Rådet for Digital Sikkerhed

September 2020

Rådet for Digital Sikkerheds holdningspapir for ansvarlig brug af data i sundhedssektoren – vi skal alle have styr på it-arkitektur og databehandlingsprocesser

Rådet for Digital Sikkerhed anbefaler både sundhedssektoren og virksomheder at have fokus på ansvarlig databehandling til gavn for udvikling af sundhedsydelser og bedre behandlinger. Der skal arbejdes målrettet på at anonymisere data i langt højere grad end hidtil. Det kan fx ske gennem udvikling af solide og gennemsigtige modeller for fremstilling af syntetisk data. Der er et stort potentiale for vækst og bedre velfærd i arbejdet med sundhedsdata, men det kræver at vi arbejder dataansvarligt, og at vi har styr på it-arkitektur og databehandlingsprocesser. Rådet har i samarbejde med professor Carsten Obel og Katrine Svendsen fra Institut for Folkesundhed, Aarhus Universitet udarbejdet nærværende holdningspapir om temaet.

Introduktion

Danmark har enorme mængder af integreret sundhedsdata af høj kvalitet. Det giver os et solidt fundament for viden om, hvilke behandlinger der virker og om udvikling af serviceydelser inden for sundhedssektoren. Det giver os også et vækstpotentiale, hvor vi som samfund kan udvikle og innovere fremtidens sundhedstilgange og løsninger baseret på kvalificeret udnyttelse af data. Det er vigtigt, at vi får truffet de rigtige designvalg og at disse overvejelser sker med fokus på Security-, Privacy-, og Ethics-by-Design (hhv. SbD, PbD og EbD).

Etiske overvejelser

Grundlæggende ser Rådet for Digital Sikkerhed, at der er en række væsentlige dataetiske overvejelser om anvendelse og mulig eksponering af persondata, til fordel for forskning og udvikling af behandling, som bør inddrages i de fremtidige rammekrav til sundhedssektorens teknologiudvikling.

Uanset at der er gode juridiske rammer for fortrolig behandling af vores data, foreslås det, at sundhedssektoren, i langt højere grad end hvad der hidtil har været traditionen, så vidt mulig anvender anonymiserede data til forskning, hvor det på ingen måde er muligt at identificere de enkelte borger[1].

Teknologiens modenhed til at understøtte Privacy-by-Design-løsninger på et højt sikkerhedsniveau og det bør prioriteres, fordi der ofte vil være tale om behandling af følsomme og personhenførbare oplysninger, men også fordi lovgivningen i dag stiller nye krav om dataminimering, Privacy-by-Design, og borgerens adgang til egne oplysninger.

Rådet for Digital Sikkerhed er af den overbevisning, at der dataansvarligt kan udvikles innovative sundhedsydelser og behandling i sundhedssektoren og erhvervslivet kan udvikle bedre medicinalprodukter til forbrugerne, også når data i højere grad anonymiseres. Videre skal borgernes kontrol med data udbygges.

Udfordringen

Der produceres og behandles enorme mængder data i sundhedssektoren. Der er flere og flere forbundne enheder og systemer, samtidig med at deling af data i sektoren øges. Det stiller store krav til håndtering af data. Samtidig produceres der ikke længere sundhedsdata alene inden for sektoren. Udbredelsen af

[1] https://alexandra.dk/sites/default/files/downloads/Anonymisation-White-Paper.pdf

"wearables", der registrerer og behandler sundhedsdata fra den enkelte, skaber nye udfordringer, hvis sundhedssektorens behandling beriges af privatopsamlede data.

Erfaringer fra foråret 2020 under COVID-19 pandemien har understeget behovet for klare, legale rammeprincipper i forhold til hvordan data og nye typer af data kan anvendes under en akut sundhedskrise[2]. Her er der særlig brug for at sætte de bedste rammer for accellereret innovation og at ikke alene myndigheder, men også forskningmiljøer og virksomheder kan agere agilt og få hurtig adgang til data i et sikkert miljø – processen omkring udvikling og idriftsættelse af Smittestop-appen har i foråret 2020 demonstreret, hvordan helt grundlæggende principper om SbD, PbD.

