsom personoplysning eller en semi-følsom personoplysning.

De følsomme personoplysninger i § 7

Oplysningerne fastsat i persondatalovens § 7 er kendt som de følsomme personoplysninger. Hvis en personoplysning falder ind under en af disse emnegrupper er den således at betragte som følsom og derfor reguleret af § 7.

Ifølge § 7, stk. 1 gælder der, at der som udgangspunkt ikke må behandles oplysninger om racemæssig eller etnisk baggrund, politisk, religiøs eller filosofisk overbevisning, fagforeningsmæssige tilhørsforhold og oplysninger om helbredsmæssige og seksuelle forhold. Der er i den juridiske litteratur bred enighed om, at en række af de nævnte oplysningstyper er sensitive og dermed tilhører privatlivsbeskyttelsens kerneområde. Oplysningstyperne er karakteriseret ved at være brede, og mange af dem indeholder både noget trivielt og noget som må betragtes som følsomt. De enkelte kategorier giver som udgangspunkt god mening, men deres afgrænsning kan undertiden give anledning til tvivl. Det giver sig selv, at oplysninger om helbredsmæssige forhold må betragtes som følsomme. Det har imidlertid været diskuteret, om den oplysning at en person er syg, uden at sygdommen er specificeret skal betragtes som en helbredsoplysning, og dermed være omfattet af § 7, stk. 1. Bestemmelsen giver i sig selv ikke noget svar herpå, men i Datatilsynets praksis antages det at være syg ikke som en helbredsoplysning, og er således ikke omfattet af persondatalovens § 7, stk. 1. Det må derfor i stedet betragtes som værende en almindelig personoplysning og er derfor omfattet af § 6, stk. 1.

Selvom opregningen af de forskellige grupper af oplysninger som udgangspunkt giver god mening, er der dog i bestemmelsen medtaget enkelte oplysningstyper, hvis sensitivitet under et dansk perspektiv ikke er ganske åbenbar, hvilket kort vil blive nærmere diskuteret. Disse oplysninger blev under den tidligere registerlovgivning ikke anset for følsomme, men at disse oplysningstyper nu er omfattet af persondatalovens § 7, stk. 1, beror på at oplysningerne var indeholdt i direktiv 95/46/EF (databeskyttelsesdirektivet), som persondataloven er baseret på, og derfor blev det besluttet også at medtage disse oplysninger i § 7 ved persondatalovens udformning.
I persondataloven, er det fx nyt i forhold til den tidligere registerlovgivning, at oplysninger om filosofisk overbevisning anses for følsomme. Der er anmeldt enkelte behandlinger til Datatilsynet, som efter den dataansvarliges skøn omfatter denne oplysningstype, men der foreligger endnu ikke nogen afgørelser som afklarer hvad der gemmer sig bag denne kategori. Der kan formentlig være

tale om oplysninger, som kan minde om en religiøs overbevisning, og således angår noget mere end det faktum at en person blot er tilhænger af en bestemt filosofisk retning. Det ville være ønskeligt med en klarere afgrænsning i persondataretten, da en kategorisering af de følsomme personoplysninger bør være forholdsvist præcis. Når afgrænsningen ikke er præcis medfører det, at den dataansvarliges behandlingsmuligheder begrænses, idet der ikke er noget specifikt at forholde sig til, og det er ikke hensigtsmæssigt.

I forhold til oplysninger om religiøs overbevisning, kan det virke naturligt at disse bør karakteriseres som følsomme. Det er dog ikke automatisk givet, at alle disse oplysninger bør opfattes således. Her i Danmark, vil en oplysning om at en person er medlem af den danske folkekirke ikke umiddelbart betragtes som følsomt, fordi dette gælder for størstedelen af befolkningen. Der kan dog være forskellige måder at betragte en oplysning om religiøst tilhørsforhold på. Oplysningen kan være følsom, enten fordi den henviser til noget indre og privat i det enkelte menneske, til tro, eller fordi kendskab til denne oplysning kan give anledning til negative reaktioner over for den enkelte. Det må lægges til grund, at det er det sidste der er udslagsgivende. Selvom tro kan være særdeles personligt, er det samtidig en oplevelse som man ofte gerne deler med andre. Under dette perspektiv er der ikke nogen grund til at skjule sådan en oplysning. Dette indebærer, at det kun er når oplysningen er om religiøs overbevisning adskiller personen fra andre personer, hvor dette kan medføre noget negativt i form af diskrimination, at den med rette kan kategoriseres som følsom. Det er således ikke selve oplysningen som er følsom, men oplysningens virkninger som er følsom.

