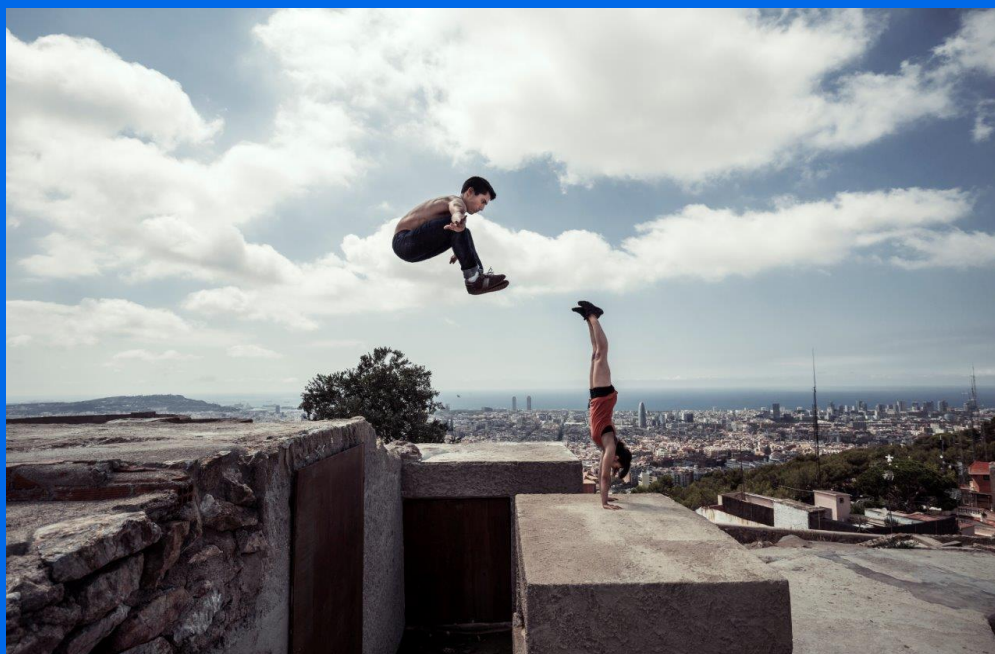


Nasjonal spesifikasjon for metadata om helsedata



HITS 1230 høringsutkast 2021

Publikasjonens tittel:

Nasjonal spesifisering for metadata om
helsedata

Rapportnummer

HITS 1230 høringsutkast 2021

Utgitt:

02/2021

Utgitt av:

Direktoratet for e-helse

Kontakt:

postmottak@ehelse.no

Postadresse:

Postboks 6737 St. Olavs plass,
0130 OSLO

Besøksadresse:

Verkstedveien 1, 0277 Oslo
Tlf.: 21 49 50 70

Publikasjonen kan lastes ned på:

www.ehelse.no

Forord

Dette dokumentet er utarbeidet av Helsedataprogrammet og beskriver hvordan metadata fra sentrale helseregistre, nasjonale medisinske kvalitetsregistre og befolkningsbaserte helseundersøkelser skal eksporteres til Nasjonal variabelkatalog, samt hvordan metadata forvaltes i Nasjonal variabelkatalog.

Målgruppen for dokumentet er dataforvaltere og tekniske utviklere i de nasjonale helsedatakildene. Spesifikasjonen setter føringer for hvordan helsedata til sekundære formål skal dokumenteres. Videre legger spesifikasjonen til rette for etableringen av en nasjonal Helseanalyseplattform ved å sikre at data som skal forvaltes på plattformen er beskrevet på en lik måte og kan utveksles maskinelt. Metadata modellen i denne spesifikasjonen vil også danne grunnlaget for å kunne spesifisere dataprodukt som skal overføres til Helseanalyseplattformen.

Spesifikasjonen bygger videre på arbeidet som ble igangsatt gjennom infrastrukturprosjektet Health Registries for Research (HRR) der de sentrale helseregistrene etablerte en felles minimumsstandard for dokumentering av registervariabler. [1] Gjennom arbeidet i Helsedataprogrammet er minimumsstandarden blitt utvidet til å inkludere dokumentering av metadata om datakildene og tilhørende datasamlinger. Spesifikasjonen er i denne versjonen også utvidet til å kunne dokumentere metadata om medisinske kvalitetsregistre og befolkningsbaserte helseundersøkelser, samt adressere noen av behovene relatert til etableringen av den nasjonale Helseanalyseplattform (HAP). Arbeidet har blitt forankret i tett dialog med berørte interessenter.

Hovedbruksområder for metadata vil være i søknadsprosessen på helsedata.no og for beskrivelse og tilrettelegging av data på HAP. Spesifikasjonen vil etter hvert revideres for å favne ytterligere behov relatert til utvikling av HAP og metadata fra flere helsedatakilder som f.eks. biobanker, samt tverrsektorielle og internasjonale behov.

Videre arbeid med begrepskoordinering på variabelnivå, bruk av kodeverk og utvikling av felles overordnede begreps- og/eller informasjonsmodeller vil foregå i nært samarbeide med Avdeling helsefaglige kodeverk og terminologi og Avdeling Internasjonale standarder og normering i Direktoratet for E-helse.

Spesifikasjonen utvikles i tråd med FAIR-prinsippene og de internasjonale standardene DCAT og SKOS.

Versjonslogg

Versjon	Dato	Endringsbeskrivelse
1.0	1.1.2020	Første versjon
1.05	20.10.2020	<p>Kapittel 2 - Logisk informasjonsmodell for metadata om helsedata:</p> <p>Oppdatert metadatamodell:</p> <ul style="list-style-type: none"> • Ny klasse for Delkilde (valgfri). Denne kan defineres i importfil 1a. Kildemetadata ved behov. • Ny klasse for Variabelgruppe (obligatorisk). Denne erstatter TemaLokalt, som utgår som egen property i variabelmetadata. Importereres som egen fil med tilhørende mapset. • Ny klasse for Instrument (valgfri). Importereres som egen fil med tilhørende mapset ved behov. • Ny klasse for Filter (valgfri). Importereres som egen fil med tilhørende mapset ved behov. <p>Kapittel 3 – Objektstruktur for helsedatakilder i Nasjonal variabelkatalog:</p> <ul style="list-style-type: none"> • Oppdatert i henhold til endringer i logisk informasjonsmodell <p>Kapittel 4.2 – Importfiler for metadata i Nasjonal variabelkatalog:</p> <p>Formatering av tekst:</p> <ul style="list-style-type: none"> • Det gis mulighet for å legge markdown-direktiver for tekstformatering i de større tekstfeltene. <p>Endringer i properties for Kildemetadata og Variabelmetadata:</p> <ul style="list-style-type: none"> • Det er innført nye properties • Enkelte tidligere properties har utgått • Noen tidligere valgfrie properties er satt til obligatoriske <p>Importfiler:</p> <ul style="list-style-type: none"> • Det er innført nye valgfrie importfiler for Variabelgruppe, Instrument og Filter. <p>Vedlegg:</p>

		<p>Nye vedlegg</p> <ul style="list-style-type: none">• Vedlegg 1: Mapping av properties i KILDEMETADATA-filen til DCAT-AP-NO og DCAT-EN• Vedlegg 2: Mapping av properties i VARIABELMETADATA-filen til SKOS-AP-NO og SKOS-EN

Innhold

Forord	3
Versjonslogg	4
1 Innledning	7
1.1 Bakgrunn.....	7
1.2 Formål.....	7
1.3 Omfang og avgrensninger	8
2 Logisk informasjonsmodell for metadata om helsedata	9
3 Objektstruktur for helsedatakilder i Nasjonal variabelkatalog	13
4 Innrapportering av metadata til Nasjonal variabelkatalog	15
4.1 Prosess for innrapportering av metadata.....	15
4.2 Importfiler for metadata i Nasjonal variabelkatalog	17
Begrepsliste	46
Referanser	48
Vedlegg 1: Mapping av properties i KILDEMETADATA-filen til DCAT-AP-NO og DCAT-EN	49
Vedlegg 2: Mapping av properties i VARIABELMETADATA-filen til SKOS-AP-NO og SKOS-EN	52

1 Innledning

1.1 Bakgrunn

Arbeidet med å definere og innrapportere standardiserte metadata fra helseregistre ble initiert gjennom infrastrukturprosjektet Health Registries for Research (HRR). [1] HRR ble avsluttet i 1. kvartal 2019, og Helsedataprogrammet viderefører dette arbeidet. Gjennom 2020 har spesifikasjonen blitt utvidet til også å inkludere metadata fra befolkningsbaserte helseundersøkelser.

Nasjonal variabelkatalog er etablert i løsningen HealthTerm, levert av CareCom AS, anskaffet av Direktoratet for e-helse.

Etableringen av Nasjonal variabelkatalog er avgjørende for utviklingen av en god variabelutforsker på helsedata.no, og vil inngå som en viktig komponent i utviklingen av en effektiv søknadstjeneste og Helseanalyseplattformen (HAP).

