

Projektdatabaser på Forskermaskinen

Projektdatabaser

Større forsknings- og analysemiljøer kan ansøge om at få oprettet en projektdatabase på Forskermaskinen, hvis de har behov for dette. En projektdatabase er en bredere adgang til grunddata, hvor brugeren (omtales herefter som datamanager) har mulighed for eksplorativt at definere nye projekter.

Projektdatabasen skal alene anvendes til at definere datagrundlaget til specifikke projekter og generere hypoteser. Der må således ikke gennemføres forskning direkte på projektdatabasen. Ligeledes må adgangen til projektdatabasen ikke anvendes til projekter eller opgaver, der ikke relaterer sig direkte til Forskermaskinen. Det vil sige projekter og opgaver, hvor databehandlingen i projektdatabasen ikke sker med henblik på oprettelse af et projekt på Forskermaskinen.

Ansøgning

For at oprette en projektdatabase, skal der udfyldes en ansøgning via Forskerservices ansøgningsskema til projekter på Forskermaskinen. Til ansøgningen skal du vedhæfte din anmeldelse af projektet til den dataansvarlige institution, samt en godkendelse.

Der skal vedhæftes en *projektbeskrivelse*. Projektbeskrivelsen skal blandt andet indeholde en beskrivelse af projektdatabasens formål samt en beskrivelse af typen af projekter, som I allerede har gennemført med udgangspunkt i registerdata. Derudover skal den indeholde en beskrivelse af typen af projekter, som I planlægger at gennemføre med udgangspunkt i projektdatabasen. Projektbeskrivelsen skal begrunde bredden af registre, population og perioder, der ønskes tilgængelige på projektdatabasen.

Der skal vedhæftes en *udtræksbeskrivelse*, der beskriver, hvordan projektdabasens population skal defineres, samt hvordan dataadgangen til de enkelte registre skal afgrænses. Derudover skal det beskrives, hvorfor de ønskede data er relevante i henhold til projektdabasens formål. Forskerservice skal til vurdering af ansøgningen kunne se en klar sammenhæng mellem typer af projekter, der skal gennemføres, og de registre og populationer der ansøges om på projektdatabasen.

Der skal vedhæftes et udfyldt *bestillingsark*, hvori det markeres hvilke variable der ønskes adgang til.

Adgang til projektdatabasen

Der skal oprettes datamanagere til projektdatabasen (max 3). De vil få tildelt brugeradgang til projektmappen og dataadgang til views.

Datamanagere får ikke automatisk adgang til projekterne under projektdatabasen. Hvis en datamanager skal oprettes som bruger på et projekt under projektdatabasen, anmodes der om dette på samme måde som for andre brugere.

Egne data i tilknytning til projektdatabasen

Der kan uploades egne data til projektdatabasen på samme vilkår som ved tilknytning af egne data til et hvert andet projekt på Forskermaskinen. Du kan læse mere om tilknytning af egne data på vores hjemmeside www.sundhedsdatastyrelsen.dk/forskerservice.

Tilladelse fra den dataansvarlige institution og databehandleraftaler

Projektdatabasen og projekterne under projektdatabasen skal anmeldes separat til din dataansvarlige institution. Anmeldelsen og godkendelsen af projektdatabasen skal vedhæftes ansøgningen til Forskerservice.

Godkendelsen skal indeholde en angivelse af hvem der er projektansvarlig for projektdatabasen. Derudover, når vi gennemgår ansøgningen, påser vi, at titel og formål i anmeldelsen er dækkende for den indsendte projektbeskrivelse.

Selve formålet skal afspejle, at en projektdatabase er en bred adgang med et bredt men afgrænset formål, som skal bruges eksplorativt og til at generere hypoteser inden for emnet.

Derudover anbefaler vi, som ved alle andre projekter på Forskermaskinen, at Forskerservice, Sundhedsdatastyrelsen fremgår som databehandler.

De enkelte projekter under projektdatabasen skal anmeldes, på samme måde som andre projekter på Forskermaskinen.

Projekter under projektdatabasen

For hvert projekt under projektdatabasen skal der ansøges på samme vis som til almindelige projekter på Forskermaskinen. Ansøgning om projekter under projektdatabasen får efterfølgende et selvstændigt FSEID. Der indgås herefter en aftale mellem dig og Forskerservice om hvilke data, der må flyttes over på det konkrete projekt under projektdatabasen.

Datamanagerne på projektdatabasen forestår selv udtræk af data til projekter i henhold til den indgåede aftale. Hvis der ønskes data på projektet under projektdatabasen, som ikke er tilgængelig i projektdatabasen, sørger Forskerservice for at danne en adgang til disse data på projektet.

