

# Innspill FHI: Høring om Nasjonal spesifisering for metadata om helsedata

## *Generell lesbarhet og innspill til forbedring*

1. Den reviderte versjonen dekker nå også innrapportering av metadata fra nasjonale medisinske kvalitetsregistre og befolkningsbaserte helseundersøkelser. Selve teksten i dokumentet omtaler fortsatt i stor grad registermiljø, registerforvalter og registre. Disse bør erstattes med mer generelle begrep som f.eks. datakilde, dataforvalter osv.
2. Begrep/property og tilhørende beskrivelsene er i høy grad register spesifikke. Det bør etableres ett sett med felles begrep for datakilder, som ikke er spesifikke for ett enkelt datakildemiljø (her registermiljø). Ett eksempel er begrep som innrapportering og innrapporterings skjema som også skal være beskrivende for spørreskjema i en helseundersøkelse.
3. For å øke lesbarheten av dokumentet bør en erstatte engelske begrep med norske, i stedet for å veksle mellom norske og engelske begrep i dokumentet. Eksempel på engelske begrep som bør erstattes: property, concept, code system osv. Målgruppen for dokumentet er ikke kjent med lokale begrep i Healthterm, og man bør ikke bruke disse begrepene i dokumentet.

## *Versjonslogg*

**Motstridende informasjon om variabelgruppe.** Nevnes først som obligatorisk klasse, mens det står "Det er innført nye valgfrie importfiler for Variabelgruppe, Instrument og Filter" lenger nede i loggen.

## *Kapittel 2*

**Instrument i Figur 1:** Hver variabel kan referere til kun 1 instrument og 1 filter. Vil det være behov for å kunne referere til flere?

**Punktliste side 12 punkt 7:** Versjon som viser hvilke variabler som kan utleveres fra en datasamling og versjon som henviser til versjon av innrapporteringskjema/spørreskjema er vidt forskjellige konsepter som ikke bør blandes sammen, men holdes adskilt i en metadatamodell. F.eks. variabler som kan utleveres kan være utledet/basert på data fra flere versjoner av et innrapporteringskjema/spørreskjema. Det er ikke umiddelbart klart hvordan dette skal modelleres, og eventuelt vises i en metadataløsning på helsedata.no

**Logisk informasjonsmodell for metadata om helsedata:** Figuren framstår som noe rotete, der det f.eks. veksles mellom piler og streker mellom klassene. Det er også uklart hva som er forskjell på Inneholder og Referer til. Foreslår at det legges til en tabell som forklarer symbolbruk i figuren.

Skal pilene mellom variabel-variabelgruppe, variabel-instrument og variabel-filter gå motsatt vei i figuren? En referanse går begge veier, så her kan man vurdere å bruke vanlige streker.

## Kapittel 3

Formålet med spesifikasjonen bør avgrenses til hvordan metadata skal overføres - altså importformat - til nasjonal variabelkatalog. De indre anliggende i HealthTerm bør håndteres i egen dokumentasjon, og fremstår som unødvendig i denne spesifikasjonen. **Kapittel tre bør da utgå.**

## Kapittel 4

### *Generelt om kapittelet*

1. Eksemplene i tabellene bør forbedres fra generelle til mer spesifikke.
2. Det bør utarbeide to sett med testfiler som en kan bruke som mal for innrapportering; ett for minimum som kun har obligatoriske, og ett for maksimum som inneholder data i alle felt.
3. Eksportformatet kunne være uavhengig i forhold til Healthterms importformat, på denne måten kan vi som leverer metadata enklere bytte ut formatet dersom vi på sikt kan levere filer på annet vis som f.eks. API. E-helse kan selv omforme til sitt interne format. På den måten blir spesifikasjonen mye ryddigere da den vil inneholde krav til færre tomme felt, og forenkle terminologi.
4. Det er unødvendig å bruke tid på å bytte ut navnet på egenskapene. For å vise støtte mot standarder kan dette presenteres i egen koblingstabell.
5. Øk bruk av klassifikasjoner/vokabular fremfor fritekst.