Gjennom etableringen av Data- og analysetjenester på Helseanalyseplattformen, vil det å overføre dataprodukt fra dagens registerforvaltning til Helseanalyseplattformen stå sentralt. Gode metadata vil være avgjørende for at Helsedataservice skal kunne forvalte dataprodukt på Helseanalyseplattformen. Videre vil datamodellen for metadata danne grunnlaget for å spesifisere dataprodukt som skal overføres til Helseanalyseplattformen. Selve dataoverføringen er nærmere forklart i *Spesifikasjon til dataoverføring til Helseanalyseplattformen*.¹

Behovet for metadata om helsedatakilder og deres innhold (variabler), er tydelig forankret i Nasjonal e-helsestrategi 2017-2022. [2]

1.2 Formål

Spesifikasjonen skal bidra til å sikre at metadata om helsedatakilder dokumenteres og tilgjengeliggjøres på en standardisert og strukturert måte slik at de kan forstås og utveksles på tvers av aktører og systemer.

Spesifikasjonen er tiltenkt å dekke informasjonsbehovet for bruk av dataen til sekundære formål som forskning, kvalitetsforbedring og annen analyse. Spesifikasjonen legger derfor føringer på hvilken informasjon som må innrapporteres og hvilken informasjon som på sikt bør innrapporteres.

Metadata vil utveksles med helsedata.no for å understøtte søknadsprosessen i Felles Søknadsskjema. Overføringen av metadata vil derfor være en forutsetning for overføring av data til Helseanalyseplattformen. Helseanalyseplattformen validerer dataene på plattformen opp mot metadataene, og koblingen mellom data og metadata er avgjørende for at dataene skal være forståelige ved utlevering.

¹ *Spesifikasjon for dataoverføring til Helseanalyseplattformen* er p.t. en prosjektleveranse fra Helsedataprogrammet i forbindelse med utviklingen av Helseanalyseplattformen og deles på forespørsel. Spesifikasjonen skal 1. januar 2021 videreutvikles og etableres som en normerende retningslinje på linje med *Nasjonal spesifikasjon for metadata om helsedata*.

1.3 Omfang og avgrensninger

Dette dokumentet er i første omgang utformet som en spesifisering. Etter hvert som spesifiseringen implementeres, metadata fra nye datakilder importeres og nye erfaringer gjøres, er målet at spesifiseringen skal forankres som en standard. Dette er en lengre prosess og avhenger av tilstrekkelig forankring i sektoren.

Spesifiseringen omfatter sentrale helseregistre, nasjonale medisinske kvalitetsregistre og befolkningsbaserte helseundersøkelser. Fremtidige versjoner av spesifiseringen vil også kunne stille krav til innrapportering av metadata fra andre helsedatakilder som f.eks. biodata og gendata.

Spesifiseringen gir ikke føringer for hvordan metadata skal struktureres og forvaltes i lokale løsninger hos den enkelte dataforvalter, kun om hvordan dette skal innrapporteres til Nasjonal variabelkatalog.

Beskrivelsene av metadataene er i stor grad basert på de internasjonale standardene DCAT [3] og SKOS [4], og Digitaliseringsdirektoratets norske versjoner av disse [5] [6], noe som gjør det langt enklere å dele og validere kvaliteten på metadata på tvers av sektorer og landegrenser. Et forslag til mapping mellom properties fra Nasjonal spesifisering for metadata om helsedata og de nevnte standardene finnes i Vedlegg 1: Mapping av properties i KILDEMETADATA-filen til DCAT-AP-NO og DCAT-EN og Vedlegg 2: Mapping av properties i VARIABELMETADATA-filen til SKOS-AP-NO og SKOS-EN.

Spesifiseringen forvaltes og videreutvikles i henhold til "Veileder for bruk av FAIR-prinsippene for helsedatakilder" [7].

2 Logisk informasjonsmodell for metadata om helsedata

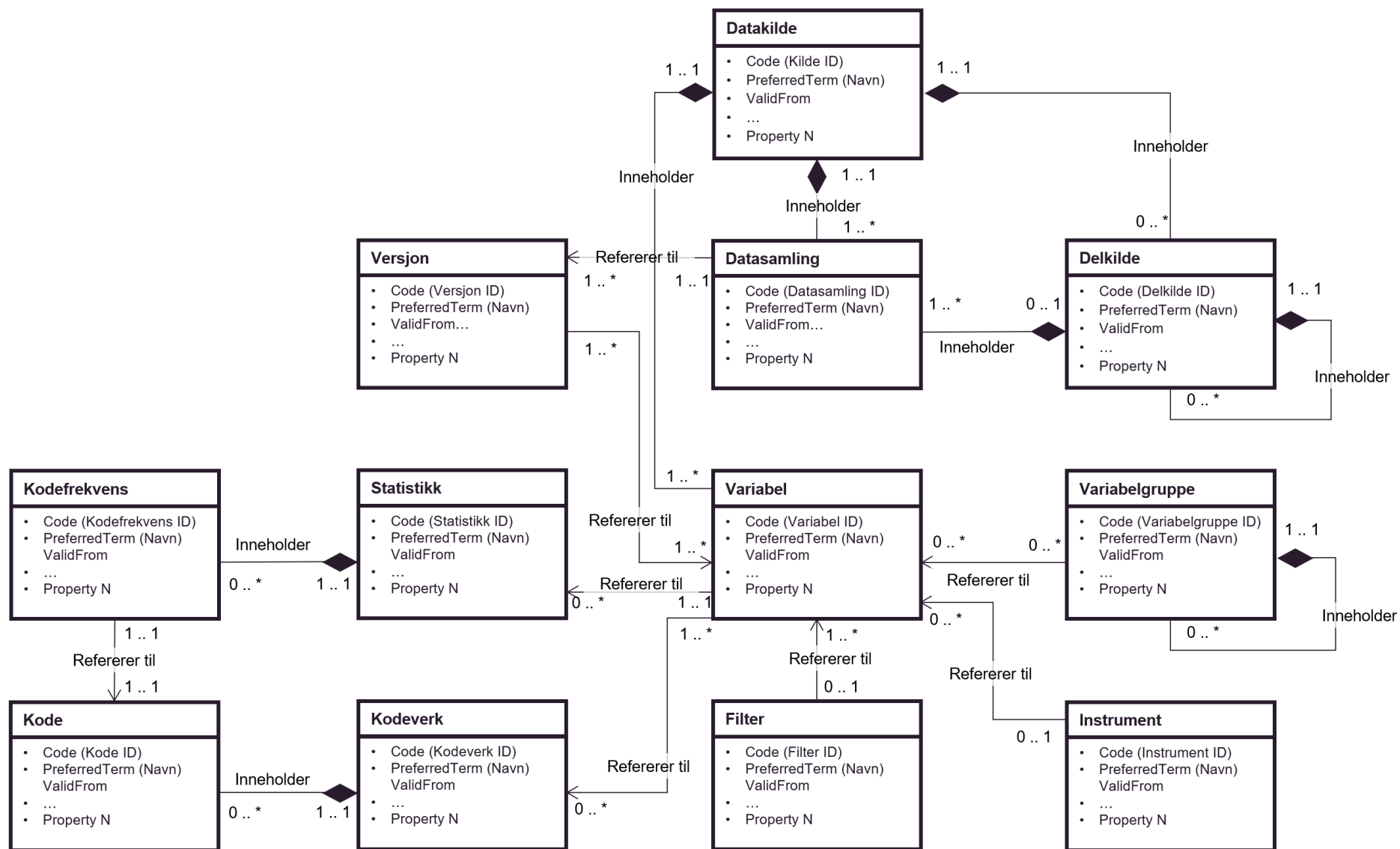
Dette kapitlet introduserer den logiske informasjonsmodellen for metadata om helsedata til sekundærbruk, som illustrert i **Figur 1**. Med logisk informasjonsmodell menes hvilke klasser som inngår i informasjonsmodellen og hvordan de logiske relasjonene mellom de ulike klassene skal håndteres. Kapittel 3 beskriver hvordan dette konkret skal håndteres i den valgte tekniske løsningen.

For å forstå den logiske informasjonsmodellen i **Figur 1** vil det være hensiktsmessig å ha kjennskap til et utvalg begrep:

- **Dataprodukt** er ikke en klasse i modellen under, men en konseptuell betegnelse på måten data skal overføres fra registerforvaltere til Helseanalyseplattformen. På et konseptuelt nivå består dataprodukt av det faktiske datasettet, samt metadata om variabler, versjon og datasamling. Sammenhengen mellom metadata og datasett som overføres til Helseanalyseplattformen er forklart nærmere i Spesifikasjon for dataoverføring til Helseanalyseplattformen.
- **Datakilde** er definert som kildene til dataene. I denne versjonen av spesifikasjonen er en datakilde ekvivalent med et helseregister, en helseundersøkelse eller et medisinsk kvalitetsregister.
- **Delkilde** kan brukes dersom datakilden har ulike deler som må beskrives ut ifra en logisk, kronologisk, juridisk eller populasjonsbasert betraktning. HUNT1, 2, 3 og 4 er eksempel på delkilder. Denne kan ha flere nivåer. For registre som skal levere dataprodukt til Helseanalyseplattformen, kan klassen Delkilde også brukes for å lage logiske grupperinger av flere dataprodukt.
- **Datasamling** er definert som et sub-register innenfor en datakilde, altså en delmengde av variablene i den aktuelle datakilden. Eksempler på datasamlinger kan være skjema i medisinske kvalitetsregistre. Klassen datasamling skal også brukes til å beskrive dataprodukter.
- **Versjon** er et teknisk hjelpemiddel som brukes til å definere hvilke variabler som inngår i ulike årganger/versjoner av de ulike datasamlingene/dataproduktene i en datakilde.
- **Variabler** er de informasjonsbærende elementene i datakilden og definerer hvilken informasjon som finnes i registrene.
- **Variabelgruppe** er en kategorisering av variabler. Strukturen kan bygges opp hierarkisk. Inntil videre begrenset til 2 nivåer.
- **Kodeverk** representerer i denne sammenhengen et verdisett. Et kodeverk kan for eksempel representere en variabls verdisett. Det kan også representere en samling verdier som brukes til å beskrive en datakilde.
- **Kode** er i denne sammenheng det samme som et svaralternativ i et verdisett.
- **Statistikk** er definert som de statistiske egenskapene ved en årgang/versjon av en variabel.

