Registerforskning

- nye muligheder og nye udfordringer

Styrelsen for Forskning
og Innovation

Ministeriet for Forskning, Innovation
og Videregående Uddannelser

Rapport fra udvalget vedrørende reorganisering og styrkelse af registerforskningen

>

Registerforskning - nye muligheder, nye udfordringer
Rapport fra Registerforskningsudvalget

Udgivet af:

Styrelsen for Forskning og Innovation

Bredgade 40
1260 København K

Telefon: +45 3344 6200
Email: fivu@fivu.dk
www.fivu.dk

Publikationen kan hentes på Styrelsen for Forskning og Innovations hjemmeside www.fivu.dk

Omslag: Techno_Background_2
Grafik: Grafixar - Morguefile

ISBN til web: 978-87-92776-87-7

>

Registerforskning - nye muligheder, nye udfordringer

Rapport fra Registerforskningsudvalget

Ministeriet for Forskning, Innovation og Videregående Uddannelser
november 2013

Indhold >

Baggrund for udvalgets arbejde 6

Indledning 8

Udvalgets anbefalinger 13
- Reorganisering af registerforskningsområdet 13
  - Udvidelse af kommissorium for Det Koordinerende Organ for Registerforskning 13
  - Etablering af Referencegruppe for Registerforskning 15
- Anbefalinger i forlængelse af reorganisering 16
  - Kortlægning af registerdata og deres tilgængelighed 16
  - Koordineret indsats i forhold til datavarehuse på velfærdsområderne 17
  - Bedre kapacitetsudnyttelse af computere ved forskerserviceenhederne 20
  - Udleveringspraksis for data til tværnationale forskningsprojekter 21
  - Projekt for udbygning af højkvalitetsdata 22
  - Finansiering af indsamling af longitudinelle surveydata 24
  - Udvikling af Multigenerationsregister 25

Bilag 27
- Bilag 1 – Kommissorium for udvalget 27
- Bilag 2 - Medlemmer af udvalget for reorganisering af registerforskningen 31

>

Baggrund for udvalgets arbejde

>

I *Dansk Roadmap for Forskningsinfrastruktur 2011* er der udtrykt behov for at reorganisere og styrke dansk registerforskning, dvs. den række af aktiviteter der er forbundet med forskeres adgang til og brug af registerdata.

> *"For yderligere at fremme den positive udvikling i forskeres adgang til registerdata søges der etableret en ny og udvidet organisering af registerforskningsområdet. Den nye organisering skal fungere som den samlende danske platform i forhold til at gøre registerdata tilgængelige og anvendelige for forskning samt styrke dialog, integration og koordination på området." (Dansk Roadmap for Forskningsinfrastruktur 2011, side 21)*

Styrelsen for Forskning og Innovation har på den baggrund nedsat et udvalg, som har fået til opgave at udarbejde forslag til fremtidig model for organisering, som kan understøtte udviklingen af registerforskningen i Danmark.

Formålet med udvalgsarbejdet er at komme med anbefalinger, der kan bidrage til at øge det videnskabelige output via registerdata og dermed styrke Danmarks position som registerforskningsnation ved at skabe den fornødne koordination og sammenhæng mellem forskere, registrerende myndigheder og dataansvarlige, og som tager højde for den digitale udvikling. Styrelsen for Forskning og Innovation har vurderet, at det er vigtigt at fastholde en samlende organisering på registerforskningsområdet på tværs af forskningsfelter, og at mål og organisering tilpasses de nye muligheder og udfordringer.

Der er i udvalgets arbejde lagt vægt på at sikre en hensigtsmæssig kontinuitet jf. det hidtidige arbejde i KOR og opbygningen af de eksisterende forskerserviceenheder, samt at den fremtidige organisering ikke bidrager til fragmentering af registerforskningsområdet.

Udvalgets arbejde har fundet sted i foråret og sommeren 2013. Udvalgets rapport blev sendt i høring den 26. september 2013 med svarfrist den 11. oktober 2013.

Det skal bemærkes, at der er afgivet høringssvar, som forholder sig til en række tematikker omkring Statens Serum Institut som forskerserviceorgan, hvilket ikke har været drøftet i løbet af udvalgsarbejdet i forsommeren 2013. Det bemærkes endvidere, at udvalget ikke har forholdt sig til etableringen af STARS* - Strategisk Alliance for Register- og Sundhedsdata, som Sundhedsministeriet annoncerede i september 2013, hvor relevante tematikker i forlængelse af bl.a. disse høringssvar kan drøftes.

Styrelsen for Forskning og Innovation vil gerne takke for de mange input, som vil blive inddraget i det videre arbejde med reorganisering af registerforskningsområdet, i dialog med de involverede organisationer.

>

Der henvises i øvrigt til høringssvarene, som kan findes på Styrelsen for Forskning og Innovation's hjemmeside.

Indledning

>

Danmark er et af de lande i verden, der registrerer flest data om sine borgere. Registre – administrative eller statistiske samlinger af person- eller virksomhedsoplysninger, primært i digital form – udgør derfor en forskningsressource, som sætter Danmark i en international førerposition med helt unikke forskningsmuligheder.

Særligt inden for de sundhedsvidenskabelige og samfundsvidenskabelige felter har anvendelsen af registre en lang tradition. Muligheden for, via det personlige CPR-nummer, at koble data fra forskellige kilder til samme individ gør det muligt at få svar på videnskabelige spørgsmål, som er tæt knyttet til væsentlige velfærdstematikker og kvaliteten og effekterne af offentlig politik. Eksempler herpå kan være: Sammenhængen mellem skole/uddannelse og social mobilitet, langtidseffekter af arbejdsmarkedsforanstaltninger eller sammenhængen mellem livsstil og sundhedstilstand.

Registerforskning er underlagt det særlige vilkår, at det ikke er forskerne selv, der indsamler og ejer de data der anvendes i forskningen. Dette forhold har gennem tiden skabt en række udfordringer og barrierer, som har virket begrænsende på den forskningsmæssige anvendelse af registerdata. Disse er relativt velbeskrevne og dokumenterede gennem mange år.

- Høje udgifter og lange ventetider for etablering af datasæt hos dataansvarlige
- Uens regler og praksis omkring adgang til data hos forskellige dataansvarlige
- Mangelfuld dokumentation og opdatering af registre og data i forhold til kvalitet og validering
- Uklar praksis for genbrug af allerede etablerede datasæt
- Vanskeligheder med at koble data fra flere forskellige kilder
- Manglende synlighed og viden om databaser og deres indhold

Der er mange årsager til, at disse barrierer er opstået. Administrativ praksis vedrørende dataregistrering samt økonomiske og tekniske begrænsninger, de lovgivningsmæssige rammer og den praksis som lovene forvaltes efter, er en del af forklaringen. Barriererne handler også om forskellige institutioners og myndigheders rolle- og ansvarsfordeling samt evne til og fokus på at tilgodese forskeres behov på tværs af store offentlige systemer under politisk og administrativ styring.

I lyset af de mange ressourcer der hvert år bruges på at registrere data, samt de store muligheder disse data rummer, har der været et stærkt fokus på at sikre en bedre forskningsmæssig udnyttelse af registerdata, og begrænse de barrierer der eksisterer. Danmarks Grundforskningsfond iværksatte allerede i 1993 et arbejde for at fremme dansk registerforskning. På den baggrund blev der etableret en filial af Danmarks Statistik i Århus, der blev indkøbt en forskermaskine, placeret på Danmarks Statistik, og der blev etableret en enhed

til arkivering af sundhedsvidenskabelige data. Endeligt blev der etableret et Center for Registerforskning, nu placeret ved Aarhus Universitet. Dette arbejde blev i 2000 fulgt op af IT- og Forskningsministeren, som nedsatte et arbejdsudvalg, der i 2001 udgav rapporten ”Registerforskning – enestående danske muligheder”.

I 2001 blev et stort skridt taget da Det Koordinerende Organ for Registerforskning (KOR) blev oprettet i forlængelse af anbefalingerne fra arbejdsudvalget. KOR har de seneste ti år arbejdet dedikeret for at forbedre forskeres adgang til registerdata, og der er på den baggrund skabt væsentlige fremskridt.

