

Vejledning til inklusion af egne data i et projekt på Forskermaskinen

På Forskermaskinen er det muligt at få inkluderet data fra andre kilder end dem, der er på Forskermaskinen. Det kan for eksempel være data, du selv har indsamlet, eller data du får videregivet fra andre. Uanset hvor disse data kommer fra, betegner Forskerservice dem som egne data. De egne data kan være personhenførbare¹ eller ikke-personhenførbare. Denne vejledning beskriver, hvordan du skal gøre for at få egne data inkluderet i dit forskningsprojekt på Forskermaskinen.

Vejledningen er delt op i fire dele:

1. Ansvar og retningslinjer for egne indsendte data
2. Tekniske krav til egne indsendte data
3. Indsendelse af personhenførbare data
4. Indsendelse af ikke-personhenførbare data

Hvis du ønsker at indsende personhenførbare data, er de tre første underpunkter relevante. Hvis du ønsker at indsende ikke-personhenførbare data, er det alene punkt 4, der er relevant. Du skal derfor starte med at undersøge, om data er personhenførbare. Hvis du er i tvivl, er dine data formentlig personhenførbare, og du bør behandle dem som sådan.

1. Ansvar og retningslinjer for egne indsendte data

Selvom du sender dine egne data ind til dit projekt på Forskermaskinen, er dataansvaret fortsat dit. Det betyder, at det er dit ansvar, at du har hjemmel til at behandle data, at data bliver behandlet i henhold til databeskyttelseslovgivningen, og at data behandles i henhold til retningslinjerne på Forskermaskinen.

Godkendelse og hjemmel til behandling af data

Dit forskningsprojekt skal være opført på den dataansvarlige institutions fortegnelse. Det er det, når du har fået en godkendelse til at gennemføre projektet fra din dataansvarlige institution.

¹ Personhenførbare data, er data, der alene eller sammen med andre data, kan identificere et individ. Det kan for eksempel være gennem et personnummer, navn, adresse, løbenummer eller blot så mange oplysninger om en person, at personen bliver identificerbar.

Når du inkluderer egne data i projektet, skal du sikre dig, at de egne data også er omfattet af fortegnelsen. Hvis de egne data ikke er omfattet af fortegnelsen, skal der foretages en ændring på fortegnelsen, og projektet skal have godkendt ændringen af den dataansvarlige institution, inden data kan inkluderes i dit forskningsprojekt på Forskermaskinen. En ændring af fortegnelsen og fornyet godkendelse af ændringen kan for eksempel være relevant, hvis inklusionen af egne data medfører ændringer i formål, populationsstørrelse eller typen af persondata.

Forskerservice forudsætter desuden, at du har hjemmel til at indhente og behandle de egne data, du anvender i dit projekt, og at denne hjemmel er til stede, inden du indsender data. Hvis du er i tvivl om, hvorvidt dit projekt har de relevante hjemler, skal du kontakte juridisk ekspertise i din egen organisation.

Særlige regler for kontakt til borgere

Bemærk, at der er særlige regler for kontakt til borgere, hvis man forud herfor har tilgået journaloplysninger eller data fra en klinisk kvalitetsdatabase. Disse regler kan have betydning for dit projekt, hvis der blandt dine egne data indgår journaloplysninger eller data fra en klinisk kvalitetsdatabase, og du på et senere tidspunkt ønsker at tage kontakt til borgerne, som indgår i dit projekt. I så fald skal du være opmærksom på, at kontakt til de pågældende borgere kræver involvering af den sundhedsperson, som har foretaget behandlingen af borgeren.

Retningslinjer for data på Forskermaskinen

Data, der skal på Forskermaskinen, skal være pseudonyme. Pseudonymitet betyder, at der ikke må være data, der indeholder CPR-numre, navne, adresser eller andre informationer, der gør data direkte personhenførbare. Disse regler gælder også for dig, når du skal inkludere egne data.

Som projektansvarlig har du derfor ansvaret for at:

- Sikre, at alle variable, der indeholder direkte personhenførbare oplysninger, er fjernet fra datasættet, medmindre de skal krypteres inden indlæsning som eksempelvis CPR-nummer og andre ID-variable.
- Sikre, at alle variable, der indeholder direkte personhenførbare oplysninger, som skal krypteres (fx CPR-nummer eller patient ID) er angivet i oplysningsskemaet. Forskerservice krypterer CPR-numre, løbenumre og andre identifikationsnøgler. Den anvendte krypteringsnøgle sikrer, at det er muligt at koble dine egne data med de øvrige data i forskningsprojektet.
- Sikre, at datasættet ikke indeholder fritekst. Hvis rådata indeholder fritekstvariable med indhold, der er nødvendige for projektet, bør dette indhold grupperes og standardiseres, inden datasættet indsendes.