- **Kodefrekvens** er definert som antall gyldige tilfeller av et svaralternativ/utfall av en variabel.
- **Instrument** kan benyttes i forbindelse med helseundersøkelser og medisinske kvalitetsregistre. Det er en betegnelse på en standardisert innsamlingsmetode for data. Dette kan være hele eller deler av et spørreskjema, eller et fysisk instrument for f.eks. måling av bentetthet, analyse av prøvemateriale etc.
- **Filter** kan benyttes for å kategorisere variabler ved hjelp av kodeverk eller egenskaper som ikke er definert som generelle properties i denne spesifikasjonen. Et filterkodeverk kan f.eks. bestå av koder for Mor, Barn og Far, der en variabel som kun benyttes for Far, knyttes til filteregenskapen Far.

En komplett oversikt over relevante fagbegrep er gjengitt i begrepslisten.



Figur 1 Logisk informasjonsmodell for metadata om helsedata til sekundærbruk

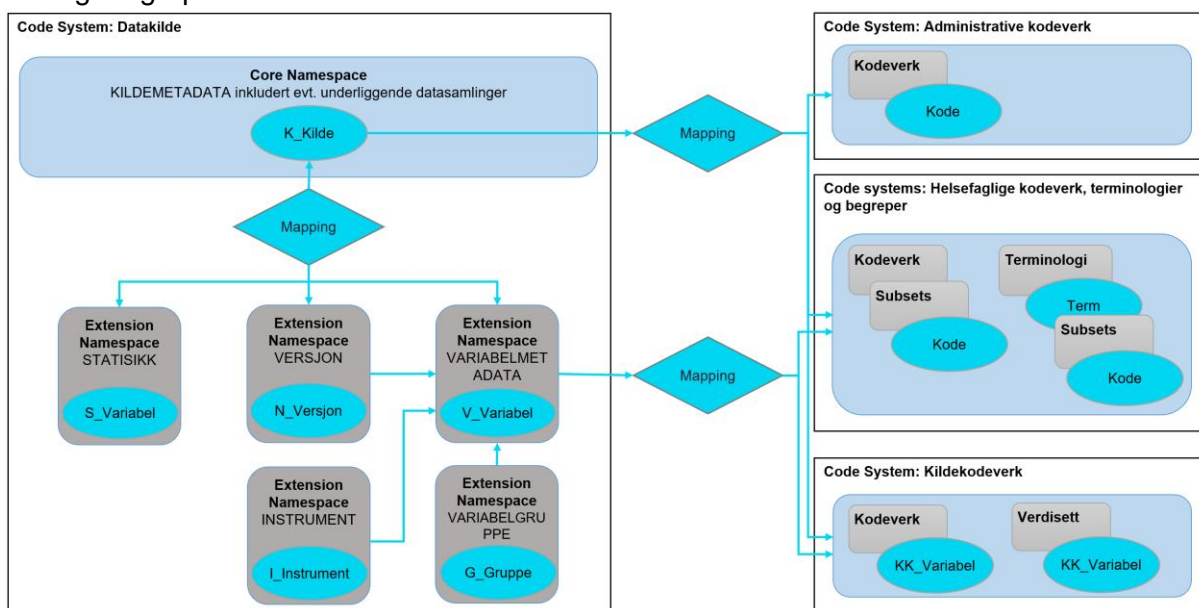
Punktlisten under forklarer informasjonslisten illustrert i **Figur 1**.

1. Hver datakilde, delkilde og datasamling er definert ut fra et sett med egne *properties*.
2. En datakilde kan *inneholde* null, én eller flere delkilder.
3. En datakilde eller delkilde *inneholder* én eller flere datasamlinger.
4. For registre som leverer dataprodukt til Helseanalyseplattformen, skal klassen *datasamling* benyttes til å dokumentere metadata om dataprodukt.
5. En datakilde *inneholder* variabler som er definert ut fra et sett *properties* om variabelen. Samme variabel kan inngå i flere datasamlinger. Det vil si at Datakilden i praksis inneholder en pool av alle variabler for alle datasamlingene under denne kilden. Derfor må en variabel være unik innenfor en datakilde.
6. Hver datasamling *referer til* en eller flere versjoner som definerer hvilke variabler som inngår i respektive versjoner av datasamlingen, og når datasamlingene er oppdaterte med nytt datagrunnlag.
7. Hver versjon *refererer til* en eller flere variabler. Det skal finnes minst en versjon som viser hvilke variabler som kan utleveres fra en datasamling. I tillegg kan det defineres versjoner som henviser til versjon av innrapporteringsskjema. Versjon brukes med dette også til å spesifisere hvilke variabler som inngår i hvilke dataprodukt for registre som skal levere dataprodukt til Helseanalyseplattformen.
8. Hver variabel kan *referere til* en eller flere variabelgrupper. Standard variabelgrupper for utvalg og filtrering kan defineres av Helsedataservice.
9. Variabler kan *referere* til kodeverk. For diskrete variabler (variabler med svaralternativ) utgjør dette kodeverket variabelens standardiserte vokabular der kodeverket *inneholder* koder som representerer alle mulige verdier (svaralternativ) en variabel kan ha.
10. Hver variabel kan også *referere* til statistikk om seg selv, som gjennomsnitt, standardavvik og årstall statistikken er gyldig for. For diskrete variabler kan statistikk *inneholde* kodefrekvens, det vil si antall utfall av de ulike mulige verdiene i variabelens standardiserte vokabular.
11. Hver variabel kan også *referere til* et instrument.

3 Objektstruktur for helsedatakilder i Nasjonal variabelkatalog

Den valgte *objektstrukturen* i Nasjonal variabelkatalog reflekterer den logiske informasjonsmodellen definert i **Figur 1** og er implementert innenfor mulighetsrommet som ligger i HealthTerm som løsning.

Figur 2 illustrerer hvordan objektstrukturen i HealthTerm skal benyttes for å modellere helsedatakilder i løsningen. Figuren er forklart nærmere i punktlisten under. Løsningsspesifikke begrep som *Code System*, *concept* og *property* er forklart nærmere i vedlagt begrepsliste.



Figur 2 Objektstruktur for helsedatakilder i HealthTerm

1. Hver enkelt datakilde importeres som et eget *Code System* som kan betraktes som en mappe der all informasjon om kilden med tilhørende variabler ligger lagret.
2. KILDEMETADATA importeres som *Code System* og blir dermed *Core Namespace* under *Code System*. KILDEMETADATA består av et *concept* med tilhørende *properties* om datakilden. KILDEMETADATA inneholder også et *concept* for hver eventuelle delkilde og for hver av datasamlingene. Hvert *concept* beskrives ved hjelp av et utvalg av de samme *properties* som er definert for kilden. I de tilfellene der en datakilde kun består av én datasamling, vil man likevel opprette ett *concept* for datasamling. Dette *conceptet* vil inneholde den samme informasjonen i de samme *propertyene* som *conceptet* for datakilden.
Hvis det er behov kan datakilden og datasamlingene deles inn i en hierarkisk struktur av delkilder. Hver delkilde defineres da som "children" av datakilden og datasamlingene blir "children" av en delkilde istedenfor selve kilden.
3. VARIABELMETADATA importeres som en *Extension* til *Core Namespace* (KILDEMETADATA). VARIABELMETADATA inneholder en oversikt over alle

variabler i datakilden modellert som egne *concepts* med tilhørende *properties*. Variablene ligger som en flat liste under VARIABELMETADATA.

