

Dato: 28. august 2018

e-Infrastruktur for registerforskning

Det Koordinerende Organ for Registerforskning (KOR) udgav i foråret 2017 rapporten *Dansk Registerforskning*, heri redegøres for udfordringer og muligheder for dansk registerforskning samt anbefalinger til fremtidens e-Infrastruktur for registerforskning.

Anbefalingerne blev godt modtaget, og hovedlinjerne har vist sig at ligge i tråd med anbefalingerne fra Sundhedsministeriets konsulentanalyse fra efteråret 2017 vedrørende infrastruktur for personlig medicin samt DeiCs behovsanalyse fra december 2017.

Udfordringer for dansk registerforskning

Dansk registerforskning er på et overordnet niveau udfordret af, at

- Datamængderne stiger, og forskerne mangler beregningskapacitet og kompetencer til at udnytte det.
- Lovgivning og folkestemninger vanskeliggør forskernes tilgang til registerdata.
- Den Internationale konkurrence stiger og der oprettes enorme, kommercielle datasamlinger, der på sigt kan udfordre de unikke, danske registre.

Herudover er der udfordringer, der kan relateres til tekniske infrastrukturer:

- Det er vanskeligt at kombinere data fra forskellige organisationer.
- Der mangler tekniske og organisatoriske muligheder for at koble registerdata til "Big Data".
- 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.
- Det er vanskeligt at opnå tilstrækkelig gennemsigtig og overskuelig datasikkerhed med mange aktører og beregningsfaciliteter.
- Data er dokumenteret uegalt og mangelfuldt.

KOR

Rigsarkivet

Rigsdagsgården 9

1218 København K

Telefon: 4171 7211

Mail: jkd@sa.dk

Web: registerforskning.dk

Sagsnr.: 16/00595

KORs erfaringer med e-Infrastruktur

KOR har initieret et meget lovende projekt i samarbejde med DeIC, Computerome Center, Danmarks Statistik, Sundhedsdatastyrelsen samt forskningsprojektet iPSYCH, der giver nyttige erfaringer, der kan bruges fremadrettet. Projektet giver forskere mulighed for at kombinere "Big Data" (f.eks. genetiske data) med registerdata fra Danmarks Statistik og snart også Sundhedsdatastyrelsen på beregningsfaciliteten Computerome, samtidig med at den dataansvarlige myndighed har fuld kontrol sine over data. Dette eksempel viser:

- 1) at national infrastruktur for registerforskning (forskerserviceenhederne) kan kombineres med national infrastruktur for life science supercomputing (Computerome) via Forskningsnettet,
- 2) at dataansvarlige myndigheder har tillid til en ekstern, fælles beregningsfacilitet,
- 3) at dataansvarlige myndigheder på tilfredsstillende vis kan etablere deres eget sikkerhedsregime på en beregningsfacilitet udenfor myndigheden selv,
- 4) at forskernes behov for beregningskapacitet og kobling af data kan tilfredsstilles,
- 5) at KOR kan samle relevante interessenter bag en fælles løsning, men det kræver koordinering, central finansiering samt forpligtende engagement fra forskningsmiljøerne.

Overordnet set har KORs skabt mulighed for at dette kunnet lykkedes ved at primært at fokusere på hvilke behov de dataansvarlige myndigheder, der skal levere data til e-Infrastrukturen, har, og først sekundært på forskernes behov for at analysere data. De dataansvarlige myndigheder skal være trygge ved e-Infrastrukturen, og samtidig skal den gøre det lettere for dem at give forskerne adgang til data, så kan vi få dem til at slutte op om e-Infrastrukturen.

Disse erfaringer er KORs udgangspunkt for samarbejde om e-Infrastruktur for registerforskning i fremtiden.

Ønsker og anbefalinger til e-Infrastruktur

Beregningsfaciliteter

- Der skal være en eller to nationale beregningsfaciliteter, hvor al forskning med følsomme data skal foregå.
- Dataansvarlige myndigheder skal via forskningsnettet levere de dataudtræk, som forskerne har fået adgang til, til denne

beregningsfacilitet. Myndighederne kontrollerer forskernes adgang til data og giver de enkelte brugere tilladelse til at tilgå data.

- Forskerne skal kunne koble data fra flere dataansvarlige myndigheder samt med egne datasæt.
- Der skal være tilstrækkelig beregningskapacitet til analyse af registerdata i kombination med "Big Data", f.eks. genetiske data.
- Beregningsfaciliteten skal have tilstrækkelig lagerkapacitet til, at data kan arkiveres så længe, der er juridiske eller faglige behov for det, eller til de skal afleveres til Rigsarkivet.