- Hvis ikke det er muligt at gruppere eller standardisere indholdet inden indlæsning, skal dette begrundes og variabelen angives i oplysningsskemaet under særlige variable.
- Sikre, at der i projektet ikke behandles flere personoplysninger end nødvendigt, i henhold til GDPR. Du bør derfor inden indsendelse sikre at mængden af oplysninger, der indsendes, samt detaljeringsniveauet i de enkelte variable ikke overstiger, hvad der er nødvendigt for at opfylde projektets formål.
 - Eksempler på variable, der ofte er mere detaljerede, end der er behov for til at opfylde projektets formål, er:
 - Timestamps (helt præcise angivelser af et tidspunkt, helt ned på sekund-niveau)
 - Variable med unødvendigt mange decimaler.

Egne data skal fremgå af den gældende aftale om dataadgang på Forskermaskinen

Det er vigtigt, at du alene indsender data, som er omfattet af den aftale, du har indgået med Forskerservice. Det betyder, at navngivning af tabeller og variable samt angivelse af særlige variable i det indsendte data skal stemme overens med det, der er angivet i Bilag C af aftalen om dataadgang. Hvis du grundet tekniske krav er nødsaget til at opdele dit datasæt i flere datasæt, så skal alle datasæt være angivet i Bilag C af aftalen om dataadgang.

Det er også vigtigt, at du indsender data i et format, der lever op til nedenstående tekniske krav.

Hvis du indsender data, der ikke lever op til aftalen eller de tekniske krav, vil det øge sagsbehandlingstiden markant og i nogle tilfælde kræve, at du fremsender nye datasæt.

Sanktion

Hvis Forskerservice ved indlæsning af data på Forskermaskinen eller efterfølgende bliver opmærksomme på, at der på et projekt ligger egne data, der indeholder navne, adresser eller ukrypterede CPR-numre, løbenumre eller andre identifikationsnøgler, anses dette som et brud på retningslinjerne for arbejdet på Forskermaskinen. Brud på retningslinjerne vil blive sanktioneret på lige fod med hjemsendelse af mikrodata.

2. Tekniske krav til egne indsendte data

Udover ovenstående krav om hjemmel, dataminimering og pseudonymitet, skal de indsendte data også overholde en række tekniske krav, så de nemt og sikkert kan indlæses på Forskermaskinen.

Du skal derfor være opmærksom på følgende:

- Filer skal indsendes som enten sas7bdat-filer eller flade filer (.csv eller .txt).
 - Anvend en delimiter, som ikke indgår som værdi i datasættet.

- Mellemrum kan ikke anvendes som delimiter.
- Forskerservice anbefaler brug af x som delimiter, da denne sjældent indgår som værdi i datasæt.
- Variabelnavne skal være SAS-kompatible og skal derfor overholde følgende kriterier:
 - Max 32 tegn lange (max 22 for variable, der skal krypteres inden indlæsning)
 - Ingen mellemrum i variabelnavnet (anvend eventuelt _ til angivelse af mellemrum)
 - Ingen specialtegn i variabelnavnet (fx ÆØÅ,.;:-`*#%&/)()
 - Labels kan ikke indlæses på Forskermaskinen, navngiv derfor dine variable med omhu.
- Indholdet i variablene skal være SAS-kompatible i engelsk version.
 - Anvend derfor punktum i decimaltal, da komma i decimaltal risikerer at blive indlæst forkert.
- Tabelnavnene skal være SAS-kompatible og skal derfor overholde følgende kriterier:
 - Maks. 18 tegn lange
 - Dette skyldes Forskerservice tilføjer præfiks (ds_ext_) og postfiks (_yymmdd dato for upload) til tabelnavnet
 - Tabelnavnet skal starte med et bogstav (A-Z)
 - Ingen mellemrum i tabelnavnet (anvend eventuelt _ til angivelse af mellemrum)
 - Ingen specialtegn i tabelnavnet (fx ÆØÅ,.;:-`*#%&/)()
- Hver tabel du indsender må maksimum fylde 100 GB.
- Hvis dit datasæt har mere end 1.024 variable skal du splitte datasættet op i tabeller med færre end 1.024 variable*.
 - Husk at inkludere relevante nøgler i alle tabeller.
 - Sørg for at alle tabeller er angivet i bilag C af aftalen.

**Der må kun være 8.192 tegn per række, så hvis der er flere tegn end dette ved 1.024 variable i datasættet, skal du splitte dit datasæt yderligere op.*

Når dit datasæt følger ovenstående guidelines, sikrer du, at data kan indlæses på Forskermaskinen uden problemer, samt at den tid Forskerservice skal bruge på at indlæse filerne minimeres.

3. Indsendelse af personhenførbare data

Det er vigtigt, at de indsendte datasæt overholder Forskerservices retningslinjer samt tekniske krav, som er beskrevet ovenfor. Derved sikrer du, at data kan indlæses på Forskermaskinen uden problemer, samt at den tid Forskerservice skal bruge på at indlæse filerne minimeres.