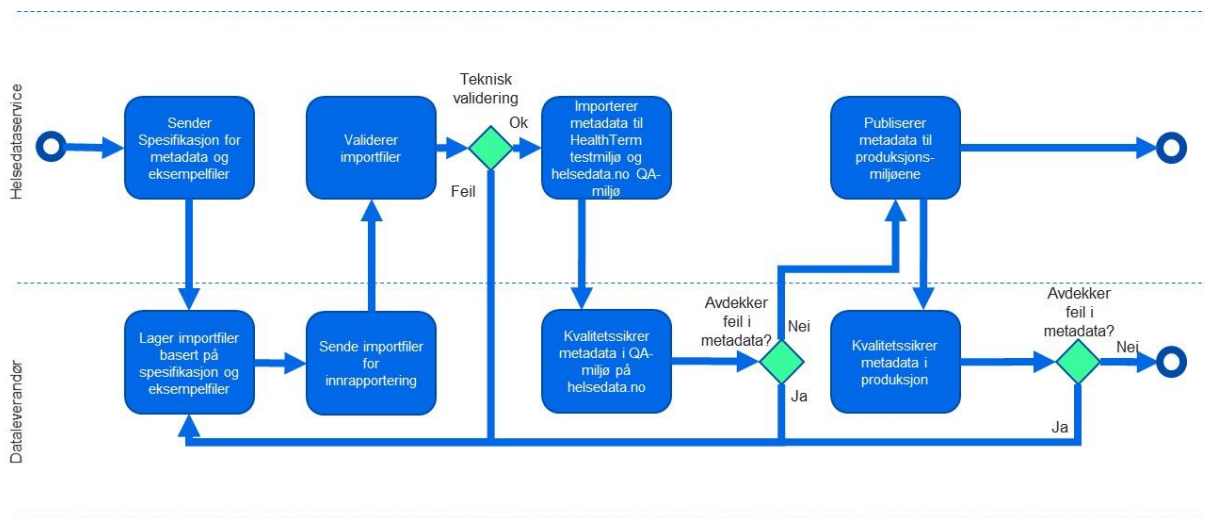
4. STATISTIKK importeres som en *Extension* til *Core Namespace*. STATISIKK inneholder statistisk informasjon om utvalgte variabler definert i VARIABELMETADATA fordelt per datasamling og VERSJON som variabelen inngår i.
5. VERSJON importeres som en *Extension* til *Core Namespace*. VERSJON brukes for å koble hvilke variabler som inngår i hvilke versjoner av en datasamling, og for å vite når datagrunnlaget er oppdatert.
6. VARIABELGRUPPE importeres som en *Extension* til *Core Namespace*. VARIABELGRUPPE brukes for å ordne variabler etter tema. VARIABELGRUPPER kan bygges opp hierarkisk. Det kan være maksimalt 2 nivåer.
7. INSTRUMENT importeres som en *Extension* til *Core Namespace*. INSTRUMENT er en betegnelse på en standardisert innsamlingsmetode for data. Det kan være hele eller deler av et spørreskjema, eller et fysisk instrument for f.eks. måling av bentetthet, analyse av prøvemateriale etc.
8. KILDEKODEVERK importeres til et eget og felles *Code System* som inneholder alle kodeverkene som anvendes i respektive datakilder og datasamlinger. Kodeverkene er definert som egne *concepts* med tilhørende *properties*. KILDEKODEVERK er strukturert ved hjelp av polyhierarkier for enkelt å kunne identifisere hvilke kilder og datasamlinger kodeverkene anvendes i og hvem som eier disse, samt et samlehierarki der alle kodeverkene med tilhørende koder (verdier/svaralternativ) er sortert i en flat alfabetisk liste. Sistnevnte for enklere å kunne identifisere kodeverk og/eller koder som kan/bør harmoniseres på tvers av kildene. Strukturen er lik den som brukes for administrative kodeverk.
9. Kodeverk som er basert på enten nasjonale administrative kodeverk, nasjonale helsefaglige kodeverk eller terminologier, eller *Subsets* av disse, kan mappes til direkte. Når man ønsker det, kan da de samme kodeverkene settes til *Retired* i KILDEKODEVERK. Nasjonale administrative kodeverk, nasjonale helsefaglige kodeverk og terminologier har egne *Customer Codes* i HealthTerm og importeres som separate *Code Systems*. Disse forvaltes av Direktoratet for E-helse, avdeling helsefaglige kodeverk og terminologi. Hvilke nasjonale kodeverk det skal mappes til direkte fra variablene er definert i kapittel 4.2.14.
10. Datasamlinger mappes til respektive VERSJONER som igjen mappes til de variablene som inngår i gjeldende versjon av datasamlingen.
11. Variablene i VARIABELMETADATA mappes til respektive KILDEKODEVERK, respektiv STATISTIKK, og evt. andre referanser som f.eks. begrepskataloger
12. For koordinerings-, harmoniserings- og berikningsformål kan variabler mappes mot flere relevante kodeverk/koder, klassifiseringer, terminologier/termer, og etter hvert f.eks. internasjonale standarder, begreps- og informasjonsmodeller som HL7 FIHR, SnomedCT, Contsys etc.

4 Innrapportering av metadata til Nasjonal variabelkatalog

4.1 Prosess for innrapportering av metadata

Metadata om datakilder, datasamlinger, versjoner, variabler og kodeverk må i størst mulig grad speile datagrunnlaget i kildene, både i nåtid og historisk. Filstrukturen er utviklet blant annet for å kunne knytte versjoner av datasamlinger og variabler opp mot metadata og statistikk, på et gitt tidspunkt og/eller i en gitt periode. I første omgang er det lagt opp til årlige oppdateringer. På sikt vil behovet for oppdatering av metadata bli styrt av oppdateringsfrekvensen for data som eksporteres til Helseanalyseplattformen.

I et fremtidig mål bilde er det ønskelig at innrapporteringsprosessen i størst mulig grad skal være automatisert. Etersom dagens tekniske løsning for Nasjonal variabelkatalog foreløpig ikke støtter automatisert innlesning av metadata, vil man i en overgangsfase legge opp til en prosess der registrene lager et sett med eksportfiler som oversendes til Helsedataservice. Det må utredes nærmere hvordan en mer automatisert prosess skal se ut. Figur 3 viser hvordan denne prosessen foregår i dag. Det forutsettes at man har kontaktet Helsedataservice og planlagt en prosess for å inngå i økosystemet rundt Helseanalyseplattformen.

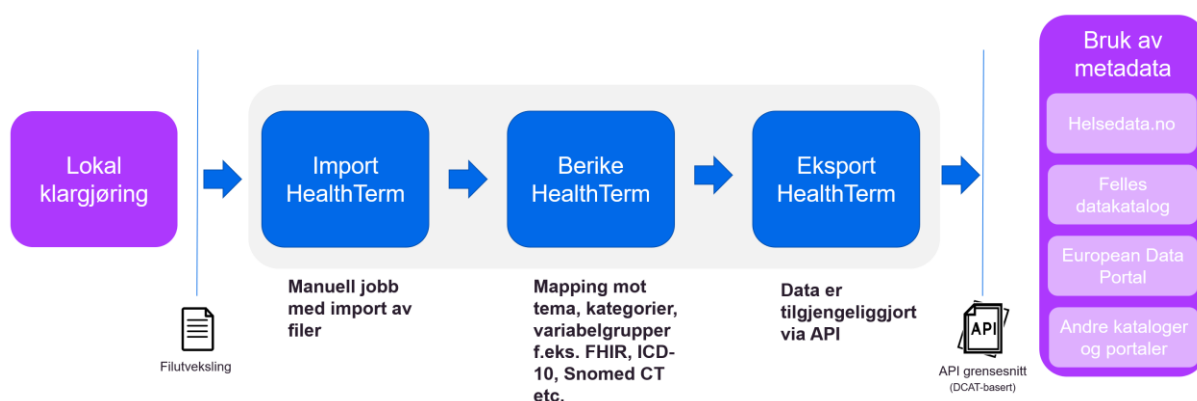


Figur 3 Stegene for manuell innrapportering av metadata til HealthTerm og publisering til helsedata.no

Ved førstegangs import av metadata er det mulig å bygge opp en endringshistorikk tilbake i tid ved å anvende propertiene ValidFrom og ValidTo. Variabler med ValidTo-dato må fjernes ved påfølgende oppdateringer. Historikk etter førstegangs import ivaretas ved at påfølgende filer importeres i kronologisk rekkefølge. ValidFrom settes til datoen innholdet gjelder fra. Dette styres i importgrensesnittet, og skal ikke angis i filen(e). HealthTerm sporer så automatisk alle endringer. Det anbefales at signifikante endringer i f.eks. kodeverk framgår av propertiene "beskrivelse" eller "kommentar", og at signifikante kvalitetsbrister i datagrunnlaget framgår av proprietien "kvalitetsnote".

Helsedataservice vil ha i oppgave å importere metadataene i Nasjonal variabelkatalog, oppdatere dem, og tilgjengeliggjøre dem via et API til eksterne kilder, i første omgang for helsedata.no men etter hvert også for andre kilder som f.eks. Felles datakatalog og European Data Portal. Arbeidet med å forvalte og tilgjengeliggjøre metadata fra Nasjonal variabelkatalog vil gjøres i nært samarbeid med registermiljøene, og metadata skal ikke eksponeres på helsedata.no eller andre kilder før registerforvalterne har fått anledning til å kvalitetssikre og godkjenne innholdet. Det legges derfor opp til et kontrollregime der registermiljøene gis anledning til å kontrollere egne variabler og metadata før disse publiseres åpent på helsedata.no.

Forvaltning av Nasjonal variabelkatalog omfatter også tilrettelegging for harmoniseringsarbeid for realisering av *Felles språk i helse- og omsorgssektoren*, [8] samt videreutvikling av mappinger og relasjoner mot andre kodeverk, terminologier og standarder. Dette er illustrert i Figur 4.



Figur 4: Prosess for import, forvaltning og eksport av variabler og metadata

For å kunne realisere den logiske informasjonsmodellen i Kapittel 2, vil registrene bli bedt om å lage inntil ni ulike filtyper. Disse er illustrert i Figur 5. Merk at en datakilde i Nasjonal variabelkatalog vil ha behov for flere mapset, derav Importfil 4x. Merk videre at det i denne versjonen av spesifikasjonen ikke forventes at registrene skal rapportere *Importfil 3: Filter* selv.



Figur 5 Importfiler til Nasjonal variabelkatalog per datakilde

De ulike importfilene vil være basert på to ulike importformater i HealthTerm:

- **11+N-formatet:** En fil med 11 obligatoriske *properties*, definert i HealthTerm, og med et ubegrenset antall tilleggs*properties* som kan legges inn i løsningen. 11+N-formatet er standard format for import av *Code Systems* og tilhørende *Extensions*. Dette formatet skal benyttes for importfilene 1a,1b,1c,1d, 1e, 1f og 2.
- **Format mappingfil:** En fil bestående av tre kolonner: løpenummer, kilde (Source) og mål (Target), der ett (eller flere) *concepts* mappes til ett (eller flere) andre *concepts*. Dette vil være standard for importfil 4x-Mapset.

