



Direktoratet for e-helse
Ehelse avdeling helsedata
Postboks 221 Skøyen

0213 OSLO

Oslo, 12.5.2021
Deres ref.: 21/121-1, Vår ref.: 21/3712
Saksbehandler: Borgny Vold

Høringssvar - Høring om Nasjonal spesifisering for metadata om helsedata

Statistisk sentralbyrå viser til brev av 17.2.2021 der det bes om uttalelse i forbindelse med utkast til revidert versjon av Nasjonal spesifisering for metadata om helsedata.

Statistisk sentralbyrå kommenterer forslaget i lys av SSBs oppgaver og ansvar - å produsere data, statistikk og analyser, og på den måten gi bidrag til det kunnskapsgrunnlaget som forvaltningen og samfunnet ellers trenger (jf. statistikklovens § 17-1, jf. <https://lovdata.no/dokument/LTI/lov/2019-06-21-32>).

Statistisk sentralbyrå anser dette som et svært viktig pågående arbeid, som vil berede grunnen for å kunne utnytte helsedata for ulike formål, herunder statistikkproduksjon. Standardiseringsarbeid er avgjørende for at dette skal kunne skje på en effektiv og sikker måte.

Statistisk sentralbyrås detaljere og konkrete merknader følger i vedlegget Detaljerte tilbakemeldinger – SSB.docx.

Med vennlig hilsen

Ann-Kristin Brændvang (sign.)
fagdirektør

Vedlegg

Kopi: Finansdepartementet

Vedlegg – vår ref. 21/3712

Nasjonal spesifikasjon for metadata om helsedata

Innspill strukturert i henhold til kapitler og underkapitler i spesifikasjonen

Forord

Versjonslogg

1. Innledning
 1. Bakgrunn
 2. Formål
 3. Omfang og avgrensninger
2. Logisk informasjonsmodell for metadata om helsedata
 - Angående «**Variabler** er de informasjonsbærende elementene i datakilden og definerer hvilken informasjon som finnes i registrene.» Foreslår å erstatte 'registrene' med 'de nasjonale helsedatakildene' for konsistens med utvidelsen av helsedatakildene i v1.1
 - Ettersom «En komplett oversikt over relevante fagbegrep er gjengitt i begrepslisten» bør det vurderes å inkludere Instrument og Filter også i begrepslisten.
 - Punkt 7 angående Figur 1 har «Versjon brukes med dette også til å spesifisere hvilke variabler som inngår i hvilke dataprodukt for registre som skal levere dataprodukt til Helseanalyseplattformen.» Sjekk at beskrivelsen for Versjon er oppdatert for utvidelsen av helsedatakilder til mer enn registre.
 - Det kunne være hensiktsmessig å søke på 'regist' i hele dokument og dobbeltsjekke at det ikke er unødvendige begrensninger til kun registre. Det er vanskelig for den enkelte leseren å vurdere dette selv alle stedene 'regist' forekommer.
 - Punkt 8 har «Hver variabel kan referere til en eller flere variabelgrupper», men figur 1 viser null eller flere. Multiplisitet i figuren og i teksten under figuren bør stemme overens.
3. Objektstruktur for helsedatakilder i Nasjonal variabelkatalog
 - Punkt 8 under Figur 2 Polyhieraki forekommer 7 steder i dokumentet. Bør det inkluderes i begrepslisten?
 - Punkt 8 under Figur 2 Det står «et samlehierarki der alle kodeverkene med tilhørende koder (verdier/svaralternativ) er sortert i en flat alfabetisk liste». Ettersom dette ikke er et hierarki, men en liste, bør det kanskje få en annet term enn «samlehierarki». Denne termen kun forekommer et sted i dokumentet så bør lett kunne endres.
4. Innrapportering av metadata til Nasjonal variabelkatalog
 1. Prosess for innrapportering av metadata
 2. Importfiler for metadata i Nasjonal variabelkatalog
 - Det finnes «**Error! Reference source not found**» til Tabell 10 under 4.2.3 Format for importfiler.
 - Angående bruk av Markdown syntax: Hvis dette brukes kun for å håndtere språk i tekst kunne det være aktuelt å se nærmere på hvordan språk håndteres i DCAT og EU Data Portal med mindre Markdown syntax er løsningsspesifikk for Health Term. Mer forklaring av syntaksvalg kunne legges i Begrepslisten.
 - Tabell 3 Nr 2 Det hadde vært fint med eksempler for Delkilde og Datasamling.
 - Tabell 3 Nr 17 Bør spesifisere format for å unngå YY.
 - Tabell 3 Nr 19 Dette bør spesifiseres som standisert vokabular for å unngå fritekst.
 - Tabell 3 Nr 23 Gitt at etter hvert noe kan inngå i EU Data Portal kan det være en fordel at allerede nå dette er obligatorisk med normal svar Norge hvis ikke noe mer detaljert er relevant.
 - Tabell 3 Nr 24 – Nr 27 Hvis Kildetype er 4 = Befolkningsbasert helseundersøkelse skal disse være tomme gitt at «Med enheter menes virksomheter»?
 - Tabell 3 Nr 37 – Nr 40. Det er først i Nr 41 at dataansvarlig er nevnt. Det bør nevnes også i Nr 37 – Nr 40 eventuelt tas med i Gruppering dvs. Kontaktinformasjon for dataansvarlig.

- Tabell 4 Nr 21 og Nr 22 Det hadde vært en fordel med standisert vokabular for dette. Disse kunne utvides ved behov. Det vil redusere behov for omkoding og øke sammenlignbarhet.
- Tabell 4 Nr 32 Kunne gjenbruke Standardisert vokabular: • 1 = Ja • 2 = Nei her heller enn lage en ny.
- Tabell 4 Nr 36 og nr 37 Beskrivelsene nevner «VariabelID». Kunne erstatte med «Code (VariabelID)» ettersom Nr 1 kun har Code og det er i Figur 1 at det kan bekreftes at Code i denne sammenheng er VariabelID.
- Tabell 5 Nr 12 og Nr 13 Anbefaler å erstatte 'code' med 'Code' for lesbarhet.
- Tabell 5 Nr 14 Det kunne være en fordel å erstatte 'Begge' med 'Utlevering og Innrapportering'
- 4.2.8 Importfil 1d- Statistikk Her står det «NB! Statistikken anvendes av brukerne for å identifisere kodingskvaliteten for variabelen, og ikke til analyse.» Dette er vesentlig informasjon som kunne med fordel også tas med i Begrepslisten og Punkt 10 under Figur 1.

Begrepsliste

Det anbefales at denne sorteres alfabetisk for å øke lesbarhet.

Vurder å inkludere begrepene: mapset og markdown i listen.

Instrument og Filter bør inkluderes her.

Angående «**Variabler** er de informasjonsbærende elementene i datakilden og definerer hvilken informasjon som finnes i registrene.» Foreslår å erstatte 'registrene' med 'de nasjonale helsedatakildene' for konsistens med utvidelsen av datakilder i v1.1

Generelt:

Det er viktig og positivt at det utvikles både

mapping av properties i KILDEMETADATA-filen til DCAT-AP-NO og DCAT-EN

og mapping av properties i VARIABELMETADATA-filen til SKOS-AP-NO og SKOS-EN

slik at spesifikasjonen i større og større grad vil bli forankret i internasjonale standarder.