

OPENs retningslinjer for hjemsendelse af
analyseresultater fra
Sundhedsdatastyrelsens forskermaskine &
Danmarks Statistiks forskerordning

Indhold

Retningslinjer gældende for alle, som har adgang til data ved Sundhedsdatastyrelsen og/eller Danmarks Statistik via OPEN's autorisation	2
Overordnede gode råd	2
Hvad må hjemsendes?	3
Tabeller	3
Figurer	4
Hvad må ikke hjemsendes?	4
Procedurer og sanktioner ved et brud på datafortroligheden	8
Procedure for hjemsendelse af filer under OPENs autorisation	9
Yderligere information fra SDS og DST	9

Retningslinjer gældende for alle, som har adgang til data ved Sundhedsdatastyrelsen og/eller Danmarks Statistik via OPEN's autorisation

Det er muligt at få hjemsendt de færdige analyser og resultater fra Sundhedsdatastyrelsens (SDS) forskermaskinen og Danmarks Statistiks (DST) forskerservice. Dette forudsætter, at alle resultater er anonymiseret, så enkeltpersoner ikke kan genkendes i data. Dette gælder uanset datatypen, der hjemsendes (programmer, logs, tabeller, figurer og analyser).

Overtrædelser af reglerne kan betyde **lukning** af data- og brugeradgange for projektet og *den autoriserede institution (herunder også lukning af alle de projekter, der benytter den autoriserede institution)!!*

Nedenfor gives en kort beskrivelse af hovedreglerne for hjemsendelse af filer, herunder hvad der menes med tabeller, figurer og mikrodata, overordnede regler og procedure ved et regelbrud samt et forslag til arbejdsmetode, der gerne skulle sikre, at de filer, du anmoder om at få hjemtaget, overholder reglerne, og dermed sikre, at OPEN's hjemsendelsesprocedure kan fungere så smidigt som muligt.

Overordnede gode råd

Nedenfor følger en liste med gode råd om, hvordan man undgår brud på reglerne om datafor-trothed og datasikkerhed:

Hjemsendte data skal som hovedregel indeholde minimum 5 observationer pr. celle

1. Skriv aldrig programmer der skriver direkte til mailmappen
2. Kontroller altid samtlige filer, der skal hjemsendes
 - a. Kontakt altid din autorisationsansvarlige i tvivlstilfælde
3. Begræns hjemsendelsen af filer til det mest nødvendige. Spørg fx dig selv:
 - a. Er det nødvendigt at hjemsende programmer og logfiler? – de opbevares sikkert på forskermaskinen
 - b. Er det nødvendigt at hjemsende foreløbige resultater?

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, familier eller andre enheder. Data på enkelt individer må ikke hjemsendes. Dette gælder også selvom det pseudonymiserede cpr-nummer eller løbenummer er fjernet. Det må med andre ord **ikke være muligt at genkende nogen personer eller enheder i det hjemsendte materiale.**

Generelt gælder nedenstående tommelfingerregler for aggregeret data

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

Det er tilladt at arbejde videre med det hjemsendte aggregerede materiale, f.eks. til at dannelse af figurer eller videre analyser i statistikprogrammer.

Tabeller

Retningslinjer for hjemsendelse af tabeller

Tabeller skal indeholde mindst 5 observationer per. celle, men aggregeringsniveauet skal altid afhænge af en konkret vurdering af, hvad tabellen indeholder

Tommelfingerreglen er, at tabeller skal indeholde **mindst 5 observationer pr celle**, i nogle tilfælde er grænsen højere. Grænsen skal derfor altid afhænge af en konkret vurdering af hvad tabellen indeholder. Kravet gælder det laveste aggregeringsniveau dvs. at hvis man kombinerer flere variabler i en tabel, så skal der mindst være fem observationer i de variabler, der kombineres.