Det er bl.a. Sundhedsloven, ISO (27001/2), HIPAA, Tekniske Minimumskrav (DIGST) og Databeskyttelsesforordningen er en generel lovgivning, der regulerer, hvorledes vores personlige data skal behandles, og hvem der har ansvaret for denne behandling. Lovgivningen definerer, hvordan der også er undtagelser for de overordnede regler for behandling af personoplysninger – fx at man på baggrund af speciallovgivning som fx sundhedsloven kan indsamle eller behandle data uden at få borgerens samtykke først, hvis det vurderes at være af bred samfundsmæssig interesse (§46, stk. 2).

Anvendelsen af persondata og retten til et privatliv skal balanceres med samfundsbehovet for udvikling af behandlinger for og diagnosticering af den brede befolkning, Derfor må hensynet til retten til privatliv nogle gange vige for hensynet til gavn for samfundet. I disse situationer skal der altid gælde et legalitetsprincip (sker indgrebet med hjemmel i national ret eller EU-ret), et proportionalitetsprincip (hensynet til retten til privatliv kontra samfundsbehovet) og et diskriminationsforbud.

Udvikling af innovative sundhedsydelser som f.eks. hurtigere og mere præcise diagnosticeringsværktøjer vil - som i den øvrige innovationsudvikling - i mange tilfælde basere sig på big data, der kan være opsamlet af IoT-devices og/eller behandlet af en kunstig intelligens. Disse store mængder af data sætter stadigt større krav til håndtering af retten til privatliv, der dog ikke må sætte en stopper for forskning og innovation i sundhedsvæsenet. Udvikling af fx sundhedsydelser på baggrund af disse data skal balanceres proportionalt med retten til privatliv, og det handler om at finde den rette grænse for, hvornår data er private, og hvornår data kan anvendes til det fælles bedste. Det er derfor centralt, at finde tekniske løsninger, der beskytter private data – særligt følsomme personoplysninger, når de anvendes til det fælles bedste i f. eks. kunstig intelligens løsninger.

I realiteten er det ikke alene i sundhedsvæsenet, at denne udfordring eksisterer. Udfordringen sættes blot på spidsen i sundhedssektoren, fordi der ofte er tale om følsomme personoplysninger i store mængder. Løsninger på udfordringen kan derfor være interessante for det resterende samfund, når der arbejdes med big data, IoT og kunstig intelligens.

I arbejdet med udvikling af nye systemer er det vigtigt, at det sker under hensyntagen til den operationelle virkelighed sundhedsvæsenet arbejder i, hvor integritet og tilgængelighed skal balanceres i forhold til fortrolighedsprincippet.

Rådet for Digital Sikkerhed finder, at balancen mellem udvikling af innovative løsninger med brug af private data, bliver nemmere at ramme, hvis data i højere grad anonymiseres, hvor det er muligt. Ved at tænke privatlivsbeskyttelse som fx anonymisering ind i selve teknologien, mindskes risikoen for sikkerhedsbrud og privatlivskrænkelser, samtidig med at formålet med løsningen realiseres.

[2] Rådet for Digital Sikkerheds holdningspapir om Corona og Persondata, Marts 2020, https://www.digitalsikkerhed.dk/s/Brug-af-persondata-til-sporing-af-coronavirus.pdf

Krav til behandling af data

De store mængder af sundhedsdata og mulighederne i anvendelsen af disse stiller krav til datahåndtering, herunder både processer og it-arkitektur, der understøtter ansvarlig brug af data.

- Ofte behandles data ved underleverandører, hvorfor det også er væsentligt, at der indføres skærpede krav til underleverandørers brug af data. Herunder at sikre tilstrækkelige sikkerheds- og styringsbestemmelser i fremtidige kontrakter for sundhedssektorens it-systemer[3].
- Så længe det er muligt at identificere den enkelte registrerede, er behandling af data underlagt GDPR, sundhedsloven, den danske databeskyttelseslov og andre speciallove, der giver vide rammer for at behandle personoplysninger.
- Logning af borgernes digitale færden og adfærd skal kun undtagelsesvist gøres til genstand for systematisk dataopsamling[4].
- Sundhedsdata skal (som alle andre personhenførbare data) vurderes ud fra nødvendigheden af:
    - Fortrolighed (transparens i og kontrol over hvem som har adgang til data)
    - Tilgængelighed (er der adgang til data og kan man give denne adgang til 3. part i en tidsbegrænset periode)
    - Integritet (kan man stole på korrektheden og oprindelsen af data)
- Fremtidens digitale sundhedssektor vil være dybt afhængig af de digitale data, hvorfor overvejelser omkring datasikkerhed, compliance og det samlede risikobillede, inkulsive trusselsbilledet, skal indtænkes i den strategiske og operationelle arkitektur.