Datamanagere danner og tilrettelægger datasæt i en mappe på projektdatabasen. Herefter skal datamanager give besked til Forskerservice, der så foretager kontrol af data og programmer. Kontrollen sikrer, at der er overensstemmelse mellem den aftalte dataspecifikation og udtrækket, der skal overflyttes. Efter kontrollen er foretaget, flytter Forskerservice data til det konkrete projekt.

Datamanagere må ikke flytte data til projekter under projektdatabasen selv om de i nogle tilfælde har adgang til dette. Dette gælder også ved eventuelle opdateringer af registerdata. Her skal Forskerservice også godkende overførslen og flytte data.

For at det er muligt for Forskerservice at foretage kontrol af, hvilke data som er overført til projekter under projektdatabasen, skal datamanagere overholde en fast fil- og mappestruktur, som beskrevet nedenfor. Der ud over skal programmer til udtræk af data skrives med SAS, og output skal være SAS-datasæt.

Mappestruktur

Projektdatabasen

I *input-mappen* placerer Forskerservice eventuelle eksterne data, som anvendes til at danne population til underprojektet.

F:\Projekter\PDB0000xxxx\FSEID0000xxx1\InputData

F:\Projekter\PDB0000xxxx\FSEID0000xxx2\InputData

I *output-mappen* placerer datamanageren de data, der skal flyttes til projektet under projektdatabasen samt de programmer, som ligger til grund for udtrækket.

F:\Projekter\PDB0000xxxx\FSEID0000xxx1\OutputData

F:\Projekter\PDB0000xxxx\FSEID0000xxx2\OutputData

Datamanagere på projektdatabasen opretter selv disse underfoldere.

Forskerservice tager kopi af data og sletter ikke data efter flytning. Datamanagere står for eventuel oprydning i mappen.

Underprojekter

For hvert projekt under projektdatabasen danner Forskerservice projektmappen og tildeler rettigheder til de brugere, som skal have adgang hertil. Når data er godkendt til projektet kopierer Forskerservice data til projektmappen for projektet og skrivebeskytter dem.

F:\Projekter\FSEID0000xxx1\InputData

Begrebsliste

Bruger

En bruger er en person, som har adgang til data på Forskermaskinen. En bruger skal godkendes af en autorisationsansvarlig, mens en projektansvarlig skal give lov til, at brugeren får adgang til data på et specifikt projekt. Brugernavnet består af 9 tegn fskxxxxyy. Brugeren oprettes i SEB-systemet og som Windows-bruger på Forskermaskinen.

Brugeradgang

En brugers adgang til data på Forskermaskinen, samt adgang til projektmappe og brugermappe på applikationsserverne.

Dataadgang

Er en tildelt adgang til at se og arbejde med alle data på projektet eller projektdatabasen.

Datamanager

En bruger med rettigheder til at arbejde på en projektdatabase. Datamanagere danner de datasæt som skal overføres til godkendte projekter under projektdatabasen. Der kan maksimalt være tre datamanagere på en projektdatabase.

Dataspecifikation

En beskrivelse af hvilke registre, tabeller og variable der gives adgang til på projektet og hvordan de afgrænses. Derudover kan du i dataspecifikationen finde information om hvilke analyseværktøjer I får tildelt adgang til, samt information om, hvem der skal tilknyttes som bruger på projektet.

Forskermaskinen

Det online beskyttede forskningsmiljø, hvor du som bruger får online adgang til data på specifikke projekter. Det miljø, hvor man som bruger får adgang til forskerdatabasen, som indeholder grundregistre og applikationsservere med SAS, R og Stata som analysesoftware.

Grundregistre

De rå registerdata som er overført fra Sundhedsdatastyrelsens datavarehus. De omfatter en række sundhedsregistre samt CPR-registeret.

Projektdatabase

En overordnet projektansøgning med et selvstændigt FSEID, der oprettes som et projekt hvor en forskergruppe ved 1-3 datamanagere har adgang til et bredere dataudsnit, der kan danne udgangspunkt for udtræk til konkrete projekter. Der må ikke gennemføres forskning direkte på data i projektdatabasen, disse skal flyttes til konkrete projekter.

View

Er en virtuel tabel, som giver brugeren mulighed for at læse opdateret data fra en grunddatatabel på Forskermaskinen. Forskerservice danner views til alle projektdatabaser, der afgrænser hvilke variable og hvilke rækker, som kan læses fra tabellen på det enkelte projekt.

Forskellen på et view og en tabel er, at et view indeholder en definition af en tabel, og at indholdet først dannes, når der laves en forespørgsel mod viewet. Views har den fordel, at data i viewet automatisk bliver opdateret, når grundregistre opdateres. Samtidig så undgår man at have mange kopier af de samme data i databasen (redundans), hvilket ville kræve en meget større kapacitet.