For så vidt angår oplysninger om fagforeningsmæssigt tilhørsforhold, forekommer det heller ikke uden videre åbenbart at der er tale om en følsom oplysningstype. Denne kategori omfatter også oplysninger som ikke identificerer den fagforening en person er medlem af, og også oplysninger om at en person tidligere blot har været medlem af en fagforening. I Danmark virker det usædvanligt at betragte en sådan oplysning som følsom. At en jurist eller økonom er medlem af DJØF, eller en IT-uddannet er medlem af PROSA, virker ikke som noget der bør karakteriseres som følsomt. Men som skrevet ovenfor, så er oplysningstypen medtaget på baggrund af at den fandtes i det bagvedliggende direktiv. Dette viser at bestemmelsen har en EU-retlig baggrund, hvor det illustreres at bedømmelserne af oplysningstyperne kan variere inden for forskellige retskulturer.

Uanset ovenstående betragtninger om logikken i at visse oplysninger skal betragtes som følsomme, betragtes de uanset dette altså som følsomme oplysninger. Disse følsomme oplysninger må derfor som udgangspunkt ikke behandles, men behandling kan dog ske, hvis betingelserne i § 7, stk. 2-7 finder anvendelse i det specifikke tilfælde.

Ifølge undtagelserne i § 7, stk. 2 kan der behandles oplysninger, der er karakteriseret som følsomme i § 7 hvis:

- den registrerede har givet sit udtrykkelige samtykke til behandling
- hvis behandlingen er nødvendig for at beskytte den registreredes eller en anden persons vitale interesser, hvor den pågældende ikke er fysisk eller juridisk i stand til at give samtykke
- hvis behandlingen vedrører oplysninger som er blevet offentliggjort af den registrerede
- hvis behandlingen er nødvendig for, at et retskrav kan fastlægges, gøres gældende eller forsvares

Derudover følger det af § 7, stk. 3-7 at der er mulighed for at behandle følsomme personoplysninger omfattet af stk. 1, når der er tale om:

- behandling af fagforeningsmæssige forhold, hvis overholdelsen er nødvendig af arbejdsretlig forpligtelse eller specifikke rettigheder
- en stiftelse, forening eller anden almennyttig organisation som har politisk, filosofisk, religiøs eller faglig art kan inden for rammerne af organisationen foretage behandling af følsomme oplysninger om sine medlemmer. Der kan dog kun ske videregivelse af sådanne oplysninger, hvis den registrerede person har givet et udtrykkeligt samtykke
- at en behandling er nødvendig med henblik på bl.a. forebyggende sygdomsbekæmpelse, eller patientbehandling mv. og behandlingen foretages af en person inden for sundhedssektoren der er undergivet tavshedspligt efter lovgivningen
- at en behandling er nødvendig af hensyn til en offentlig myndigheds varetagelse af sine opgaver på det strafferetlige område
- at en behandling af oplysninger sker af grunde, der vedrører hensynet til vigtige samfundsmæssige interesser, og tilsynsmyndigheden giver tilladelse hertil

De semi-følsomme personoplysninger i § 8

Oplysningerne fastsat i persondatalovens § 8 er kendt som de semi-følsomme personoplysninger. I medfør af § 8, stk. 1 må der for den offentlige forvaltning ikke behandles oplysninger om strafbare forhold, væsentlige sociale problemer og andre rent private forhold end de i § 8, stk. 1 nævnte, medmindre det er nødvendigt for varetagelsen af myndighedens opgaver.

Oplysningerne indeholdt i § 8, stk. 1 kan klassificeres som en opsamling af oplysninger der burde have været medtaget i persondatalovens § 7. Ved udformningen af persondataloven, som er baseret på databeskyttelsesdirektivet (95/46/EF), blev oplysningerne i artikel 7 som § 7 er baseret på anset for at være udtømmende. Dvs. det var derfor ikke muligt at medtage de oplysninger som tidligere var at finde i § 9, stk. 2 i lov om offentlige myndigheders registre om strafbare forhold, væsentlige sociale problemer og andre rent private forhold i persondatalovens § 7, da disse ikke var at finde i artikel 7 databeskyttelsesdirektivet. Derfor valgte man i stedet at lave en slags ”opsamlingsbestemmelse” i § 8 for disse oplysninger og reglerne for behandling af dem. Hvis en personoplysning derfor er omfattet af denne emnegruppe, er den at betragte som ”semi-følsom” og er således reguleret af persondatalovens § 8.

Indholdet af betegnelsen ”strafbare forhold” giver sig selv: her er der tale om informationer om personer, der vedrører forhold for hvilke der har kunnet idømmes straf efter fx straffeloven.
Som eksempler på oplysninger om væsentlige sociale problemer, kan bl.a. nævnes oplysninger om langvarig arbejdsløshed, eller at en person modtager førtidspension, idet tilkendelse af førtidspension er betinget af en varigt nedsat erhvervsevne pga. fysisk eller psykisk invaliditet eller sociale forhold.
Som eksempler på andre oplysninger om rent private forhold kan nævnes oplysninger om andre foreningsmæssige forhold end fagforeningsmæssige forhold som nævnt i § 7, stk. 1. Derudover kan der være tale om oplysninger om interne familieforhold såsom familiestridigheder og opdragelsesmåde og lignende. Endvidere vil oplysninger om selvmordsforsøg, separationsansøgninger, skilsmissebegæringer mv. antages at være at betragte som private forhold.

