

Hørings svar - Høring om Nasjonal spesifisering for metadata om helsedata

Til: Direktoratet for e-helse
Deres ref: Sak 21/121
Fra: Helse Midt-Norge
 Kontakt ved evt. spørsmål: ellen.royneberg@helse-midt.no
Kopi:
Dato:

Innhold

1	Innledning/bakgrunn	2
2	Overordnet tilbakemelding.....	2
2.1	Praksisen på overføring av metadata.....	2
2.2	Forventet gevinst ved å ta i bruk modellen.....	2
2.3	Variabler som ikke finnes lagret i kilden	2
2.4	Lesbarhet.....	2
3	Logisk informasjonsmodell	2
4	Konsekvenser av endringer.....	3
4.1	Kommentarer til punkter under 1a: kildemetadata.....	3
4.1.1	Emneord (16).....	3
4.2	Kommentarer til punkter under 1b: variabelmetadata	3
4.2.1	ValidFrom (7), ValidTo (8).....	3
4.2.2	NavnEngelsk (12), BeskrivelseEngelsk (14).....	3
4.2.3	Identifiseringsgrad (28)	3
4.2.4	AnbefaltForDigitaltInnsyn (32)	3
4.2.5	Tillegg (40-N)	3
4.2.6	Utgått variabel: TemaLokalt	4
5	Forslag til veien videre.....	4

1 Innledning/bakgrunn

Det vises til brev datert 23. februar 2021 med invitasjon for å gi skriftlig innspill til *Høring om Nasjonal spesifisering for metadata om helsedata*. Dette notatet er et samlet innspill på spørsmålene i høringsbrevet fra helseforetakene i Helse Midt-Norge, HEMIT, Helseplattformen AS og Nasjonalt servicemiljø region Midt.

2 Overordnet tilbakemelding

Generelt er Helse Midt-Norge positive på at det gjøres et forsøk på å få på plass en nasjonal spesifisering på dette området. Det vil være nyttig å ha en felles logisk informasjonsmodell for avlevering av data til Helseanalyseplattformen. I dette kapittelet vil vi gi noen overordnede innspill på standarden. Noen av innspillene går utover selve innholdet i høringsbrevet, men er knyttet til hvordan bruk av modellen vil være i praksis.

2.1 Praksisen på overføring av metadata

Høringsbrevet sier ikke noe om hvordan det er tiltenkt at data skal overføres til HAP. I piloten som HEMIT har vært delaktig i, har det ikke blitt utarbeidet en bærekraftig måte å overføre data på. I dagens praksis sendes metadata til en personlig e-postadresse. Dette er en sårbar praksis og det etterlyses derfor tilgang til en API eller et filområde hvor det er mulig å sende data. Dette er viktig ut fra en sikkerhetsmessig og prinsipiell vurdering.

2.2 Forventet gevinst ved å ta i bruk modellen

Det etterlyses en beskrivelse av hva som vil være gevinsten av å ta i bruk denne standarden for alle parter. Modellen er nå utvidet, og det er uklart hva disse ekstra elementene vil bidra med.

2.3 Variabler som ikke finnes lagret i kilden

I de fleste kvalitetsregister er det vanlig å beregne data ut fra registrerte data. Et eksempel på dette er å beregne tiden på *Time to needle*. *Time to needle* er tiden som går fra du kommer inn på sykehuset og til du får behandling. Denne beregningen bør gjøres i datavarehus ut fra variablene i kvalitetsregisteret. Det finnes ikke metadata for denne typen data, det er i den beskrevne modellen kun mulig å få metadata fra kildevariablene. Hvordan er det tiltenkt å samle data som er generert i kilden inn i Helseanalyseplattformen?

2.4 Lesbarhet

Dokumentet er komplisert og forutsetter inngående teknisk kunnskap for å forstå det. Dokumentet fremstår som rettet mot teknisk tilbydere, men selv med relevant teknisk kunnskap er det krevende å forstå og sette seg inn i modellen. Komplexiteten er for høy for ansatte i kvalitetsregistre og servicemiljøet. Det bør også vurderes om det vil være nyttig at modellen er lesbar for f.eks. forskere eller andre som potensielt skal benytte dataen. Slik modellen fremstår nå vil terskelen være høy.

3 Logisk informasjonsmodell

Informasjonsmodellen er forståelig, men endrer noen av forutsetningene for hvordan metadata skal forvaltes. Endringer i denne har konsekvenser for de som skal starte med metadataarbeidet, og for de som allerede har definert metadata. Endringer og konsekvenser av disse omtales i kapittel 4, fordelt på de ulike egenskapene som vi har valgt å kommentere spesifikt.

