

Retningslinjer for hjemsendelse af analyseresultater fra Danmarks Statistiks forskerordning

Danmarks Statistik stiller vores mange værdifulde data til rådighed for forskere og andre analytikere, som har et velbegrundet ønske om at analysere data til gavn for det danske samfund. Det er vigtigt, at dette sker på en måde, hvor både datasikkerhed og datafortrolighed er i top. Den enkelte borger og virksomhed har krav på, at fortrolige data behandles med den største omhyggelighed. De generelle regler for hvordan fortrolige data skal behandles er fastlagt dels i Lov om behandling af personoplysninger (Persondataloven) og dels i Danmarks Statistiks datafortrolighedspolitik, der kan læses på følgende LINK:

<http://www.dst.dk/da/TilSalg/Forskningservice/Dataadgang.aspx>

Datafortrolighed er en helt central forudsætning for, at forskerordningen kan eksistere. Alle datasæt der gives adgang til via forskerordningen er fortrolige, jf. Forvaltningslovens § 27, stk. 3, og Straffelovens § 152. Derfor forpligter I jer som forskere – gennem skriftlige aftaler med Danmarks Statistik - til kun at hente generelle resultater og *ikke* individdata ud fra forskermaskinerne på Danmarks Statistik.

Nedenfor gives en beskrivelse af hovedreglerne for hjemsendelse af filer, herunder hvad der menes med mikrodata og detaljerede tabeller, proceduren ved et brud på datafortroligheden, samt nogle gode råd til hvordan man undgår brud – dvs. hjemsender data man ikke må.

Hvad må hjemsendes?

Som hovedregel må der kun hjemsendes analyseresultater, aggregerede tabeller eller figurer, hvor det ikke er muligt at identificere enkelte enheder, dvs. hverken personer, husstande, familier, virksomheder og andre enheder med af-identificerede løbenummer. Dette gælder også selvom det af-identificerede løbenummer er fjernet. **Det må med andre ord ikke være muligt at genkende nogen enkeltpersoner eller -virksomheder i det hjemsendte materiale.**

Generelt gælder nedenstående tommelfingerregler for aggregerede data.

Hjemsendte resultater bør som udgangspunkt have sådan en form, at de umiddelbart kan anvendes i en publikation.

Men det er tilladt at arbejde videre med det hjemsendte *aggregerede* materiale, fx til at dannelse af figurer eller videre analyser i statistikprogrammer.

For tabeller gælder nedenstående:

Tabeller bør indeholde mindst 3 observationer pr. celle

For tabeller er tommelfingerreglen, at tabeller bør indeholde mindst 3 observationer, men aggregeringsniveauet skal altid afhænge af en konkret vurdering af

hvad tabellen indeholder. For yderligere information henvises til ”[Datafortrolighedspolitik i Danmarks Statistik](#)”, kapitel 3.

Max og min og percentiler som fx median må kun hjemsendes hvis der ikke er fare for identifikation af enkeltindivider. Ingen personer må derfor kunne genkendes i det hjemsendte output.

Funktioner som min og max henviser ofte til enkeltindivider, men de kan hjemsendes, hvis det er sikret at ingen personer kan genkendes i materialet.

Figurer

Figurer må kun hjemsendes, hvis de ikke indeholder identificerbar information. Man skal ved figurer være særligt opmærksomme på såkaldte outliers/ ekstreme værdier.

Det skal derudover sikres, at figurerne ikke indeholder indlejrede mikrodata. Sidstnævnte kan fx, sikres ved at hjemsende figuren i pdf-format.

Hvad må ikke hjemsendes?

Ingen form for mikrodata må hjemsendes

Mikrodata defineres som data, der fx indeholder nedenstående:

- Datasæt – eller dele af datasæt - med oplysninger på individniveau - dvs. hvor der i datasættet findes en række for hvert individ (enkeltrekords), også i de tilfælde, hvor man har fjernet det af-identificerede løbenummer fx personnummeret. Dvs. et datasæt på individniveau bestående af baggrundsoplysninger som fx indkomst, uddannelse og socioøkonomisk status, men hvor det af-identificerede personnummer for hver person er fjernet
- Af-identificerede nøglevariabler som personnumre, CVR-numre, arbejdsstedsnumre, adressekoder mv. betragtes altid som mikrodata, da de angiver et unikt nummer, der henviser direkte til en enkelt person eller virksomhed. Selv om der fx blot hjemsendes af-identificerede personnumre uden anden information er dette stadig ikke tilladt at hjemsende. Dette gælder også eksterne data, dvs. data som man selv har indsendt til Danmarks Statistik og som skal indgå i et konkret projekt

Pas på med programfiler og logfiler

Programmer og logfiler kan også indeholde mikrodata fx:

- Programkode der indeholder mikrodata skrevet ind i en betingelse
- Logfiler der lister observationer. Nogle procedurer lister fx et lille udsnit af det datasæt der er anvendt til analysen

En hjemsendt logfil der lister af-identificerede personnumre betragtes som et regelbrud – også selv om disse personnumre ikke indeholder anden information.