Disse ”semi-følsomme” oplysninger, må som udgangspunkt heller ikke behandles, men ligesom i § 7 om følsomme personoplysninger, er der naturligvis også undtagelser til denne hovedregel. Disse undtagelser findes i §, 8, stk. 1-6.

Først og fremmest følger det af § 8, stk. 1, at sådanne oplysninger ikke må behandles, medmindre det er nødvendigt for myndighedens opgaver. Derudover følger det af § 8, stk. 2 at der gerne må ske *videregivelse* af disse oplysninger, hvis:

- den registrerede har givet sit udtrykkelige samtykke til videregivelse
- videregivelsen sker til varetagelse af offentlige eller private interesser der klart overstiger hensynet til de interesser der begrunder hemmeligholdelse, herunder hensynet til den som hemmeligholdelse angår
- videregivelse er nødvendig for udførelse af en myndigheds virksomhed eller påkrævet for en afgørelse som myndigheden skal træffe
- videregivelsen er nødvendig for udførelsen af en persons eller en virksomheds opgaver for det offentlige

Endelig følger det af § 8, stk. 3-6 at:

- forvaltningsmyndigheder der udfører opgaver inden for det sociale område, må kun videregive oplysninger som nævnt i både § 8, stk. 1 og § 7, stk. 1 hvis der er givet samtykke eller videregivelse er et nødvendigt led i sagens behandling eller nødvendig for at en myndighed kan føre tilsyns- eller kontrolopgaver
- private må behandle oplysninger som nævnt i § 8, stk. 1, hvis den registrerede har givet sit udtrykkelige samtykke hertil. Derudover kan der ske behandling hvis det er nødvendigt til varetagelse af en berettiget interesse, som klart overstiger hensynet til den registrerede
    - Private må dog ikke videre give disse oplysninger uden den registreres samtykke, men videregivelse kan dog ske uden samtykke når det sker til varetagelse af offentlige eller private interesser, herunder hensynet til den pågældende selv, der klart overstiger hensynet til de interesser der begrunder hemmeligholdelse.
- et fuldstændigt register over straffedomme, må kun føres for en offentlig myndighed

Personnummer som personoplysning
Reglerne for behandling af et personnummer fremgår af § 11, stk. 1 hvorefter offentlige myndigheder gerne må behandle oplysninger om personnummer med henblik på entydig identifikation eller som journalnummer.
Reglerne for hvornår private må behandle et personnummer fremgår af § 11, stk. 2, hvorefter personnummeret må behandles:

- når det følger af lov eller bestemmelser fastsat i lov,
- når den registrerede har givet samtykke hertil,
- når behandlingen finder sted til bl.a. videnskabelige eller statistiske formål eller hvis der er tale om videregivelse som et naturligt led i den normale drift af en virksomhed, og når videregivelse er af afgørende betydning for at sikre en entydig identifikation af den registrerede eller videregivelsen kræves an en offentlig myndighed
    - Uanset dette, må der ikke ske offentliggørelse af et personnummer uden udtrykkeligt samtykke.

Et personnummer er en personoplysning, men i sig selv udgør personnummeret ikke som sådan en følsom personoplysning. Personnummerets store styrke er, at det kan give en tydelig identifikation af alle borgere. De første 6 cifre i personnummeret angiver en persons fødselsdato, mens de sidste 4 cifre er løbenumre, hvoraf man af det sidste der fungerer som et kontrolciffer kan se personens køn. Informationen der udspringer af personnummeret, er altså alder og køn, og dermed personoplysninger der er at betragte som almindelige, og dermed omfattet af § 6, stk. 1.

Det er dog ikke nummeret i sig selv der giver anledning til betænkeligheder, men personnummerets anvendelse og funktion som giver anledning til databeskyttelsesretlige overvejelser. I fx Norge og Sverige er personnummeret ikke at betragte som en fortrolig oplysning, men personnummeret har en særegen status i Danmark, hvor det bliver brugt som indgangsnøgle til mange forskellige registre. I princippet kan man via personnummeret finde alle oplysninger om en person i alle de systemer, hvor personnummeret anvendes. Derudover kan andre personers personnumre anvendes til identitetstyveri, hvilket er meget udbredt i andre lande, og som ligeledes forekommer i Danmark.

Som udgangspunkt betragtes personnummeret altså ikke som en følsom personoplysning. På grund af personnumrets særegne status i Danmark som adgangsmiddel til mange tjenester, og almindelige menneskers opfattelse af at personnumret er noget meget privat, må det dog lægges til grund, at personnummeret skal behandles forsigtigt. Derfor må det på trods af behandlingsbetingelserne i persondatalovens § 11 behandles varsomt, og må derfor ikke offentliggøres til videreanvendelse i digitale tjenester.
Dette må også siges at være gældende, selvom der gives et udtrykkeligt samtykke til behandling, idet det ikke vil være klart for den registrerede, hvor langt samtykket vil række. Hvis et personnummer først bliver offentliggjort til videreanvendelse i digitale tjenester, er det ikke til at forudse i hvilket omfang personnummeret kan anvendes, og derfor er det således udelukket at anvende det i datasæt der skal offentliggøres.

