



Dato: 17. maj 2017

## Dansk registerforskning

Danske forskere mener, at der er behov for bedre it-infrastruktur, høj datasikkerhed, bedre datadeling og transparens i brug af data. Desuden er det tid til faglighed, samarbejde og politisk vilje.

Mulighederne for og med registerforskning er bedre end nogensinde før, hvis registerdata kobles med f.eks. biologiske data, elektroniske patientjournaler og andre data fra administrative systemer eller transaktionsdata.

Men det kræver, at myndigheder, universiteter og forskere samarbejder om at skabe det rette grundlag. Der stilles bl.a. høje krav til datasikkerhed og beregningskapacitet. Det skal sikres juridisk og teknisk, at de enkelte dataansvarlige organisationer har kontrol over, hvilke forskere der får adgang til data, og hvad det bruges til, samtidig med at forskerne får de rette muligheder for at alle relevante datakilder kan analyseres samlet.

Forskerne anbefaler bl.a., at der oprettes en One-Stop-Shop, som skal lette adgang til data, mindske risiko for fejl og øge sikkerhed og transparens.

Forskerne skal have adgang til faciliteter med tilstrækkelig beregningskapacitet til at håndtere de voksende datamængder og nye analysemetoder. Det anbefales desuden at indføre en "Code of Conduct" for, hvordan både databehandlere og forskere skal omgås data.

Det er meldingen fra de danske forskere, som Det Koordinerende Organ for Registerforskning har talt med som baggrund for notatet "Dansk registerforskning".

Ifølge forskerne kræver det administrative omstruktureringer og klare politiske prioriteringer at kunne høste det fulde udbytte af registerforskningens potentiale. Hvis vi sammen får styr på dette, vil det hurtigt komme borgere og velfærdssamfund til gode.

### KOR

Rigsarkivet

Rigsdagsgården 9

1218 København K

Telefon: 4171 7211

Mail: [jkd@sa.dk](mailto:jkd@sa.dk)

Web: [registerforskning.dk](http://registerforskning.dk)

Sagsnr.: 2016/00595

## Indhold

Dansk registerforskning .....	1
Baggrund for rapport .....	3
Historien bag de danske registre .....	3
Dansk registerforsknings bidrag til samfundet .....	4
Udfordringer for dansk registerforskning .....	5
Ønsker til en god infrastruktur for registerforskning .....	6
En One-Stop-Shop frem for kontakt til flere myndigheder .....	6
Sikker behandling af data fra flere organisationer .....	6
Kobling af registre og andre datatyper .....	7
Fokus på beregningsfaciliteter og kompetencer .....	7
Den bedste sikkerhed .....	7
Arkivering af data .....	8
Lovgivning .....	8
Optimal udnyttelse af økonomiske og faglige ressourcer .....	9
Konklusion: .....	10

## Baggrund for rapport

På opfordring fra Styrelsen for Forskning og Uddannelse har Det Koordinerende Organ for Registerforskning<sup>1</sup>, (KOR) samlet og udarbejdet et forslag til rammerne om en ny infrastruktur for registerforskning.

Notatet er en opgørelse af, hvordan forskere mener, at dansk registerforskning fungerer på nuværende tidspunkt samt, hvad der er af muligheder og ønsker for nærmeste og fjern fremtid.

Det er en grundlæggende præmis og forståelse fra forskernes side, at registerforskning er til for borgerne og for vort samfunds udvikling.

Forskningen foregår med baggrund i de data, som borgerne stiller til rådighed, hvorfor forskerne skal have fælles mål og nytte for øje.

Uanset om forskernes forslag går på konkrete eksempler og problematikker i det daglige arbejde eller positionering internationalt set, er udgangspunktet at producere mere viden til gavn for befolkning og kommende generationer. De danske forskere udtrykker stor agtelse for borgernes data. Sikkerhed vægtes højt – samtidig med et håb om, at endnu bedre sikkerhed for data og brug deraf ikke medfører unødvendige restriktioner, der vil skade både registerforskningen og i sidste ende muligheden for at vort samfund kan udnytte det store potentiale.

Det er vigtigt at slå fast, at det ikke er i registerforskernes interesse at se data på individniveau – det er på baggrund af mange, såkaldte pseudo-anonymiserede data, at der forskes. Tidligere tiders læk af borgernes data, hvor uheldige og tilfældige de end har været, kan ikke tilskrives registerforskere.

