

Direktoratet for e-helse
Postboks 221 Skøyen
0213 OSLO

Deres ref.: 21/121
Vår ref.: 21/6245-2
Saksbehandler: Even Gunnarson Anderssen
Dato: 25.05.2021

Nasjonal spesifisering for metadata om helsedata - høringsvar

Det vises til Direktoratet for e-helses ekspedisjon av 23. februar 2021, der det bes om innspill til utkast til revidert versjon av Nasjonal spesifisering for metadata om helsedata. Det vises også senere dialog med seniorrådgiver Truls Korsgaard om forlenget høringsfrist.

Vi vil innledningsvis takke for anledningen til å komme med innspill til arbeidet med en revidert versjon av den nasjonale spesifiseringen.

Direktoratet for e-helse ber i høringsbrevet om tilbakemelding på "generell lesbarhet". Når det gjelder dette konstaterer vi at utkastet til revidert versjon av Nasjonal spesifisering for metadata om helsedata forutsetter at leserne har inngående kunnskap om fagområdet. Dette er naturlig, sett hen til at den primære målgruppen for dokumentet er dataforvaltere og tekniske utviklere hos de nasjonale helsedataskildene. Vi ser imidlertid at en mer tilgjengelig form kunne bidratt til å gi dokumentet et noe bredere anvendelsesområde og derigjennom til en økt forståelse av betydningen av gode metadata. Vi oppfatter for øvrig at den reviderte spesifiseringen ikke bør være et "endelig produkt", men et dokument som utvikles og endres i takt med endringer og behov på området. Den videre utviklingen av spesifiseringen bør derfor, i likhet med arbeidet med utkastet til revidert spesifisering, skje i nært samarbeid med dataforvaltere og tekniske utviklere hos de ansvarlige for de nasjonale helsedataskildene.

Vi finner grunn til å gjøre oppmerksom på at det i Helse- og omsorgsdepartementets høringsnotat om forslag til forskrift om nasjonal løsning for tilgjengeliggjøring av helsedata punkt 2.11 er beskrevet at et "dataprodukt" består av et datasett, inkludert metadata, og en beskrivelse av relevante bruksområder. I utkastet til revidert spesifisering er ikke "bruksområder" tematisert. Det vil imidlertid, som følge av den planlagte etableringen av en nasjonal løsning for tilgjengeliggjøring av helsedata fra helseregistre, være behov for å avklare hvordan informasjon om "bruksområder" skal struktureres og overføres.

Overføring av registerspesifikk informasjon og kunnskap om opplysningene som inngår i dataprodukter, fra de dataansvarlige for helseregistrene til Helsedataservice, er en forutsetning for at Helsedataservice skal kunne forstå og forvalte (herunder tilgjengeliggjøre) opplysningene i dataproduktene på en god måte. Slik vi oppfatter det, favner ikke utkastet til revidert spesifisering om alle begreper og sammenhenger mv. som er avgjørende for å sikre at

HelseDirektoratet

Avdeling helseregistre

Even Gunnarson Anderssen

Postboks 6173 Torgarden, 7435 TRONDHEIM • Besøksadresse: Holtermanns veg 70, Trondheim • Tlf.: (+47) 47 47 20 20

Org.nr.: 983 544 622 • postmottak@helsedir.no • www.helsedirektoratet.no

opplysningene i et dataprodukt forstås og forvaltes riktig. Gode metadata i dataprodukter vil kunne bidra til å redusere behovet for andre former for kunnskaps- og kompetansedeling, men ikke eliminere dette.

Det fremgår av høringsbrevet at de mer konkrete høringsinnspillene ønskes strukturert i henhold til kapitler og underkapitler i spesifikasjonene, og ikke direkte i dokumentet. Vi har derfor systematisert våre innspill i en tabell, som viser hvilke konkrete kapitler, underkapitler, tabeller og figurer mv. innspillene er knyttet til.