Ikke mindst har etableringen af forskerserviceenheder, delvist finansieret via af de særlige registerforskningsmidler på finansloven, i henholdsvis sundhedssystemet (tidligere ved Sundhedsstyrelsen, nu Statens Serum Institut) og på Danmarks Statistik medført en række betydelige forbedringer. Generering af data til forskere er blevet professionaliseret væsentligt, godkendelsesprocedurerne er blevet smidiggjort, og ikke mindst er leverancetiderne for data faldet betydeligt. Samtidig er prisen for levering af datasæt faldet markant. Prisfaldet skyldes dels en effektivisering af datadannelsen og dels en ændret prispolitik, således at udregningen af pris nu er direkte baseret på ressourceforbruget for konstruktion af datasæt og ikke datasættets størrelse. Samtidig er der etableret mulighed for en sikret ekstern elektronisk adgang til data og computerkapaciteten er udvidet væsentligt. Udbygningen af højkvalitetsdatadokumentation i Danmarks Statistik, finansieret via registerforskningsbevillingen har samtidig medført en kvalitativ styrkelse af forskeranvendelsen på de mest centrale databaser.

Som en direkte effekt af disse tiltag har der siden 2000 været en markant udvikling:

- Forskere kan via en sikret elektronisk adgang tilgå data i Danmarks Statistik
- Forskerne har fået adgang til nye, vigtige datamaterialer som for eksempel lægemiddeldatabasen, der alene i 2012 anvendtes i mere end 500 forskningsprojekter
- Den gennemsnitlige tid fra bestilling af datasæt til data er tilgængelig er faldet fra cirka 50 dage til nu 20 dage både hos Statens Serum Institut og ved Danmarks Statistik
- Ved Danmarks Statistik er antallet af igangværende forskningsprojekter mange-doblet (fra godt 100 til 1100)
- Antallet af forskningsmiljøer der har direkte elektronisk adgang til Danmarks Statistiks registre er steget fra omkring 50 til nu over 350
- Forskerne kan i dag arbejde med data over hele befolkningen

Registerforskningen i Danmark har således gennem mange år været igennem en meget gunstig udvikling, hvor der skridt for skridt er sket en række markante forbedringer af betingelserne for at drive forskning baseret på registre i Danmark. Denne udvikling har været understøttet af de særlige registerforskningsmidler på finansloven og KOR's arbejde og dialog med dataansvarlige institutioner. Effekten har været en forbedret forskningsmæssig udnyttelse af de værdifulde data, som registreres i Danmark.

Nye muligheder – Nye udfordringer

Behovet for at gennemtænke den nuværende organisering af registerforskningsområdet skyldes, sin succes til trods, at der inden for de seneste ti år er sket en kolossal udvikling på det digitale område som på den ene side rummer store udfordringer og kompleksiteter, og på den anden side også rummer meget store muligheder for registerforskningen.

Mængden af data der registreres i dag, er betydeligt større end den var for ti år siden, hvilket især har været understøttet af en væsentlig øget digitalisering inden for den offentlige administration - eksempelvis ved etablering af elektroniske patientjournaler, etablering af online administrative systemer og digitalisering af analoge arkivdata, samt i privat regi, en meget hastig udvikling i den elektroniske registrering af data – mobiltelefoni, økonomitransaktioner, søgehistorik på internettet, etc. Udviklingen i mængden af data må forventes at stige markant i de kommende år.

Digitaliseringen rummer et overordentlig stort forskningsmæssigt og innovativt potentiale. Udnyttelsen af store datamængder (Big Data) ved at gøre offentlige og privatregistrerede data tilgængeligt og anvendelige vil skabe en række muligheder for udvikling af samfundet og i forhold til innovation og kommercialisering i private virksomheder.

Men samtidig er der udfordringer i forhold til kvaliteten af de data der registreres og tilgængeligheden af forskellige datakilder. Registerdata indsamles i udgangspunktet ikke med forskning for øje. Digitaliseringsbølgen er drevet af myndigheders umiddelbare administrative behov og politiske fokus. Risikoen er, at særligt forskningsrelevante data ikke indsamles, eller at data ikke indsamles med de krav til dokumentation og kvalitet som forskningen efterspørger. På den baggrund er der behov for en bedre dialog med de registrerende myndigheder og de dataansvarlige institutioner omkring rammerne for registerforskningen.

Koblingen af data fra flere forskellige kilder stiger markant i disse år. Muligheden for at koble data via CPR-nummer og elektronisk adgang til forskellige databaser har medført, at forskere i stigende grad vil få betydelige muligheder for at koble forskellige datasæt på tværs af registre og forskningsområder. Af samme grund flyder grænserne mellem

forskningsområderne ud. Eksempelvis anvendes social- og sundhedsdata i stigende omfang samlet. Surveydata kobles til social- og sundhedsdata, og grænsen mellem de kliniske discipliner, biologien og epidemiologien er blevet mere og mere flydende. Koblingen af data fra forskellige kilder rummer udfordringer, i forhold til udvikling af nye søgeredskaber og forskningsmetoder, samt kompetencer til at opbygge avancerede datasæt. Dertil kommer et markant stigende behov for computerkapacitet i forhold til at bevare og behandle meget store datamængder og foretage komplicerede søgninger og beregninger. I det hele taget bliver anvendelsen af computere, og dermed e-Science kompetencer en stadig vigtigere del af forskningsprocessen.

Datafortrolighed og databeskyttelse kræver et skærpet fokus i forlængelse af en øget anvendelse af registerdata. De store mængder af nye data fra forskellige kilder, der vil kunne samkøres og stilles til rådighed for forskerne, samt oprettelsen af nye dedikerede forskningsregistre i forskellige forskningsmiljøer, betyder at det er afgørende for mulighederne for at drive registerbaseret forskning, at befolkningen, dvs. dem som data vedrører, kan føle sig helt sikre på, at personhenførbar data om dem ikke slipper ud.

Ikke mindst stiller lovgivningen krav om, at data som udgangspunkt ikke offentliggøres eller videregives i personhenførbar form. Et centralt punkt er her, at der fra et register, der ofte er dannet med bidrag fra forskellige datakilder, ikke kan ske videregivelse af personhenførebare oplysninger, at kun særligt autoriseret personale har adgang til identificerbare oplysninger, samt at data kun anvendes til analyser og statistik, og ikke til enkeltudtræk af personer. Derudover forudsættes det, at oplysninger, der er behandlet i statistisk eller videnskabeligt øjemed, kun kan videregives ske efter forudgående tilladelse fra Datatilsynet.

De nye tiltag, der her lægges op til for at fremme dansk registerforskning, må nødvendigvis gennemføres under særlig hensyntagen til den eksisterende databeskyttelseslovgivning og inden for de rammer, der såvel nationalt som internationalt sættes af den relevante lovgivning, herunder navnlig persondataloven, sikkerhedsbekendtgørelsen, sundhedsloven, cpr-loven og lov om videnskabsetisk behandling af sundhedsvidenskabelige forskningsprojekter. Det bemærkes, at et forslag til en ny forordning for behandling af personfølsomme oplysninger, der pt. forhandles på plads på EU-plan, ligeledes vil have indflydelse på rammerne for dansk registerforskning.

Herudover er det vigtigt, at dataansvarlige kan sikre sig imod, at den reelle risiko for udslip øges, hvilket kræver et øget fokus på systemmæssige og organisatoriske midler til at sikre, at data holdes under kontrol. Derudover er det vigtigt, at manglende overholdelse af lovgivning kan opdages og sanktioneres.

Opbygning af national og europæisk forskningsinfrastruktur er et andet område som har udviklet sig markant inden for de seneste 7-8 år. Ministeriet for Forskning, Innovation og Videregående Uddannelse har over en femårig periode, via Forskningsinfrastrukturpuljen, udmøntet omkring en milliard kroner til national forskningsinfrastruktur og dansk deltagelse i europæisk forskningsinfrastruktur. Derudover har universiteterne på eget initiativ løbende investeret i forskningsinfrastruktur. En række af disse investeringer relaterer sig direkte til registerforskningen, og andre har en berøringsflade hertil.