## Historien bag de danske registre

Danmark har en lang tradition for registrering af befolkningen, egentlig kvalificeret personregistrering startede med påbud om at føre kirkebøger for hele landet allerede i 1646. Etableringen af CPR-registret i 1968 indebar personlig identifikation af høj kvalitet og gjorde det muligt at indsamle informationer til forskningsprojekter fra forskellige registre samt at følge befolkningsgrupper over et helt livsforløb.

Dette har medført, at dansk sundheds- og samfundsforskning er unik og af højeste, internationale kvalitet. Det er et faktum, at de ambitioner og den forudseenhed, som lå bag oprettelsen af CPR-Registret i 1968, ikke kan

---

<sup>1</sup> Med bidrag fra programleder, professor Søren Brunak, Novo Nordisk Foundation Center for Protein Research og professor Lars Pedersen, Klinisk Epidemiologisk Afdeling, Aarhus Universitet.

genskabes eller kopieres af andre lande. Det er også derfor, at det anbefales, at vi udviser rettidig omhu for registrene – både med hensyn til opretholdelse, udvikling samt brug af data.

## Dansk registerforsknings bidrag til samfundet

Registerforskning har hver dag betydning for den enkelte borger og velfærdssamfundet. Viden fra registrene hjælper til at forbedre den enkelte borgers sundhed og velfærd. Registerforskning leverer ”byggesten” af viden og fakta til brug for fortsat udvikling af politik og indsatser og er med til at sikre social og geografisk lighed og øget lighed i sundhed.

Af konkrete eksempler på samfundsnytte kan nævnes **forskning i p-piller**, hvor det med registerforskning kunne påvises, at visse præparater udsatte kvinder for ca. dobbelt så stor risiko for blodpropper. Forskning på baggrund af koblinger af data om ledere, undervisere og elevkarakterer har vist, hvilke **ledelsesformer der virker bedst** i uddannelsessystemet. Registerforskning har afklaret, at **kræft ikke smitter ved bloddonation**, ligesom forskningen har redegjort for risiko for kræft i forskellige erhvervsgrupper. På samme måde har registerforskning bidraget til **tidlig identifikation af børn i risiko for autisme og ADHD** og **målretning af forebyggelsesindsatser**. For blot at nævne nogle eksempler.

Der er løbende behov for at vurdere infrastrukturen for dansk registerforskning, som fremført i rapporten, ”Dansk Roadmap for Forskningsinfrastruktur 2011”, med en uddybning om registerforskningen i ”Registerforskning – nye muligheder og nye udfordringer” i 2013. En række forhold gør det nødvendigt igen at kaste et kritisk blik på infrastrukturen:

1. Antallet af digitale data stiger eksplosivt. Det giver mulighed for nye former for registerforskning og stiller større krav til datasikkerhed og beregningskapacitet. Udfordringen vokser med tiden.
2. Lovgivning om deling af data skal revideres, så den tager højde for EU-dataforordningen og anvendelse af Big Data. Det er vigtigt, at lovene tager hensyn til de særlige danske vilkår for registerforskning.
3. Den internationale konkurrence tiltager inden for registerforskning. F.eks. er der i USA opbygget sundhedsdatabaser med flere end 140 mio. personer, hvor data kan analyseres online. I England kan forskningsinstitutioner tegne abonnemeter på at anvende store sundhedsdatabaser såsom ”The Clinical Practice Research Datalink”. Initiativerne udfordrer dansk forsknings position, hvilket har direkte indvirkning på borgernes udbytte af registerforskning. Det er en

præmis for den fremmeste forskning, at danske forskere er i en international attraktiv liga mht. at tilegne sig ny viden og samarbejde med verdenseliten inden for forskning samt tiltrække forskningsmidler.

## Udfordringer for dansk registerforskning

Forskerne ser et stort potentiale for registerforskning i årsager til om f.eks. sygdomsbilleder eller sociale forhold kan tilskrives arv eller miljø. Dette kan give afgørende informationer til brug for forebyggelse og behandling. Denne type forskning kræver, at de biologiske markører for f.eks. sygdomseksposering kan kombineres med oplysninger om f.eks. patientforløb, uddannelse og indkomst fra eksisterende registre. Den nuværende forskningsinfrastruktur understøtter ikke denne type forskning.