Kapittel/figur/tabell	Tema/område	Side	Kommentar
Kap. 1.2	Formål	Side 7	I lys av det pågående arbeidet med sikte på tilrettelegging for deling av helsedata i EU, jf. Joint Action InFact og Joint Action TEHDAS, kan formålsbeskrivelsen vurderes endret slik at det fremgår at spesifikasjonen skal bidra til å sikre at metadata om helsedatakilder dokumenteres og tilgjengeliggjøres på en standardisert og strukturert måte slik at de kan forståes og utveksles mellom aktører og systemer, også på tvers av landegrenser.
Figur 1	Datamodell	Side 11	Statistikk, f.eks. beregning av kompletthet, er nyttig på variabler, men bør kunne knyttes til spesifikke datasamlinger/delmengder av data (f.eks. år/sector). I spesifikasjonen er statistikk kun koblet til variabel. Uklart om det vil være mulig å angi kompletthet på delmengder. Hva med for eksempel kompletthets-/dekningsgradsmål? For variabler vil det uansett være avgjørende å kunne knytte statistikkmålene til en delmengde av data (Aktuelle dimensjoner for NPR er f.eks. sektor/periode/rapporteringsenhet)
Figur 1	Datamodell	Side 11	Uklart hva versjon skal inneholde og hvordan dette skal brukes. Når brukes versjon og når brukes filter? Hva er versjon vs. delkilde og årgang? Behov for eksempler.
Figur 1	Datamodell	Side 11	Tungt å ivareta historiske data slik modellen er definert. F.eks. hvis

			gyldighet på variabel endrer seg og er ulik i flere perioder - vil det da måtte produseres en komplett versjonfil for hver gyldighetsperiode? Ønskelig med eksempel som illustrerer hvordan dette skal ivaretas.
Figur 5	Filformat	Side 16	Flatt filformat som skal matche en datamodell med relasjoner kompliserer løsningen med behov for mange id'er for å koble data. Det bør være et mål at overføring skal automatiseres, og det bør etableres et filformat som er bedre egnet for koblede data. (F.eks. XML eller JSON).
Figur 4	Filoverføring	Side 16	Det angis at filutveksling skal benyttes, men det er ikke beskrevet hvordan dette skal gjøres. Det bør være et mål at overføring av data skal automatiseres og en tjeneste for dette bør som et minimum være foreslått og beskrevet i dokumentet.
Kap. 4.1.1	Oppdatering	Side 17	Hvordan angi at en metadatafil som sendes er ny eller oppdatert?
Tabell 3	Kildemetadatafil	side 25	Savner informasjon om hva som er telle-enhet i filgrunnlaget (pasient/ opphold/ vedtak/ kommune/ ...)
Tabell 3	Kildemetadatafil	Side 25	Felt nr. 25 og 26. Hva er telleenheten?
Tabell 3	Kildemetadatafil	Side 25	Felt nr. 24: Definisjonen av enheter er "Virksomhet". Dette er ikke hensiktsmessig for alle register, f.eks. KPR.
Tabell 3	Kildemetadatafil	Side 25	Felt nr. 24, 25, 26: Ingen av feltene for kompletthet/dekningsgrad er obligatoriske. Bør være føringer for at minst ett slikt felt er fylt ut. Usikker på om det er lurt å legge føringer for hvordan kompletthet skal måles eller om dette er vurderinger som kan overlates til innrapporterende enhet med beskrivelse av metode (teller og nevner). Eventuelt vurdere underliggende klasse for dekningsgradsmål med en-til-mange-relasjon? For et tjenestebasert register vil tilslutningsgrad kanskje måles på enhetsnivå mens for en

			folkehelseundersøkelse vil det måles på individnivå. Se også kommentarer til felt 26 om måleenhet for dekningsgrad.
Tabell 3	Kildemetadatafil	Side 25	Felt nr. 26: Feltet er avgrenset til å gjelde pasienter. Dette er ikke hensiktsmessig for alle tilfeller, men avhengig av ønsket måleenhet. F.eks. kan antall episoder, hendelser, kartlegginger, henvisninger osv. være en annen avgrensning avhengig av hva som er dataenheten. Bør man ha to nivåer for dekningsgrad - individer og hendelser/tilfeller?
Tabell 3	Kildemetadatafil	Side 23 og 24	Felt 13 og 15: Felt for engelsk oversettelse er angitt som obligatoriske. For KPR og NPR er ikke dette informasjon som eksisterer, og vil ikke bli tilgjengelig ved overlevering av første versjon av dataprodukter.
Kap. 2	Versjon og variabler	Side 12	I flg. pkt. 2 må variabler være unike innenfor en datakilde. Betyr det at variabler med ulike egenskaper (eks gyldighet, hendelser) må være representert ved flere kontekstspesifikke variabler? Eller er det tenkt at man kan bruke versjon for å ivareta dette? Uklart hva versjon er tenkt å være. Se også tabell 4 felt 7
Kap. 4.1	Siste avsnitt	Side 15	Spesifikasjonene adresserer i liten grad spørsmål knyttet til "metadata om kvalitet". Det er beskrevet på side 15 at <i>"Det anbefales at signifikante endringer i f.eks. kodeverk framgår av proprietene "beskrivelse" eller "kommentar", og at signifikante kvalitetsbrister i datagrunnlaget framgår av proprietien "kvalitetsnote"."</i> Det bør være et mål at spesifikasjonen på sikt utvides til også å inkludere kvalitetsmetadata som gir informasjon om f.eks. følgende aspekter ved datakvalitet: <ul style="list-style-type: none"> - Nøyaktighet: Er data finmasket nok til å beskrive det man ønsker å karakterisere?