Adgangen til at offentliggøre personoplysninger i forhold til PSI-loven
Som ovenfor nævnt, følger det af PSI-loven at anden lovgivning respekteres, ligesom direktivet som loven er baseret på har undtagelser til den normale adgang til at offentliggøre data, når det kommer til personoplysninger. Nærmere bestemt følger det af PSI-direktivet, og af PSI-loven at princippet om videreanvendelse ikke er automatisk, når retten til beskyttelse af personoplysninger står på spil.

Som udgangspunkt, så giver databeskyttelseslovgivningen ikke mulighed for, at offentlige instanser kan offentliggøre personoplysninger, som er indsamlet til andre normalt administrative formål. Der gælder her en formålsbegrænsning for personoplysninger, som angiver at personoplysninger kun skal benyttes til det formål de er indsamlet. Det følger af § 5, stk. 2 i persondataloven, at senere behandling af personoplysninger ikke må være uforenelig med de formål til hvilket oplysningerne oprindeligt blev indsamlet. Det vil sige, at i disse tilfælde er det pga. formålsbegrænsningen ikke muligt at videreanvende personoplysninger som led i initiativer for videreanvendelse for åbne data.

Som anført ovenfor under afsnittene om de respektive personoplysninger, følger det af persondataloven, at der er adgang til at behandle personoplysninger når betingelserne herfor er opfyldt. Men denne adgang til behandling strækker sig ikke automatisk til at kunne offentliggøre de pågældende personoplysninger til videreanvendelse i digitale tjenester i medfør af PSI-loven.

Denne normale adgang der er til behandling vil nemlig oftest vedrøre en intern behandling, som vil være foreneligt med det formål som personoplysningerne er indsamlet til. I forhold til videreanvendelse af åbne data skal man for det første huske på, at det ikke er en selvfølgelighed at personoplysninger udstillet på en offentlig platform vil blive brugt til de formål de er indsamlet til. Derudover skal man huske på, at åbne data er ensbetydende med at man giver adgang til data til en meget stor kreds af personer, og det er ikke hensigtsmæssigt at de får adgang til personoplysninger som kan henføres til specifikke personer, netop fordi det er ikke på forhånd er givet hvilket formål oplysningerne bliver frigivet til, og ej heller til hvilken kreds af personer. Selv et udtrykkeligt samtykke fra en registreret person, kan vise sig ikke at være tilstrækkeligt, idet den samtykkende person ikke på forhånd kan vide hvad der reelt samtykkes til. Dermed kan et samtykke komme til at stille den registrerede i en situation denne slet ikke havde forventet ved afgivelse af samtykket.
Derudover gælder, at hvis personoplysninger bliver behandlet i strid med deres oprindelige formål, eller bliver behandlet i strid med deres behandlingsbetingelser, så kan dette få konsekvenser for både den dataejer som i første gang udstiller personoplysningerne, da dette udgør en behandling. Ligeledes kan det få konsekvenser for dem som efterfølgende videreanvender og derved behandler de pågældende oplysninger. Såfremt den kommende revidering af databeskyttelsesdirektivet som ventet kommer til at indføre yderst høje bødetakster for persondatabehandling i strid med dets bestemmelser, kan det hurtigt blive en meget dyr affære for både dataejere såvel som senere videreanvendere.

Som udgangspunkt må det derfor lægges til grund, at de datasæt som bliver gjort offentligt tilgængelige til videre brug i digitale tjenester ikke må indeholde personoplysninger, og datasæt indeholdende personoplysninger skal derfor renses for disse før de udstilles. Alternativt kan datasæt indeholdende personoplysninger jf. nedenfor anonymiseres til en sådan grad, at de ikke længere kan siges at indeholde personoplysninger, og dermed ikke længere være omfattet af databeskyttelseslovgivningen.
For så vidt angår personoplysninger som allerede er gjort offentligt tilgængelige, kan det følge af lovbestemmelser at disse personoplysninger ikke må videreanvendes. Fx følger det af lov om bygnings- og boligregistrering (BBR) at oplysninger om ejeres økonomiske forhold ikke må

videreanvendes, selvom disse oplysninger allerede kan findes ved en søgning på nettet. Dette bør således haves for øje når man påtænker at offentliggøre data.
Modsat kan det også ved lov være besluttet, at visse personoplysninger godt kan offentliggøres til videreanvendelse i digitale tjenester.

