Beskæftigelses-
ministeriet

Bilag til datastrategi: Databeskyttelse – anonymisering af personoplysninger til statistik

December 2020

Forord

Beskæftigelsesministeriet vil være i front med at bruge data og digitale muligheder, mens den enkeltes persondata samtidig beskyttes.

I Beskæftigelsesministeriet indsamles en lang række informationer om både borgere og virksomheder. Det sker eksempelvis i forbindelse med beskæftigelsesindsatsen eller via tilsynsindsatsen.

Det er en ambition, at Beskæftigelsesministeriet arbejder struktureret og ambitiøst efter at gøre data så åbent så muligt og samtidig så lukket som nødvendigt, så data i videst mulig omfang kan komme samfundet til gavn.

Når data udstilles åbent, er der store krav til, at den enkeltes ret til privatlivets fred er sikret. Åbne data skal være anonymiseret i et omfang, der medfører, at ingen kan genkendes ud fra statistikken. Beskæftigelsesministeriet vil have særlig opmærksomhed på den identifikationsrisiko, der er ved kombination med andre statistikker.

Derfor har Beskæftigelsesministeriet udarbejdet en række principper, som skal følges, når ministeriet åbner data op. Principperne skal understøtte, at ingen enkeltpersoner kan genkendes. Samtidig skal principperne sikre en ensartet tilgang på tværs af ministeriet og bidrage til en velunderbygget vurdering af, at Beskæftigelsesministeriets åbne data er tilstrækkeligt anonymiseret.

Rammen for principperne

Principper skal give regler og retningslinjer på tværs af Beskæftigelsesministeriets koncern for anonymiseringen af oplysninger om personer og om virksomheder. Principperne for anonymisering anvendes ved udstilling af åbne data fra Beskæftigelsesministeriet herunder fx spørgsmål fra Folketinget med respekt for øvrig lovgivning.

Beskæftigelsesministeriets principper afspejler regler for anvendelse af persondata, der er fastsat i EU's databeskyttelsesforordning og i den danske databeskyttelseslov.

Principper for anonymisering

Beskæftigelsesministeriet behandler mange oplysninger om personer og virksomheder i Danmark i forbindelse med varetagelse af beskæftigelsesindsatsen, forsk-

ning i arbejdsmiljø og tilsynet med arbejdsmiljø på arbejdspladser i Danmark.

For at sikre legitimiteten omkring myndighedsvaretagelsen, er det vigtigt at beskytte oplysninger om borgere og virksomheder i opgørelser, som Beskæftigelsesministeriet udstiller.

Beskæftigelsesministeriet har derfor opstillet følgende principper for anonymisering af data:

- Beskæftigelsesministeriet tilstræber, at data anonymiseres, således at det sikres, at ingen personer eller medarbejdere i virksomheder kan identificeres ud fra den konkrete opgørelse eller i kombination med andre oplysninger. Beskæftigelsesministeriet er særligt opmærksom på, at statistikbank-løsninger udgør en særlig udfordring i forhold til at sikre tilstrækkelig anonymisering, da brugeren har mulighed for at sammensætte data på et utal af måder i en statistikbank.
- Beskæftigelsesministeriet tilstræber en effektiv anonymisering af data gennem automatiserede processer, for på den måde at sikre en hurtig og løbende opdatering af data på ministeriets hjemmesider samt minimere risiko for fejl.
- Beskæftigelsesministeriet benytter primær diskretionering samt aggregering til at anonymisere data. Dertil tilstræber Beskæftigelsesministeriet at benytte sekundær diskretionering og afsøger løbende nye metoder til anonymisering af data.
- Beskæftigelsesministeriet vurderer ethvert datasæt forud for offentliggørelse eller udlevering i forhold til datatypens følsomhed, sandsynligheden for identifikatiom og konsekvensen af identifikation.
- Beskæftigelsesministeriet benytter minimumsregler for diskretionering, der afhænger af datatypens følsomhed. Der skal ved det enkelte datasæt foretages en konkret vurdering af datatypens følsomhed, sandsynligheden for identifikation samt konsekvensen af identifikation og på den baggrund træffes beslutning for diskretionering.

I nedenstående gennemgås væsentlige definitioner. I næste afsnit præsenteres beskæftigelsesministeriets minimumsregler for diskretionering.

Figur 1: Væsentlige definitioner relateret til anonymisering

Anonymiseret data

- Data er anonymiseret, når det ikke er muligt at henføre en [personlig] oplysning til en identificerbar eller identificeret fysisk person [eller virksomhed].
- Oplysninger, der er gjort anonyme, sådan at ingen fysiske personer kan identificeres ud fra oplysningerne eller i kombination med andre oplysninger, er ikke omfattet af databeskyttelsesreglerne. Det er i den forbindelse en betingelse, at anonymiseringen er uigenkaldelig.

Aggregering

- Aggregering er et redskab til et opnå anonymisering og er en proces, hvor oplysninger enten grupperes, summeres eller kombineres.
- Aggregering er med til at forhindre, at en fysisk person kan udskilles fra en mængde af personer og dermed blive identificeret.