Der er via den Danske Roadmap for Forskningsinfrastruktur peget på en række nye behov for forskningsinfrastruktur. På det samfundsvidenskabelige område er det særligt udfordringer vedrørende adgang til kobling mellem register- og spørgeskemadata, som efterspørges, samt prioritering og finansiering af longitudinelle undersøgelser – herunder dansk deltagelse i de europæiske projekter ESS (European Social Survey), CESSDA (Council of European Social Science Data Archives) og SHARE (Survey of Health, Ageing and Retirement in Europe). På det sundhedsvidenskabelige område blev der i roadmapprocessen efterspurgt en koordineret anvendelse og opbygning af scannings- og karakteriseringsværktøjer, adgang til beregningsfaciliteteter samt kompetenceopbygning i forhold til databehandling, samt etablering af et multigenerationsregister. Derudover blev der efterspurgt en lettere anvendelse af driftsmæssigt genereret sundhedsdata og bedre koordination på biobankområdet.

Der er dog samtidig taget initiativer til at imødekomme nogle af disse behov. Ministeriet for Forskning, Innovation og Videregående Uddannelser har bidraget med midler til dannelsen af Danmarks Nationale Biobank i regi af Statens Serum Institut, og Ministeriet for Sundhed og Forebyggelse har iværksat en udvikling, der har til formål at styrke anvendelsen af driftsmæssigt genererede sundhedsdata og sikre en bedre national koordinering, også af biobankområdet.

Alt i alt tegner der sig et billede, hvor der er taget en række væsentlige initiativer, men hvor der samtidig er behov for nye strategiske overvejelser om registerforskningen i Danmark og behov at få skabt den bedst mulige koordination af det mangesidige område.

Udvalgets anbefalinger >

I dette kapitel præsenteres en række anbefalinger der efter udvalgets vurdering vil bidrage til en betydeligt forbedret forskningsanvendelse af registerdata.

Nogle af anbefalingerne er af organisatorisk karakter og vil kunne iværksættes inden for en kort tidshorisont. Disse anbefalinger er beskrevet under afsnittet Reorganisering af registerforskningsområdet.

Efterfølgende er der beskrevet en række anbefalinger, som forudsættes igangsat efter at en reorganisering har fundet sted. En del af disse anbefalinger, skitserer de udfordringer, som bør behandles i regi af den nye organisering, mens andre er anbefalinger til konkrete udviklingsaktiviteter, som bør igangsættes.

Skal Danmark drage nytte af de forskningsmæssige muligheder vi har, er der behov for yderligere finansiering – dels i form af engangsinvesteringer til kvalitetssikring og dokumentation og opbygning af nye værktøjer for registerforskningen, og dels til et generelt løft af den finanslovsbevilling til registerforskning, som de seneste ti år har medført en dokumenteret effektiviseringsgevinst i udnyttelsen af registerdata til forskning.

Reorganisering af registerforskningsområdet

Udvidelse af kommissorium for Det Koordinerende Organ for Registerforskning

Udvalget anbefaler, at mandatet for Det Koordinerende Organ for Registerforskning revideres med det formål at skabe et bredere kommissorium, der matcher udviklingen inden for området.

Det Koordinerende Organ for Registerforskning har igennem de sidste ti år bidraget til en meget positiv udvikling i forhold til vilkårene for forskeres adgang til registerdata. Etableringen og drift af Forskerserviceenheder baseret på aftaler formuleret mellem KOR og Danmarks Statistik og nu Statens Serum Institut, har medført en betydelig effektiviseringsgevinst i forhold til adgang til registerdata. Det er på den baggrund vigtigt fortsat at sikre kontinuitet i forhold til den udvikling KOR har bidraget til – herunder sikre en fortsat udvikling af Forskerserviceenheder ved Danmarks Statistik og Statens Serum Institut i overensstemmelse med forskernes behov.

Det anbefales desuden, at det nye Koordinerende Organ for Registerforskning, som det nuværende, fortsat består af anerkendte forskere, der arbejder med registerdata og udpeges på baggrund af indstilling fra Det Fri Forskningsråd. Det nye organ bør dog have flere medlemmer end de nuværende syv for at tilgodese bredden i registerforskningen og skabe større sammenhæng og interaktion til forskermiljøerne.

For at styrke dialogen mellem serviceleverandører og KOR anbefaler udvalget, at de centrale dataansvarlige gives mulighed for at deltage i KOR's møder. Dette drejer sig på nuværende tidspunkt om Danmarks Statistik, Statens Serum Institut samt Statens Arkiver / Dansk Data Arkiv.

Endeligt anbefaler udvalget, at sekretariatsfunktionen styrkes, således at det, i modsætning til i dag, sikres at der til stadighed er tilstrækkelige ressourcer til at løfte de nye opgaver som udviklingen på området tilsiger. Finansieringen kan for eksempel ske gennem et løft i finanslovsbevillingen til registerforskning.

En væsentlig drivkraft for reorganisering af registerforskningsområdet er at skabe bedre dialog og koordinering mellem forskere, registrerende myndigheder og dataansvarlige, samt iværksætte tiltag der skridt for skridt kan forbedre rammerne for forskningen, men det kræver at der etableres en faglig sekretariatsfunktion, som både har kompetencer og ressourcer til at understøtte samarbejdet og koordinationen mellem dataansvarlige og forskere. Opgaveporteføljen for sekretariatet forventes at være:

- Medvirke til at virkeliggøre anbefalingerne i denne rapport
- Sekretariatsbetjening af KOR og dennes referencegruppe (se næste anbefaling)
- Fremsætte forslag til initiativer for KOR og referencegruppen og efter behov tage ansvar for fremdrift i initiativer, herunder fundraising og projektledelse
- Sikre sammenhæng mellem forskellige initiativer på området
- Aktivt kommunikere information til forskersamfundet
- Konsekvensanalysere effekterne af ny lovgivning og praksis på området for registerdata
- Gennemføre en kortlægning af de væsentligste datakilder for forskningen

Sekretariatet bør ideelt have en bred kontaktflade i forskersamfundet og ressourcer til at gå i dialog med forskere og formidle ny viden til forskersamfundet. Sekretariatet bør ligeledes have en tilstrækkelig velfunderet faglig indsigt i registerdata, samt de legale rammer for at lagre og anvende data i forskningssammenhænge, samt et solidt overblik over eksisterende digitale og analoge dataressourcer.

Udvalget har drøftet fordele og ulemper vedrørende to principielt forskellige modeller for forankring af sekretariatet. I den ene model placeres det faglige sekretariat ved den institution, hvorfra formanden kommer. I den anden model placeres det faglige sekretariat fast forankret ved en relevant institution. Den endelige beslutning af model må efter udvalgets opfattelse afhænge af ambitionsniveau og muligheder omkring sekretariatets opgaver og bemanding.

Etablering af Referencegruppe for Registerforskning

Udvalget anbefaler, at der etableres en Referencegruppe for Registerforskning, bestående af repræsentanter for de offentlige myndigheder og institutioner der ejer, indhenter og anvender data. Dertil kommer eventuelt de private registerejere, som måtte have interesse i at gøre deres data tilgængelige for forskningen.

Referencegruppen etableres med det formål at skabe samarbejde og synergi omkring registeranvendelse mellem de aktører, der repræsenterer registerejere, samt dataansvarlige og –anvendere, herunder forskere via KOR. Referencegruppen skal først og fremmest øge den gensidige informationsdeling og koordination på området for registeranvendelse. Gruppen skal ligeledes løbende drøfte mulighederne for at iværksætte tiltag, der kan bidrage til at øge mængden, tilgængeligheden og kvaliteten af registerdata og relaterede forskningsdatabaser.

Grundlaget for referencegruppen er den fælles interesse forskere og registrerende myndigheder har i, at forbedre udnyttelsen og kvaliteten af registerdata. De registerdata forskere udnytter, har først og fremmest et anvendelsesformål for offentlige institutioner og myndigheder i administrative systemer til statistik og monitorering og for at skabe bedre vidensgrundlag for politiske beslutninger. Det må forventes, at det offentlige velfærdssystem i stadig stigende grad vil blive mødt med krav om at basere sin virksomhed på dokumentation og evidens.

