

Respons på «Nasjonal spesifikasjon for metadata om helsedata»

Bakgrunn

Dette dokumentet er en respons fra Helse Nord IKT (HNIKT) på «Nasjonal spesifikasjon for metadata om helsedata» som er på høring fra eHelse-direktoratet. HNIKT er leverandør av nasjonale kvalitetsregistre, og har vært involvert i utvikling av pilot-løsninger for metadatautveksling.

Overordnet tilbakemelding

Spesifikasjonen er veldig detaljert og omfattende, og virker unødvendig kompleks. Selv om vi har arbeidet med disse tingene i en årrekke må vi innrømme at spesifikasjonen oppleves som tung, og vi må rett som det er tilkalle menneskelig hjelp for å oppnå forståelse. Det er nok ikke enklere for dataforvaltere som presenteres for dette for første gang.

Likevel vil vi berømme arbeidet som er nedlagt, og vi synes prosjektet er spennende, samt at vi virkelig ser fram til at registrene våre blir en del av dette.

Videre mener vi at den siste oppdateringen av dokumentet klargjør og retter opp i problematiske og merkelige ting ved den opprinnelige versjonen. Ting som at nesten alle propertyene opprinnelig var definerte som String(4000) var nesten komisk. Dette er ryddet opp i, og formatbeskrivelsene fremstår nå som mer ryddig enn tidligere. Det virker dog noe merkelig at norsk og engelsk brukes om hverandre i formatbeskrivelsene (Blant obligatoriske *properties* er *PatientFriendlyTerm* og *Beskrivelse*.) I denne siste versjonen av dokumentet er det blitt mye ryddigere, siden det ser ut til at kun HealthTerms 11 nå er på engelsk, og kun *property* gjenstår som engelsk kolonneoverskrift. Ellers applauderer vi at propertyen *TemaLokalt* er tatt ut. Den fikk vi aldri kontroll på.

Noen konkrete eksempler

Formatbeskrivelsene av importfilene har gjennomgått en stor forbedring siden dokumentets første versjon. Der har kommet inn et knippe nye properties som virker fornuftige, men *Identifiseringsgrad* kan bli litt vel for enkel for enkeltvariabler. I mange tilfeller er det en kombinasjon av to eller flere variable som utgjør en identifiseringsrisiko, selv om enkeltvariablene er risikofrie. Vi ser selvsagt at å introdusere multivariabel identifiseringsgrad høyner kompleksiteten i et dokument der kompleksiteten er for høy i utgangspunktet.

Propertyen *ValidFrom* innser vi kan bli litt vrien å håndtere etter formålet. Det står at denne ikke skal være representert i importfila, da den blir satt automatisk ved import til HealthTerm. Det vi ser for oss som et potensielt problem med denne er at registre av og til introduserer variabler med tilbakevirkende kraft. Det vil si at en nyintrodusert variabel kan ha et regelsett som også definerer verdier i eksisterende database uten at bruker må legge disse inn. Variabelen framstår dermed som om den har vært i registeret siden tidenes morgen. Ville det ikke da bli korrekt at *ValidFrom* speiler dette?

Registrene vi jobber med kan ha ganske omfattende databaser, med store mengder av variabler. Det pålegges registerledelsen veldig mye arbeid for å legge inn metadatabeskrivelser av alle variablene, uten at det framkommer godt nok hvilken gevinst de har av dette arbeidet. Hvilken gevinst har registrene av å gjøre dette egentlig? Og det virker i overkant brutalt å kreve at propertyer som *NavnEngelsk* og *BeskrivelseEngelsk* som være obligatoriske. Dette gjør arbeidsoppgaven enorm.

Til slutt savner vi en protokoll for selve filoverføringen fra registrene til mottaker. For våre registre fungerer dette i dag slik at en person som har tilgang til helsenettet manuelt logger inn i registeret, og laster ned alle relevante filer (både .xlsx- og .csv-filene) Disse pakkes så inn som epost-vedlegg og sendes manuelt til bruker utenfor helsenettet som tar dem ut av eposten for å mate dem inn i HealthTerm. Et API hadde vel gjort seg her.

Oppsummering

Dette dokumentet er et lovende skritt i riktig retning for en viktig og fremtidsrettet tjeneste. De største ankepunktene er høy kompleksitet og mangel av overføringsprotokoll.