### Der peges især på fire udfordringer for registerforskningen:

- 1) Det er vanskeligt at kombinere data fra forskellige organisationer.
- 2) Der mangler tekniske og organisatoriske muligheder for at koble registerdata til Big Data.
- 3) Den nuværende infrastruktur for registerforskning er baseret på fælles forskermaskiner i forskerserviceenhederne. Store forskermiljøer, der har behov for mere beregningskapacitet, må købe deres egen server. Denne organisering medfører, at flere institutioners servere ikke udnyttes fuldt ud, mens der andre gange ikke er tilstrækkelig beregningskapacitet. Der er ingen mulighed for at dele beregningskapacitet og omkostninger.
- 4) Det er vanskeligt at opnå tilstrækkelig gennemsigtig og overskuelig datasikkerhed med mange aktører og beregningsfaciliteter.

Udfordringen er også, at Danmarks Statistik (DST) ikke udleverer data fra de socioøkonomiske registre til beregningsfaciliteter uden for DST. Man råder desuden ikke selv over tilstrækkelig beregningskapacitet til at analysere store datamængder.

Med de nuværende muligheder og regler vil forskere ikke kunne kombinere variabler fra de veldefinerede populationsbaserede registre om sundhed og socioøkonomiske forhold med store, ofte ustrukturerede datamængder (Big Data) f.eks.; konteringsoplysninger, selvrapporterede sundhedsdata, Enterprise Ressource Planning eller biologiske data.

Hvis dette skal kunne lade sig gøre, kræver det for det første beregningsfaciliteter, som ikke er til stede hos de organisationer, der administrerer adgangen til registrene.

For det andet kræver det yderst specialiseret og fagspecifik datamanagement

og statistisk ekspertise, som kun i mindre grad er til stede i nævnte organisationer.

## Ønsker til en god infrastruktur for registerforskning

En One-Stop-Shop frem for kontakt til flere myndigheder

For at overholde lovkrav om sikker anvendelse af registerdata skal forskere søge tilladelse til at anvende data hos de organisationer og myndigheder, der hver især har ansvar for tilladelse til anvendelse af forskellige data.

Det er en vanskelig proces både for den enkelte forsker og for de organisationer, der administrerer tilladelserne. Hertil kommer, at det er tidskrævende, hvilket er problematisk, hvis det drejer sig om potentielt livstruende situationer og mønstre. Forskerne oplever derudover betydelig variation og tilfældigheder i sagsbehandling og adgang til data.

Derfor er der behov for samordning mellem de ministerier, der er ressortministerier for de forskellige organisationer.

Der bør oprettes en One-Stop-Shop, hvorfra anmeldelserne kvalitetssikres og fordeles til rette dataansvarlige organisation eller myndighed (Datatilsynet, Danmarks Statistik, Sundhedsdatastyrelsen, Styrelsen for Patientsikkerhed, RKKP m.fl.). Det vil lette arbejdet for forskerne og samle relevante kompetencer ét sted frem for spredt på forskningsinstitutionerne. Endelig vil det højne datasikkerheden, da der vil ske færre fejl via ensartet sagsbehandling og et transparent workflow.

Sikker behandling af data fra flere organisationer

Forskerne har brug for adgang til data, der pt. ligger spredt hos flere forskellige organisationer bl.a. Sundhedsdatastyrelsen, Danmarks Statistik, Rigsarkivet, biobanker, regioner, kommuner og universiteter.

Når forskere skal koble data fra flere organisationer, skal de have adgang til personidentificerbare data for at gøre det.

Det skal foregå med optimal datasikkerhed, og de dataansvarlige organisationer skal have kontrol med, hvem der har adgang til data og med hvilket formål.

**Tre mulige scenarier vil gøre dette sikkert og transparent:**

- At der indgås aftale om, at dataansvarlige organisationer kan overføre relevante dataudtræk til en beregningsfacilitet, hvor der er gensidig tillid til datasikkerheden. I så fald skal der etableres en sikker måde at sende data mellem dataansvarlige organisationer og beregningsfaciliteten.
- At dataansvarlige organisationer har hele registre til at ligge i krypteret kopi på beregningsfaciliteter, hvor den dataansvarlige selv står for bl.a. brugerstyring, logning og hvilke forskere, der har adgang til hvilke data.

- At der bygges en infrastruktur, hvor data ligger hos de dataansvarlige organisationer, men hvor forskerne kan etablere virtuelle datasæt bestående af data fra flere forskellige organisationer.  
Herved flyttes beregninger og workflows på en sikker måde derhen, hvor data er – i stedet for at flytte data.