Al arbejde med mikrodata skal foregå på forskermaskinerne

Det er tilladt at danne logfiler, der lister af-identificerede personnumre – fx som kontrol af et program, men en sådan fil SKAL, i lighed med alle andre filer der indeholder mikrodata, blive på forskermaskinen og må IKKE hjemsendes.

Detaljerede tabeller bør ikke hjemsendes

Tabeller, der indeholder celler mindre end tre observationer bør ikke hjemsendes.

Indeholder tabellerne færre observationer bør det overvejes at aggregere disse tabeller yderligere, alternativt blænde celler med få observationer før tabellen hjemsendes. Afblanding kan fx ske ved at slette cellens indhold, eller erstatte det med en markering såsom "n<3".

Derudover anvendes for erhvervsstatistik en yderligere fortrolighedsregel for økonomiske variabler (fx omsætning eller værditilvækst), det såkaldte dominanskriterium. Det betyder, at hvis det største eller de to største firmaer/faglige enheder i en tabelcelle tilsammen udgør en dominerende andel, dvs. mere end 80 pct. af værdien i cellen, så træder dominanskriteriet i kraft og cellen skal blændes. For yderligere information henvises til "[Datafortrolighedspolitik i Danmarks Statistik](#)", kapitel 4.

Vær opmærksom på at man heller ikke af en tabels marginale totaler, procenter eller rater må kunne udlede om en given information stammer fra for få observationer. Fx må der ikke af en tabels summer kunne udledes hvor mange personer, der er blevet diskretioneret. Man må med andre ord ikke kunne "regne tilbage".

Modsat mikrodata, er det, når det gælder tabeller ikke muligt at lave helt faste regler for hvad der må hjemsendes og hvad der ikke må. Det er med andre ord tilladt at hjemsende tabeller, figurer mv. der her en større detaljeringsgrad end de tommelfingerregler som vi anbefaler, hvis det er sikret, at det ikke er muligt at identificere enkeltpersoner i det hjemsendte output. Fx er det OK at hjemsende medianen, selv om det reelt er mikrodata, hvis det er sikret, at ingen person/virksomheder kan genkendes i materialet. Det er dog vigtigt at understrege at al hjemsendelse sker på eget ansvar.

Procedurer og sanktioner ved et brud på datafortroligheden

Hvis et brud på datafortroligheden konstateres, så skelnes der mellem hjemsendelse af detaljerede tabeller og hjemsendelse af mikrodata.

Hvis vi konstaterer, at der er hjemsendt mikrodata, så lukker vi for institutionens adgang og følger proceduren beskrevet nedenfor.

Hvis vi konstaterer, at der er hjemsendt detaljerede tabeller, så lukker vi IKKE for adgangen, men kontakter forskeren for en redegørelse, vejleder forskeren i reglerne for hjemsendelse af data og lukker herefter sagen. Kun hvis der er tale om meget detaljerede tabeller som har karakter af mikrodata, dvs. langt de

fleste celler indeholder 1-taller, så kan vi overveje at behandle sagen som en hjemsendelse af mikrodata.

Sanktioner rammer både den enkelte forsker og den autoriserede institution, der ejer et projekt

Kernen i forskerordningen er, at Danmarks Statistik autoriserer konkrete institutioner til at anvende data til nærmere afgrænsede analyseprojekter. Det er vigtigt, at vi kan have fuld tillid til de institutioner, vi autoriserer. Danmarks Statistik har et godt kendskab til de autoriserede institutioner, mens kendskabet til de enkelte forskere er langt bedre i de autoriserede institutioner end på Danmarks Statistik. Derfor er det afgørende vigtigt, at vi kan stole på, at institutionerne står 100 % inde for de forskere, som de ønsker at give adgang til deres projekter. Institutionerne skal stå inde for at de forskere, der arbejder på deres projekter har de nødvendige kompetencer, og at de er godt bekendt med reglerne om datafortrolighed mv.