Anbefalinger

Der er behov for at sikre metoder og principper for behandling af personhenførbare oplysninger i datasæt med det formål at lave statistiske undersøgelser, der kan bidrage til fx bedre behandlinger. Skemaet herunder kategoriserer forskellige typer af data, samt beskriver hvilke krav der i Rådet for Digital Sikkerheds optik skal stilles til behandling af data.

Type af data	Krav til brug af standarder og it-arkitektur i forbindelse med databehandling
Identificerbare data Data hvor fx cpr. nr. fremgår koblet på ex sygdoms og behandlingshistorie Der kan være både identifikatorer og kvasi-identifikatorer, der gør data identificerbare. Identifikatorer er attributter der unikt kan identificerer individet fx cpr. nr. Kvasi-identifikatorer er attributter der i fællesskab kan identificere et individ, fx postnummer og fødselsdato.	• Der skal foretages logning af alle anvendelser af personoplysninger • Principperne og kravene i GDPR og speciallovgivningen gælder fuldt ud. • Det er rådets anbefaling, at data i denne form skal begrænses til der, hvor der er brug for identifikation.
Anonymiserede data Her fjernes muligheden for at identificere data ved at fjerne identifikatorer og kvasi identifikatorer. Fx ved at generalisere adresser til områder eller fødselsdato til intervaller. På den måde kan data om den enkelte hverken udskilles af selve datasættet eller ved sammenkobling med andre datasæt.	• Da der er tale om data som er anonyme, er det ikke muligt at gendanne identitet fra data; men det forudsætter, at anonymiseringen sker således, at der skabes irreversibilitet. Det er derfor vigtigt, at ALLE datasæt, som er identitetsbærende, maskeres / forandres irreversibelt. • Anonymisering er ikke trivielt, og der har været eksempler, hvor anonymiseringen ikke har været stærk nok. Det kræver derfor grundighed og en vis viden at anonymisere data, samt evt. en risikovurdering.

[3] Rådet for Digital Sikkerheds holdningspapir om styring af leverandørsikkerhed, juni 2018
https://www.digitalsikkerhed.dk/s/leverandrsikkerhedv097Juni2018.pdf

[4] Rådet for Digital Sikkerheds positionspapir for udvidet logning af danskerne internettrafik, februar

Billedmateriale fx røntgenbilleder vanskeligt at anonymisere, men man kan fjerne tilknyttede informationer så man så man fx fjerner navne CPR-nr.	• Anonymisering skal benyttes, hvor det er muligt at fjerne personhenførbare data, samtidig med at det forskningsmæssige aspekt bevares. [5]
Pseudonymiserede data Identifikatorer og kvasi-identifikatorer erstattes med en anden og tilsyneladende tilfældig værdi, hvorved det er vanskeligere at knytte datasættet til den registreredes originale identitet. Det er stadig muligt at udskille og sammenkoble enkeltpersoners identitet på tværs datasæt.	• Man kan pseudonomisere i (automatiserede) mapningstabeller, der kan automatiseres – herved kan også laves fysisk adskillelse af databaser. • Kryptering af identifikationsoplysninger med en hemmelig nøgle, hvorved man kan finde tilbage til de oprindelige identiteter, hvorimod brugerne af datasættet kun kan se den kodede værdi af identifikationsoplysningerne.
Syntetiske data Et konstrueret datasæt, der bevarer karakteristika fra rigtige data, uden individer kan identificeres	• Som ved anonymisering kræver det grundighed og en vis viden at generere syntetiske data, samt evt. en risikovurdering • Det vil ofte være muligt at benytte syntetiske data i stedet for de rigtige datasæt så det test- og forskningsmæssige aspekt bevares.