Indeholder tabellerne færre observationer, bør det overvejes, at aggregere tabellerne yderligere, alternativt blænde celler med få observationer før tabellen hjemsendes. Afblænding kan f.eks. ske ved at slette cellens indhold, eller erstatte det med en markering såsom "n<5". **Vær opmærksom på**, at man heller ikke af en tabels marginale totaler, procenter eller rater må kunne udlede om informationen stammer fra få observationer, f.eks. må der ikke af en tabels summer kunne udledes, hvor mange personer, der er blevet "lagt sammen". **Det må ikke være muligt** at "regne tilbage" og nå en værdi under 5. Det kan derfor anbefales at sammenlægge kategorier, så denne problematik undgås.

Detaljeret tabeller og resultater bør ikke hjemsendes

Det er også tilladt at hjemmesende værdier, der beskriver en fordeling, som min., maks. og median, der ofte henviser til enkeltindivider forudsat, at der **ikke** er risiko for identifikation af enkeltindivider. Indeholder materialet outliers/ekstreme værdier kan disse eventuelt fjernes inden hjemsendelse.

Maks og min og percentiler, som f.eks. median må kun hjemsendes, hvis der ikke er fare for identifikation af enkeltindivider. Ingen personer må kunne genkendes i det hjemsendte output

Figurer

Retningslinjer for hjemsendelse af figurer

Figurer må kun hjemsendes, hvis de ikke indeholder information, som kan identificere enkeltindivider – vær særlig opmærksomme på såkaldte outliers/ekstreme værdier

Det skal derudover sikres, at figurerne ikke indeholder indlejrede mikrodata. Sidstnævnte kan f.eks., sikres ved at hjemsende figuren i pdf-format eller forskellige billede formater.

Hvad må ikke hjemsendes?

INGEN form for mikrodata må hjemsendes

Mikrodata defineres som data, der f.eks. indeholder nedenstående:

- Datasæt – eller dele af datasæt - med oplysninger på individniveau, dvs. hvor der i datasættet findes en række for hvert individ (enkeltekords), også i de tilfælde, hvor man har fjernet det pseudonymiserede løbnummer fx det pseudonymiserede personnummer. Dvs. datasæt bestående af observationer på enkeltindivider med fx fødselsdato, kommune, køn, diagnose eller ATC kode
- Pseudonymiserede nøglevariabler såsom personnumre og ydernumre betragtes altid som **mikrodata**, da de angiver et unikt nummer, der henviser direkte til en enkelt person eller praktiserende læge. Derfor er det heller ikke tilladt at hjemsende pseudonymiserede personnumre alene. Dette gælder også egne data, dvs. data som man selv har indsendt til Sundhedsdatastyrelsen og som skal indgå i et konkret projekt
- Dette **gælder også eksterne data**, dvs. data som man selv har indsendt til SDS eller DST og som skal indgå i et konkret projekt

Pas på med programfiler og logfiler

Programmer og logfiler kan også indeholde mikrodata, jf. nedenstående eksempler på programmer og logfiler, der **ikke** må hjemsendes

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

Bemærk at hvis man sender en logfil hjem, der lister pseudonymiserede personnumre er det et brud på reglerne – også selv om disse pseudonymiserede personnumre ikke indeholder anden information.

Al arbejde med mikrodata skal foregå på forskermaskinerne

Al arbejde med mikrodata skal derfor foregå på SDS' eller DST's forskermaskinen. Det er tilladt at danne logfiler, der lister pseudonymiserede personnumre – f.eks. 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**.

Tabel 1: Eksempel på mikrodata

Pnr	Køn	Indkomst	Bopæl	Alder	Pnrb
1	M	1 500 000	København	40	15
2	M	1500	Aarhus	45	22
3	K	1000	Bornholm	50	37
4	M	500	Bornholm	50	65
4	M	250	København	40	87
5	M	150	Aarhus	20	
6	K	150	Odense	25	
7	M	750	Herning	35	19
8	K	500	København	60	74
9	M	250	Odense	65	74
10	K	150	Aarhus	20	
11	K	50	Bornholm	15	
12	M	0	København	10	

Tabel 1 indeholder oplysninger om 12 personernes køn, indkomst, bopæl, alder og en reference til personens barn. Oplysningerne henviser direkte til et individ, og **må derfor ikke hjemsendes, men skal blive på forskermaskinen**.