Diskretionering

- Diskretionering er en metode til opnå anonymisering, hvor bestemte oplyseninger udelades, sløres eller fjernes.
- Diskretionering skal hindre, at en fysisk person kan identificeres.

Primær diskretionering

- En proces, hvor oplysninger i en celle/gruppe automatisk fjernes, fordi et antalskriterium ikke er opfyldt.

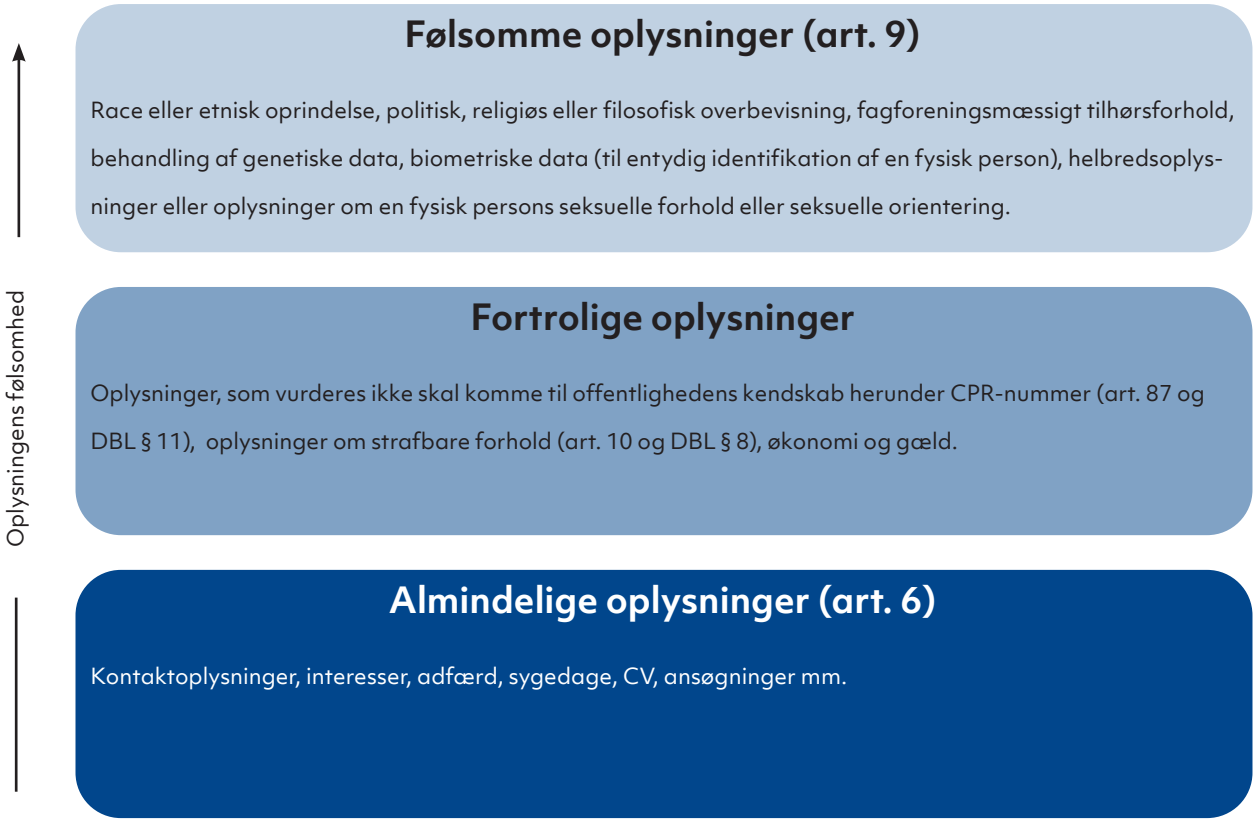
Sekundær diskretionering

- En proces, hvor der tages stilling til, om yderligere celler skal diskretioneres, så det ikke er muligt at rekonstruere/udlede de oplysninger, der blev fjernet ved den primære diskretionering.

Kilde: Beskæftigelsesministeriet

Beskæftigelsesministeriet opererer endvidere med tre datatyper, som illustreret i nedenstående figur 2 og uddybet yderligere i det efterfølgende:

Figur 2: Tre datatyper



Kilde: Beskæftigelsesministeriet

Almindelige personoplysninger

Almindelige personoplysninger omfatter alle oplysninger, der ikke er karakteriseret som særlige kategorier af oplysninger, såsom følsomme personoplysninger.

Virksomhedsoplysninger behandles som personoplysninger, da personligt ejede virksomheder sidestilles med personer, og Beskæftigelsesministeriet på nuværende tidspunkt ikke har mulighed for at adskille disse fra øvrige virksomheder.

Følsomme personoplysninger

Følsomme personoplysninger er udtrykkelig afgrænset i databeskyttelsesforordningen til følgende: Race og etnisk oprindelse, politisk overbevisning, religiøs eller filosofisk overbevisning, fagforeningsmæssige tilhørsforhold, genetiske data, biometriske data med henblik på entydig identifikation, helbredsoplysninger, seksuelle forhold eller seksuel orientering.