Selvom forskerne kan bidrage med at højne kvaliteten og tilgængeligheden af registerdata, er det i sidste ende de registrerende myndigheder, der må vurdere hvorvidt, det er muligt at tilgodese forskernes behov, i forhold til de administrative og tekniske ressourcer, der er forbundet med at registrere, opbevare og sikre adgang til data. Forskernes behov vil til enhver tid være underlagt de offentlige organisationers registrerings- og dokumentationspraksis.

En vigtig pointe er, at forskerne ikke kun er brugere af registerdata, men at forskerne også er drivende for en positiv udvikling i både kvalitet, tilgængelighed og i sidste ende udnyttelse af data for en lang række andre brugere. Eksempelvis har udviklingen af højkvalitetsdata i samarbejde mellem KOR og Danmarks Statistik været en stor succes, ikke blot for forskerne, men også for andre brugere med behov for højkvalitetsdata. Desuden er ansvaret for den nationale sundhedsdokumentation blevet overdraget til Statens Serum Institut. Formålet er blandt andet at udvikle sundhedsdokumentationen i forhold til videnskabelige standarder og behov i tæt tilknytning til aktive forskningsmiljøer.

Referencegruppen for Registerforskning skal inddrage de væsentligste ministerier på området, herunder Sundhedsministeriet, Undervisningsministeriet, Social- og Integrationsministeriet, Erhvervs- og Vækstministeriet, Beskæftigelsesministeriet samt Uddannelsesministeriet og Danske Universiteter. Herudover bør Finansministeriet eventuelt ved Digitaliseringsstyrelsen samt Økonomi- og Indenrigsministeriet inddrages i Referencegruppen.

Samarbejdsformen mellem KOR, dataansvarlige og referencegruppen skal afklares i samarbejde med de involverede partnere.

Anbefalinger i forlængelse af reorganisering

Kortlægning af registerdata og deres tilgængelighed

Udvalget anbefaler, at der gennemføres en kortlægningsanalyse af datakilder. I første omgang for at skabe et grundlag, der på konkret plan vil gøre det muligt at drøfte muligheder og udvikling af området.

I flere sammenhænge, senest ved ”*Workshop om fremtiden for dansk for dansk registerforskning*” afholdt i november 2011, blev det nævnt, at forskerverdenen savner overblik over de datakilder, der ikke er umiddelbart tilgængelige via de store statistikdatabaser. Det er centralt, at forskerne forholdsvis enkelt kan få et overblik over, hvilke databaser der eksisterer, og som kan bidrage til at løse deres forskningsspørgsmål. Mange værdifulde registerdata efterlades uanvendt af forskere, fordi data er svært tilgængelige.

I nogle tilfælde er udfordringen, at meget få personer har kendskab til eksistensen af en database. Her er en af opgaverne at sørge for, at disse databaser bliver gjort offentligt tilgængelige og sikret for eftertiden. I langt de fleste tilfælde er problemet imidlertid ikke manglende viden om eksistensen af en database, men eksakt viden om databasens konkrete indhold og under hvilke forudsætninger data er blevet registreret. Skal en forsker anvende en database til sin forskning, kræver det dokumentation omkring databasens standard og indhold, således at det er muligt at vurdere, hvorvidt den er anvendelig i forhold til at svare på specifikke forskningsspørgsmål.

Analysen skal tage udgangspunkt i at kortlægge eksisterende datakilder, herunder dokumentation og adgang. Analysen skal have særligt fokus på kortlægning af svært tilgængelige data, herunder eventuelt data genereret i privat regi – eksempelvis adoptionsdata, transaktionsdata, etc. Et udgangspunkt vil være at se på de databaser der hvert år indleveres til Statens Arkiver af offentlige myndigheder, og som pt. i oversigtsform ligger på en samlet liste ved Statens Arkiver.

På længere sigt vil analysen kunne danne grundlag for etablering af en hjemmeside, hvor forskerne via en søgemaskine, kan få overblik over relevante datakilder, deres metadatastruktur, dokumentation samt dataansvarlig aktør og adgangsvilkår på en let og smidig måde. I den forbindelse er der behov for at etablere et samlende overblik, som gør det muligt for forskerne at afsøge de muligheder for at anvende data der eksisterer.

Koordineret indsats i forhold til datavarehuse på velfærdsområderne

Udvalget anbefaler, at der i forlængelse af en reorganisering af registerforskningsområdet, herunder etablering af en referencegruppe med inddragelse af de relevante ressortministerier, påbegyndes et arbejde for at sikre en tværgående koordinering i forhold til de datavarehuse, som pt. er under opbygning i forskellige ministerier. Der er de senere år kommet betydelig fokus på anvendelse af registerdata til styring og analyse, og der er på flere ministerområder igangsat projekter for opbygning af datavarehusløsninger i eget regi - bl.a. for at sikre større viden om effekter af indsatser.

Etableringen af datavarehuse for monitorering og analyse inden for arbejdsmarked, sundhed, uddannelse og socialforhold i regi af ressortministerier rummer store perspektiver for et øget samarbejde mellem forskningen og analytikere, men også for et øget koordinationsbehov, blandt andet for at sikre ensartethed i begrebsanvendelse og i forhold til dokumentationsstandarder.

På sundhedsområdet er der en lang tradition for at anvende registerdata og på opbygning af datavarehuse for sundhedsdata. Der arbejdes øjeblikket med konkrete initiativer som led i Regeringens Vækstplan for sundheds- og Velfærdsløsninger, for at skabe et endnu bedre grundlag for den forskningsmæssige udnyttelse af sundhedsdata, herunder en national strategi for adgang til sundhedsdata og for etableringen af én effektiv indgang til de nationale sundhedsdata. Dette igangsættes i 2013, og vil inddrage relevante offentlige og private aktører.

I forhold til sundhedsområdet rummer en tværgående tilgang et stort potentiale, hvor Danmark grundet sine dataressourcer og ekspertise inden for sundheds- og biovidenskabelig forskning står meget stærkt. Ikke mindst vil bedre adgang til data om diagnostik, behandling, rehabilitering og forebyggelse baseret på bl.a. kliniske databaser understøtte udviklingen af dansk klinisk og epidemiologisk forskning. Ligeledes er der inden for den integrerede biovidenskab et stigende fokus på at anvende registerdata inden for sundheds- og socialområdet tværgående med biologisk forskningsdata.

Danmark investerer pt. væsentlige ressourcer i opbygning af forskningsinfrastruktur og e-Science redskaber til anvendelse inden for biologi, fysiologi, medicin og bioteknologi. Der er tale om en bred vifte af værktøjer inden for molekylær billeddannelse, avanceret mikroskopi, single- og multimodale scannningsværktøjer, biobanker og ikke mindst etablering af

europæiske forskningsinfrastrukturer, herunder neutronstrålefaciliteten European Spallation Source i Lund og bioinformatik projektet ELIXIR. Opbygningen af disse faciliteter og tilhørende databehandlingsfaciliteter vil få stor betydning for studier på subcellulært og atomart niveau - for studier af gener, metabolisme og proteiner, af receptorer, membraner og proteinstrukturer, for modellering og visualisering af organer og biologiske processer.

Adgang til sundhedsdata må derfor betragtes som én delkomponent i et samlet kompleks af forskellige typer af forskningsinfrastruktur inden for den integrerede biovidenskab og sundhedsforskning. Mulighederne for at omsætte forskning til translationel udnyttelse i sundhedssektoren samt udvikling af eksempelvis nye lægemidler i samarbejde mellem industri, sundhedsvæsen og forskningssystem vil i væsentligt grad være afhængig af en helhedsstrategisk satsning. Denne må afspejle sig i udmøntningen af initiativerne i regeringens vækstplan og kompetenceopbygningen og investeringer i forskningsinfrastruktur i forskningsverdenen.

På beskæftigelsesområdet har Beskæftigelsesministeriet gennem en lang årrække haft adgang til et datavarehus, hvori der indgår en lang række centrale data om personer, der er omfattet af beskæftigelsesindsatsen i jobcentre, a-kasser og andre aktører. Oplysningerne indsamles på individniveau og omfatter dels data om beskæftigelsesindsatsen som Beskæftigelsesministeriet selv er ansvarlig for at indsamle direkte hos de forskellige aktører i beskæftigelsesindsatsen dels centrale data indhentet fra en lang række andre offentlige myndigheder.

