

Introduktion til Forskningservice i Danmarks Statistik

1. Mikrodata til forskning

Danmarks Statistik stiller data til rådighed til forsknings- og analyseprojekter. Nedenfor redegøres for hovedprincipperne bag denne servicevirksomhed.

Til langt de fleste af nedenstående punkter findes der en mere omfattende beskrivelse på Danmarks Statistik hjemmeside. Se nærmere på:

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

| | |
|---|---|
| <i>Danmarks Statistikbank</i> | Danmarks Statistik råder over meget betydelige mængder af data som grundlag for statistikproduktionen og publiceringen af denne. Danmarks Statistik stiller en stor del af de indsamlede data <u>gratis</u> til rådighed gennem Danmarks Statistikbank og denne vil (sammen med publikationerne) ofte være et godt udgangspunkt for den, der ønsker at danne sig et indtryk af mængden og mangfoldigheden af data. Statistikbanken findes på følgende link: http://www.statistikbanken.dk/statbank5a/default.asp?w=1024 |
| <i>Ingen identifikation</i> | Principperne bag Danmarks Statistikbank er, at det ikke skal være muligt at identificere enkeltpersoner eller virksomheder. Tallene er derfor aggregeret til et vist niveau. |
| <i>Forskere har behov for mikrodata</i> | Når det drejer sig om anvendelse af registerdata til forskning og analyse, vil der i mange tilfælde være behov for adgang til data på såkaldt mikroniveau, dvs. data for enkelte personer eller virksomheder. |
| <i>Unikke muligheder</i> | Danmarks Statistik ønsker at gøre det muligt for forskere at arbejde med mikro-data, da datasamlingerne udover at være utroligt omfattende er helt unikke ikke blot i et dansk, men også i et internationalt perspektiv. |
| <i>Krav om fortrolighed</i> | Den enkelte borger og virksomhed har krav på, at data behandles fortroligt. Data afgivet til Danmarks Statistik enten direkte fra borger/virksomhed eller indirekte fra en anden offentlig myndighed må kun anvendes til statistik og analyse, aldrig til administrativ sagsbehandling. Deraf følger også, at Danmarks Statistik ikke kan udlevere oplysninger som direkte gennem personnummer, virksomhedsnummer, adresse mv. kan henføres til en person eller en virksomhed. Reglerne er fastlagt dels i databeskyttelsesforordningen og dels i Danmarks Statistiks datasikkerhedspolitik. Findes på følgende LINK: http://www.dst.dk/da/TilSalg/Forskningservice/Dataadgang.aspx |
| <i>Pseudonomiserede data</i> | Det samme princip følger Danmarks Statistik for såkaldt pseudonomiserede data, dvs. data hvor personnummer, virksomhedsnummer eller anden identifikationsnøgle er helt fjernet eller erstattet med et tilfældigt løbenr. Årsagen er, at data som altovervejende hovedregel er meget detaljerede, og at en identifikation ikke kan udelukkes ved at fjerne personnummeret eller andre nøgler. |

Mikrodata til forskningsbrug udleveres derfor ikke, men til gengæld tilbydes forskere m.v. ansat ved autoriserede forsknings- og analysemiljøer on-line adgang til de nødvendige pseudonomiserede mikrodata, som lægges på en forskningsserver på Danmarks Statistik.

2. Organisation

Særligt kontor- FSE Danmarks Statistik har organiseret sin service overfor forskere, analytikere m.fl. i et særligt kontor: Forskningsservice (FSE), der organisatorisk er placeret i Personstatistikafdelingen.

Henvendelse om oparbejdelse af datasæt og om brugen af Danmarks Statistiks forskerordning skal ske til en af medarbejderne i dette kontor. Oversigt over medarbejdere i FSE kan ses her:

<http://www.dst.dk/da/TilSalg/Forskningservice/medarbejdere-iforskningservice.aspx>

Kom tidligt Det er vigtigt at rette henvendelse til FSE så tidligt som muligt i projektforløbet for at få et overblik over datamulighederne og en beregning af omkostningerne. Henvendelsen kan udmærket ske på basis af en foreløbig (kort) projektbeskrivelse.

