

Helse- og Omsorgsdepartementet
Postmottak@hod.dep.no
www.regjeringen.no

SINTEF Konsernstab
Postadresse:
Postboks 4760 Torgarden
7465 Trondheim

Besøksadresse:
Strindvegen 4
7034 Trondheim

Sentralbord: 73593000
Direkte innvalg: 93058509

info@sintef.no

Foretaksregister: NO 948 007 029 MVA

Deres ref.:
Deres ref

Vår ref.:
54/ABG/RRE

Prosjekt / Sak:
Høringsuttalelse – rapport fra
Helsedatautvalget

Dato
2017-12-01

Høringsuttalelse til Rapport fra Helsedatautvalget

Helse- og omsorgsdepartementet har sendt på høring Helsedatautvalgets rapport *Et nytt system for enklere og sikrere tilgang til helsedata* fra 30. juni 2017, med høringsfristen 1. desember 2017.

SINTEF takker for denne muligheten til å komme med tilbakemelding til Helsedatautvalget. Kommentarene er forankret i SINTEF sine forskningsmiljøer for helse og velferd og digitalisering.

Departementet har invitert til å gi innspill til utvalgets:

- Vurdering av utfordringer ved dagens system
- Beskrivelsen av og vurderingen av pågående arbeid
- Utvalgets forslag til tiltak for å sikre enklere tilgang til helsedata

SINTEF deler utvalgets syn på hva som er utfordringene ved dagens system og synes beskrivelsene av utfordringene er gjenkjennelige og utfyllende. SINTEF mener imidlertid at det fortsatt bør skilles mellom forskning og offentlig planlegging jf. statistikkloven. Det foreligger i dag forbud mot å tilby annet enn helt anonyme data til statistisk bruk i offentlig forvaltning, mens det kan tilbys indirekte identifiserbare data til statistisk bruk i forskning.

I det følgende gir vi særlig innspill til det tredje punktet.

Kommentarene tar hensyn til bruk av helsedata til statistikk, planlegging, helseanalyser, kvalitetsforbedring, forskning, innovasjon og næringsutvikling. SINTEF støtter i prinsippet utvalgets forslag til en radikalt annen teknisk løsning for tilgjengeliggjøring av helsedata. Den nye løsningen leverer ikke helsedata ut på minnepinner, disker eller andre former for distribuering, men lar brukeren få komme *inn* – inn på analyseplattformen. Direktoratet for e-helse jobber med utviklingen av en slik analyseplattform. Vi mener at det er nødvendig med særskilte tiltak for å sikre at den nye løsningen skal tilfredsstillende følgende krav: 1) kvalitet i analysearbeidet, 2) relevans og kvalitet i forskningsprosjektene som bruker helsedata, 3) sikkerhet og 4) lik tilgang for alle relevante brukere av helsedata.

SINTEFs kommentarer til Utvalgets anbefalinger og forslag til tiltak:

Forenklede regler for behandling av søknader om tilgang til helsedata

Forhåndsgodkjenning i REK

Et flertall i utvalget foreslår at man fjerner kravet om forhåndsgodkjenning i REK for forskning som bruker helsedata. Dette er et ledd i å forenkle prosessen med å få tilgang på helsedata og dermed sørge for raskere og mer forutsigbar tilgang og mer effektiv utnyttelse av helsedataene. Flertallet fremholder at 1) risikoen ved denne type forskning er så liten at den ikke nødvendiggjør etisk forhåndsvurdering som den REK står for, og 2) det er kun medisinsk og helsefaglig forskning som faller inn under Helseforskningsloven som i dag har krav om forhåndsgodkjenning av REK, og helseforskningen bør ikke ha dette særskilte kravet.