Sett fra en utviklers perspektiv er det en tidkrevende prosess å sette seg inn i modellen og å mappe den opp mot egen modell. Det er mange entiteter og flere av dem ligner på hverandre, noe som er med å øke kompleksiteten. Det vil kreve betydelig jobb å implementere modellen for hver enkelt kilde som skal avlevere i henhold til modellen. Dette vil igjen føre til at vedlikeholdskostnadene vil være betydelige og risikoen for feil vil øke sammen med kompleksiteten av modellen.

HEMIT har vært med i pilotering av første versjon av modellen og er lagt ned betydelig arbeid i å forstå og implementere modellen. Erfaringen fra piloten sier oss at det vil være utfordrende for andre som ikke har

deltatt i piloten å tolke modellen og gjøre nødvendig implementasjon. Med bakgrunn i dette anbefales det at direktoratet ser på muligheten for å legge kompleksiteten på mottaker siden ved å f.eks. ta i bruk en flatere modell fra avsendersiden. Modellen slik den foreligger nå vil kreve et betydelig apparat hos Direktoratet for e-helse for å bistå i implementeringer, ta imot spørsmål og bidra avsender med hjelp/råd/avklaringer.

4 Konsekvenser av endringer

4.1 Kommentarer til punkter under 1a: kildemetadata

4.1.1 Emneord (16)

Det er en fordel at emneord er introdusert. Vi har diskutert en tilsvarende funksjon på variabelnivå for å kunne "tagge" variabler etter emne. Fordelen med emneord for kildemetadata er at det er realistisk at disse blir fylt ut og blir nyttige for forskere.

4.2 Kommentarer til punkter under 1b: variabelmetadata

4.2.1 ValidFrom (7), ValidTo (8)

Hvis disse egenskapene beskriver tidsperioden data finnes tilgjengelig for, oppfyller det behovet til dokumentasjon av registeret. Vi har støtt på problemer i flere registre hvor det er diskrepans mellom gyldighetsperioden som hentes fra databasen (MRS) og tidsperiode for faktisk innhold i variabler. Hvordan bør dette løses?

4.2.2 NavnEngelsk (12), BeskrivelseEngelsk (14)

Vi er kritiske til innføringen av disse som obligatoriske egenskaper, da det vil innebære betydelig jobb for registrene. Hvis disse skal fylles med gode data, mener vi at det vil det innebære omfattende faglig kvalitetssikring og profesjonell oversetterkompetanse for å ivareta tilsvarende kvalitet for de engelske navnene/beskrivelsene som de norske.

4.2.3 Identifiseringsgrad (28)

Vi stiller spørsmål ved nytten av denne egenskapen. Et punkt som beskriver om det er spesielt sensitive data kan være hensiktsmessig, men da som en varselampe for helt entydige individuelle variabler som adresse eller lignende. For dataminimering er det ofte sammensetningen av variabler som avgjør identifiseringsgrad, og ikke enkeltvariabler. Vi forutser problemer med å kategorisere variabler i én spesifikk kategori (ingen, lav, middels, høy), da dette vil variere fra sak til sak.

4.2.4 AnbefaltForDigitaltInnsyn (32)

Her ønskes det en klargjøring av hva egenskapen innebærer. Er dette for å forberede seg til forespørslers om innsyn i egne data fra enkeltpersoner til Helsedataservice

4.2.5 Tillegg (40-N)

Det er en fordel at man kan lage tilleggsegenskaper. Det er behov for tilpasninger av metadata for kvalitetsregistre og vi håper denne kan ivareta disse behovene.

4.2.6 Utgått variabel: TemaLokalt

Vi forstår at TemaLokalt går ut av spesifikasjon for variabelmetadata. Vi har i pilotprosjektet brukt denne som en underinndeling, men ser at denne ikke har fungert godt, og støtter derfor fjerningen.

4.2.7 Generell kommentar – Skille variabler ut fra skjema

En variabel kan være i ulike skjema og forsker vil trolig ikke vite hva de skal velge. Vår forståelse er at helsedataservice skal bruke denne oversikten til å levere ut data. F.eks. så finnes komplikasjoner både i

hovedskjema, oppfølgingsskjema og i pasientforløpet i hjerteinfarktregisteret. Hvilke variabler som skal brukes kommer an på problemstillingen. Det er noe usikkerhet knyttet til om denne løsningen kan være et godt nok hjelpemiddel slik at helsedataservice velger rett.