I hver autoriseret institution er der derfor udpeget en ansvarlig chef, der har ansvaret for - og fører tilsyn med - at forskere og analytikere, der er tilknyttet deres projekter, er bekendt med Danmarks Statistiks datafortrolighedsregler.

Af de aftaler Danmarks statistik indgår med de autoriserede institutioner og deres forskere, fremgår det bl.a. at en alvorlig overtrædelse af datafortrolighedsreglerne vil medføre, at forskere, der overtræder reglerne kan udelukkes fra at anvende Danmarks Statistiks forskerordninger permanent eller i en periode. Ligeledes gælder det, at den institution, der ejer det konkrete projekt, kan blive udelukket fra at bruge ordningen. En overtrædelse af datafortrolighedsreglerne kan altså få konsekvenser både for den person, der overtræder datafortrolighedsreglerne og for de projekter og forskere, der er knyttet til den autoriserede institution, som ejer pågældende projekt. Begrundelsen for dette er, at en alvorlig overtrædelse af datafortrolighedsreglerne, tyder på, at den involverede autoriserede institution ikke har taget sit ansvar tilstrækkeligt alvorligt. Det er et problem for hele forskerordningen. Derfor er sanktionen ikke kun rettet mod den enkelte forsker, men mod den autoriserede institution, der har ansvaret for at pågældende forsker kender reglerne og efterlever reglerne.

Hjemsendelse af detaljerede tabeller

Ved hjemsendelse af tabeller hvor vi vurderer at der er fare for identifikation af enkeltindivider, bliver forskeren der har hjemsendt filen, samt den autorisationsansvarlige for den institution som forskeren er tilknyttet kontaktet og bedt om at sende en redegørelse for hjemsendelsen samt bekræfte, at filen er slettet. Ligeledes vil der blive bedt om en redegørelse for hvordan man fremadrettet vil undgå sådanne hjemsendelser. Ved enkeltstående overtrædelser vil det for forskeren og institutionen normalt ikke få konsekvenser for adgangen til mikrodata men ved gentagne regelbrud, hvor der er sendt tabeller hjem hvor der er fare for identifikation af enkeltpersoner, kan denne praksis føre til lukning for institutionens adgang til mikrodata og en videre behandling af sagen svarende til hjemsendelse af mikrodata.

Hjemsendelse af mikrodata

Hvis Danmarks Statistik konstaterer, at en forsker har hjemsendt mikrodata, lukkes der omgående for forskernes adgang til mikrodata og for samtlige projekter, der hører under den autoriserede institution, der ejer det projekt, hvorfra der er hjemsendt mikrodata. Hvis en forsker hjemsender mikrodata på et projekt, der ikke er ejet af vedkommendes egen institution, men af en anden autoriseret institution, lukkes der for den institution, der ejer projektet – ikke for forskerens egen institution. Det datasikkerhedsmæssige ansvar for en forsker påhviler altså den projektejende institution. For den forsker, der har hjemsendt mikrodata, lukkes adgangen til alle projekter uanset hvilken institution de hører under.

Derefter bliver forskeren der har hjemsendt filen, samt den ansvarlige for institutionens autorisationsaftale kontaktet. De bliver bedt om øjeblikkeligt, at slette samtlige filer der overtræder reglerne for hjemsendelse af mikrodata samt sende en bekræftelse på, at alle filer der overtræder reglerne er slettet. Dette gælder både data beliggende på harddiske, mail kontoer og hvor disse data ellers måtte være opbevaret.

Danmarks Statistik vil derudover udbede sig en redegørelse for regelbruddet, samt omfanget af hjemsendte filer der bryder reglerne. Ansvar for redegørelsens udarbejdelse påhviler den autorisationsansvarlige for institutionen. Efterfølgende vil Danmarks Statistik bede institutionen om snarest muligt at sende en plan for, hvordan den fremadrettet vil undgå yderligere regelbrud. Alle adgange vil forblive lukket indtil både redegørelsen og planen er modtaget.

Herefter beslutter Danmarks Statistiks direktion hvilken sanktion institutionen skal have. Sagsbehandlingen for kan tage op til 8 arbejdsdage.

Som udgangspunkt vil sanktionen være, at der lukkes for adgangen til mikrodata for alle forskere og projekter tilknyttet institutionen i minimum en måned, samt for den berørte forsker al adgang til mikrodata i samme periode. Ved gentagne brud vil adgangen blive lukket i en længere periode eller i særlig alvorlige tilfælde permanent.