			<ul style="list-style-type: none"> - Konsistens: Fremkommer data i forventet format og legger til rette for kryss-referanser med samme resultat? - Tidsriktig: Er data oppdatert og fanger opp ev. endringer fra avsender? - Kompletthet: Graden av kompletthet i en gitt periode - Relevans: Møter data kravene til kvalitet for ulike typer bruksområder?
Tabell 3	Kildemetadafil		Savner informasjon om relevante løpenummer/ koblingsnøkler
Tabell 3 og 4	Kildemetadafil og Variabelmetadafil		Mye fritekst med stort rom for fortolkning av hva det spørres etter. Kunne mer av informasjonen ha vært kodet? Eks. felt 23 i tabell 3: 'GeografiskAvgrensning'. Generelt behov for mer presise definisjoner og eksempler på bruk
Tabell 3	Kildemetadafil	Side 25	Felt 26 - Hvordan kan man vite noe om egenskaper ved/ behandling for ved pasienter det ikke er rapportert om?
Tabell 3	Kildemetadafil	Side 25	Felt 22 - Antar dette feltet er tenkt å skulle gi en beskrivelse av populasjonen? Kan tydeliggjøres i beskrivelsen
Tabell 3	Kildemetadafil	Side 25	Felt 28 - Savner mer tydelige føringer for hvilke kriterier man skal evaluere etter
Tabell 3	Kildemetadafil	Side 27	Felt 46 - Hva betyr akkumulert i denne sammenhengen? For NPR vil data akkumuleres gjennom året (med revideringer av tidligere perioder), også lukkes produksjonen endelig per år.
Tabell 4	Variabelmetadafil	Side 28	Felt 3 - Hva betyr polyhierarki i denne sammenhengen? Behov for bedre beskrivelse og eksempler på bruk
Tabell 4	Variabelmetadafil	Side 29	Felt 10 - Hva betyr pasientvennlig?
Tabell 5	Versjon	Side 34	Felt 14 - Hva skal informasjonen i dette feltet brukes til? Er informasjonselementet hensiktsmessig plassert i modellen?
Tabell 6	Statistikk	Side 37	Felt 14 og 15 - Hva er telleenheten, og hvordan vurderes dette for variabler

			med 0 - mange relasjon til en overliggende klasse? Eks. tilstandskoder. Hva representerer da Antall som mangler innrapportering? (felt 15) Må kunne relateres til en telleenhet i en datasamling. Må kunne beskrive hvordan antall/ andel er beregnet
Tabell 4	Variabelmetadata	Side 29	Felt 15 - Hva er forskjellen på beregnet og avledet? Vi har ikke det skillet, alt som ikke foreligger slik det er innrapportert, er avledet
Kap. 4.1.1	Oppdatering av innrapportert metadata	Side 17	Spesifikasjonen beskriver at ved oppdatering eller feilretting skal nye komplette filer sendes inn. Det er ikke ønskelig å sende alle komplette årssett bakover i tid, dette med tanke på at mengden metadata med tiden vil konstant være økende. For mest praktisk og ytelsesmessig best løsning er det ønskelig å datamodellen ivaretar at det kan sendes informasjon om kun det som er nytt eller endret.
	Begrepsliste	Side 46	Det er beskrevet at "Variabelgruppe" er hierarkisk, men begrenset til 2 nivåer. Det bør vel legges til rette for hierarki med flere nivåer?

Ta gjerne kontakt dersom det skulle være behov for supplerende opplysninger knyttet til noen av våre innspill, eller for å diskutere disse.

Vennlig hilsen

Jan Arild Lyngstad e.f.
divisjonsdirektør

Olav Isak Sjøflot
avdelingsdirektør

Dokumentet er godkjent elektronisk