## Kapittel 4.1 Prosess for innrapportering av metadata

**Kommentar til tekst «I første omgang er det laqt opp til årlige oppdateringer»:** I denne første perioden bør det legges opp til at datakilder, i tillegg til årlig oppdatering, kan sende oppdatering etter behov. Dette er viktig for at det ikke skal gå for lang tid til f.eks. metadata for nye datasamlinger blir tilgjengelig på helsedata.no. For helseundersøkelser vil dette være aktuelt når data fra nye spørreskjema blir tilgjengelig for forskning. Datakilder må og kunne sende rettinger på feil som oppdages i metadata.

**Kontaktinformasjon:** Det vil være nyttig om det settes opp en liste over hvem som skal kontaktes vedrørende ulike problemstillinger, enten i dette kapitlet eller i starten av dokumentet. Eksempler på problemstillinger: Importfiler til helsedata.no, infosider på helsedata.no, søknadsskjema på helsedata.no.

**Figur 3:** Viktig at HDS også etablerer rutiner for å fange opp feil i egen innlasting til HealthTerm/helsedata.no, og sørger for at feil som oppdages knyttet til ett register også blir rettet for andre registre.

**Kapittel 4.1.1:** Kunne vært nyttig med en referanse til «Spesifikasjon for dataoverføring til HAP» for å minne om relasjonen mellom data og metadata.

## Kapittel 4.2.2 Navnestandard for importfiler

**Første eksempel i nederste punktliste i kapitlet:** Skal det stå «Filen mapper kildemetadata og...»?

## Kapittel 4.2.3 Format for importfiler

**Første kulepunkt:** Variabler er vel ett av flere Concepts i modellen? Viser også til generell kommentar om begrepsbruk og generell lesbarhet.

## Kapittel 4.2.5 (Tabell 3: Format for importfil 1a-Kildemetadata)

Gruppering/ beskrivelse	Nr	Property	Innspill
Healthterm spesifikke felt	1-11	Code-ClinicalFriendlyTerm	I selve tabellen kommer det ikke tydelig frem at dette er healthterm spesifikke begrep (Ref. kap. 4.2.3). Dette bør gjøres tydeligere i tabell 3 til tabell 9.

Gruppering/ beskrivelse	Nr	Property	Innspill
<b>Obligatoriske felt felles for datakilde, delkilde og datasamling</b>	21, 22, 29, 30, 31, 33, 34, 37, 38, 39, 40	Kildetype, InklusjonsOgEksklusjonskriterier, formaal, lovverk, forskrift, dataansvarlig, databehandler, hjemmeside, e-post, telefonnummer, adresse	Det er uklart ut i fra tabellen om obligatoriske felt som er felles for klassene datakilde, delkilde og datasamling er obligatoriske for alle tre klasser. Det vil være kostbart å holde feltene ajour for alle klasser, og feltene er heller ikke like relevante. Disse bør derfor ikke være obligatoriske på alle nivå, men kun obligatoriske dersom de er ulike fra foreldrenoden. Eksempler på obligatoriske felt: kildetype, InklusjonsOgEksklusjonskriterier, formaal, lovverk, forskrift, dataansvarlig, databehandler, hjemmeside, e-post, telefonnummer, adresse.
<b>Administrative formål</b>	1	Code	Det er ikke klart ut i fra tekst hvem som administrere disse unike kodene, om det finnes et register av disse og hvordan de tildeles. Burde kunne bruke alle Unicode tegn uavhengig av bokstavstørrelse.
<b>Tittel og beskrivelse</b>	4	PreferredTerm	Beskrivelsen blir mer tydelig dersom eksempelet viser til samme datakilde. Foreslår å endre tekst til f.eks. "Den norske mor, far og barn undersøkelsen", "MoBa Hovedskjema" og "Spørreskjema 1" for henholdsvis datakilde, delkilde og datasamling.
<b>Administrative formål</b>	7	ValidFrom	Hva skjer om ValidFrom endres i senere overføringer, for eksempel dersom den var satt feil ved første overføring?
<b>Tittel og beskrivelse</b>	13	NavnEngelsk	Ikke alle helseundersøkelser har et offisielt engelsk navn. Det blir derfor kunstig å ha dette som et påkrevd felt. Det virker også tungvint å ha et eget felt ved «NavnEngelsk». Det er en fare for at dette gir en veldig lite objektorientert struktur som er vanskelig å forvalte og å bygge videre på i fremtiden. Dette gjelder alle «Engelske»-språk felt i tabell 3 til tabell 9.
<b>Tittel og beskrivelse</b>	16	Emneord	På sikt vil mulighet til å velge blant felles nasjonale vokabular for emneord, filtre og instrumenter kunne gi fordel for brukere av helsedata.no.
<b>Opphav og ferskhet</b>	20	Innsamlingsmetode	Det fremstår som om dette er et fritekstfelt. Dette bør være fra et styrt vokabular. Eksempel: <a href="https://ddialliance.org/Specification/DDI-CV/ModeOfCollection_3.0.html">https://ddialliance.org/Specification/DDI-CV/ModeOfCollection_3.0.html</a>