### Kobling af registre og andre datatyper

Der genereres stadig større datamængder i forbindelse med administration, forskning og diagnosticering i den offentlige sektor samt af private aktører. Hvis forskere får mulighed for at koble disse Big Data med registre, vil de ved hjælp af analysemetoder såsom data-mining og machine learning kunne skabe nye resultater. Big Data adskiller sig fra de klassiske registre ved, at datamængderne er så store, at de ikke kan analyseres i den nuværende infrastruktur for registerforskning. Det kræver juridisk, organisatorisk og teknisk nytænkning at kombinere disse data.

### Fokus på beregningsfaciliteter og kompetencer

For at få fuldt udbytte af de voksende datamængder og nye analysemetoder, skal forskerne have adgang til beregningsfaciliteter, der kan rumme opgaven. Denne type faciliteter kræver samarbejde mellem højt specialiserede kompetencer samt store investeringer med relativ kort afskrivningstid. Det anbefales, at alle organisationer, som udleverer data til forskningsbrug, ikke opbygger parallelle infrastrukturer, men satser på få, fælles faciliteter med den nødvendige sikkerhed, økonomi og tekniske kompetencer. I dag er forskellige initiativer sat i gang bl.a. det Nationale Genom Center, High Performance Computer (HPC) anlæggene; Computerome ved DTU, Abacus ved Syddansk Universitet og GenomeDK på Aarhus Universitet.

### Den bedste sikkerhed

Data skal opbevares og analyseres i et miljø, hvor forskerne og befolkningen er sikre på, at personfølsomme oplysninger opbevares korrekt og sikkert. Derfor skal data opbevares på beregningsfaciliteter med adgangskontrol og uden mulighed for download af personhenførbare oplysninger. De data, forskerne ser, skal være sikret således, at direkte identifikation af individer ikke er mulig. Det gøres ved at pseudonymisere CPR-numre og fjerne andre variable såsom navn og bopæl. Det skal dog være muligt for f.eks. certificerede datamanagere i forskningsinstitutionerne at opnå adgang til hele personhenførbare populationer i registre for at kunne foretage den nødvendige kobling på tværs af registre og databaser eller Big Data.

Der skal laves en certificeringsordning, så opbevaring af og beregninger på personfølsomme data kun må foregå i miljøer, der lever op til klare sikkerhedskrav.<sup>2</sup>

Der skal vedtages en Code of Conduct for hvordan både dataansvarlige, databehandlere og forskere skal omgås data på de certificerede beregningsfaciliteter.<sup>3</sup>

Hvis data analyseres i få, certificerede faciliteter, vil man have styr på forskernes aktiviteter samt ensartede retningslinjer for håndtering af personfølsomme data. Dette vil imødekomme bekymringer, der rejses angående transparens i brug af data mht. formål og forskningens interessenter.

For at få adgang til data skal den enkelte forskningsinstitution være autoriseret. Det skal kun være etablerede, forskningsinstitutioner med klar, organisatorisk forankring og en ansvarlig chef, der kan autoriseres. I autorisationen lægges vægt på institutionens erfaringer med registerforskning, kendskab til datafortrolighedsregler samt kompetencer inden for datamanagement.

### Arkivering af data

Der skal udarbejdes klare retningslinjer for, hvordan data og beregninger arkiveres, så oplysningerne ikke kompromitteres. Men data og beregninger skal stadig være tilgængelige, da man ikke kender langtidsværdien af sådanne data til f.eks. intergenerational forskning. Desuden skal resultaterne af hensyn til videnskabelig dokumentation kunne genskabes mange år efter.

### Lovgivning

Forskernes adgang til personfølsomme data er bl.a. reguleret af persondataloven og sundhedsloven, der opererer med det såkaldte need-to-

---

<sup>2</sup> En sådan certificering kan bl.a. indeholde krav om:

- Lovgivning vedr. behandling af personhenførbare data.
- ISO27001 og ISO27002 certificeringer: Internationale standarder for informationssikkerhed.
- Health Insurance Portability and Accountability Act (HIPAA) certificering: USA's standard for transport, behandling og deling af sundhedsdata.
- Security Technical Implementation Guide (STIG): Standardisering af sikkerhedsprotokoller for netværk, servere og computere, der øger den samlede sikkerhed.
- En fælles referencearkitektur for informationssikkerhed: Fælles visioner, værdier og principper for sikker og effektiv håndtering af personoplysninger.
- Fælles standard for dataudveksling.

<sup>3</sup> Det kan f.eks. være inspireret af NordForsk projektet Tryggves Code of Conduct for secure service providers <https://wiki.neic.no/wiki/File:Tryggve-CodeofConduct.pdf>



know princip. Det er formuleret meget bredt, hvorfor det i praksis er op til en forhandling mellem forskerne og den dataansvarlige organisation, hvorvidt en forsker kan få adgang til en given population eller samling af variable.