SINTEF støtter mindretallet i utvalget sitt syn med hensyn til forhåndsgodkjenning. Vi mener at sekundærforskning på helsedata er av en slik karakter at **ekstern forhåndsgodkjenning er nødvendig** for å sikre god kvalitet i forskningen og dermed også gi legitimitet til bruk av helsedata. REK sine vurderinger omfatter mer enn risiko. REK gjør en helhetlig vurdering også av kvalitet og forsvarlighet, om kvaliteten i forskningsprotokollene er slik at en kan forsvare bruk av sensitive helsedata fra en stor del av befolkningen. Forskningen er avhengig av tillit i befolkningen. Om etisk vurdering legges til forskningsinstitusjonene, slik flertallet foreslår, kan dette ha uheldige konsekvenser som også innebærer forskjell i vurderingene. Små forskningsenheter vil ikke ha ressurser til egne etiske utvalg med tilstrekkelig kompetanse og erfaring, og slike mindre enheter kan bli presset ut av denne type forskning. SINTEF mener at alle former for forskning på helsedata skal være gjenstand for en ekstern forskningsetisk vurdering.

Når det gjelder sen igangsetting av prosjekter pga saksbehandlingstid i REK og hos databehandlingsansvarlige, opplever ikke SINTEF at REK bruker lang tid på behandling av søknader, men at det er de mange databehandlingsansvarlige som representerer et problem her.

Siden REKene kan ha noe ulik praksis i sine vurderinger og ikke alltid tilstrekkelig kompetanse til å gjøre alle de nødvendige etiske vurderingene, kan det være en mulig løsning å **legge den etiske godkjenningen til den nye tilgangsførvalteren**. Over tid vil dette organet få omfattende erfaring med å gjennomføre slike vurderinger og dermed bygge opp solid kompetanse som sikrer bedre forutsigbarhet og kvalitet i vurderingene. En slik nasjonal løsning gjør at forskningsmiljøene har lik tilgang til kompetanse om etiske vurderinger. Det er også lite økonomisk hensiktsmessig at denne kompetansen skal bygges opp i den enkelte virksomhet

Én tilgangsførvalter

SINTEF støtter flertallet av utvalgets medlemmer i anbefalingen av en nasjonal tjeneste med én tilgangsførvalter til helsedata for alle godkjente formål. Det kan være hensiktsmessig og innebære en stor forenkling og bedre tilgjengelighet sammenlignet med dagens løsning. Vi vil videre fremholde at direktoratstrukturen kan være velegnet for å ivareta en slik felles registerportal, Helsedirektoratet forvalter allerede NPR, IPLOS og KUHR (KPR), og Direktoratet for e-helse kan også være aktuell forvalter av den foreslåtte tilgangsportalen.

Tilgang til data og kompetanse om registrene

En bør unngå å legge tilgangsførelsen til en institusjon som selv driver forskning på registerdata og som selv utvikler helseregistre.

Det bør sikres at løsningen som velges ikke gir konkurransevridende effekter for noen forskningsmiljø på bekostning av andre. For eksempel må alle miljø ha samme tilgang til hele datagrunnlaget, slik at det ikke skapes situasjoner som gir fortrinn.

Vi kan ikke se at utvalget har vurdert følgende: Hvordan tenkes den kunnskapen som i dag ligger i de ulike registrene forvaltet? Vil kunnskapen overføres til helsedataløsningen, eller vil den bli værende i de ulike registrene? Registereier forvalter data og benytter data blant annet til utforming av nasjonal statistikk, og har derfor inngående kunnskap om de ulike variablene som inngår i registrene. Forskere vil, erfaringsmessig, ha behov for tett dialog med registereier i ulike faser av forskningsprosjekt, og det vil være avgjørende at forskere har tilgang til fagpersoner med inngående kunnskap om de ulike registre. Hvordan vil helsedataløsningen bli bemannet med registerdatakompetanse?

SINTEF mener at både eksisterende og nye registre som etableres må ha tilstrekkelig metadata slik at brukerne på en best mulig og effektiv måte, og uten å være helt avhengige av registereier, for å ta i bruk helsedataene i egen forskning. Det finnes mange muligheter for å bygge opp slike metadatasett, også av eksisterende datasett der lite eller ingen metadata finnes gjennom semantiske teknologier og ontologier.