Gruppering/ beskrivelse	Nr	Property	Innspill
<b>Innhold og kvalitet</b>	23	GeografiskAvgrensning	Denne er oppført som obligatorisk i vedlegg til revidert spesifikasjon, men ikke obligatorisk i revidert spesifikasjon. Hva er riktig?
<b>Innhold og kvalitet</b>	26	Dekningsgrad	Beskrivelse av Dekningsgrad kan med fordel generaliseres litt mer, eller det kan tydeliggjøres at teksten knyttet til pasient/sykdom beskriver et eksempel.
<b>Kontaktinformasjon</b>	37-40	Hjemmeside, e-post, telefonnummer, adresse	Det vil være kostbart å holde alle feltene ajour, og det vil være tryggere at kun hjemmeside er påkrevd. Spesielt i en overgangsfase med årlig oppdatering av metadata vil det være risiko for utdaterte metadata.
<b>Kontaktinformasjon</b>	41	OrgnrDataansvarlig	Feltet «OrgnrDataansvarlig» bør være påkrevd.

#### Kapittel 4.2.6 (Tabell 4: Format for importfil 1b-Variabelmetadata)

Gruppering/ beskrivelse	Nr	Property	Innspill
		Generell kommentar	Det er ikke klart hvordan HAP skal avklare hvilke kodeverk/svaralternativ som er gyldig for en variabel i en gitt periode. Hva vil skje dersom en «utgått» kode fortsatt brukes av en datakilde.
<b>Tittel og beskrivelse</b>	4, 12, 13, 14	PrefferdTerm, NavnEngelsk, Beskrivelse og Beskrivelse Engelsk	Ikke alle helseundersøkelser har engelsk tekst. Feltene burde ikke være påkrevd.
<b>Tittel og beskrivelse</b>	4, 12	PrefferdTerm, NavnEngelsk	MoBa spørreskjema variabler har ikke beskrivende navn, men variabelnummerering (databasereferanse), spørsmålsnummer og spørsmålstekst (ledetekst). MoBa inneholder et høyt antall variabler, og det vil være svært krevende å navngi variabler med beskrivende navn. Ved påkrevd navn må disse genereres automatisk basert på f.eks. databasereferanse.
<b>HealthTerm</b>	5	ConceptType	Burde denne være tom for variabel, da alle variabler ifølge beskrivelsen uansett er «Regular»? Definisjonen av ConceptType burde kopieres til begrepslisten. Viser også til generell kommentar om begrepsbruk og generell lesbarhet.