Datasikkerhed

- Der skal være gensidig tillid til sikkerheden på beregningsfaciliteterne blandt både myndigheder, forskere og befolkning
- Beregningsfaciliteterne skal garantere den fysiske datasikkerhed, mens de dataansvarlige myndigheder selv kan definere et sikkerhedsregime, der tilfredsstiller deres datasikkerhedsprocedurer i relation til forskerne.
- Fælles beregningsfaciliteter skal hindre spredning af data udenfor de sikrede faciliteter.

Dataadgang

- Forskere skal søge om adgang via én indgang – en såkaldt "One-Stop-Shop" – hvor forskere kun skal definere f.eks. formål og studiepopulation én gang.
- En del af en "One-Stop-Shop" skal være en ansøgningsportal, der skal fordele ansøgninger til de dataansvarlige myndigheder f.eks. SDS, DST, Regionernes Kliniske Kvalitetsudviklingsprogram, National Videnskabsetisk Komité, Rigsarkivet, og Styrelsen for Patientsikkerhed.
- Fælles arbejdsgange kan aflaste de dataansvarlige myndigheder, så de ikke hver især f.eks. skal vurdere om et forskningsprojekt er "af væsentlig samfundsmæssig betydning".
- På sigt skal der tilknyttes en egentlig forskerstøtteenhed til "One-Stop-Shop", der kan hjælpe forskerne med at følge op på ansøgninger, koble data etc. Denne enhed kan f.eks. forankres i DeiC. Samtidig sikres ensartethed i sagsbehandling, opbygning af kompetencer samt monitorering af service fra dataansvarlige myndigheder.

Datadokumentation

- Data skal være tilstrækkeligt dokumenteret, for at forskere kan anvende dem uden risiko for fejlslutninger.
- Et fælles dokumentationssystem skal sikre, at datadokumentation er tilgængelig for forskere og sagsbehandlere fra dataansvarlige myndigheder.
- Dokumentation af de data forskerne arbejder med skal udarbejdes af dataansvarlige myndigheder i samarbejde med forskerne.
- Datadokumentationen skal bygge på erfaringerne fra KORs projekt for høj kvalitetsdokumentation.

Økonomi

- Infrastruktur for registerforskning (forskerserviceenhederne i SDS og DST) finansieres af KOR via finanslovsbevilling fra SFU. Forskernes institutioner betaler for sagsbehandling og programmering.
- Beregningsfaciliteter finansieres af SFU og Universiteterne via DeiC.
- One Stop Shop finansieres af universiteterne via DeiC.
- Datadokumentation finansieres af dataansvarlige myndighed og SFU og udarbejdes via KOR.

Konkret forslag om etablering af e-Infrastruktur for analyse af følsomme data

KOR vurderer, at den helt store drivkraft indenfor forskningsrettet infrastruktur bliver personlig medicin og "Big Register Data". Det skal sikres, at den infrastruktur, der bygges til formålet, ikke alene tager hensyn til den kliniske anvendelse, men også til forskningsbrug. For at opnå dette skal Uddannelses og Forskningsministeriet og forskerne engagere sig i arbejdet, og UFM har efter KORs mening nogle konkrete erfaringer at byde ind med.

Forskermaskinerne og forskerservice i DST og SDS er dels finansieret af SFU finanslovsbevilling (1/3) og dels af forskernes brugerbetaling. KOR monitorerer kvaliteten i deres arbejde, finansierer deres infrastruktur og tager initiativer på tværs af de to forskerservices, f.eks. aftale om dataudveksling mellem DST og SDS, samt udvikling af fælles sikkerhedsprogrammer.

Der er oprettet forbindelse mellem forskermaskinerne og Computerome, så forskere kan kombinere registerdata fra forskerservice med "Big Data" på Computerome, dette i kombination med at organisationerne tilbydes

tilstrækkelig kontrol over data, samt mulighed for at implementere eget sikkerhedsregime.

DeiC har i samarbejde med Danmarks Tekniske Universitet og Københavns Universitet har etableret Computerome som life science HPC, der har opbygget de nødvendige kompetencer til at levere HPC til infrastruktur for personlig medicin samt til at behandle alle typer følsomme data.

Infrastrukturene er bundet sammen af Forskningsnettet, der drives i regi af DeiC og universiteterne.

SFU, DeiC og KOR bør på denne baggrund tilbyde at levere denne del af forskningsinfrastrukturen for personlig medicin samt sikre at den også kan anvendes af andre fagområder. Desuden skal de tilbyde de dataansvarlige myndigheder adgang til en sikker infrastruktur med tilstrækkelig og skalerbar beregningskapacitet på Computerome.

KOR faciliterer samarbejdet med de dataansvarlige myndigheder. Herunder snitfalder til bl.a. beregningsfaciliteter, sikkerhedsprogrammer, bestillingssystemer og opdatering af grunddata.

DeiC leverer via Computerome den nødvendige beregningskapacitet og sikker transport af data via Forskningsnettet.

SFU bør sikre politisk opbakning fra relevante myndigheder.