Anonymisering af personoplysninger

Såfremt man ønsker at offentliggøre et datasæt som indeholder personoplysninger, kan dette lade sig gøre hvis datasættet gennemgår en anonymiseringsproces, således at der ikke længere er genkendelige personoplysninger i datasættet. I det følgende vil anonymisering blive kort berørt, herunder hvad man skal være opmærksom på og hvilke teknikker man bl.a. kan anvende. Der er dog kun tale om en overfladisk gennemgang, så hvis man ønsker en mere dybdegående gennemgang af hvad anonymisering er og hvordan man gør, er det nødvendigt at kigge i den øvrige litteratur om anonymisering, og de tekniske vejledninger der findes hertil

Hvis et datasæt med personoplysninger er korrekt anonymiseret, og enkeltpersoner ikke længere er identificerbare, har oplysningerne ikke længere karakter af personoplysninger, og den europæiske databeskyttelseslovgivning finder ikke længere anvendelse. I Danmark vil det sige, at oplysningerne ikke længere er omfattet af persondataloven. Da oplysningerne ikke længere er omfattet af persondataloven, er loven ikke til hinder for offentliggørelse af de pågældende data, og datasættet kan således som udgangspunkt udstilles til videreanvendelse i digitale tjenester.
Det skal dog haves for øje, at selvom oplysningerne ikke længere har karakter af personoplysninger, kan der være en risiko for at et datasæt som er anonymiseret, kan kombineres med at andet datasæt på en sådan måde at de oprindelige personoplysninger kan genskabes. I det følgende ses der nærmere på med hvilke metoder personoplysninger kan anonymiseres, og hvilke ting man bør være opmærksom på i forhold til anonymisering og genkendelse af de oprindelige oplysninger.

Nærmere definition af anonymisering

Anonymisering er omtalt i betragtning 26 i direktiv 95/46/EF, som persondataloven er baseret på. Heri står der, at "beskyttelsesprincipperne gælder ikke oplysninger, som er gjort anonyme på en sådan måde, at den registrerede ikke længere kan identificeres". For at oplysninger er anonymiseret ifølge betragtning 26, skal der fjernes tilstrækkelige elementer fra dem, så den

registrerede person ikke længere kan identificeres. Det vil sige, at oplysningerne skal være behandlet på en sådan måde, at "de ikke længere kan bruges til at identificere en fysisk person, ved hjælp af alle de hjælpemidler der med rimelighed kan tænkes bragt i anvendelse, for at identificere den pågældende person, af den registeransvarlige eller af tredjemand". Dette betyder, at behandlingen af oplysninger til "anonymiseret form" skal være uigenkaldelig, og resultatet af anonymisering som en teknik der anvendes på personoplysninger, bør ifølge direktivet være ligeså permanent som sletning af de pågældende personoplysninger. Der sættes dermed en meget høj standard i direktivet, for hvor uigenkaldelig en anonymisering skal være, førend at den kan siges at være fyldestgørende.

Lovligheden ved anonymisering

Når personoplysninger anonymiseres, er udgangspunktet at personoplysningerne skal være indsamlet og behandlet i overensstemmelse med persondataloven. Ifølge persondatalovens § 5, stk. 2 skal indsamlingen ske "til udtrykkeligt angivne og saglige formål, og senere behandling må ikke være uforenelig med disse formål". Selve anonymiseringsprocessen af de pågældende personoplysninger, dvs. behandlingen af personoplysningerne for at anonymisere dem må betragtes som "senere behandling", og denne behandling skal derfor opfylde kravet om forenelighed, dvs. anonymiseringen må ikke være uforenelig med indsamlingen af oplysninger der er sket til udtrykkeligt angivne og saglige formål.

Det er givet, at anonymisering som et led i den senere behandling af personoplysninger kan anses for foreneligt med behandlingens oprindelige formål, såfremt anonymiseringsprocessen med sikkerhed frembringer oplysninger jf. ovenstående definition, hvorfra det ikke det er muligt at genskabe de oprindelige personoplysninger. Det er således ikke i strid med persondatalovens § 5, stk. 2 om forenelighed, at anonymisere personoplysninger ved en senere databehandling.

Muligheden for at identificere personer i anonymiserede data

Anonymiserede oplysninger er som nævnt anonyme oplysninger som tidligere henviste til en identificerbar person, men hvor identifikation ikke længere er mulig. Aggregering og anonymisering skal ske så tidligt som muligt og udføres af den registeransvarlige, eller af en tredjemand som handler på den registeransvarliges vegne, og som besidder de nødvendige specialiserede færdigheder. Det må ikke overlades til en videreanvender, at udføre anonymiseringen – dvs. man må ikke udstille personoplysninger, og så stille det som en betingelse

i en licens at datasættet skal anonymiseres førend at det må videreanvendes. Det skal være anonymiseret, inden det udstilles på en åben platform.

Ved anonymisering af oplysninger, må de registeransvarlige tage stilling til muligheden og risikoen for identificering ved deres valg af anonymiseringsmetode. I den henseende er der nogle vigtige faktorer der skal tages i betragtning. Som nævnt ovenfor under definition af anonymisering, er oplysninger anonymiseret når de er behandlet på en sådan måde, at de ikke længere kan bruges til at identificere en person, ved hjælp af alle de hjælpemidler der med rimelighed kan tænkes bragt i anvendelse.