Gruppering/ beskrivelse	Nr	Property	Innspill
<b>Administrative formål</b>	7	ValidFrom	Hva skjer om ValidFrom endres i senere overføringer, for eksempel dersom den var satt feil ved første overføring?
<b>Tittel og beskrivelse</b>	13, 14	Beskrivelse og Beskrivelse Engelsk	For MoBa erstatter ledetekst i stor grad behovet for feltet beskrivelse. Med et høyt antall variabler i MoBa vil det være svært krevende å lage en beskrivelse (norsk og engelsk) for hver enkel variabel. Her foreslås det å legge til en tekst i kolonnen Description for «Beskrivelse» og «Beskrivelse Engelsk»: Dette feltet er ikke obligatorisk dersom feltet ledetekst brukes.
<b>Opphav</b>	15	Opprinnelse	Standardisert vokabular bør inkludere papir og digitale spørreskjema. Kan disse inkluderes i punkt 5 eller bør det inngå som et eget punkt? (Forslag: 5 = Direkte fra skjema (eks. hentet direkte fra papir- eller digitalt skjema) Standardisert vokabular bør også inneholde verdien “imputert verdi”.  Ønsket bruk av standardisert vokabular dersom flere av alternativene er gyldige kan gjerne spesifiseres. Skal man prioritere fra 1-6? Eksempel: Dersom verdien er avledet basert på manuelt kodete opplysninger, skal da avledet velges?
<b>Skjema</b>	29	Ledetekst	Det mangler ett felt for Engelsk ledetekst. Denne er viktig for MoBa som har behov for å ha metadata tilgjengelig på engelsk.
<b>Skjema</b>	29, 30, 31	Ledetekst, SporsmaalNummer, SkjemaVersjon	Disse feltene er veldig relevante for befolkningsbaserte helseundersøkelser og variabler fra spørreskjema. Beskrivelsen bør oppdateres til å omfatte spørreskjema og ikke kun «registrets innregistreringsskjema» og «innrapporteringsskjema». Se generelle innspill om bruk av felles begrep.

Gruppering/ beskrivelse	Nr	Property	Innspill
<b>Administrative formål</b>	32	AnbefaltForDigitaltInnsyn	<p>Innsyn er et kompleks tema der problemstillinger vil kunne variere mellom ulike datakildene. Her må man ha en løsning på plass som dekker behov for innsyn for den enkelte datakilden før det er aktuelt å levere detaljerte metadata ift. innsyn (Annet enn 2=Bør ikke inngå).</p> <p>Eksempel: Både for MFR og MoBa vil anbefaling om hvorvidt en variabel bør inngå i et innsyn være avhengig av hvem som ber om innsyn, dvs. mor, far eller barn.</p> <p>For helseundersøkelser der samtykke ligger i bunn, er det ekstra viktig at løsning for innsyn oppleves som sømløs og profesjonell for å unngå negativ omtale. Det er viktig at deltager i helseundersøkelser kan henvende seg i en portal for å få innsyn i egne opplysninger, inkludert egne opplysninger som ikke omfattes av dataprodukter på HAP.</p>
<b>Administrative formål</b>	36,37	Erstatter, ErstattesAv	Hva er variabelID her?
<b>Administrative formål</b>	39	Synlig	<ul style="list-style-type: none"> <li>- Dette feltet bør være obligatorisk. Datakilden bør ha et forhold til om variabelen skal være tilgjengelig på helsedata.no og/eller i HealthTerm.</li> <li>- Hva er default verdi dersom denne variabelen er tom? Dette bør stå spesifisert i beskrivelsen (<i>ikke relevant dersom variabelen endres til obligatorisk</i>).</li> <li>- Det er noe uklart ut i fra teksten om variabelen vil være tilgjengelig både på helsedata.no og via API-grensesnittet i HealthTerm dersom variabelen settes til «JA». Mulig det er en skrivefeil i beskrivelsen og at det skal stå «og» i teksten i stedet for «eller».</li> <li>- Det bør spesifiseres at variabler markert med synlig = nei heller ikke vil være tilgjengelig for utlevering.</li> </ul>