Fortrolige personoplysninger

Fortrolige oplysninger er en særlig kategori af oplysninger, der ikke nævnes udtrykkeligt i databeskyttelsesreglerne, men hvor oplysningerne vurderes af en karakter, som ikke skal komme til offentlighedens kendskab, jf. forvaltningslovens § 27. Fortrolige oplysninger omfatter fx CPR-nummer, strafsbare forhold samt efter omstændighederne oplysninger om indtægts- og formueforhold, arbejds-, uddannelses- og ansættelsesmæssige forhold.

Minimumsregler for diskretionering

Beskæftigelsesministeriet diskretionerer altid ud fra et forsigtighedsprincip. Ministeriet har derfor som koncern defineret minimumsregler for diskretionering af forskellige datatyper, afhængig datatypens følsomhed. Disse regler skal anvendes i forbindelse med udstilling og udlevering af data som åbne data.

Minimumsreglerne gælder for alle niveauer og kombinationer af det udstillede data. Reglerne skal således være opfyldt, uanset om data vises på sit mest detaljeret niveau eller aggregerede niveau og uanset hvilke og hvor mange fordelingsvariable, der er valgt.

Beskæftigelsesministeriets minimumsregler for primær diskretionering fremgår af nedenstående tabel 1:

Diskretioneringsregel	Almindelige personoplysninger	Fortrolige personoplysninger	Følsomme personoplysninger
Antal observationer i en celle	Mindst 3 observationer i en celle. Reglen vedrører oplysninger opgjort i ant al uanset hvordan oplysningen er opgjort.	Mindst 3 observationer i en celle. Reglen vedrører oplysninger opgjort i antal uanset hvordan oplysningen er opgjort.	Mindst 5 observationer i en celle. Reglen vedrører oplysninger opgjort i antal uanset hvordan oplysningen er opgjort.
Beregnede værdier i en celle	Beregnede værdier skal som minimum være baseret på 3 observationer.	Beregnede værdier skal som minimum være baseret på 3 observationer.	Beregnede værdier skal som minimum være baseret på 5 observationer.
Cellens population	Oplysningen skal som minimum være baseret på en population bestående af 10 observationer.	Oplysningen skal som minimum være baseret på en population bestående af 25 observationer.	Oplysningen skal som minimum være baseret på en population bestående af 50 observationer.

De enkelte enheder kan hæve grænsen for diskretionering, hvis det vurderes, at konsekvensen af at blive identificeret er høj. Fx kan grænsen for følsomme personoplysninger hæves til mindst 10 observationer i population på mindst 100.

Både før og efter data diskretioneres, bør der tages stilling til aggregeringsniveauet. Et højere aggregeringsniveau gør det mindre sandsynligt, at der er behov for diskretionering.

Dertil tilstræbes en sekundær diskretionering, der i de fleste tilfælde er en manual proces. I denne proces tager man stilling til, om yderligere celler skal diskretioneres, så det ikke er muligt at rekonstruere/udlede de oplysninger, der blev fjernet ved den primære diskretionering, eller identificere enkeltpersoner eller virksomheder.

I det efterfølgende præsenteres nogle forskellige eksempler.

Eksempel 1

I Arbejdstilsynet diskretioneres anmeldte erhvervssygdomme , hvor diskretioneringsgrænserne er fastsat til mindst 100 ansatte på en p-enhed og mindst 10 anmeldelser i en tabelcelle. For opgørelse af arbejdsulykker sker der diskretionering ud fra minimumsgrænsen på 5 observationer ud af en population på mindst 50.

Eksempel 2

Arbejdstilsynet opgør tilsynsbesøg og tilsynsreaktioner på mange underbrancher fordelt på geografi, hvilket kan resultere i en tabelcelle med fx 8 reaktioner i en underbranche i en given kommune. Ved nærmere kontrol af antallet af virksomheder i den branche beliggende i den angivne kommune, vil man se, at der er tale om 2 virksomheder. Der vil da være tale om identifikationsrisiko, som skal diskretioneres yderligere.

Implementering af principper

De enkelte enheder under Beskæftigelsesministeriet udarbejder med afsæt i de fælles principper egne retningslinjer for diskretionering af deres respektive data. I Beskæftigelsesministeriets enheder er typen af data meget forskellig, og denne tilgang er derfor nødvendig for at opnå tilstrækkelig konkrete og operationelle retningslinjer for diskretionering.

Retningslinjerne indeholder konkret stillingtagen til enhedens dataområder og vurdering af særlige risikoområder, hvor der skal være ekstra bevågenhed forud for udstilling eller udlevering af særtræk til fx folketingsbesvarelser eller aktindsigter.

Revidering af principper

Principperne tages op til revision årligt eller såfremt, der sker ændring i regler, som forudsætter ændringer. Det sker via et møde på medarbejderniveau på tværs af koncernen.

Sekretariatsfunktionen går på omgang mellem enhederne. På baggrund af mødet lægges en sag til den afdelingschef i Beskæftigelsesministeriets departement, som er ansvarlig for analyse og statistik.

Beskæftigelsesministeriet
Holmens Kanal 20
1060 København K

T +45 72 20 50 00
E bm@bm.dk
W bm.dk