11+N-formatet og mappingformatet er nærmere dokumentert i CareCom sin egen dokumentasjon av løsningen og kan deles av Helsedataservice på forespørsel.

4.1.1 Prosess for oppdatering av innrapporterte metadata

Oppdateringer av metadata gjøres ved at den eller de filene der ny informasjon er tilgjengelig sendes inn på nytt. Også feilretting gjøres ved å sende inn nye komplette filer. En oppdatering kan for eksempel være å legge til en ny variabel, fjerne en variabel som er utgått, endre navn på en variabel eller endre innholdet i en *property*. Historikk ivaretas av sporingsfunksjonaliteten i HealthTerm. Dato for når en endring ble gjeldene kan overstyres i importgrensesnittet.

4.2 Importfiler for metadata i Nasjonal variabelkatalog

Dette kapittelet beskriver i detalj kravene til de ulike importfilene definert i Figur 5 og relaterer de ulike importfilene til klassene i den logiske informasjonsmodellen illustrert i Figur 1.

4.2.1 Filtyper for importfiler

Alle importfiler skal være i Excel-format (*.xlsx eller *.xls). Mapping-fil skal være i CSV-format (*.csv) med komma (',') som separasjonstegn, ikke semikolon (;').

4.2.2 Navnstandard for importfiler

For å sikre sporbarhet skal importfilene definert i Figur 5 gis et standardisert navn.

Filtyperne 1a, 1b, 1c, 1d, 1e, 1f og 2 navngis i henhold til følgende standard:

DATAKILDE_Importfiltype-Importfilnavn_Årssett*_Versjon*_Dato sendt (YYYYMMDD)

*Versjon brukes for kilder som oppdateres og/eller endres flere ganger i året

*Årssett er siste komplette årgang som er inkludert i datagrunnlaget

Eksempler:

- DAR_1a-Kildemetadata_2018_v128_20190130
(Importerer kildemetadata om Dødsårsaksregisteret som er gyldig for 2018)
- NPR_1b-Variabelmetadata_2017_v17.b_20180123
(Importerer variabler med metadata fra NPR gyldige for 2017)

Filtype 4x navngis i henhold til følgende standard:

DATAKILDE_Importfiltype-Importfilnavn_FRA-TIL*_Årssett_Versjon_Dato sendt (YYYYMMDD)

*FRA-TIL. Mapset skal navngis på en slik måte at det framgår hva det mappes fra og hva det mappes til. Følgende forkortelse skal brukes:

- V for variabelmetadata
- K for kildemetadata, delkilder og tilhørende datasamlinger
- N for versjon
- S for statistikk
- I for Instrument
- G for variabelgruppe
- KK for kildekodeverk
- AK for administrative kodeverk
- HK for Helsefaglige kodeverk og andre masterkodeverk (i henhold til tabell 11)
- F for filter
- For andre mappinger brukes en egendefinert logisk forkortelse

Eksempler

- KREG_3a_MAPSET_K-N_2018_V34d_20190402
(Filen mapper KILDEKODEVERKET og tilhørende datasamlinger i Kreftregisteret til VERSJON 2018)
- KREG_3b_MAPSET_N-V_2018_V34d_20190402
(Filen mapper fra VERSJON til de variablene i Kreftregisteret som inngikk i VERSJON 2018)
- KREG_3c_MAPSET_V-KK_2018_V34d_20190402
(Filen mapper variablene i Kreftregisteret til KILDEKODEVERK gyldige i 2018)
- KREG_3d_MAPSET_V-S_2018_V34d_20190402
(Filen mapper variablene i Kreftregisteret til STATISTIKK for i 2018)

4.2.3 Format for importfiler

Importfilene for metadata i HealthTerm er spesifisert i **Error! Reference source not found.** til Tabell 10. Følgende begrep brukes for å definere importsesifikasjonen.

- **Concept** er det grunnleggende dataobjektet i HealthTerm. Variabler er derav definert som et *concept*.
- **Property** inneholder egenskaper ved, og metadata om, et *concept* – en variabel. De første 11 standard *propertiene* er systemspesifikke og begrepene tilsvarer de som vises i HealthTerm. Begrepene som anvendes på de resterende *propertiene* er definert i samarbeid med registerforvalterne og med utgangspunkt i aktuelle standarder. Informasjonsinnholdet importeres som tekst, og har en maksimal lengde på 4000 tegn for hver property.
- **Obligatorisk** spesifiserer om informasjonsinnholdet er obligatorisk eller valgfritt.

- **Beskrivelse** inneholder beskrivelser av informasjonen i kolonnene.

Importfilene skal definere både struktur og innhold i datakilden. Strukturen etableres ved hjelp av mor-barn relasjoner mellom de ulike konseptene som importeres. **Rot-noden** i hele hierarkiet defineres ved hjelp av 1. linje i importfil 1a: Kildemetaddata. Den identifiseres ved en unik kode for kilden, for eksempel DAR. Dette konseptet skal ikke ha noen verdi i ParentCode.

Denne koden vil selv være ParentCode for **Toppnodene** i hver av hovedgrenene i kodesystemet. Hovedgrenene er Kildemetaddata, Versjon, Variabelmetaddata, Statistikk, Variabelgruppe og Instrument. Disse toppnodene vil defineres ved hjelp av 1. linje i de enkelte importfilene, med unntak av kildemetaddata, der toppnoden defineres av 2. linje i importfil 1a. Toppnodene skal defineres som beskrevet i tabell 1. *Code* – feltet i toppnoden vil være ParentCode til alle de etterfølgende radene.

Tabell 1: Hierarkisk oppbygging av importfilene					
Filtype	Code	Parent Code	Hierar chy	PreferredTerm	ConceptType
1a: Kildemetaddata	DAR			Dødsårsaksregisteret	GROUPING
1a: Kildemetaddata	K_DAR	DAR		KILDEMETADATA	REGULAR
1b: Variabelmetaddata	V_DAR	DAR		VARIABELMETADATA	GROUPING
1c: Versjon	N_DAR	DAR		VERSJON	GROUPING
1d: Statistikk	S_DAR	DAR		STATISTIKK	GROUPING
1d: Statistikk	S_DAR_2017. DAR	S_DAR		DØDSÅRSAKSREGISTERET 2017	GROUPING
1e: Variabelgruppe	G_DAR	DAR		VARIABELGRUPPE	GROUPING
1f: Instrument	I_DAR	DAR		INSTRUMENT	GROUPING

Importfil 1d: Statistikk skal inneholde et ekstra nivå i hierarkiet. Nivået skal angi statistikkår og datasamling. Det er *code*-verdien på dette nivået som skal være ParentCode for alle underliggende statistikkverdi-rader.

For **Importfil 2: Kildekodeverk** brukes det som kalles polyhierarki. Før importen etableres det en grunnstruktur i kodeverket. Denne inneholder en felles toppnode KK_KODEVERK, med tre grener. Alle, Eier og Kilde. Denne strukturen defineres som en del av systemadministrasjon før import. Innholdet i Importfil 2 skal utformes på grunnlag av denne strukturen. Det vil si at hver kode må gjentas for hver gren i hierarkiet.

Tabell 2: Polyhierarkisk oppbygging av Importfil 2: Kildekodeverk				
Code	ParentCode	Hierarchy	PreferredTerm	ConceptType
KK_KODEVERK*		KILDEKODEVERK	KILDEKODEVERK	GROUPING
KK_ALLE*	KK_KODEVERK	KILDEKODEVERK	ALLE	GROUPING

KK_EIER*	KK_KODEVERK	KILDEKODEVERK	EIER	GROUPING
KK_KILDE*	KK_KODEVERK	KILDEKODEVERK	KILDE	GROUPING
KK_<Eier>**	KK_EIER	KILDEKODEVERK	<Kortnavn>	GROUPING
Eksempel: Variabelen KJONN_K Tilhører datakilden KK_DAR (Dødsårsaksregisteret) har eieren KK_FHI og inngår også i strukturen «Alle»				
KK_FHI	KK_EIER	KILDEKODEVERK	FHI	GROUPING
KK_DAR	KK_FHI	KILDEKODEVERK	DØDSÅRSAKSREGISTERET	GROUPING
KK_DAR	KK_KILDE	KILDEKODEVERK	DØDSÅRSAKSREGISTERET	
KK_DAR.KJONN_K	KK_ALLE	KILDEKODEVERK	Kjønn	
KK_DAR.KJONN_K	KK_DAR	KILDEKODEVERK	Kjønn	
*Grunnstruktur som KILDEKODEVERK importeres inn i defineres som en del av systemadministrasjon før import. **Eiere defineres som en del av grunnstrukturen. Eksempel KK_FHI, KK_HDIR, KK_KREFT etc.				