Et projektindstillingsskema kan downloades via følgende link:

http://www.dst.dk/~media/Kontorer/13-Forskning-ogmetode/Indstilling_eksempel_003.docx?la=da

Det er en god ide, at få et prisoverslag fra FSE inden der ansøges om finansiering.

Tabeller? FSE's medarbejdere udarbejder som hovedregel ikke egentlige tabeller til brug for forsknings- og analyseprojekter. Sådanne leverancer kan efter behov aftales med DST Consulting i Danmarks Statistik, consulting@dst.dk eller 39 17 35 44.

Interviewundersøgelser? Danmarks Statistik kan endvidere også tilbyde at udføre interviewundersøgelser for kunder. Disse undersøgelser aftales særskilt med DST Survey i Danmarks Statistiks Interviewservice, DSTSurvey@dst.dk eller 39 17 32 55.

3. Autorisation

Kun autoriserede miljøer Danmarks Statistik giver kun adgang til mikrodata for autoriserede forsknings- og analysemiljøer. Kun mere varige forsknings- og analysemiljøer med en ansvarlig chef og flere forskere/analytikere kan autoriseres.

Forsknings- og analysemiljøer autoriseres på grundlag af en konkret vurdering. Der kan være visse begrænsninger i private organisationers adgang til data, specielt hvad angår virksomhedsdata.

Ikke autorisation til udenlandske institutioner Danmarks Statistik autoriserer ikke udenlandske forskningsinstitutioner, da det ikke er muligt effektivt at garantere for datasikkerhed mv. Men udenlandske forskere kan få adgang til data, hvis de er tilknyttet et autoriseret dansk forskningsmiljø.

Studerende

Ph.d.-, bachelor- og specialestuderende kan få adgang til mikrodata på lige fod med forskere, forudsat at den pågældende studerende er tilknyttet et autoriseret forskerinstitut samt har en vejleder tilknyttet projektet. Vejlederen skal være autoriseret under forskerordningen.

Modeldatasæt Danmarks Statistik kan udarbejde særlige modeldatasæt til anvendelse i undervisningen. Disse datasæt er små og indholdet ”scramblet”, således at oplysningerne ikke kan genfindes i registrene. Data er beregnet på øvelser i statistisk metode og lignende. Oplysninger om hvilke modeldatasæt, der kan stilles til rådighed, fås ved henvendelse til Forskningsservice.

De detaljerede regler, autorisationsaftaler mv. fremgår af Forskningsservice’s hjemmeside, se følgende link:

<http://www.dst.dk/da/TilSalg/Forskningsservice/Dataadgang/Autorisering>

4. Rammeaftalen mellem Styrelsen for Forskning og Uddannelse

Et væsentligt økonomisk grundlag for forskerserviceordningen er en rammeaftale mellem Styrelsen for Forskning og Uddannelse (SFU) og Danmarks Statistik.

Hvem er omfattet? Rammeaftalen omfatter ydelser til offentlige forskere, dvs. ansatte i offentlige forsker- og analysemiljøer (universiteter, forskningscentre, ministerier, styrelser mv.) samt forskere ansat hos alment velgørende fonde i Danmark. Den private sektor (interesseorganisationer og konsulentvirksomheder m.v.) er derimod ikke omfattet af aftalen.

Rammeaftalen med SFU kan findes på Forskningsservice’s hjemmeside.

5. Projektet

Projektbeskrivelse Medarbejdere i Forskningsservice giver råd og vejledning til forsknings- og analyseprojekter. Den endelige beregning af tilbud på et datasæt kan ske, når den endelige projektbeskrivelse er formuleret.

“Need to know” Adgang til data gives under det såkaldte ”need to know” princip, dvs. at man kan få adgang til de data, der er påkrævet til forskningsformålet. Forsknings- og analysemiljøerne skal derfor dokumentere, at det ønskede dataindhold og projektbeskrivelsen står i rimeligt forhold til hinanden.