Data indsamles løbende og omfatter både historiske og aktuelle oplysninger, der anvendes til analyse og statistikformål i forbindelse med understøtning af den beskæftigelsespolitiske indsats samt betjening af folketing og minister.

Såvel historiske som aktuelle data om beskæftigelsesindsatsen udstilles løbende i Beskæftigelsesministeriets statistikbank *www.jobindsats.dk* med henblik på at skabe fuld gennemsigtighed i beskæftigelsesindsatsen og mulighed for benchmarking mellem de beskæftigelsespolitiske aktører. Data er endvidere samkørt i Beskæftigelsesministeriets forløbsdatabase DREAM, hvor den ugentlige arbejdsmarkedsstatus for alle personer, der gennem årene har været omfattet af beskæftigelsesindsatsen, er opgjort. DREAM er opgjort tilbage til 1992 og kan bl.a. anvendes til forskningsformål.

Data vedr. beskæftigelsesindsatsen som Beskæftigelsesministeriet er ansvarlig for at indsamle videresendes til Danmarks Statistik, hvor data blandt andet kan tilgås til forskningsformål via Danmarks Statistik's forskeradgang.

På uddannelsesområdet fremgår det af Regeringens 2020 plan, at der skal etableres et fælles datavarehus for uddannelsessektoren, der indeholder data om

studerendes forløb, omkostninger, resultater mv. i uddannelsessystemet med henblik på at give en bedre indsigt i effekterne af den uddannelsespolitiske indsats, uddannelseskvaliteten og anvendelsen af de statslige tilskud til uddannelse.

Ministeriet for Børn og Unge blev ansvarlig for at gennemføre projektet. Realisering af et datavarehus omfattende alle uddannelser under Ministeriet for Børn og Unge er vurderet til at strække sig over 4-5 år. I en fase 2 inddrages de øvrige uddannelsesområder under andre ministeriers ressort – herunder uddannelser under Uddannelsesministeriet.

På socialområdet er der i foråret 2013 igangsat en undersøgelse af mulighederne for at etablere en samlet datavarehusløsning. Der er de senere år igangsat flere initiativer for at forbedre dokumentationen af sociale indsatser, herunder nye indberetninger til Danmarks Statistik om indsatsen i forhold til ældre, handicappede og udsatte børn. Det er en udvikling, der også understøttes af en øget anvendelse af digitale sagsbehandlingssystemer, såsom elektroniske omsorgsjournaler og digitale sagsbehandlingssystemer i forhold til udsatte børn (DUBU) og handicappede (DHUV).

I kommunalt regi er der udviklet et samlet nøgletalssystem (FLIS), der giver de deltagende kommuner adgang til egne data fra kommunernes fagsystemer. FLIS systemet vil løbende blive udvidet med flere fagområder. Ligeledes er der i regi af Kommunernes og Regionernes Analyse (KORA) etableret et nøgletalssystem (ECO) som giver et videnbaseret grundlag for at vurdere en kommunes rammebetingelser, finansielle dispositioner og økonomiske balance, udgiftspolitik, serviceniveau med mere.

Tværgående udfordringer
Udvalget anbefaler at der i forlængelse af opbygning af ministerielle datavarehuse sikres en tværgående koordinering. Konkret bør der tages hånd om at sikre mere aktuelle data og forbedre datakvaliteten, herunder sikre anvendelse af ensartede definitioner mv. Uden koordinering er risikoen, at ”man ikke taler samme sprog” – eksempelvis at data med samme betegnelse ikke udtrykker samme værdi, fordi de er genereret i forskellige systemer.

Ligeledes kan information være registreret på en måde, der ikke umiddelbart gør data anvendelige til statistiske analyser. En problemstilling i den sammenhæng er, at der ikke på alle områder er tale om landsdækkende systemer. I den sammenhæng er kortlægning af datakilder, som foreslået af udvalget, et vigtigt redskab til at afdække svært tilgængelige data.

Samarbejde mellem forskere og ministerier omkring datadokumentation og begrebsstandarder og algoritmer vil ikke blot styrke forskningsanvendelsen, men også øge datakvaliteten og dermed nytten for de ministerielle analyseenheder betydeligt.

I regi af Referencegruppen for Det Koordinerende Organ for Registerforskning bør det drøftes, hvordan og på hvilke vilkår forskere kan få adgang til databaser i ministerielle datavarehuse, samt hvordan der sikres en tværgående adgang til data for at se på sammenhæng mellem sundhed, beskæftigelse, sociale forhold og uddannelse.

For at sikre en optimal udnyttelse af den konkurrencefordel Danmark har, anbefaler udvalget, at forskere i videst muligt omfang sikres adgang til data for hele befolkningen. Der bør ses på muligheden for at registre med stor efterspørgsel tilbydes via forskerserviceenhederne. På den måde undgås den situation, at registerbrugere skal igennem flere forskellige systemer, med uensartet adgangsvilkår og dokumentationssystemer for at koble data. Eksempelvis stiller Danmarks Statistik i forvejen gennem forskerordningen data på velfærdsområdet til rådighed for forskere og analysemiljøer via PSD – den personstatistiske database – der rummer data fra samtlige områder i personstatistik; befolkning, arbejdsmarked, uddannelse og velfærd suppleret med yderligere data på individniveau, og disse data kombineres ofte med forskellige sundhedsdata efter forskernes behov.

Det er i den sammenhæng vigtigt, at forskerserviceenhederne fortsat tilgodeser den tværgående forskning ved løbende at give forskere og analytikere, der både anvender data ved Statens Serum Institut og data ved Danmarks Statistik mulighed for at få leveret datasæt på tværs af de to dataansvarlige aktører på en omkostningseffektiv og smidig måde.

Udvalget har endvidere drøftet muligheden for at finde en løsning, hvor Danmarks Statistik oparbejder grunddata, som fx. køn, etnicitet, højst fuldførte uddannelse, forældres uddannelse, arbejdsmarkedsstatus, og stiller dem til rådighed for de store datavarehusløsninger, og hvor ansvaret for sikkerheden omkring disse oplysninger kunne indgå i den samlede sikkerhed omkring de store datavarehuse, under forudsætning af, at de statistisk indsamlede data sikres mod spredning. Mulighederne for at etablere et sådan initiativ må videre behandles i samarbejde med de involverede parter. Det bemærkes yderligere, at formål og juridisk grundlag for et sådan initiativ, herunder forenelighed med persondataloven, skal afklares.

Bedre kapacitetsudnyttelse af computere ved forskerserviceenhederne

Udvalget anbefaler, at forskerserviceenhederne i samarbejde med de danske universiteter og forskningsinstitutioner, samt eventuelt andre analytikere gennemfører et projekt for at skabe en bedre kapacitetsudnyttelse på de computere der anvendes ved forskerserviceenhederne.

Aktuelt er det en udfordring under den nuværende registerforskningsbevilling, at sikre tilstrækkelig kapacitet på Forskermaskinen ved Danmarks Statistik. Derfor har mange forskningsinstitutioner og andre organisationer installeret

egne servere på Danmarks Statistik parallelt med Forskermaskinen. Pt. er der installeret over 70 separate servere ejet af forskellige forskningsmiljøer. Dette for at sikre, at der er tilstrækkelig beregningskapacitet tilstede for de enkelte institutioner uafhængigt af de perioder med spidsbelastning, som der er på Forskermaskinen. Dette er dog både en meget uøkonomisk måde at skaffe beregningskapacitet på og en meget dårlig ressourceudnyttelse. De servere som forskergrupper finansierer på Danmarks Statistik har samlet set en kapacitet, der er langt større end behovet, og som rigeligt vil kunne håndtere spidsbelastningsperioder.

Derfor burde der kunne skabes en bedre og mere omkostningseffektiv model for anskaffelse og drift af beregningsfaciliteter på Danmarks Statistik og på Statens Serum Institut. I stedet for, at institutionerne hver især finansierer egen beregningskapacitet, kunne der laves en fælles IT-arkitektur og langsigtet investeringsplan for forskermaskinerne på henholdsvis Statens Serum Institut og Danmarks Statistik. Dette vil sikre en langt mere effektiv computerudnyttelse og en stabil kapacitetsudvidelse for færre penge.