Først og fremmest bør de registeransvarlige fokusere på de konkrete hjælpemidler der skal bruges til at fjerne anonymiseringen, herunder de udgifter og den viden der er nødvendigt for at bringe disse hjælpemidler i anvendelse, og vurdere sandsynligheden herfor. De registeransvarlige bør afveje anonymiseringsindsatsen og omkostningerne i forhold til at der kommer flere og flere billige tekniske hjælpemidler til identificering til rådighed, og den stadig større adgang til andre datasæt på baggrund af at offentligheden åbner op for data. Endvidere kan identificeringsrisikoen stige med tiden pga. udviklingen inden for IT og hurtigere og billigere computerkraft.
For det andet er de hjælpemidler der tages i anvendelse, hjælpemidler der kan anvendes af "den registeransvarlige eller af enhver anden person", dvs. den registeransvarliges egne hjælpemidler skal også tages i betragtning. Det betyder, at hvis en registeransvarlig ikke sletter originale identificerbare oplysninger i et datasæt og den registeransvarlige udleverer en del af datasættet, er det resulterende datasæt stadig personoplysninger og beskyttet af databeskyttelseslovgivningen da der kan ske identifikation af personoplysninger ved sammenkobling med de stadigt eksisterende ikke-anonymiserede personoplysninger. Det er kun hvis den registeransvarlige anonymiserer dataene til et niveau, hvor enkelte hændelser eller personer ikke længere kan identificeres, at det resulterende datasæt kan betragtes som anonymt. Hvis dette gøres effektivt, kan ingen parter udskille bestemte personer, sammenkoble flere poster i et datasæt eller mellem forskellige datasæt eller udlede nogle identificerende oplysninger.

Af de anonymiseringsteknikker der bliver gennemgået nedenfor, randomisering og generalisering, er der ulemper ved begge, men de kan også begge afhængigt af omstændighederne og konteksten

være egnede til at opfylde det ønskede formål uden at krænke registrerede personers privatliv. I forhold til krænkelse af privatliv og identifikation, skal det bemærkes at identificering ikke kun omhandler muligheden for at finde en specifik persons navn eller adresse, men også den potentielle identificerbarhed ved at udskille et bestemt individ, sammenkoble informationer med andre informationer og udlede oplysninger.

Databeskyttelseslovgivningen, dvs. persondataloven finder anvendelse uanset den registeransvarliges eller modtagerens hensigter. En tredjemand der behandler et datasæt, som er anonymiseret af den oprindelige registeransvarlige kan gøre dette lovligt uden at skulle tage højde for databeskyttelseslovgivningen, da datasættet pga. anonymiseringen ikke er omfattet af persondataloven, da der ikke længere er tale om personoplysninger. Men tredjemanden skal dog tage hensyn til konteksten og omstændighederne ved de anonymiseringsteknikker den registeransvarlige har anvendt, og hvordan sådanne anonymiserede data skal anvendes. Tredjemanden skal i forbindelse med behandlingen af datasættet og videreanvendelsen af dette, tage stilling til muligheden for identificering og skal ligesom den registeransvarlige tage stilling til de hjælpemidler der kan anvendes til identificering. Hvis der er en uacceptabel risiko for identificering ved tredjemands anvendelse af datasættet, vil datasættet ikke kunne anses for anonymiseret og behandlingen vil derfor igen falde ind under databeskyttelseslovgivningens område.

Anonymiseringsteknikker

I forhold til anonymisering af personoplysninger, findes der forskellige metoder som er mere eller mindre robuste. Ved anvendelsen af anonymiseringsteknikker, skal de registeransvarlige tage højde for tre risici der er forbundet med anonymisering: Udskilning, som er muligheden for at udskille nogle eller alle de poster i et datasæt som identificerer en bestemt person. Sammenkobling, som er muligheden for at koble mindst to poster vedrørende den samme registrerede eller en gruppe registrerede sammen. Udledning, som er muligheden for med stor sandsynlighed at udlede værdien af en attribut i et datasæt fra værdierne af et sæt andre attributter.

En løsning som kan beskytte mod disse tre risici vil være robust over for genidentificering ved hjælp af de mest sandsynlige og rimelige hjælpemidler som den registeransvarlige og en tredjemand kan anvende. Det må dog understreges, at forskning har vist at ingen teknikker som

sådan er uden ulemper. Der er som udgangspunkt to overordnede fremgangsmåder for anonymisering, som er baseret på henholdsvis randomisering og på generalisering, og under disse er der nogle øvrige teknikker som vil blive berørt i det følgende.