Forskerne oplever stor variation i sagsbehandlingen for dataadgang, og nogle dataansvarlige tenderer til en meget rigid fortolkning af, hvad der er nødvendigt at udlevere til forskeren i forhold til et givent formål. Der er dermed risiko for ringere forskning, bl.a. fordi der ikke kan kontrolleres for fejlkilder. Dette kan også skabe vanskeligheder for forskningsprojekter, hvor der søges efter ukendte sammenhænge mellem mange parametre eller ved anvendelse af data-mining eller machine learning-metoder.

En løsning er, at forskere autoriseres til adgang til hele, pseudonymiserede populationer på beregningsfaciliteten, samt at det tillades, at datamanagere kan udføre datamanagement og indholdsanalyser uden forudgående godkendelse af et specifikt forskningsprojekt. Først når det indledende analysearbejde er foretaget, er forskeren forpligtet til at anmelde projektet til de dataansvarlige organisationer og gøre nøje rede for variabler og populationer. Kun hvis denne tilladelse er opnået, vil forskeren kunne downloade sine resultater fra beregningsfaciliteten.

Der skal desuden være klare retningslinjer for, hvordan danske forskere kan indgå i komparative forskningsprojekter, der anvender data fra flere lande, samt hvordan data udveksles internationalt.

### Optimal udnyttelse af økonomiske og faglige ressourcer

Forskningsinstitutionerne bør have en fælles infrastruktur, som sikrer tilstrækkelig beregningskapacitet, der kan allokere regnekraften, hvor der er behov for det, samt for at udnytte stordriftsfordele ved bl.a. drift og indkøb. Herved vil forskningsinstitutionerne kunne udnytte økonomiske og faglige ressourcer optimalt. En sådan infrastruktur bør kunne finansieres af f.eks. universiteter og forskningsinstitutioner, Danish e-Infrastructure Cooperation (DeIC), Uddannelses- og Forskningsministeriet og eventuelt via private fonde. Den nuværende infrastruktur for registerforskning er decentralt finansieret og langt fra omkostningseffektiv. Forskermaskinerne ved Danmarks Statistik og Sundhedsdatastyrelsen har begrænset beregningskapacitet, hvorfor mange større forskningsinstitutioner køber deres egen server for at sikre tilstrækkelig beregningskapacitet.

Der er pt. oprettet flere end 70 eksternt finansierede servere ved Danmarks Statistik og fire ved Sundhedsdatastyrelsen, og antallet er stigende. Mange af disse servere benyttes ikke fuldt ud, og det er omkostningstungt at drifte mange maskiner.

## Konklusion

**En god infrastruktur** for registerforskning vil optimalt samle sagsbehandlingen for adgang til data i en **One-Stop-Shop**, hvor de ansatte har den faglige ekspertise og ikke er tilknyttet bestemte forskningsmiljøer. Dette vil gøre adgangen smidigere, hindre fejl og øge datasikkerheden.

**Forskerne skal have adgang til data** på få, certificerede beregningsfaciliteter, der samler omkostninger, tekniske kompetencer og tilstrækkelig beregningskapacitet samt garanterer datasikkerhed på højeste niveau.

**Forskerne skal** inden for rammerne af disse beregningsfaciliteter kunne få adgang til registre fra flere forskellige dataansvarlige organisationer såsom Danmarks Statistik, Sundhedsdatastyrelsen og biobankerne samt Big Data og kobling til andre typer data.

**Det skal sikres både juridisk og teknisk**, at de enkelte dataansvarlige organisationer har fuld kontrol over, hvilke forskere der har adgang til data, og hvad data bruges til, samtidig med at forskerne får mulighed for at koble til data andre steder fra. Samtidig skal interessenterne udarbejde en code of conduct for dataadgang, -sikkerhed og -analyse.

**Beregningsfaciliteterne skal certificeres** efter klare krav til it-arkitektur, datasikkerhed, adgangsprocedurer og kompetencer. Dette skal understøtte brugerstyring, logning og kontrol af de resultater, som forskerne downloader. Beregningsfaciliteterne skal fungere som lukkede, sikre miljøer, hvor data ikke kan flyttes fra, og hvor kun forskere med den rette certificering kan få adgang til faciliteterne.

**Samarbejde mellem ministerier, offentlige organisationer og forskerne** skal styrkes, således at faglighed og viden er udgangspunkt for en klar plan og strategi for at styrke registerforskning i Danmark.