6. Registerstrukturen

Registerbegrebet Danmarks Statistik har organiseret sine datasamlinger i såkaldte registre. Betegnelsen ”register” er egentlig forældet og stammer fra den tidlige lovgivning om persondata. Under denne måtte Danmarks Statistik kun opbevare personhenførbare data i særlige datasæt (registre) med et ganske bestemt indhold beskrevet i registerforskriften. Denne administrative procedure er helt ændret med den nye Persondatalov. Registerejeren anmelder ikke længere konkrete datasæt, men derimod de behandlinger, der sker af data. Danmarks Statistik har indtil nu opretholdt den traditionelle betegnelse ”Register” for de konkrete afgrænsede datasæt, som anvendes i statistikproduktionen indenfor de forskellige emneområder, fx befolkningsstatistik, arbejdsmarkedsstatistik mv.

7. Forskernes egne data/eksterne registre

Egne data kan inddrages Det er muligt at inddrage andre registerdata i et projekt end de data, som Danmarks Statistik – som registerejer – råder over. Der kan være tale om egne datasamlinger fra specifikke surveys eller om registre fra andre myndigheder. Sådanne datasæt kan være fra registre ejet af Sundhedsdatastyrelsen (fx Landspatientregistret eller Lægemiddelstatistikregistret) eller ejet af andre offentlige myndigheder. Forskeren skal som hovedregel selv indhente data fra eksterne kilder.

Godkendelse af Lægemiddeldata Kopier af nogle af de mest benyttede registre såsom Landspatientregistret, det Medicinske fødsels- og dødsårsagsregister samt Lægemiddelstatistikregistret findes dog i Danmarks Statistik efter aftale med de pågældende myndigheder. I forbindelse med udtræk af data fra Lægemiddeldatabasen gælder det særlige, at Sundhedsdatastyrelsen skal godkende anvendelsen af lægemiddeloplysninger i projektet. Der lægges i den forbindelse afgørende vægt på, at projektet vurderes at have et væsentligt samfundsrelevant sigte. Godkendelse fås ved at indsende en projektbeskrivelse til Danmarks Statistik, og herefter søger vi for indsendelsen og godkendelsen fra Sundhedsdatastyrelsen. Ansøgningen kan ikke indsendes direkte til Sundhedsdatastyrelsen. Data findes rent fysisk hos Danmarks Statistik.

Registerdata til forskningsprojekter er som hovedregel ikke omfattet af tilladelser fra de etiske komiteer, idet undersøgelserne ikke indebærer personkontakt.

Eksterne data Det er en selvfølge, at de eksterne data skal afleveres til Danmarks Statistik med de nødvendige nøgler (Personnummer m.fl.). Danmarks Statistik vil, efter sammenkobling med andre data, pseudonomisere alle oprindelige nøgler. Det er en forudsætning at det indsendte datasæt ikke indeholder personhenførbare variable. Det kan f.eks. være åbne spørgsmål og svar i en survey. Forskeren skal med andre ord have ”luget ud” i variablene. Eller bede den eksterne leverandør om at gøre det. Når en medarbejder i FSE modtager et eksternt datasæt kigges dette igennem for personhenførbare variable. Disse vil enten blive slettet (åbne spørgsmål i surveys, variable, der ligner nøgler mv.). Eller de vil blive pseudonomiseret. Hvis medarbejderen kan se, at det kræver meget tid at behandle et eksternt datasæt, vil forskeren blive bedt om at indsende et nyt.

8. Dokumentation

En tilstrækkelig og korrekt dokumentation af dataindholdet i de enkelte registre er af afgørende betydning, ikke blot for statistikproduktionen generelt, men naturligvis også for de forskningsresultater, der fremkommer ved forskernes anvendelse af data.

TIMES det bærende dokumentationssystem Danmarks Statistik dokumenterer alle væsentlige registre på Internettet dels i varedeklarerer og dels i det såkaldte TIMES system.

Højkvalitetsdokumentationen For de hyppigst anvendte variable er der blevet udformet en særlig dybdegående og kvalitetssikret dokumentation til registerområderne og variablerne, specielt hvad angår databrud og tidsserier. Denne særlig grundige dokumen-

tation er blevet til i samarbejde med Videnskabsministeriet, og det Koordinerende organ for Registerforskning (KOR). Denne dokumentation har vi kaldt Høj kvalitetsdokumentation, og kan findes på følgende link:

<http://www.dst.dk/TilSalg/Forskningservice/Dokumentation/hkt4Forsker.aspx>

TIMES 3 og TIMES 4

Information og dokumentation om øvrige registre og variable findes i vores dokumentationssystem TIMES. I TIMES 4 findes dokumentation for de variable, der publiceres og anvendes i dag, mens TIMES 3 dokumenterer historiske variable.