En helseanalyseplattform som felles portal og tilgangstjeneste

Rapporten fra Helsedatautvalget gjør grundige analyser av viktige faktorer som vil være knyttet til en helseanalyseplattform som felles portal og tilgangstjeneste. SINTEF vil peke på noen potensielle problemstillinger knyttet en helseanalyseplattform. Hvis portalen er ment som en fullstendig erstatning for dagens ordning med utlevering av data, og at denne skal inneholde de data og analyseverktøy som skal kunne benyttes for å gjøre analysene så utgjør dette et brudd med dagens praksis som kan ha store konsekvenser for den potensielle nytten av dagens registre.

- 1) **Tilgang til data og kobling til andre datakilder** Hvilke data gis det tilgang til gjennom dette systemet og er det anledning til å koble disse til andre datakilder? Begrensninger på hvilke data det gis tilgang til vil ha konsekvenser for hvilke forskningsspørsmål som kan besvares. Å på forhånd avgrense datatilgangen ut i fra en antakelse om at man vet hva som er de sentrale problemstillingene og de sentrale variablene, vil potensielt utelukke viktig fremtidig forskning. Hvis det er begrensninger på de data som skal tilbys gjennom denne portalen, bør det være løsninger som gjør det mulig å koble på data fra andre kilder. Interoperabilitet mellom ulike datakilder må sikres.
- 2) **Organisering av data** Hvordan er data som det gis tilgang til gjennom portalen organisert? Jf. kommentar om metadata ovenfor. Ulike forskningsspørsmål krever ulik datastruktur i analysene. Tverrsnittsundersøkelser vil for eksempel kreve en annen datastruktur enn tidsserieanalyser. Begrensninger på hvordan data kan organiseres vil altså begrense de problemstillingene som det kan forskes på. Er dataene beriket med tilstrekkelig metadata, vil dette betraktelig hjelpe automatisering av transformasjon av data mellom ulike datastrukturer.

- 3) **Datakvalitet** Hvis data det gis tilgang til gjennom portalen er tilrettelagte data, hvordan håndteres usikkerhet knyttet til datakvalitet? I dagens løsning vil data som utleveres være rådata, eller lite bearbejdede data som så må gjennom en prosess med rensing og bearbejding. Her kan for eksempel observasjoner med opplagte feilregistreringer lukes vekk. Basert på dette tilretteleggingsarbeidet kan det lages statistikk over andeler som utelukkes og på hvilke kriterier. Dette er sentrale mål på datakvalitet. Hvis data det gis tilgang til gjennom portalen allerede er bearbejdet vil det ikke være mulig å fremstille troverdige mål på datakvalitet. Igjen kan mekanismer og rutiner for å berike datasett med metadata avhjelpe og forsterke kvaliteten av data gjennom bruk. Det skal med dagens teknologier for digitalisering være unødvendig at hver gang helsedata benyttes så må dataene gjennomgås på nytt uten historikk for å sikre kvalitet.
- 4) **Analyseverktøy** Hva slags analyseverktøy tilbys gjennom portalen? Det eksisterer i dag et utall analyseverktøy med ulik funksjonalitet som ofte er tilpasset spesielle former for analyser. Hvordan skal en slik portal kunne tilby samtlige av disse verktøyene? Skal denne portalen kunne tilby STATA, SAS, R, Excel, Matlab, ARCGIS, og alle andre sentrale analyseverktøy som benyttes i forskning på disse helseregisterdata? Begrensninger på dette betyr at en rekke problemstillinger ikke vil kunne besvares. En begrensning i hva slags analyseverktøy som kan benyttes vil også ha store konsekvenser for forskere som har brukt år på å utvikle syntaks og programmeringsspråk knyttet til spesifikke analyseverktøy for å undersøke aktuelle problemstillinger. Portalløsningen for slike helsedata må ha en arkitektur som sikrer at ulike analyseverktøy skal kunne utnyttes sømløst og uavhengig om verktøyet tilbys fra portalen eller om at "lokale" verktøy kan knyttes til datasett som det gis tilgang gjennom portalen.
- 5) **Informasjonssikkerhet** Hvordan skal en sørge for at sikkerheten ivaretas? Samling av helsedata og endringer i teknologien som brukes for tilgang til helsedata vil også kunne medføre nye potensielle sårbarheter, og risiko for uautorisert tilgang til helsedata. Flertallet i utvalget mener at nytten av en innføring av helseanalyseplattformen er større enn personvernulempene, mens et mindretall i utvalget mener at informasjonen om de tekniske løsningene i helsedataplattformen på rapporttidspunktet ikke er tilstrekkelig til å kunne konkludere med at ulempene oppveies av nytteverdien (jmf. kapittel 7.2 Personvern og innbyggertjenester). Informasjonssikkerhet og personvern er kritisk for en vellykket innføring av helseanalyseplattformen, noe som krever et større fokus på akkurat disse aspektene enn det ser ut som at dette utvalget har hatt anledning til, gitt utvalgets sammensetning og brede mandat. SINTEF vil derfor stille seg bak mindretallets vurderinger på dette området og foreslå en nærmere utredning som inkluderer en risiko- og sårbarhetsanalyse basert på en grundigere gjennomgang av forskjellige teknologiske og organisatoriske løsninger for helseanalyseplattformen, jf. også General Data Protection Regulation (GDPR) som trer i kraft i EU/EØS den 25. mai 2018. Basert på denne risiko- og sårbarhetsanalysen vil man være i stand til å stille konkrete krav til løsningen når det gjelder informasjonssikkerhet og personvern, både ved design og drift av løsningen. Herunder at noen kan ta ut data som bryter til kravet om anonymitet.
- 6) **Fremtidige behov** I hvilken grad evner en slik portal å tilby løsninger for fremtidens helsedataanalyser? Det er i dag en rivende utvikling innen kunstig intelligens og maskinlæring innen helseområdet. SINTEF bidrar i arbeidet med å utvikle denne teknologien – en teknologi som vil bli stadig viktigere i utnyttelsen av store datamengder i fremtiden. Tilgangen til helseregisterdata er sentralt i disse analysene hvor analyseverktøyene utvikles parallelt med