Kilde: Pseudonymiseringsprincipper for sundhedsdata til statistikproduktion, Lakeside, November 2016

Rådet for Digital Sikkerheds holdning

Politikerne bør stille krav til et passende niveau af sikkerhed og dataansvarlighed i behandling af sundhedsdata – herunder krav om Privacy- og Security-by-Design /default. Sikkerhed skal bygges ind fra starten af designprocessen, og der skal være fokus på dataminimering og proportionalitet.

Syntetisk data

Arbejdet med syntetisk data er stadig i sin vorden, og derfor udestår fortsat et arbejde med at definere begrebet. Der foregår forskning på området om anvendelse af syntetisk data.

Rådet for Digital Sikkerhed mener, der skal prioriteres ressourcer til yderligere at arbejde med syntetisk data og udvikle modeller til anvendelse af syntetisk data. Der er et stort potentiale i at anvende sådanne data til at skabe vækst og innovation i sundhedssektoren på dataansvarlig vis.

Det vil ofte være muligt at benytte syntetiske data i stedet for de rigtige datasæt, så det test- og forskningsmæssige aspekt bevares. Ved at konstruere syntetisk data bevares karakteristika fra rigtige data uden at individer kan identificeres. Det afhænger dog af algoritmen, der genererer de syntetiske data ud fra det oprindelige datasæt, hvor godt karakteristika bevares.

Som ved anonymisering kræver det grundighed og en vis viden at generere syntetiske data, samt evt. en risikovurdering. Med syntetiske data kan der ofte arbejdes med sundhedsdata til udvikling og test af algoritmer og machine learning-modeller indenfor eksisterende lovgivning. Man kan eksempelvis forestille sig, at man kan arbejde med udvikling af algoritmer i syntetiske data, som efterfølgende kan appliceres på data der ikke er syntetiske.

[5] Identifikation skal være mulig indtil relevante datakilder er koblet med hinanden. Det kan laves sådan, at det kun er maskinen og ikke forskeren, der kan se den korrekte koblenøgle, og forskeren bare kan bestille koblingen udført (kommando/script), og denne udføres korrekt, så kan visningen af data sagtens være uden personidentifikation

Rådets anbefalinger

1. Der skal arbejdes målrettet på at anonymisere data i langt højere grad end hidtil. Er data ikke anonymiserede bør ejerskabet af sundhedsdatadata ligge hos borgerne. Borgeren skal have indsigt i, og kontrol over data, der vedrører vedkommende.
2. Anvend pseudonomiserede eller anonymiserede data, hvor det er muligt og sørg for at udvikle/ anvende solide og gennemsigtige modeller til anonymisering og generering af syntetiske data.
3. Sørg for at indbygge Privacy-by-Design /default og Security-by-Design.
4. Reglerne for samtykke og undtagelser der regulerer det offentliges adgang til og brug af personhenførbare data bør tydeliggøres, også i forskningsmæssige scenarier.
5. Under stadig hensyntagen til den operationelle virkelighed skal sundhedsvæsenet systemer designes, så der balanceres mellem integritet, tilgængelighed og fortrolighed, samtidige med at der skabes transparens og effektivt tilsyn og kontrol (f.eks. via logning) med adgang til systemerne.
6. Der bør politisk prioriteres ressourcer til yderligere at arbejde med syntetisk data og udvikle modeller til anvendelse af syntetisk data. Der er et stort potentiale i at anvende sådanne data til at skabe vækst og innovation i sundhedssektoren på dataansvarlig vis.

På vegne af bestyrelsen i Rådet for Digital Sikkerhed

Gert Læssøe Mikkelsen, Aexandra Instituttet

Christian Wernberg-Tougaard, KPMG

Morten Rosted Vang, DI

Anette Høyrup, Forbrugerrådet Tænk

Kim Larsen, Systematic

Michael Lind Mortensen, Bankdata

Ole Kjeldsen, Microsoft

I samarbejde med

Carsten Obel, Institut for Folkesundhed, Aarhus Universitet

Katrine Svendsen, Institut for Folkesundhed, Aarhus Universitet