Konkret anbefaler udvalget, at KOR henvender sig til DEIC – Danish e-Infrastructure Cooperation, for at undersøge mulighederne for, at de fornødne IT-investeringer på forskerserviceenhederne løftes gennem DEIC. DEIC er den organisation i Danmark som på vegne af danske forsknings- og uddannelsesinstitutioner håndterer forskeres adgang til e-infrastruktur – herunder indkøb af hardware.

I samarbejde med DEIC kunne forskerserviceenhederne ligeledes forsøge at få lavet bedre licensaftaler for det programmel som anvendes ved forskermaskinerne. Både forskerserviceenheden ved Statens Serum Institut og ved Danmarks Statistik har oplevet voldsomme prisstigninger på licenser. Dette på trods af, at de forskere der anvender programmerne, allerede er omfattet af en licensaftale via deres hjeminstitution.

Udleveringspraksis for data til tværnationale forskningsprojekter

Udvalget anbefaler at der tages skridt til at sikre, at danske registerdata mere effektivt kan anvendes i forbindelse med internationale forskningsprojekter med særligt fokus på Norden og Europa.

Udfordringen er, at en række væsentlige forskningsspørgsmål er vanskelige at svare på, fordi der, på baggrund af danske data, ikke er tilstrækkeligt stort datagrundlag til at udlede sammenhænge med tilstrækkelig statistisk sikkerhed. I takt med at forståelsen af sociale sammenhænge eksempelvis mellem levevis, sociale forhold, sygdom, sundhed og genetik udvikles, opstår der et behov for at fokusere og indsnævre udvælgelsen af det individdata, der indgår i undersøgelser, men derved stiger også behovet for at sikre et tilstrækkeligt stort datagrundlag.

Inden for sundhedsområdet er der imellem de nordiske lande en lang tradition for forskningssamarbejde, og det er her muligt at arbejde og udveksle ikke blot aggregerede data, men også individbaserede data med henblik på f.eks. fælles kliniske forsøg og forskningsprojekter med involvering af biologiske materialer etc. Inden for samfundsvidenskaberne ville der kunne opnås fordele ved, at man kunne udveksle data på lignende vis.

Der er ikke en principiel juridisk barriere for at udlevere registerdata over landegrænserne inden for EU til forskningsformål, hvis sikkerhedskravene kan opfyldes. På den baggrund bør den nuværende praksis ses efter for at finde en teknisk og/eller organisatorisk løsning i tæt samarbejde med Danmarks Statistik, der vil muliggøre en mere smidig anvendelse af danske registerdata i nordiske og europæiske forskningssamarbejder, under forudsætning af at den relevante lovgivning overholdes.

Pt. er der en forholdsvis restriktiv praksis i forhold til data forvaltet af Danmarks Statistik, hvor data fra flere forskellige lande kun kan udnyttes såfremt disse lægges samlet på Danmarks Statistik's forskermaskine. I de tilfælde hvor dette ikke kan lade sig gøre, kan danske data ikke udnyttes og muligheden for at sammenligne danske data med data fra andre lande besværliggøres eller mistes helt.

Det skal dog bemærkes, at Danmarks Statistik konstruktivt har bidraget til løsninger på nordisk plan, dels ved at facilitere de ønsker, der har været fremsat fra forskere om tværnordiske datasæt, og dels ved at deltage i et arbejde i nordisk regi, hvor man søger at skabe grundlag for at give lettere adgang til registerdata på tværs af de nordiske lande. Visionen er, at sikre adgang til registerdata på grundlag af de nordiske landes 25 millioner indbyggere, hvilket vil rumme store muligheder for forskning inden for sundhedsområdet samt komparativ samfundsforskning.

Det skal understreges at den restriktive politik for dataadgang forskerne oplever, sker ud fra et forsigtighedsprincip i overensstemmelse med det ansvar Danmarks Statistik og andre nationale dataansvarlige har i forhold til at forvalte personhenførbare data, jf. også bl.a. bestemmelserne i persondataloven, der bygger på EU-direktiv 95/46/EF. I sidste ende er forskningen dybt afhængig af at nyde befolkningens tillid i forhold til anvendelsen af personidentificerbare data, og særligt at der ikke sker uautoriseret dataudlevering. I den sammenhæng må arbejdet med en ny EU-forordning om databeskyttelse følges nøje.

Projekt for udbygning af højkvalitetsdata

For forskere er det afgørende at få let og billig adgang til højkvalitetsdata (HKD). Fra forskersiden er der et klart ønske om, at der tilføres ressourcer til at styrke opbygningen af højkvalitetsdata, således som KOR har samarbejdet med Danmarks Statistik om gennem flere år. Udviklingen af højkvalitetsdata på Danmarks Statistik har været en meget stor succes, som har bidraget til at

nyttiggøre statistikdata ikke blot for forskere, men også for de offentlige myndigheder og institutioner som anvender statistisk data i deres arbejde. På den baggrund er det vigtigt at fortsætte HKD-udbygningen ved Danmarks Statistik, og igangsætte HKD-arbejdet ved Statens Serum Institut.

Det er afgørende, at data dokumenteres tilbage i tid med særlig fokus på at dokumentere databrud. Der bør medtages variabler for så lang en tidsperiode som muligt. Herved kan der gennemføres panelanalyser, der i international sammenhæng vil være helt unikke.

At data er veldokumenterede er med til at øge og sikre kvaliteten af forskningen, idet risikoen for fejltolkninger mindskes, ligesom arbejdet med registerdata vil blive lettere og vil kunne bidrage til en betydelig samfundsmæssig effektiviseringsgevinst. Et stort antal analytikere og forskere skal hver bruge væsentligt færre arbejdstimer på at sætte sig ind i en variabel, hvis den er veldokumenteret og alt relevant information er samlet et sted. Veldokumenterede data vil i sig selv øge produktiviteten i forskningsservice-enhederne, da forskere og analytikere har mindre behov for at få vejledning inden bestilling af data og ved efterfølgende brug af data.

Fælles konstruktion af nye beregningsalgoritmer/komplicerede variabler på områder, hvor mange analytikere og forskere gennemfører analyser, vil også bidrage til en betydelig tidsbesparelse. På arbejdsmarkedsområdet er der mange registerbrugere, som hver især bruger ganske megen tid på at konstruere algoritmer for eksempelvis at definere, hvorvidt en given person i arbejdstyrken er i beskæftigelse i en given uge eller måned. Ved at samle udviklingen af sådanne algoritmer ét sted, vil der blive sparet et stort antal arbejdstimer, og kvaliteten af forskning og analyse bliver væsentligt forøget, fordi alle anvender det samme datagrundlag. Andre eksempler kan være konstruktion af pendlingsoplande, definition af disponibel indkomst, hvilke diagnoser for hospitalsindlæggelser og koder for medicinforbrug, der bedst kan bruges til at afgrænse en given sygdom mv.

Udover at konstruere helt nye variabler er der behov for at reparere på nogle af de variabler, hvor der er databrud over tid. Reparation af databrud kan være ganske tidskrævende for den enkelte forsker, og der er stor risiko for, at forskellige forskere gør det forskelligt. Danmarks Statistik har i forbindelse med databrud i flere tilfælde lavet konsistente opgørelser over tid. Det vil være en stor fordel, hvis dette princip kunne etableres i alle de tilfælde, hvor der forbindelse med overgangen til en ny opgørelsesmetode er vanskeligt at føre den nye opgørelsesmetode tilbage i tid. Der bør desuden etableres feedback mekanismer, således at der sker en hurtigere vidensudveksling om og opfølgning på rapporterede fejl i HKD baseret på brugernes erfaringer med de enkelte variabler. Heri ligger en betydelig effektiviseringsgevinst, som dog kræver, at der i forskersamfundet skabes en forståelse af vigtigheden af dette.

De senere år har KOR haft mulighed for at investere mellem en halv og en hel million årligt til HKD, svarende til mellem 20 og 40 variabler. Sammenlagt er nu godt 600 variable højkvalitetsdokumenterede ved Danmarks Statistik. Baseret på erfaringerne med højkvalitetsdokumentation har prisen for dokumentation af en enkelt variabel ligget på omkring 25-30.000 kroner. Dette skal ses i forhold til at der på henholdsvis Danmarks Statistik og Staten Serum Institut sammenlagt, efter et forsigtigt skøn, er op mod 15.000 variabler som kunne højkvalitetsdokumenteres, omend det skal understreges, at det ikke er målet, at alle variabler skal være højkvalitetsdokumenterede.