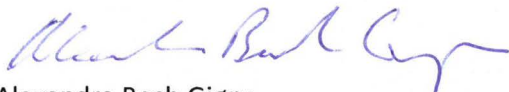
gjennomføringen av analysene. Hvis en portalløsning legger begrensninger på muligheten for å drive slike utviklingsprosjekt vil dette kunne ha store konsekvenser for de nasjonale forskningsmiljøenes evne til å føre an i utviklingen av denne teknologien.

Helseanalyseplattformen vil være verdifull for næringslivet, og en må avklare på hvilke premisser næringslivet skal gis mulighet til å ta ut data og hvorvidt norsk næringsliv skal ha noen fortrinn. Norge har store og til dels unike underlagsdata som vil ha stor verdi i utviklingen av kunstig intelligensverktøy. Vi mener det er viktig at en også grundig vurderer løsninger for å kvalitetssikre og ta imot verdifulle data generert av den enkelte borger.

- 7) **Morgendagens løsninger** Det er spesielt to forhold som enten mangler eller som må forsterkes betydelig for at et nytt system for tilgang til helsedata skal lykkes: 1) Gjennom GDPR implementeringen kommer det prinsipielt en grunnleggende endring i forhold til den enkeltes rett til selv å disponere og bestemme over sine egne data, bla. en rett til fritt å overføre egne data til andre tjenester (inkludert kommersielle tjenester) i maskinlesbar form. Dette gjelder også den enkeltes rett til selv å velge tjeneste som skal ha tilgang til data og hvordan disse skal kobles mot andre data enn de dataene som allerede finnes i helseregistre. 2) Dette endrer også hele landskapet rundt helsedata på måter vi i dag ikke klarer å forutse. Andre datakilder kan f.eks. være data fra wearables, loggede data fra internettaktivitet, pasientjournaler, medisinbruk på apotek osv. Vi vil kunne se at borgerne selv velger å la kommersielle aktører få tilgang til sine helsedata, og vi kan få helsedata om større populasjoner drevet som rein forretning (også omtalt som data market places). Denne typen utfordringer som omtalt i 1) og 2) har utvalget ikke håndtert.

Med vennlig hilsen

SINTEF



Alexandra Bech Gjørvi

Konsernsjef