Randomisering

Anonymiseringsmetoden randomisering, består af en gruppe forskellige teknikker som ændrer nøjagtigheden af dataene i et datasæt, så forbindelsen mellem data og en person fjernes. Hvis dataene i et datasæt er unøjagtige nok, kan de ikke længere henvise til en bestemt person. Randomisering kan beskytte mod udledningsangreb og kan kombineres med generaliseringsteknikker for at øge beskyttelsen personoplysninger, men randomisering reducerer ikke i sig selv de enkelte posters individualitet. De forskellige teknikker af randomisering vil i det følgende blive kort beskrevet:

Tilførsel af støj

Tilførsel af støj anvendes særligt, når informationer i et datasæt kan få store konsekvenser for specifikke personer. Teknikken består i, at ændre oplysninger i et datasæt så de ikke er helt nøjagtige, mens den overordnede fordeling opretholdes. En behandler af et datasæt vil antage, at værdierne i datasættet er nøjagtige, men dette vil kun være tilfældet til en vis grad. Hvis tilførsel af støj bruges effektivt, vil en specifik person ikke kunne identificeres, og en tredjemand vil ikke kunne reparere dataene eller finde ud af hvordan dataene er ændret.

Det er stadig muligt at udskille poster relateret til en person, ligesom det er muligt at sammenkoble poster relateret til den samme person, mens udledningsangreb også er mulige men der vil være mindre chancer for at disse lykkes. For at mindske risikoen for udskilning, sammenkobling og udledning, skal tilførsel af støj normalt kombineres med andre anonymiseringsteknikker, såsom fjernelse af åbenlyse værdier og identifikatorer.

Permutation

Permutation består i, at værdierne i et datasæt ombyttes, så nogle af værdierne sammenkobles kunstigt med forskellige registrerede personer. Dette er gavnligt hvis det er vigtigt at bevare den nøjagtige fordeling af enkelte værdier i et datasæt. Denne form for randomisering kan betragtes som en særlig afart af tilførsel af støj. Da det til tider kan være vanskeligt at generere konsekvent

støj, og det muligvis ikke er nok at ændre værdierne en smule i et datasæt for at beskytte privatlivets fred, kan permutation være et alternativ som ændrer værdierne i et datasæt ved at ombytte dem fra en post til en anden. Ombytningen vil betyde, at værdiernes interval og fordeling er den samme, men det vil sammenhængen mellem værdierne og personerne i datasættet ikke være.

Ligesom ved tilførsel af støj, er der ikke garanti for at permutation sikrer anonymisering, og permutation bør derfor altid kombineres med fjernelse af åbenlyse værdier og identifikatorer.

Generalisering

Den anden gruppe af anonymiseringsteknikker er generalisering, som går ud på at "generalisere", eller udvande de registrerede personers værdier, ved at ændre på målestoksforholdet eller størrelsen. Dvs. man kan anvende en region i stedet for en by, eller en måned i stedet for en uge. Denne metode kan være effektiv til at undgå at udskille specifikke personer, men anonymiseringen fungerer ikke altid, og det kræver specifikke fremgangsmåder for undgå sammenkobling og udledning

Aggregering og k-anonymitet

Formålet med denne teknik er at forhindre, at en registreret person bliver udskilt, ved at gruppere personen med mindst *k* andre personer. For at sikre, at en person ikke bliver udskilt bliver værdierne i et datasæt generaliseret så meget, at hver enkelt person har samme værdi. Når man sænker klasseinddelingen af personer fra f.eks. en by til et land bliver et større antal registrerede omfattet af datasættet. Fødselsdatoer kan generaliseres til intervaller eller måneder, og andre værdier som fx indtægt, vægt, højde og lignende kan også inddeles i intervaller. Dette bør gøres, når sammenhængen mellem forskellige specifikke værdier i et datasæt kan medføre til direkte identifikation.

Ved anvendelse af denne anonymiseringsmetode bør det ikke længere være muligt at udskille en person i en gruppe af k-brugere, da generaliseringen har medført at personerne nu har samme værdi i datasættet. Der er dog stadig mulighed for sammenkobling, og den største ulempe ved denne metode er at den ikke forhindrer udledningsangreb

L-diversitet/t-nærhed

L-diversitet bygger oven på teknikken k-anonymitet for at sikre imod udledningsangreb. Dette sker ved at sørge for, at hver enkelt attribut i et datasæt har mindst *l* forskellige værdier i hver klasse.

Det vigtigste formål er at sikre imod, at en angriber med baggrundsviden om en bestemt registreret, altid vil have en betydelig usikkerhed når denne forsøger at udlede noget fra de informationer der er at finde i datasættet. L-diversitet er en god metode til at beskytte data mod udledningsangreb, når værdierne i datasættet er godt fordelt.
T-nærhed er en videreudvikling af l-diversitet, hvis formål er at frembringe klasser som ligner den oprindelige fordeling af værdier i datasættet. Til det formål sættes en yderligere begrænsning på klasserne, idet der ikke kun skal være *l* forskellige værdier i hver klasse, men hver enkelt værdi skal også være repræsenteret så mange gange, som det er nødvendigt for at spejle den oprindelige fordeling af hver enkelt værdi.
Denne anonymiseringsteknik beskytter ligesom k-anonymitet mod udskilning, og den største forbedring i forhold til k-anonymitet er, at det ikke er muligt længere at anrette udledningsangreb med 100 % sikkerhed. Der er dog stadig en risiko for sammenkobling.