Hvis forskeren selv gør Danmarks Statistik opmærksom på, at vedkommende har hjemsendt mikrodata, vil det blive betragtet som en formildende omstændighed, og det vil muliggøre at dataadgangen vil kunne åbnes igen, så snart Danmarks Statistik har fået klarhed over overtrædelsens omfang og karakter.

En vejledende oversigt over sanktioner fremgår af oversigten i Tabel 1.
Tabel 1. Vejledende oversigt sanktioner ved regelbrud.

		Sanktion mod institution			Sanktion mod forsker	
		1. gang for institution	2. gang for institution	Gentagne gange for institution	1. gang	Flere gange
Hjemsendelse af individdata						
a.	Teknisk uheld, uden vilje til at gøre dette	1 måneds karantæne for institution (kan være en formildende omstændighed hvis institutionen fx selv underretter om fejlen)	2 måneders karantæne for institution	3 måneders karantæne for institution	1 måneds karantæne fra alle forskningsprojekter (kan være en formildende omstændighed hvis institutionen fx selv underretter om fejlen)	3 måneders karantæne fra alle forskningsprojekter
b.	Bevidst handling, ville se på dem (fejlsøgning)	2 måneders karantæne for institution	3 måneders karantæne for institution	3 måneders karantæne for institution	2 måneders karantæne fra alle forskningsprojekter	Permanent udelukkelse
c.	Bevidst, forsøg på at identificere	3 måneders karantæne for institution	3 måneders karantæne for institution	Permanent udelukkelse	Permanent udelukkelse	Kan ikke forekomme
Password eller adgang videregivet af autoriseret person						
		1. gang for institution	2. gang for institution	Gentagne gange for institution	1. gang	Flere gange
a.	Sjusk	1 måneds karantæne for institution	2 måneders karantæne for institution	3 måneders karantæne for institution	1 måneds karantæne fra alle forskningsprojekter	3 måneders karantæne fra alle forskningsprojekter
b.	Med vilje	3 måneders karantæne for institution	3 måneders karantæne for institution	Permanent udelukkelse	3 måneders karantæne fra alle forskningsprojekter	Permanent udelukkelse

Hvis der sker brud på datafortroligheden

Hvis der ved en fejl sendes mikrodata eller detaljerede tabeller hjem så kontakt Danmarks Statistik hurtigst muligt.

Orienter altid snarest Danmarks Statistik ved et regelbrud

Beskriv i mailen hvornår data er sendt hjem samt omfanget af hjemsendelsen. Adgangen til mikrodata vil fortsat blive lukket og der skal sendes en redegørelse og plan for forebyggelse, men det betragtes som en formildende omstændighed, at institutionen selv rapporterer regelbruddet.

Orienteres der hurtigt om et regelbrud betragtes det som en formildende omstændighed

Gode råd

Nedenfor følger en liste med gode råd til hvordan man undgår brud på reglerne om datafortrolighed og datasikkerhed:

For forskeren, der har adgang til mikrodata:

1. Kontroller altid samtlige filer, der skal hjemsendes
 - a. Kontakt altid Forskningsservice i tvivlstilfælde
2. Begræns hjemsendelsen af filer til det mest nødvendige.
 - a. Er det fx nødvendigt at hjemsende programmer og logfiler?
– de opbevares sikkert under forskerordningen
 - b. Er det nødvendigt at hjemsende foreløbige resultater?

For den ansvarlige for autorisationen:

1. Lav en plan der sikrer at alle forskere er orienteret om reglerne vedrørende datafortrolighed og datasikkerhed under forskerordningen
 - a. Herunder også at både den sidst tilkomne og den udenlandske forsker har tilstrækkelig viden om reglerne!
 - b. At der i miljøet findes en kontaktperson ved tvivlstilfælde
2. At alle autoriserede forskere/analytikere du har ansvar for og har underskrevet forskeraftaler for kender proceduren ved et regelbrud
3. At alle autoriserede forskere er bekendt med de ”fælder” der kan være med at arbejde med mikrodata – fx at logfiler kan indeholde mikrodata
4. Lav evt. kontrolforanstaltninger ved hjemsendelse af resultater
5. Begræns evt. antallet af forskere pr. projekt, der må sende output hjem
6. Jævnligt at kontrollere at kun de relevante forskere har adgang til mikrodata til institutionens projekter og orientere Forskningsservice hvis forskere skal have deaktiveret deres adgang