På trods af den klare effektiviserings- og forskningsgevinst der ligger i HKD, er der en kadence, hvormed det er muligt at opbygge HKD data, som er svær at forcere. På den baggrund anbefaler udvalget, at der afsættes en fast årlig bevilling som prioriteres af KOR til anvendelse til HKD på Statens Serum Institut og Danmarks Statistik med en beløbsstørrelse svarende til HKD for 100-125 variabler. Denne kadence skønnes at være en hensigtsmæssig kadence for HKD-opbygning, i forhold til det træk på ressourcepersoner i forskerverdenen og i forskerserviceenhederne som opbygningen af HKD kræver.

Finansiering af indsamling af longitudinelle surveydata

Udvalget anbefaler, at det Nationale Udvalg for Forskningsinfrastruktur drøfter, hvordan og i hvilket omfang Danmark kan deltage i longitudinelle surveydataindsamlinger, eventuelt i samarbejde med relevante ressortministerier.

Indsamling af longitudinelle surveydata (data indsamlet over lange tidshorisonter) har igennem en årrække været finansieret via forskningsrådsbevillinger. Imidlertid er det selvstændige virkemiddel for databaser som finansierede indsamling af longitudinelle data nu afskaffet i regi af Forskningsrådet for Samfund og Erhverv.

Hvis der ikke findes en finansieringsløsning for indsamling af longitudinelle data, vil social surveydata fra Danmark fremover ikke indgå i de surveydatabanker, som Danmark i mange år har deltaget i - herunder European Social Survey samt Survey of Health, Ageing and Retirement in Europe. Da indsamling af surveydata ofte har til formål at give et øjebliksbillede, eksempelvis befolkningens holdninger til indvandrere i Europa, og måle udvikling over tid, er det vigtigt at der med jævne mellemrum indsamles data. Står Danmark uden for en dataindsamlingskørsel kan dette datasæt ikke reetableres. Konsekvensen er, at danske og europæiske forskere går glip af muligheden for at sammenligne danske data med data fra andre lande.

På den baggrund bør Det Nationale Udvalg for Forskningsinfrastruktur drøfte muligheden for finansiering af dansk deltagelse i surveydataindsamlinger, og eventuelt på en case by case basis tage stilling til, hvorvidt Danmark bør deltage. I forbindelse med gennemførelsen af longitudinelle undersøgelser og andre undersøgelser, hvor behovet for langvarig opbevaring af personhenførbar

forskningsdata er påkrævet, bør anvendelsen af mulighederne for arkivering af registerdata i Statens Arkiver ved Dansk Data Arkiv styrkes.

Udvikling af Multigenerationsregister

CPR-registret indeholder information om alle personer bosiddende i Danmark fra 1968 og frem. Registret har vist sin værdi i forbindelse med talrige store befolkningsundersøgelser af årsagerne til ophobning af sygdomme i familier, samt alle slags forskning, hvor familieforhold spiller en rolle.

For alle personer født i Danmark 1960 eller senere indeholder CPR-registret oplysning om identiteten af mor og i langt de fleste tilfælde også far. For personer født før 1950 er denne information imidlertid kun tilgængelig i meget få tilfælde. Blandt alle personer født i Danmark og som er registreret i CPR, er der over to millioner født 1920-1959, hvortil der ikke er oplysning om identiteten af mor. Konsekvensen er, at man for langt de fleste sygdomme, kun kan beskrive den familiære ophobning af sygdomme eller hændelser indenfor to generationer.

En udvidelse af generationsoplysningerne i CPR-registeret vil give dansk forskning en unik mulighed for at studere familiære ophobning af sygdomme eller andre hændelser indenfor 3 eller flere generationer. Muligheden for at koble befolkningen sammen i familier kan blive et afgørende løft i de unikke muligheder vi har for befolkningsbaseret forskning. Dette er påvist i både Island og Sverige, hvor man i dag har de bedst udbyggede slægtskabsoplysninger i elektroniske landsdækkende registre.

I Island har koblingen imellem detaljerede beskrivelser af slægtskabsforhold og genetiske data været uvurderlig for udvikling af nye metoder indenfor populationsgenetik såvel som udforskning af genetiske sygdomsårsager. I Sverige har oprettelsen af et multigenerationsregister betydet helt nye muligheder for både sundheds- og samfundsvidenskabelig forskning.

Danmark har nogle fordele i forhold til begge disse lande: Vi har en befolkning der er ca. 20 gange større end den islandske, med deraf følgende langt bedre muligheder for studier med tilstrækkelig statistisk styrke, og i forhold til Sverige er mulighederne for kobling til biologiske data betydeligt bedre i Danmark.

Konkret vil etablering af et multigenerationsregister kræve indtastning af oplysninger fra kirkebøger og folketællinger, om alle personer født i Danmark 1920-1959 indeholdende alle relevante oplysninger om personen samt personens forældre. Statens Arkiver har allerede foretaget registreringer af mange kirkebøger og folketællinger. Disse registreringer bør inddrages så data ikke skal indtastes på ny, og den konkrete afgrænsning af multigenerationsregisteret må fastlægges, når omfanget af materialer i Statens Arkiver kendes. Også oplysninger fra Danmarks Statistiks fødsels- og dødsfaldsregistre bør inddrages. Jf. Danmark Statistik er der 2.991.452 fødsler i

den omhandlede periode, og det skønnes at oplysningerne kan indtastes på ca. 2 minut pr person.

I 2010 stillede Københavns Universitet, Statens Arkiver og Danmarks Statistik forslag til infrastrukturpuljen om et multigenerationsprojekt, som blev positivt bedømt men ikke opnåede støtte. Det anbefales at projektet fra 2010, samt interessenterne bag det, inddrages i det videre arbejde med udvikling af et multigenerationsregister.

Baseret på de indtastede oplysninger og oplysninger om fulde navne i CPR-registret skal der udvikles software, som kan linke relevante oplysninger i kirkebøgerne og folketællinger med tilsvarende oplysninger i CPR, og som kan danne pseudo-personnumre for personer, der ikke findes i CPR. Det skønnes at udviklingen af et sådant software vil tage ca. 1 år. Der bør herefter gennemføres et pilotstudie og et udviklingsprojekt for software til identifikation af mors og fars personnummer.

Samtidig bør der gennemføres et arbejde, hvor fokus er på, hvorledes relevante data fra kirkebøger kan digitaliseres for de hen ved 3 millioner individer, der er tale om, både i forhold til økonomi og teknisk løsning. Samtidig er det centralt at det juridiske myndighedsansvar afklares i forhold til etablering af et multigenerationsregister. Det må i den forbindelse overvejes, om der skal etableres et særligt retsgrundlag for registreringen, ligesom spørgsmålet om oplysningspligt i forhold til de registrerede må overvejes.

Bilag 1 – Kommissorium for udvalget

Kommissorium for Udvalget vedrørende reorganisering og styrkelse af dansk registerforskning

I *Dansk Roadmap for Forskningsinfrastruktur 2011*, er der peget på behovet for en reorganisering og styrkelse af dansk registerforskning, dvs. den række af aktiviteter, der er forbundet med forskeres adgang til og brug af registerdata.

> *"For yderligere at fremme den positive udvikling i forskeres adgang til registerdata søges der etableret en ny og udvidet organisering af registerforskningsområdet. Den nye organisering skal fungere som den samlende danske platform i forhold til at gøre registerdata tilgængelige og anvendelige for forskning samt styrke dialog, integration og koordinatuion på området." (Dansk Roadmap for Forskningsinfrastruktur 201, side 21)*

Styrelsen for Forskning og Innovation nedsætter på den baggrund et udvalg, som får til opgave at udarbejde forslag til fremtidig model for organisering og finansiering, herunder også udpegning af tekniske løsningsmodeller med videre, som kan understøtte udviklingen af registerforskningen i Danmark.