TIMES4 findes på følgende adresse:

<http://www.dst.dk/Statistik/dokumentation/times.aspx>

TIMES3 findes på følgende adresse:

<http://www.dst.dk/Statistik/dokumentation/times3.aspx>

Alle dokumentationssystemerne er hierarkisk opdelt i emneområder/statistik områder. Efterhånden som der klikkes frem på siderne vil informationer om de enkelte variable fremkomme, herunder værdisæt for de enkelte variable, beskrivelser af eventuelle databrud etc. Derudover hjælper Forskningservice naturligvis gerne med yderligere råd og vejledning.

9. Hvordan etableres data?

Data til godkendte projekter etableres af medarbejdere i Forskningservice, evt. med støtte fra et eller flere statistikkontorer.

Data pseudonomiseres af Forskningservice, idet identifikationsnøgler erstattes af et neutralt løbenr. Som oftest opsplittes et datasæt i flere, fx med data fra forskellige registre i hver sit datasæt, men naturligvis med en fælles nøgle, som gør det muligt for brugeren at flette data sammen efter den konkrete problemstilling.

Data placeres fysisk på forskningsmaskinerne i Danmarks Statistik.

10. Forskningsmaskinerne

Forskere med adgang til mikrodata afvikler p.t. udelukkende deres kørsler på en af Danmarks Statistiks forskermaskiner eller på maskiner hostet af Danmarks Statistik. Alle maskinerne er adskilt fra Danmarks Statistiks alm. it-konfiguration. Forskerne konkurrerer således ikke med statistikproduktionen om ressourcerne (men selvfølgelig med hinanden). Bemærk at maskinerne befinder sig i Danmarks Statistik, og at forskeren på intet tidspunkt har data placeret på sin egen computer. Forskerne kan alene afvikle kørsler på Danmarks Statistik og kan kun hjemtage resultater af disse kørsler i form af fx tabeller eller anden summeret statistik.

Danmarks Statistik yder ikke bistand til forskernes programmering, men vejleder naturligvis mht. tekniske problemer med opkobling mv. Det forudsættes dog, at den enkelte er selvhjulpne (eller får hjælp af en kollega) et meget langt stykke af vejen.

Større software pakker til rådighed

Der er en række softwareprodukter til rådighed, hvoraf de væsentligste er: SAS, STATA, SPSS, WPS og R.

Andre produkter kan også blive tilgængelige, men forskere kan ikke forvente, at andre produkter kan installeres uden videre, blandt andet fordi licensaftaler kan være så dyre, at budgettet for driften af maskinerne ikke kan bære udgiften. Vi er i Forskningsservice åbne overfor forskernes eventuelle ønsker om særlige behov. Ønsker herom kan rettes til en medarbejder i Forskningsservice.

11. Den elektroniske forskerordning

On-line adgang

De forskere, der får godkendt et projekt, skal anvende Danmarks Statistiks on-line ordning.

Forskeren får ved projektets start udleveret et password, der giver forskeren adgang til de aftalte data på Danmarks Statistiks forskermaskiner. En detaljeret vejledning til hvordan man får adgang til Forskermaskinerne, findes på følgende hjemmeside:

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

Passwords er strengt fortrolige

Alle passwords er strengt fortrolige og forskeren må ikke give uautoriserede personer adgang til data. Overtrædelse heraf vil føre til midlertidig eller varig inddragelse af forskerens adgang til projekter samt midlertidig eller varig inddragelse af institutionens autorisation. Forskningsservices sekretær varetager alle opgaver vedr. kontrakter, passwords, tokens etc.

Output hjemtages som e-post

Tilladt output fra forskerens kørsler lægges i en særlig "udbakke" som tømmes ca. hvert 5. minut. Forskeren modtager det pågældende output som en e-post. Den enkelte e-post meddelelse må ikke overstige 2 MB. Det er dog muligt at zippe filerne. Se yderligere information om hjemsendelse af filer i Nyt fra Forskningsservice, 1. kvartal 2014:

<http://www.dst.dk/ext/370058889/0/forskning/Nyt-fra-Forskningservice-1-kvartal-2014--pdf>

Al post, der hjemsendes af forskerne registreres (logges) af Forskningsservice, og underkastes løbende kontrol mht. overholdelse af datasikkerheden. Overholdes datasikkerheden ikke, kan forskeren eller Institutionen i værste fald miste deres autorisation til data adgang.