Du anmoder om tilføjelse af egne data ved at udfylde og indsende et oplysningsskema via ansøgnings- eller henvendelsesformularen, afhængig af om de egne data skal tilføjes et nyt projekt eller et igangværende projekt.

På Sundhedsdatastyrelsens hjemmeside under [Egne data på Forskermaskinen](#) finder du oplysningsskemaet, som du skal udfylde med vigtige oplysninger om det datasæt, der skal indlæses på Forskermaskinen:

- **Tabelnavn**

Angiv tabelnavnet på dit datasæt. Når du uploader din tabel, skal du anvende den navngivning, som du har angivet i skemaet.

Ved indlæsning tilføjer Forskerservice præfiks (ds_ext_) og postfiks (_yymmdd dato for upload) til tabelnavnet, og dit tabelnavn kan derfor maks. være 18 tegn. Det endelige tabelnavn bliver ds_ext_**tabelnavn**_yymmdd.

- **Særlige variable og variabeltype**

I oplysningsskemaet skal du angive særlige variable i dit datasæt. Særlige variable er følgende:

- **Direkte personhenførbare variable, som skal krypteres ved indlæsning**

Ved direkte personhenførbare oplysninger, forstås oplysninger hvis formål det er at identificere en person.

Eksempler på personhenførbare oplysninger er:

- CPR-nummer
- Projektdeltagers løbenummer
- Patient-ID eller et unikt prøve-ID
- Ydernummer

Direkte personhenførbare oplysninger, der ikke kan bruges i krypteret form som f.eks. navne og adresser, må ikke indgå i datasæt, der skal indlæses på Forskermaskinen.

'Variabeltype' udfylder du med 'Personhenførbare'.

- **Variable der skal anvendes til afgrænsning af din dataadgang**

Du skal angive de variable, der skal anvendes, når Forskerservice afgrænser din dataadgang. Eksempler på afgrænsningsvariable er:

- CPR-nummer på din population eller populationer hvis du indsender forskellige populationer
- Indeksdato/inklusionsdato, hvis denne skal anvendes til at definere en udtræksperiode

'Variabeltype' udfylder du med typen på afgrænsningsvariablen. Det kan til eksempel være 'Population' eller 'Indeksdato' i eksemplerne ovenfor.

- **Fritekstvariable (hvis nødvendigt)**
Fritekstvariable er sjældent anvendelige i registerforskning. Derudover øger fritekstvariable risikoen for, at der gemmer sig direkte personhenførelse oplysninger, som et navn eller et CPR-nummer i variablen.

Fritekstvariable i dine datasæt skal derfor som udgangspunkt kategoriseres inden indsendelse til Forskerservice. Hvis det ikke er muligt at kategorisere indholdet, inden du indsender data, skal du sende en begrundelse for dette. Derudover skal variablen angives under 'Særlige variable'.

'Variabeltype' udfylder du med 'Fritekst'.

Upload egne data

Når Forskerservice har behandlet og godkendt din anmodning om indsendelse af dine data, har Forskerservice lov til at modtage dine data.

Forskerservice benytter ShareFile® til deling af data, hvilket gør det nemmere og mere sikkert at sende dine data til os.

Anmodning om uploadlink

Indsendelse af nye egne data

Hvis du skal indsende nye egne data enten til et nyt eller igangværende projekt, kræver det, at der indgås en ny eller opdateret aftale om dataadgang, hvor de nye egne data fremgår.

Forskerservice sender et uploadlink umiddelbart efter modtagelse af accept af aftale, medmindre aftalen skal i ekstern høring. I så fald vil linket blive fremsendt, så snart Forskerservice har modtaget positivt svar på høringen.

Opdatering af eksisterende egne data

Hvis der er tale om opdatering af eksisterende egne data, som allerede fremgår af gældende aftale om dataadgang, skal du indsende en henvendelse og anmode om at få fremsendt et uploadlink.

Upload af data

Uploadlinket sendes altid til den projektansvarlige og kan kun anvendes i 7 dage. Linket kan videresendes til en anden, hvis det ikke er dig som projektansvarlig, der selv skal uploade data. Du skal dog sikre, at den der uploader data er bekendt med navngivning og oplysninger om egne data, som angivet i bilag C af aftale om dataadgang, samt om de tekniske krav angivet ovenfor.

Ved at følge linket vil du komme til Sundhedsdatastyrelsens upload-side, hvor du kan uploade dine data til Forskerservice. Når du uploader data, bliver datafilen automatisk krypteret. Det er derfor ikke nødvendigt, at du krypterer datafilen.

Uploader du mere end én datafil, må du meget gerne pakke filerne i én zip-komprimeret mappe.

4. Indsendelse af ikke-personhenførbare data

Du kan indsende filer med ikke-personhenførbare data via vores henvendelsesformular, som du finder på vores hjemmeside under [Kontakt](#).