Pseudonymisering
Pseudonymisering fungerer normalt ved, at man erstatter én værdi i en post med en anden. En fysisk person kan derfor blive indirekte identificeret, så hvis pseudonymisering anvendes alene, giver det ikke et anonymt datasæt. Pseudonymisering vil derfor som udgangspunkt gøre det vanskeligere at knytte et datasæt til en registreret, men hvis pseudonymisering anvendes alene vil det ikke give et anonymt datasæt. Det kan således være en nyttig sikkerhedsforanstaltning, men bør ikke betragtes som en metode til anonymisering.

Pseudonymisering er medtaget i denne fremstilling, for at gøre opmærksom på at det ikke er en effektiv metode til anonymisering. Registeransvarlige har ofte den antagelse, at det er tilstrækkeligt at fjerne eller erstatte flere attributter i et datasæt for at gøre dette anonymt, men der eksisterer mange eksempler på at dette ikke er tilfældet. Det vil ikke forhindre en tredjemand i at identificere en registreret person, hvis der alene er tale om at ændre id'et, hvis der stadig er andre identifikatorer i datasættet, eller hvis andre værdier kan identificere person. Det er vigtigt at huske på, at en angriber allerede kan have noget information på forhånd, der kan hjælpe med at identificere en registreret person. Hvis der fx alene er tale om at et navn er ændret til en talværdi, mens andre attributter er kendt af angriberen, så vil denne kunne identificere den registrerede.
I flere situationer kan det være lige så nemt at identificere en registreret person i et pseudonymiseret datasæt som i de oprindelige originale data. Det er derfor nødvendigt at der

gøres en større indsats for at anonymisere et datasæt, ved enten at generalisere, randomisere, eller slette de originale data. Anvendelse af pseudonymisering alene er således ikke tilstrækkeligt.

Opsamling på anonymisering

Ovenfor er forskellige anonymiseringsteknikker kort beskrevet, samt hvilke fordele og ulemper der er ved de respektive teknikker. Som det er beskrevet, kan et anonymiseret datasæt stadig udsætte registrerede personer en risiko for at blive identificeret. Anonymiseringsteknikker kan sikre beskyttelsen af privatlivets fred for den enkelte registrerede person, men kun hvis anonymiseringen anvendes korrekt. Det betyder, at kravene til konteksten og formålet med anonymiseringsprocessen skal være klart beskrevet for at opnå et ønsket anonymiseringsniveau. Ved anvendelse af anonymiseringsteknikker, skal de registeransvarlige overveje de begrænsninger nogle af teknikkerne har, samt tage hensyn til de formål anonymiseringen skal forfølge når et datasæt offentliggøres.

Ingen af de teknikker der er beskrevet ovenfor, opfylder i sig selv kriterierne for en effektiv anonymisering hvor der ikke kan ske udskillelse, sammenkobling eller udledning. Nogle af de beskrevne risici kan fjernes helt eller delvist ved brug af en given teknik, så man skal være omhyggelig med at få afklaret hvordan en teknik anvendes i den pågældende situation, samt hvordan en kombination af disse kan bruges som metode til at sikre datasættet imod angreb.

Det er derfor op til de registeransvarlige at danne sig et overblik over til hvilket formål dataene skal anvendes, og i den kontekst få afklaret hvilke eller hvilke anonymiseringsteknikker der er optimale i den givne situation, i forhold til risikoen for af-anonymisering af de anonymiserede data, og i forhold til hvilke hjælpemidler der kan bruges til af af-anonymisere og dermed genkende personoplysninger. I den sammenhæng er det vigtigt at gøre opmærksom på, at registeransvarlige ikke bør offentliggøre data uden efterfølgende at overvåge dem. På grund af risikoen for identifikation bør de registeransvarlige regelmæssigt finde nye risici og evaluere risiciene igen, vurdere om kontrollen med de konstaterede risici er tilstrækkelig, og overvåge og kontrollere risiciene. Dette er en god måde at sikre imod, at et datasæt bliver af-anonymiseret pga. risici der ikke var taget højde for ved den oprindelige anonymiseringsproces.

Kilder

- Blume, Peter – Databeskyttelsesret 4. udgave (2013)
- Blume, Peter, Behandling af persondata – en kritisk kommentar (2003)
- Blume, Peter – Persondataretten i en brydningstid (2014)
- Korfits Nielsen, Kristian & Waaben, Henrik – Lov om behandling af personoplysninger – kommenteret (2008)
- Artikel 29-gruppens udtalelser:
    - Udtalelse nr. 4/2007 om begrebet personoplysninger
    - Udtalelse nr. 6/2013 om videreanvendelse af åbne data og PSI
    - Opinion 03/13 on purpose limitation – findes kun på engelsk
    - Udtalelse nr. 5/2014 om anonymiseringsteknikker