12. Et typisk projektforsløb

Nedenfor er et typisk forskningsprojektforsløb skitseret:

Trin 1: Godkendelse af et projekt

En projektindstilling indsendes til Forskningsservice fra den dataansvarlige institution. Projektbeskrivelsen indeholder en formålsbeskrivelse, en præcis beskrivelse af populationen og hvilke registre, der ønskes adgang til via forskerordningen, og en beskrivelse af eventuelle eksterne data, som skal tilknyttes projektet. Den dataansvarlige institution skal bekræfte at den konkrete behandling er lovlig i henhold til databeskyttelsesforordningen artikel 6 inden projektindstillingen kan godkendes af Forskningsservice.

Forskningservices projektindstillingsskabelon kan downloades fra Forskningsservices hjemmeside her:

<https://www.dst.dk/da/TilSalg/Forskningservice/Dataadgang/Projektprættelse>

Trin 2: Dataudtræksbeskrivelse og pris

Forskeren kommer med et udspil til hvilke variable og tidsserier, der ønskes, hvorefter en detaljeret udtræksbeskrivelse udarbejdes. Når der er enighed om dataudtræksbeskrivelsen, udregnes et endeligt pristilbud.

Trin 3: Kontrakt

Der udarbejdes en kontrakt på projektet. Arbejdet med dataudtræk påbegyndes normalt efter modtagelse af underskrevet kontrakt.

Trin 4: Forskeraftaler

Når projektet er godkendt i Forskningservice, underskrives en forskeraftale med hver enkelt deltagende forsker. I den proces oprettes adgangen rent fysisk til forskerserveren, og forskeren får tilsendt passwords. Som udgangspunkt oprettes forskeren i et windowsbaseret miljø, hvor der er mulighed for at benytte SAS, SPSS, STATA, GAUSS, R m.v.

Trin 5: Adgang til data

Gives når forskningservice har oparbejdet de projektspecifikke data. Data leveres i udgangspunktet som SAS data-sæt.

I forbindelse med at indviddata bliver stillet til rådighed for forskere, pseudonomiseres data, således at det ikke er muligt at identificere personer, firmaer, adresser, lægepraksisser mv. Når der indsendes materiale til benyttelse under forskerordningen, vil materialet ligeledes undergå samme pseudonomisering.

Trin 6: Sletning og tilbagelevering af oplysninger

Når analysearbejdet er afsluttet slettes alle personfølsomme data og kopier af data. Sletning af data sker kun på anmodning fra den dataansvarlige institution. Forskningservice tilbageleverer ikke mikrodata.

13. Indsendelse af fortrolige data

Fortrolige data kan leveres til Danmarks Statistik på følgende måde:

1. Data kan sendes via sikker e-post med kryptering
2. Data kan uploades
3. Data kan leveres personligt i Danmarks Statistiks reception
4. Data kan sendes anbefalet til Danmarks Statistik.

Sikker e-post

Hvis data sendes via sikker e-post med kryptering, er det en forudsætning at institutionen har hentet Danmarks Statistiks certifikat. Certifikatet skal hentes på www.dst.dk/certifikater - vælg forskerpost@dst.dk – vcr-formatet anbefales til Outlook. De krypterede data skal sendes til forskerpost@dst.dk med angivelse af projektnummer og kontaktperson i Danmarks Statistik i beskedfeltet. Danmarks Statistik yder ikke support på kryptering og Digital Signatur, der henvises i stedet til www.digitalsignatur.dk

Upload

Under Danmarks Statistiks forskerordning er det muligt at uploade eksterne filer med data til anvendelse i et eksisterende forskningsprojekt. Det er kun data der må uploades. Programmer, dokumentation mv. kan, som sædvanligt, sendes som til den FSE-medarbejder, der er ansvarlig for projektet. Se under følgende link for mere information om upload:

<http://dst.dk/ext/3942639397/0/forskning/Vejledning-til-upload-af-data--pdf>