Tabel 2: Eksempel på mikrodata

Køn	Indkomst	Bopæl	Alder
M	1 500 000	København	40
M	1500	Aarhus	45
K	1000	Bornholm	50
M	500	Bornholm	50
M	250	København	40
M	150	Aarhus	20
K	150	Odense	25
M	750	Herning	35
K	500	København	60
M	250	Odense	65
K	150	Aarhus	20
K	50	Bornholm	15
M	0	København	10

I tabel 2 er identifikationsvariablerne pnr og pnr b fjernet. Selvom om variablerne med pseudonymiserede personnumre er fjernet, er der stadig tale om mikrodata. De tilbageværende oplysninger kan henføres til et enkelt individ, og må **ikke hjemsendes, men skal blive på forskermaskinen.**

Tabel3: Eksempel på mikrodata

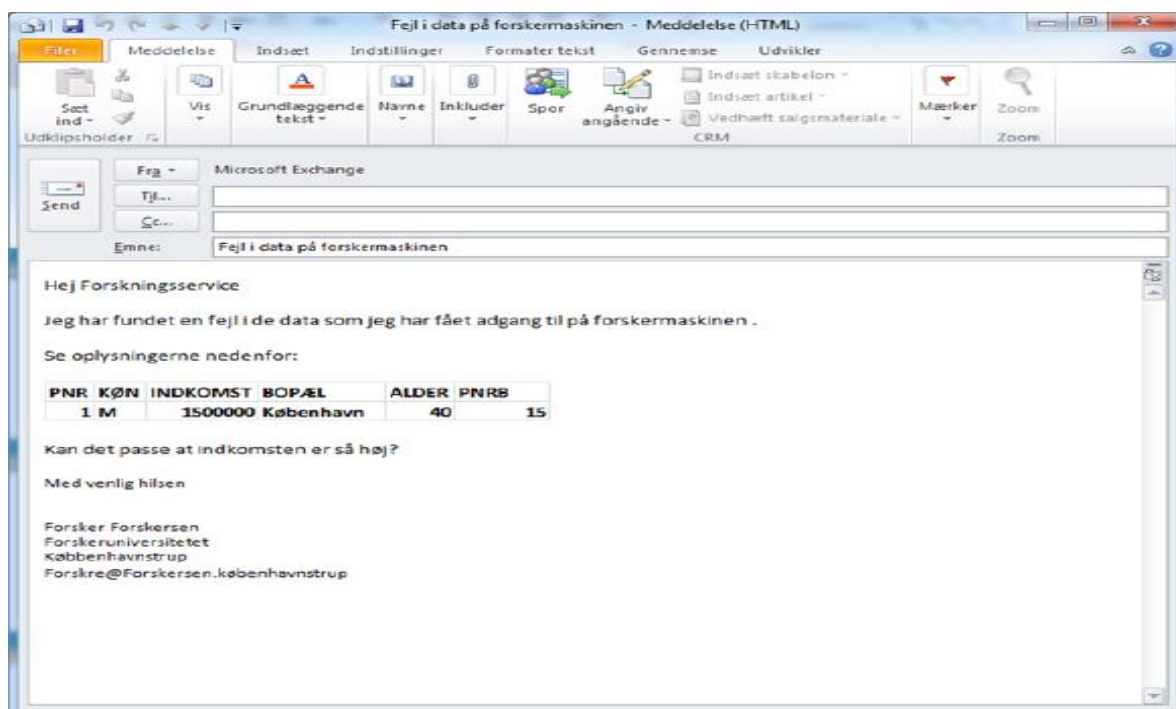
Pnr
1
2
3
4
4
5
6
7
8
9
10
11
12

I tabel 3 er al information fjernet på nær det pseudonymiserede personnummer. Det er heller ikke tilladt at hjemsende denne fil, da den stadig indeholder individdata. Det er uden betydning, at alle de øvrige informationer er fjernet. Der er stadig individdata, da oplysningerne kan henføres til et enkelt individ, og **data må ikke hjemsendes, men skal blive på forskermaskinen.**

Alle tre ovenstående tabeller indeholder altså mikrodata og må derfor **ikke hjemsendes!**

I forbindelse med f.eks. fejlfinding er det ikke tilladt at sende mikrodata på mail

I forbindelse med f.eks. fejlfinding er det **ikke tilladt at sende mikrodata på mail**. Det gælder også selvom mailen sendes til en medarbejder i Forskerservice hos SDS eller DST jf. nedenstående eksempel.