### Kapittel 4.2.7 (Tabell 5: Format for importfil 1c-Versjon)

Gruppering/ beskrivelse	Nr	Property	Innspill
		Generell kommentar	Tabellen mangler kolonnen gruppering.
-	8	ValidTo	<p>“Versjoner skal ikke ha overlappende gyldighetsdatoer mellom revisjoner”: Dette vil være vanskelig å oppfylle, ettersom vi ikke vet når neste versjon skal sendes.</p> <p>Det bør ikke være obligatorisk å sette en sluttdato for en versjon frem i tid. Det må være mulig å sette ValidTo til gyldig per dagsdato.</p>

### Kapittel 4.2.8 (Tabell 6: Format for importfil 1d-Statistikk)

Gruppering/ beskrivelse	Nr	Property	Innspill
		Generell kommentar	Tabellen mangler kolonnen gruppering.
		Generell kommentar	Kunstig begrensning på de utvalgte statistikkene som kan gis. Ikke klart hva man gjør hvis man skal levere Skewness, Kurtose, Median absolute deviation. Ikke klart hvordan man beskriver usikkerhet i oppsummerende statistikk. Vil dette kunne dekkes av «mulige tilleggsproperties»?
		Generell kommentar	Det hadde vært en fordel med noen eksempler på bruk av statistikk. Tenker spesielt på hvordan statistikkår i Code, ValidFrom, ValidTo og SisteOppdaterteAarssett spiller sammen – også sett opp mot StatistikkType = «Akkumulert» i kildemetadata. Generelt viktig å ha i bakhodet at en versjon kan inneholde mange årganger.
-	8	ValidTo	<p>“Statistikk skal ikke ha overlappende gyldighetsdatoer”: Dette vil være vanskelig å oppfylle, ettersom vi ikke vet når neste statistikk skal sendes.</p> <p>Det bør ikke være obligatorisk å sette en sluttdato for statistikk frem i tid. Det må være mulig å sette ValidTo til gyldig per dagsdato.</p>



### Kapittel 4.2.9 (Tabell 7: Format for importfil 1e-Variabelgruppe)

Gruppering/ beskrivelse	Nr	Property	Innspill
		Generell kommentar	Tabellen mangler kolonnen gruppering.
		Generell kommentar	Variabelgruppe bør ha mulighet for å legge til ytterlige felt ved behov.
-	4	PreferredTerm	Spesifisere at PreferredTerm ikke kan inneholde spesialtegn, dersom dette er tilfelle. Viser til erfaring med bruk av “/”.
-	4, 12, 13, 14	PrefferedTerm, Beskrivelse, NavnEngelsk, BeskrivelseEngelsk	En del metadata i Moba er kun tilgjengelig på engelsk. MoBa inneholder et stort antall tema, og det vil være krevende å oversette disse til norsk. Det kan bli behov for å bruke engelsk metadata i feltene PrefferedTerm og Beskrivelse.
-	15	Filter	Er det noen relasjon mellom propertien Filter og klassen Filter?

### Kapittel 4.2.10 (Tabell 8: Format for importfil 1f-Instrument)

Gruppering/ beskrivelse	Nr	Property	Innspill
		Generell kommentar	Tabellen mangler kolonnen gruppering.
		Generell kommentar	Instrument bør ha mulighet for å legge til ytterlige felt ved behov.
		Generell kommentar	<p>Det er ikke umiddelbart klart hvordan rike metadata for instrument skal håndteres. Eksempel på rike metadata:</p> <ul style="list-style-type: none"> <li>- <i>Enkelt instrument: Gitt at for individ X har operatør O målt verdi V med instrumentet som har serienummer S, instrumentet har type betegnelse T med måleområde [M_I, M_U] og senterkalibrering C, instrumentet ble sist kalibrert DTG, mot standardreferanse U, avviket før kalibrering var delta.</i></li> <li>- <i>Lab analyse: Gitt at for individ X har operatør O målt verdi V med instrumentet som har serienummer S, instrumentet har type betegnelse T med måleområde [M_I, M_U] og senterkalibrering C, instrumentet ble sist kalibrert DTG, mot standardreferanse U, avviket før kalibrering var delta. Prøven ble prosessert i batch B, lå i brønn b, analytisk null hadde id N_id, og avvik fra analytisk null i denne kjøringen var N_delta.</i></li> </ul>
		Generell kommentar	Det bør være et felt for unik identifikator (DOI) til et instrument.