4.2.4 Formatering av tekst

Properties med beskrivende tekst kan formatteres ved hjelp av Markdown syntaks. Dette brukes kun i propertiene «Beskrivelse», «BeskrivelseEngelsk», «Kommentar», «KommentarEngelsk», «InklusjonsOgEksklusjonskriterier», «OpprinnelseUtledning» og «OpprinnelseUtledningEngelsk»

Formatet støtter bare et utvalg formatteringskoder:

Overskrift:	#Tittel ##Kapittel ###Avsnitt	Eks: Tilsvarende HTML <H3>, <H4> og <H5>
Linjeskift:	 	Eks: Første linje tekst. Andre linje tekst.
Avsnitt:	En blank Linje Dvs. To linjeskift	Eks: Første linje tekst Nytt avsnitt.
Bold:	** <tekst> **	**Bold tekst**
Link:	[Tekst](Url)	Eks : Se [Denne](http://helsedata.no) Lenken
Punktliste:		

Nasjonal spesifisering for metadata om helsedata

Ordnet	1. 2. 3.	Eks: 1. Første 2. Andre 3. Tredje
Uordnet	' - '	Eks: - Første - Andre - Tredje

4.2.5 Importfil 1a-Kildemetaddata

Importfil 1a-Kildemetaddata brukes til å beskrive både datakilden, eventuelle delkilder og underliggende datasamlinger, på et overordnet nivå ved hjelp av et sett standardiserte *properties*. For registre som skal levere dataprodukt vil importfilen også brukes til å definere dataprodukt i form av datasamlinger og evt logiske grupperinger av dataprodukt i form av delkilder. Filen er basert på HealthTerm 11+N importformat og har følgende spesifikasjoner.

Datakilden, eventuelle delkilder og datasamlinger representeres ved en Parent/Child-struktur i en og samme fil.

I påfølgende tabell er det en kolonne for Gruppering. Den skal ikke inkluderes i importfilen og er lagt til i spesifikasjonen for å øke lesbarheten. De 11 første propertyene (fet skrift) er posisjonbestemte som følge av HealthTerm sitt 11+N importformat og er derfor ikke plassert sammen med andre properties i samme gruppering.

Enkelte properties har ulike verdier og beskrivelser for datakilde, delkilde og datasamling. Der det ikke er eksplisitt spesifisert så gjelder verdiene og beskrivelsene for alle de tre kildemetadatypene.

Tabell 3: Format for importfil 1a-Kildemetaddata.				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
Administrative formål	1	Code	Ja	Unik navnestreng. Store bokstaver, uten mellomrom og æ, ø, å. Datakilde: Prefiks K for kilde, underscore, kortnavn på datakilden. F.eks. K_NPR. Delkilde: Prefiks K for kilde, underscore, kortnavn på datakilden, punktum, og navn på delkilde. F.eks. K_HUNT.HUNT1 Datasamling: Prefiks K for kilde, underscore, kortnavn på datakilden, punktum, og navn på datasamling. F.eks. K_HUNT.HUNT1_QUESTIONAIRE1
	2	ParentCode	Ja	Datakilde: Skal være lik kortnavnet på datakilden, f.eks. NPR. Delkilde: Skal være lik Code for tilhørende datakilde. Datasamling: Skal være lik Code for tilhørende datakilde. Hvis datasamlingen tilhører en delkilde, skal delkildens Code benyttes.

Tabell 3: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
HealthTerm	3	Hierarchy	Nei	Skal være tom.
Tittel og beskrivelse	4	PreferredTerm	Ja	Kort beskrivende navn. F.eks. "Norsk pasientregister", "MOBA Basisundersøkelse" eller "Spørreskjema1" for henholdsvis datakilde, delkilde og datasamling.
HealthTerm	5	ConceptType	Nei	Skal være tom
Administrative formål	6	PresentationOrder	Nei	Datakilde: Skal være tom. Delkilde, datasamling: Brukes dersom man ønsker å presentere delkildene eller datasamlingene i en egendefinert rekkefølge som avviker fra alfabetisk.
	7	ValidFrom	Ja	Brukes for å angi fra når en datakilde, delkilde eller datasamling er gyldig. Settes ved første import. Format: YYYYMMDD
	8	ValidTo	Nei	Skal være tom.
Tittel og beskrivelse	9	Synonym	Nei	Skal være tom (inntil videre)
	10	PatientFriendlyTerm	Nei	Skal være tom (inntil videre)
	11	ClinicalFriendlyTerm	Nei	Skal være tom (inntil videre)
	12	KortNavn	Ja	Forkortelse av navnet.
	13	NavnEngelsk	Ja	Engelsk navn.

Tabell 3: Format for importfil 1a-Kildemetaddata.				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
	14	Beskrivelse	Ja	En utførlig beskrivelse. Skal være forståelig også for innbyggere. Se eksempler på helsedata.no
	15	BeskrivelseEngelsk	Ja	En utførlig beskrivelse på engelsk. Skal være forståelig også for innbyggere.
	16	Emneord	Nei	Beskrivende emneord (tag). Skilles ved å bruke #Emneord – mellomrom – #NyttEmneord, osv. F.eks. #Emneord1 #Emneord2
Opphav og ferskhet	17	Opprettet	Ja	Året datakilden, delkilden eller datasamlingen ble opprettet og begynte å samle inn data.
	18	SistOppdatert	Ja	Dato for siste oppdatering av KILDEMETADATA Format: YYYYMMDD
	19	Frekvens	Nei	Datakilde, delkilde: Skal være tom. Datasamling: Oppdateringsfrekvens brukes for å beskrive hvor ofte datagrunnlaget oppdateres på Helseanalyseplattformen.
	20	Innsamlingsmetode	Nei	Kort beskrivelse av metode for innsamling av data.
Innhold og kvalitet	21	Kildetype	Ja	Kategorisering av datakilden. <u>Standardisert vokabular:</u> <ul style="list-style-type: none"> • 1 = Sentralt helseregister² • 2 = Nasjonalt medisinsk kvalitetsregister • 3 = Annet medisinsk kvalitetsregister • 4 = Befolkningsbasert helseundersøkelse

² <https://www.fhi.no/div/datatilgang/om-sentrale-helseregistre/>

Tabell 3: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
				<ul style="list-style-type: none"> • 5 = Biobank • 6 = Annen datakilde
	22	InklusjonsOgEksklusjonskriterier	Ja	En faglig overordnet beskrivelse av hvilke kriterier som stilles for at en person skal være registrert i datakilden, delkilden eller datasamlingen.
	23	GeografiskAvgrensning	Nei	Beskrivelse av opptaksområdet for innbyggerne som inngår i datakilden, delkilden eller datasamlingen.
	24	Tilslutningsgrad	Nei	Andel enheter av totalt antall relevante enheter som rapporterer inn til datakilden, delkilden eller datasamlingen. Med enheter menes virksomheter. Oppgis som desimaltall. (Kompletthet i form av dekningsgrad på institusjons- eller enhetsnivå)
	25	BeregningTilslutningsgrad	Nei	Hvordan tilslutningsgraden er beregnet. F.eks. hva er nevneren i brøken?
	26	Dekningsgrad	Nei	Andel pasienter som er rapportert inn sett opp mot totalt antall pasienter som er blitt behandlet for samme sykdom i en periode (individnivå). Oppgis som desimaltall. (Kompletthet i forhold til totalt antall hendelser)
	27	BeregningDekningsgrad	Nei	Hvordan dekningsgrad er beregnet. F.eks. sammenstilt med data fra NPR.
	28	Kvalitetsnote	Nei	Overordnet tekstlig beskrivelse av kvaliteten på datagrunnlaget.
<i>Juridisk informasjon</i>	29	Formaal	Ja	Kort beskrivelse av formålet registeret er opprettet for. Dersom mulig, en kopi av formålet slik det står skrevet i forskrift.
	30	Lovverk	Ja	URL til lovverk på lovdata.no. F.eks. URL til Lov om helseregistre og behandling av helseopplysninger.
	31	Forskrift	Ja	URL til forskrift på lovdata.no. F.eks. URL til Forskrift om innsamling og behandling av helseopplysninger i Medisinsk fødselsregister.

Tabell 3: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
	32	HjemmelTilgjengeliggjoring	Nei	URL til paragraf med hjemmel for tilgjengeliggjoring.
	33	Dataansvarlig	Ja	Navn på virksomheten eller enheten som eier og har det juridiske dataansvaret for datakilden, delkilden eller datasamlingen.
	34	Databehandler	Ja	Navn på virksomheten eller enheten som faktisk behandler datakilden. Er det samme som Dataansvarlig med mindre denne har satt ut databehandlingen til ekstern tredjepart.
	35	GradAvPersonidentifikasjon	Nei	Høyeste grad av identifisering det er mulig å søke om fra denne datakilden, delkilden eller datasamlingen. <u>Standardisert vokabular:</u> <ul style="list-style-type: none"> • 1 = Direkte personidentifiserbare opplysninger • 2 = Indirekte personidentifiserbare opplysninger • 3 = Anonyme opplysninger
	36	JuridiskNote	Nei	Relevant juridisk informasjon som ikke dekkes av andre properties. F.eks. informasjon om reservasjonsrett eller hvilke andre kilder data kan sammenstilles med.
<i>Kontaktinformasjon</i>	37	Hjemmeside	Ja	Lenke til hjemmeside.
	38	Epost	Ja	E-postadresse for henvendelser om datakilden, delkilden eller datasamlingen.
	39	Telefonnummer	Ja	Telefonnummer for henvendelser om datakilden, delkilden eller datasamlingen.
	40	Adresse	Ja	Adresse for henvendelser om datakilden, delkilden eller datasamlingen.
	41	OrgnrDataansvarlig	Nei	Organisasjonsnummer til dataansvarlig.