Baggrund

Registerforskningen i Danmark har gennem de seneste ti år været igennem en meget gunstig udvikling. Siden det Koordinerende Organ for Registerforskning (KOR) blev oprettet i 2001, er det sket ganske betydelige fremskridt i adgangen til registerdata. Ikke mindst har etableringen af Forskerserviceenheder i Sundhedsstyrelsen og på Danmarks Statistik medført en række betydelige forbedringer: Udleveringen af data er blevet professionaliseret væsentligt, godkendelsesprocedurerne er blevet smidiggjort, der er etableret online adgang til databaserne, og ikke mindst er leverancetiderne for data og prisen for datasæt faldet betydeligt, ligesom dokumentation af et stort antal registervariabler er forbedret markant. Denne udvikling havde næppe været mulig uden de særlige registerforskningsmidler på finansloven og KOR's arbejde og tætte dialog med registerejerne. Effekten har været en markant forbedret udnyttelse af de værdifulde data som registreres i Danmark.

I samme periode er der ikke mindst på det digitale område sket en markant udvikling som på den ene side rummer store udfordringer og kompleksiteter, og på den anden side også rummer meget store potentialer for registerforskningen. Mængden af data der registreres i dag er betydeligt større end den var for ti år siden, hvilket især har været understøttet af en øget digitalisering (for eksempel i forbindelse med etablering af nye nationale og

internationale forskningsinfrastrukturer, etablering af elektroniske patientjounaler, digitalisering af analogt arkivdata, etc.).

Digitaliseringen rummer et overordentlig stort forskningsmæssigt potentiale, men samtidig udfordringer i forhold til dokumentation og i forhold til tilgængelighed i form af e-infrastruktur, ligesom det selvfølgelig må holdes for øje, at det altid vil være forbundet med stor usikkerhed at vurdere samfundets og forskernes fremtidige behov og efterspørgsel efter data. Koblingen af data fra flere forskellige kilder stiger markant i for disse år. Muligheden for at koble data via CPR og nemmere adgang til databaser, har medført at forskere i stigende grad vil få betydelige muligheder for at koble forskellige datasæt på tværs af registre og forskningsområder. Af samme grund flyder grænserne mellem forskningsområderne ud. Eksempelvis anvendes social- og sundhedsdata i stigende omfang samlet. Surveydata kobles i stigende omfang til registerdata, og grænsen mellem de kliniske discipliner og epidemiologien er blevet mere og mere flydende.

Alt i alt tegner der sig et billede, hvor der er behov for nye strategiske overvejelser om registerforskningen i Danmark og behov at få skabt den bedst mulige og sammenhængende koordination af det mangesidige område.

Formål og præmisser

Formålet med udvalgsarbejdet er at styrke Danmarks position som registerforskningsnation og det nationale udbytte af de danske registerdata, ved bl.a. at

- øge det videnskabelige output baseret på registerdata
- øge omfanget af registerdata og tilgængeligheden af allerede eksisterende data
- højne kvaliteten af registerforskningen via bl.a. registerdokumentation af høj kvalitet
- øget effektiviteten på registerforskningsområdet ved anvendelse af de mest hensigtsmæssige tekniske platforme, etablering af så vidt muligt konsistente dataserier over tid, algoritmer for komplekse variabler, der anvendes af mange forskere mv.

der efterspørges et forslag til en samlende struktur, der skaber den fornødne koordination og sammenhæng i systemet, og som tager højde for den digitale udvikling

Udvalget skal som minimum behandle følgende forhold:

- Tilgængeligheden, anvendeligheden og kvaliteten af registerdata **(** Overblik, Godkendelsesprocedurer, Uens metadata og bedre dokumentation, Opbygning af Højkvalitetsdata etc)

- Metodiske og teknologiske udviklingsmuligheder med henblik på at maksimere anvendelsen registerdata på en så omkostningseffektiv måde

>

som muligt (Online adgang, Data mining, E-infrastruktur, E-science metodikker, Deklaration på databaser, Uddannelse, etc)

Udvalgets arbejde skal primært fokusere på registerbaseret forskning baseret på data, der kan knyttes til fysiske personer, og hvor data typisk ikke er indsamlet med forskning for øje, eller forskningsdata baseret på personoplysninger hvor personidentifikationen bevares over længere tid. Udvalget skal som minimum inddrage registerdata i form af

- Data opbevaret ved statslige forvaltningsmyndigheder
- Data indsamlet og/eller opbevaret ved kommunale og regionale myndigheder
- Data baseret på biologiske prøver – Biobankdata
- Forsknings- og surveydata der kan kobles til registerdata

Der lægges endvidere vægt på at udvalgets forslag sikrer en hensigtsmæssig kontinuitet (jf. det hidtidige arbejde i KOR og opbygningen de eksisterende forskerserviceenheder) og at den fremtidige organisering ikke bidrager til fragmentering af registerforskningsområdet.

Udvalgets sammensætning

Udvalget sammensættes således:

- Kontorchef Peter Sloth, Styrelsen for Forskning og Innovation (formand)
- Analyse- og Forskningschef Hans Hummelgaard, KORA, Det Nationale Institut for Kommuners og Regioners Analyse og Forskning
- Sektordirektør Mads Melbye, Statens Serum Institut, Sektor for National Sundhedsdokumentation og Forskning
- 1 medlem som udpeges efter indstilling af Danmarks Statistik
- 1 medlem som udpeges efter indstilling fra Digitaliseringsstyrelsen
- 1 medlem som udpeges af Datatilsynet
- 4 medlemmer som udpeges efter indstilling fra Det Koordinerende Organ for Registerforskning
- 4 repræsentanter som udpeges af relevante ressortministerier – Sundhedsministeriet, Social & Integrationsministeriet, Beskæftigelsesministeriet samt Børn & Undervisningsministeriet

Udvalget skal efter behov konsultere relevante myndigheder, registerejere, forskningsinstitutioner, virksomheder samt relevant teknisk bistand.

Sekretariatsfunktionen varetages af Styrelsen for Forskning og Innovation. Formanden/sekretariatet kan anmode udvalgets medlemmer om at tilvejebringe eller udarbejde skriftligt materiale, ligesom der kan afholdes udgifter til køb af tjenesteydelser til brug for udvalgets arbejde.

>

Udvalgets forslag skal være færdiggjort og afleveret med udgangen af maj 2013.

>

Bilag 2 - Medlemmer af udvalget for reorganisering af registerforskningen

- Peter Sloth (Formand), Kontorchef, Kontoret for Forskningsinfrastruktur, Styrelsen for Forskning, og Innovation
- Hans Hummelgaard, Analyse- og Forskningschef KORA - Det nationale Institut for Kommuners og Regioners Analyse og Forskning (Formand for Det Koordinerende Organ for Registerforskning)
- Lisbeth B. Knudsen, Professor, Institut for Sociologi og Socialt Arbejde, Aalborg Universitet
- Henrik Toft Sørensen, Professor, Klinisk Epidemiologisk Afdeling, Institut for Klinisk Medicin, Aarhus Universitet
- Marianne Simonsen, Lektor, Institut for Økonomi, Aarhus Universitet
- Preben Bo Mortensen, Professor, NCRR, National Centre for Register-based Research, Aarhus Universitet
- Mads Melby, Sektordirektør, Sektor for Sundhedsdokumentation og Forskning, Statens Serum Institut
- Henrik Kyvsgaard, Afdelingschef, Børneområdet og Analyser, Ankestyrelsen, p.v.a. Social- og Integrationsministeriet
- Julie Caroline Preuthun, Fuldmægtig, Digitaliseringsstyrelsen
- Kristian Ørnsholt, Kontorchef, Kontor for Uddannelsesanalyse, Departementet, Ministeriet for Børn- og Undervisning
- Marie Rønde, Chefkonsulent, Sygehuspolitik, Ministeriet for Sundhed og Forebyggelse
- Lars Thygesen, Afdelingschef Salg & Marketing, Danmarks Statistik
- Birgit Kleis, Kommitteret, Datatilsynet
- Klaus Henrik Langager, Kontorchef Statistik, Analyse og Overvågning, Arbejdsmarkedsstyrelsen, p.v.a. Beskæftigelsesministeriet

Sekretariat

- Troels Tvedegaard Rasmussen, Specialkonsulent, Kontoret for Forskningsinfrastruktur, Styrelsen for Forskning, og Innovation
- Peter Uffe Meier, Chefkonsulent, Kontoret for Forskningsinfrastruktur, Styrelsen for Forskning, og Innovation