Gruppering/ beskrivelse	Nr	Property	Innspill
		Generell kommentar	Det er uklart hvordan systemet kan skille mellom variabler som inneholder beregnet score fra et sett med spørreskjema-variabler (definert av et instrument) og selve spørreskjema-variablene. Vil man kunne skille disse basert på feltet opphav og kobling til instrument-tabell?
		Generell kommentar	<p>I spørreskjema og telefonintervju brukes det ofte en innledende tekst eller «anchoring vignettes» for å etablere en skala eller sette «kontekst» for påfølgende spørsmål. Det er ikke umiddelbart klart hvordan dette er tenkt støttet.</p> <p>Eksempel fra spørreskjema:</p> <div style="border: 1px solid black; padding: 5px;"> <p><i>Enheter alkohol</i>  <i>For å sammenligne ulike typer alkohol spør vi etter det vi kaller alkoholenheter (= 1,5 cl ren alkohol). En alkoholenhet tilsvarer:</i>  <i>1 flaske rusbrus / cider</i>  <i>1 glass (1/3 liter) øl</i>  <i>1 vinglass rød eller hvitvin</i>  <i>1 hetvinsglass, sherry eller annen hetvin</i>  <i>1 drammeglass brennevin eller likør</i></p> </div>
-	4, 12	PrefferdTerm, Beskrivelse	Det mangler felt for NavnEngelsk og BeskrivelseEngelsk. Dette er viktig for MoBa som har behov for å ha metadata tilgjengelig på engelsk.
	4, 12	PrefferdTerm, Beskrivelse	En del metadata i Moba er kun tilgjengelig på engelsk. MoBa inneholder et større antall spørreskjemainstrument, og det vil være krevende å oversette disse til norsk. Det kan bli behov for å bruke engelsk metadata i feltene PrefferdTerm og Beskrivelse.
	15	Filter	Er det noen relasjon mellom propertien Filter og klassen Filter?

### Kapittel 4.2.11 (Tabell 9: Format for importfil 2-Kildekodeverk)

Gruppering/ beskrivelse	Nr	Property	Innspill
		Generell kommentar	Tabellen mangler kolonnen gruppering.
		Generell kommentar	Kildekodeverk bør ha mulighet for å legge til ytterlige felt ved behov.
-	8	ValidTo	Det bør ikke være obligatorisk å sette en sluttdato for et kildekodeverk frem i tid. Det må være mulig å sette ValidTo til gyldig per dagsdato.
-	4, 19	PreferredTerm, NavnEngelsk	MoBa har et høyt antall verdisett/kodeverk fra spørreskjema som ikke har beskrivende navn. Det vil være svært arbeidskrevende å navngi disse med beskrivende navn.
-	4, 12, 13	PreferredTerm, Beskrivelse, Kommentar	Det mangler felt for NavnEngelsk, BeskrivelseEngelsk og KommentarEngelsk. Dette er viktig for MoBa som har behov for å ha metadata tilgjengelig på engelsk.

### Kapittel 4.2.14 Administrative og helsefaglige kodeverk

**Kobling mot masterkodeverk:** Fungerer koblingen mot masterkodeverk som tenkt, eller er dette et fremtidsmål?

Det er tilfeller der fritekst har blitt kodet med lokale utvidelser av helsefaglige kodeverk fra tabell 11, skal heller ikke disse kodene innrapporteres?

12.mai 2021

Turid Solberg (avd. helseundersøkelser) , Guri Sundgot Halvorsen (avdeling helseregistre) og Yngve Klakegg (avd. IT systemer Bergen)