Tabell 3: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
	42	HERidDatakilde	Nei	<p>Datakilde: Unik identifikator som tildeles av NHN, Brukes for identifisering av datakilden innenfor helsenettet. Kan brukes til spørringer.</p> <p>Delkilde, datasamling: Skal være tom.</p>
	43	HERidDataansvarlig	Nei	Unik identifikator som tildeles av NHN, Brukes for identifisering av datakilden innenfor helsenettet. Kan brukes til spørringer.
Administrative formål	44	Kildemetadatatype	Ja	<p>Angir hvilken type kildemetadata som beskrives.</p> <p><u>Standardisert vokabular:</u></p> <ul style="list-style-type: none"> • 1 = Datakilde • 2 = Delkilde • 3 = Datasamling
	45	ErlSamsvarMed	Nei	Referanse til evt. standard(er), spesifikasjon(er) etc. som datakilden, delkilden eller datasamlingen er i henhold til, f.eks. DCAT-AP-NO.
	46	StatistikkType	Nei	<p>Hvorvidt medfølgende statistikk er akkumulert eller årsbasert.</p> <p><u>Standardisert vokabular:</u></p> <ul style="list-style-type: none"> • 1 = Akkumulert • 2 = Årsbasert
Tillegg	47-N	Mulige tilleggsproperties		<p>Gir mulighet for å legge til ytterligere properties som f.eks. registerspesifikke tilleggsproperties, properties for kategoriseringsformål etc.</p> <p>Disse kan defineres etter hvert som ønsker og behov oppstår.</p>

4.2.6 Importfil 1b-Variabelmetadata

Importfil 1b-Variabelmetadata brukes til å beskrive variablene med utgangspunkt i et sett standardiserte *properties*. Filen er basert på HealthTerm 11+N importformat.

I påfølgende tabell er det en kolonne for Gruppering. Den skal ikke inkluderes i importfilen og er lagt til i spesifikasjonen for å øke lesbarheten.

Tabell 4: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
Administrative formål	1	Code	Ja	Unik navnestreng for en variabel som også benyttes for å koble metadata med data. Store bokstaver, uten mellomrom og æ, ø, å. Prefiks V for variabel, underscore, kortnavn på datakilden, punktum, og samme variabelnavn som i kilden, f.eks. V_DAR.ALDER_AAR.
	2	ParentCode	Ja	Forkortelse for kildenevnet med prefiks V for Variabel og underscore, f.eks. V_DAR. "ParentCode" for toppnoden skal være lik kortnavnet på datakilden, f.eks. DAR.
HealthTerm	3	Hierarchy	Nei	Brukes dersom man ønsker å presentere variablene i flere kontekster, også kalt polyhierarkier.
Tittel og beskrivelse	4	PreferredTerm	Ja	Kort beskrivende navn på den enkelte variabel, f.eks. Kjønn. PreferredTerm på toppnoden skal ha navn lik "VARIABLEMETADATA" med store bokstaver.
HealthTerm	5	ConceptType	Nei	Brukes for å skille mellom, og kunne søke på, forskjellige typer concepts; Regular, Grouping og Record. "Regular" forteller at et concept er et vanlig concept med tilhørende properties, f.eks. en variabel. "Grouping" forteller at et concept kun er til for å gruppere regulære concepts, f.eks. kapitler, tema, etc. "Record" er et concept som brukes for å legge til utfyllende informasjon på et moder-concept. Et "Record"-concept kan ikke ha barn under seg, men er ellers som et normalt concept.
Administrative formål	6	PresentationOrder	Nei	Brukes dersom man ønsker å presentere variablene i en egendefinert rekkefølge som avviker fra en alfabetisk.
	7	ValidFrom	Ja	Brukes for å angi fra når en variabel er gyldig. Settes ved første import. Forutsetning for å kunne tilrettelegge for søk på gyldige variabler på en gitt dato eller endringer på en variabel i en tidsperiode.

Tabell 4: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
				Med «gyldig» menes at det finnes data tilgjengelig for variabelen. Nye avledede variabler kan være gyldig tilbake i tid. Format: YYYYMMDD
	8	ValidTo	Nei	Brukes for å angi at en variabel ble ugyldig fra en gitt dato. Brukes kun ved initial import når man ønsker å få med historikk på utgåtte variabler. Variabler med ValidTo-dato må fjernes fra importfilen ved neste gangs import for å unngå feil i HealthTerm Format: YYYYMMDD
Tittel og beskrivelse	9	Synonym	Nei	Brukes dersom man ønsker å legge til synonymer til et variabelnavn. Krever en ny rad i importfilen dersom mer enn ett synonym legges til.
	10	PatientFriendlyTerm	Nei	Pasientvennlig navn på variabelen.
	11	ClinicalFriendlyTerm	Nei	Helsefaglig begrep på variabelen.
	12	NavnEngelsk	Ja	Engelsk navn på variabelen.
	13	Beskrivelse	Ja	Utførlig beskrivelse av variabelen. Se eksempler på helsedata.no.
	14	BeskrivelseEngelsk	Ja	Utførlig beskrivelse av variabelen på engelsk.
Opphav	15	Opprinnelse	Ja	Om variabelen er hentet direkte fra helsetjenesten, er beregnet, avledet eller kodet. <u>Standardisert vokabular:</u> <ul style="list-style-type: none"> • 1 = Beregnet (F.eks. basert på en formel) • 2 = Avledet (F.eks. en kategorisering) • 3 = Kodet (F.eks. manuell koding basert på tekst)

Tabell 4: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
				<ul style="list-style-type: none"> • 4 = Direkte fra fagsystem/EPJ (eks. hentet direkte fra NPR-melding) • 5 = Direkte fra skjema (eks. hentet direkte fra elektronisk skjemaløsning) • 6 = Annet
	16	OpprinnelseUtledning	Nei	Hvordan beregnede, avledede og kodete variabler er utledet. Ikke aktuell for variabler hentet direkte fra helsetjenesten eller en elektronisk skjemaløsning
Innhold og kvalitet	17	Kommentar	Nei	Kommentar til variabelen. Dette kan f.eks. være når variabelen gikk over fra å være basert på ICD-9 til ICD-10 eller andre endringer av betydning.
	18	KommentarEngelsk	Nei	Engelsk oversettelse av Kommentar.
	19	Kvalitetsnote	Nei	Overordnet tekstlig beskrivelse av kvaliteten på datagrunnlaget som variabelen representerer, f.eks. komplettethet i form av fullstendighet (kodingskvalitet) og/eller kodekvalitet.
	20	KvalitetsnoteEngelsk	Nei	Engelsk oversettelse av Kvalitetsnote.
	21	MaaleType	Nei	Hva variabelen måler. F.eks. tid, alder, vekt og puls.
	22	MaaleEnhet	Nei	Hvilken måleenhet variabelen er basert på. F.eks. år, mnd., kg, gram, slag/minutt.
	23	DataType	Ja	Datatype som spesifisert i FHIR Primitive Types. <u>Standardisert vokabular:</u> <ul style="list-style-type: none"> • 1 = String • 2 = Integer • 3 = Decimal • 4 = Boolean • 5 = Time • 6 = Date

Tabell 4: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
				<ul style="list-style-type: none"> • 7 = Datetime • 8 = Uri • 9 = Base64binary
	24	Lengde	Nei	Maksimal lengde på variabelen
	25	Presisjon	Nei	Antall desimaler
	26	GrenseLav	Nei	Nedre grenseverdi
	27	GrenseHoy	Nei	Øvre grenseverdi
Juridisk informasjon	28	Identifiseringsgrad	Nei	<p>Registerets klassifisering av bidrag til risiko for identifisering av individ. Nyttig informasjon for forsker når prosjekt må vurdere tiltak for dataminimering.</p> <p><u>Standardisert vokabular:</u></p> <ul style="list-style-type: none"> • 1 = Ingen • 2 = Lav • 3 = Middels • 4 = Høy
Skjema	29	Ledetekst	Nei	Dette er teksten som står i registrets innregistreringsskjema ved hver variabel. Ledetekst er primært aktuell for variabler som er registrert via et skjema/en skjemaløsning.
	30	SporsmaalNummer	Nei	Spørsmålsnummer i spørreskjema.
	31	SkjemaVersjon	Nei	Vise hvilke(n) versjon(er) av innrapporteringskjema denne variabelen finnes i.
Administrative formål	32	AnbefaltForDigitaltInnsyn	Ja*	Anbefaling til HDS om hvorvidt variabelen bør inngå i et digitalt, automatisert innsyn via helsenorge.no.

Tabell 4: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
				<p><u>Standardisert vokabular:</u></p> <ul style="list-style-type: none"> • 1 = Bør inngå • 2 = Bør ikke inngå <p>*Gjelder kun for variabler som skal inngå i dataprodukt som skal overføres til Helsanalyseplattformen.</p>
	33	TabellvennligNavn	Nei	Alternativ kolonneoverskrift for eksport eller utskrift av data.
	34	TabellvennligNavnEngelsk	Nei	Alternativ kolonneoverskrift for eksport eller utskrift av data for engelskspråklige.
	35	DatabaseReferanse	Ja	API-streng/URL/SourceID/TechName/unikt navn på klasse og/eller attributt i datakildens database. Denne referansen vil bli benyttet ved forespørsel om henting av variabelen fra kilden.
	36	Erstatter	Nei	VariabelID som er erstattet av denne variabelen.
	37	ErstattesAv	Nei	VariabelID til variabelen som evt. erstatter denne variabelen.
	38	StartdatoHistorikk	Nei	<p>Dato for start reell historikk.</p> <p>Hvis angitt er dette den første datoen man har historikk på metadata.</p> <p>Dette for å dokumentere at det kan være periode der variabelen er i bruk, men at man ikke kjenner hvilke endringer som er gjort.</p> <p>Hvis ikke angitt, regnes ValidFrom som start på historikk</p>
	39	Synlig	Nei	<p>Definerer om variabelen skal presenteres på helsedata.no eller gjøres tilgjengelig via API-grensesnittet i HealthTerm.</p> <p><u>Standardisert vokabular:</u></p> <ul style="list-style-type: none"> • 1 = Ja • 2 = Nei

Tabell 4: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Obligatorisk	Beskrivelse
Tillegg	40-N	Mulige tilleggsproperties	Nei	Gir mulighet for å legge til ytterligere properties som f.eks. registerspesifikke tilleggsproperties, properties for "harmoniseringsarbeid", properties for kategoriseringsformål, properties med referanser til begrepskataloger, terminologier, standarder etc. Skal også gi mulighet for generelle eller registerspesifikke filtrerings mekanismer.