Figur 1. Eksempel på mikrodata på mail, der ikke må sendes

Procedurer og sanktioner ved et brud på datafortroligheden

Konstaterer SDS eller DST et brud på datafortroligheden, skeles der mellem hjemsendelse af detaljeret tabeller og resultater og hjemsendelse af mikrodata.

Hjemsendes mikrodata **lukkes** institutionens adgang til forskermaskinen, imens de nærmere omstændighederne omkring hjemsendelsen undersøges. Det gælder både for den bruger, som har hjemsendt mikrodata og for alle andre brugere fra den autoriserede institution.

Hjemsendes detaljeret tabeller og resultater **lukkes der ikke** for adgangen, men forskeren kontaktes for en redegørelse for hjemsendelsen samt bekræftelse på at filen er slettet. Forskeren vejledes i reglerne for hjemsendelse af data, hvorefter sagen lukkes. Er der tale **om hjemsendelse af meget detaljeret tabeller**, hvor de fleste celler indeholder 1-taller, og derfor har karakter af mikrodata, vil SDS og DST overveje at behandle sagen som hjemsendelse af mikrodata.

Der vil i hvert tilfælde blive foretaget en konkret vurdering fra forskerservice omkring alvorligheden af bruddet, som brugeren har foretaget. Det betyder, at det i mildere tilfælde kan føre til en skriftlig advarsel til brugere og autorisationsindehaveren. **Ved gentagne overtrædelser der har ført til skriftlig advarsel fører det til udelukkelse af brugeren.**

Såfremt vilkårene omkring datafortrolighed og sikkerhed, der er specificeret i aftalerne, misligholdes, **udelukkes brugeren fra at anvende enten SDS' eller DST's Forskermaskiner i fremtiden.**

Procedure for hjemsendelse af filer under OPENs autorisation

OPEN har en gatekeeperfunktion, der betyder, at den enkelte forsker ikke selv kan hjemsende filer. Gatekeeperen skal godkende, at retningslinjerne for hjemsendelse overholdes.

Det er dog stadig **dit ansvar** at ovenstående overholdes, også selv om OPEN måtte vælge at følge din anmodning og hjemsende data til dig.

Ønsker du at hjemsende aggregerede data (tabeller, figurer eller lignende), skal du gøre følgende

1. Samle data i en eller flere filer, som lægges i en i øvrigt tom folder
2. Sikre dig at data overholder SDS/DST's krav til data, der hjemtages (se ovenfor)
3. Sende en mail til OPEN.registry@rsyd.dk, hvori du oplyser:
 - a. SDS projektnummer (FSEID nr) eller DST projektnummer
 - b. Hvilken mappe de ønskede filer ligger i
 - c. Navn på de pågældende filer

Gatekeeperen hos OPEN vil derefter gennemse filerne og forestå selve hjemsendelsen indenfor 1-2 arbejdsdage, hvis filernes vurderes at overholde reglerne.

Yderligere information fra SDS og DST

For yderlig information henvises til følgende dokumenter på DST hjemmeside

- [Hjemsendelse af filer](#)
- [Hjemsendelse af filer - tilladte filtyper](#)
- [Retningslinjer for hjemsendelse af analyseresultater](#)

[Datafortrolighedspolitik i Danmarks Statistik](#)

For yderlig information henvises til følgende dokumenter på SDS' hjemmeside

<https://sundhedsdatastyrelsen.dk/-/media/sds/filer/forskerservice/databehandling-pa-forskermaskinen.pdf?la=da>

https://sundhedsdatastyrelsen.dk/-/media/sds/filer/forskerservice/retningslinjer-for-hjemsendelse-af-analyseresultater_2018_07_03.pdf?la=da