4.2.7 Importfil 1c-Versjon

Importfil 1c-Versjon brukes til å identifisere hvilke variabler som inngår i f.eks. et årssett av en datakilde/datasamling/dataprodukt. En variabel vil kunne inngå i flere datasamlinger i forskjellige perioder. Hver versjon gis en unik *code* og ID, og en angivelse av gyldighetsperiode. En variabls tilhørighet til en versjon defineres ved hjelp av *Mapsets*.

Filen er basert på HealthTerm 11+N importformat og har følgende spesifikasjoner.

Tabell 5: Format for importfil 1c-Versjon			
Nr	Property	Obligatorisk	Beskrivelse
1	Code	Ja	Unik navnestreng for en versjon. Store bokstaver, uten mellomrom og æ, ø, å. Prefiks N, kortnavn på datakilden, punktum, kortnavn på datasamlingen, underscore og versjonsnummer/ID. N_NPR.AMK_2018
2	ParentCode	Ja	Forkortelse av datakildenavn med prefiks N for versjon, f.eks. N_NPR. "ParentCode" for toppnoden skal være lik kortnavnet på datakilden, f.eks. NPR.
3	Hierarchy	Nei	Skal være tom for VERSJON
4	PreferredTerm	Ja	Kort beskrivende navn på Versjonen, f.eks. Somatikk. Toppnoden skal ha navn lik VERSJON.

Tabell 5: Format for importfil 1c-Versjon			
Nr	Property	Obligatorisk	Beskrivelse
5	ConceptType	Nei	Skal være tom for VERSJON
6	Presentation Order	Nei	Skal være tom for VERSJON
7	ValidFrom	Ja	Brukes for å angi fra når en versjon er gyldig. Den første versjonen av en datasamling vil ha ValidFrom lik den datoen datagrunnlaget er gyldig fra.
8	ValidTo	Ja	Brukes for å angi at en versjon ble, eller vil bli, ugyldig fra en gitt dato. Versjoner skal ikke ha overlappende gyldighetsdatoer mellom revisjoner.
9	Synonym	Nei	Skal være tom for VERSJON
10	PatientFriendly Term	Nei	Skal være tom for VERSJON
11	ClinicalFriendlyTerm	Nei	Skal være tom for VERSJON
12	Datasamling	Ja	Kortnavn på datasamlingen, som angitt i code f.eks. AMK
13	Versjon	Ja	Versjon, som angitt i code. Dette kan være årgang, frys-versjon, skjemaversjon eller liknende
14	Forvaltningstype	Nei	Hvorvidt dette datasettet benyttes til utlevering av data, eller om det er en beskrivelse av de innrapporterte dataene. Mulige verdier: <ul style="list-style-type: none"> • 1 = Utlevering • 2 = Innrapportering • 3 = Begge

Tabell 5: Format for importfil 1c-Versjon

Nr	Property	Obligatorisk	Beskrivelse
15	DatabaseReferanse	Nei	API-streng, URL eller unikt navn på klasse, tabell eller view i datakildens database. Sammen med databasereferanse på variabelmetadata skal dette utgjøre eksakt referanse til dataelement i databasen.
16	DataFra	Nei	Beskriver fra når det finnes data fra versjonen på Helseanalyseplattformen. Format YYYYMMDD.
17	DataTil	Nei	Beskriver til når det finnes data fra versjonen på Helseanalyseplattformen. Format: YYYYMMDD
17-N	Mulige tilleggsperties	Nei	Gir mulighet for å legge til ytterligere properties. Disse kan defineres etter hvert som evt. ønsker og behov oppstår.

4.2.8 Importfil 1d-Statistikk

Importfil 1d-Statistikk brukes til å presentere statistikk for datakildens variabler. Statistikken beskrives ved hjelp av et sett standardiserte *properties*, og fordeles på eksempelvis årssett. Filen er basert på HealthTerm 11+N importformat og har følgende spesifikasjoner.

NB! Statistikken anvendes av brukerne for å identifisere kodingskvaliteten for variabelen, og ikke til analyse.

I denne versjonen av denne spesifikasjonen er det ikke obligatorisk å produsere statistikk.

Tabell 6: Format for importfil 1d-Statistikk			
Nr	Property	Obligatorisk	Beskrivelse
1	Code	Ja	Unik navnestreng for et statistikk-concept. Store bokstaver, uten mellomrom og æ, ø, å. Prefiks S for statistikk, underscore, årstall, kortnavn på datakilden, punktum, datasamling, punktum, variabelnavn, punktum KodeverkLokalID, f.eks. S_NPR_2017.BUP.KJONN.1
2	ParentCode	Ja	Forkortelse av datakildens navn med prefiks S for statistikk som toppnode f.eks. S_NPR. "ParentCode" for toppnoden skal være lik kortnavnet på datakilden, For å splitte underliggende datasamlinger på årssett brukes i tillegg et årstall etterfulgt av kortnavnet på datasamlingen, f.eks. S_NPR_2017.BUP.
3	Hierarchy	Nei	Skal være tom for STATISTIKK
4	PreferredTerm	Ja	Kort beskrivende navn på den enkelte variabel, f.eks. Kjønn. "PreferredTerm" på toppnoden skal ha navn lik STATISTIKK med store bokstaver.
5	ConceptType	Nei	Skal være tom for STATISTIKK
6	PresentationOrder	Nei	Brukes dersom man ønsker å presentere statistikken på verdier/koder i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk.
7	ValidFrom	Ja	Brukes for å angi fra når statistikken er gyldig. Forutsetning for å kunne tilrettelegge for søk på statistikk på en gitt dato eller endringer i en valgt tidsperiode f.eks. utvikling over flere år. Format: YYYYMMDD
8	ValidTo	Ja	Brukes for å angi til når statistikken er gyldig. Forutsetning for å kunne tilrettelegge for søk på statistikk på en gitt dato eller endringer i en valgt tidsperiode f.eks. utvikling over flere år. Statistikk skal ikke ha overlappende gyldighetsdatoer. Format: YYYYMMDD
9	Synonym	Nei	Skal være tom for STATISTIKK

Tabell 6: Format for importfil 1d-Statistikk			
Nr	Property	Obligatorisk	Beskrivelse
10	PatientFriendly Term	Nei	Skal være tom for STATISTIKK
11	ClinicalFriendly Term	Nei	Skal være tom for STATISTIKK
12	Kommentar	Nei	En mer utførlig beskrivelse av statistikken ved behov.
13	SisteOppdaterte Aarssett	Ja	Året statistikken gjelder for. Brukes for enklere å kunne filtrere på årssett. Kompletterer "ValidFrom" og "ValidTo".
14	GyldigeTilfeller	Ja	Antall rapportert inn totalt
15	ManglendeTilfeller	Nei	Antall som mangler innrapportering
16	Desimaler	Nei	Antall desimaler avrundet til
17	Prosentandel	Nei	Dersom beregnet i kilden
18	MIN	Nei	Minimum
19	MAX	Nei	Maksimum
20	AVG	Nei	Gjennomsnitt
21	STD	Nei	Standardavvik
22	MED	Nei	Median
23	KodeverkLokalID	Nei	Lokal ID for kodeverket og koden, f.eks. 1 eller 2 for hhv Mann og Kvinne.
24-N	Mulige tilleggspproperties	Nei	Gir mulighet for å legge til ytterligere properties ved behov

4.2.9 Importfil 1e-Variabelgruppe

Importfil 1e-Variabelgruppe brukes til å beskrive variabelgrupper med utgangspunkt i et sett standardiserte *properties*. Filen er basert på HealthTerm 11+N importformat.

Tabell 7: Format for importfil 1e-Variabelgruppe			
Nr	Property	Obligatorisk	Beskrivelse
1	Code	Ja	Unik navnestreng for en variabelgruppe. Store bokstaver, uten mellomrom og æ, ø, å. Prefiks G for variabelGruppe, underscore, kortnavn på datakilden, punktum, og samme variabelnavn som i kilden, f.eks. G_HUNT.EXPOSURE
2	ParentCode	Ja	Forkortelse for kildenavn med prefiks G for variabelGruppe og underscore, f.eks. G_HUNT. "ParentCode" for toppnoden skal være lik kortnavnet på datakilden, f.eks. HUNT. Eller referere til overliggende variabelgruppe i et hierarki.
3	Hierarchy	Nei	Brukes ikke for variabelgrupper.
4	PreferredTerm	Ja	Kort beskrivende navn på den enkelte variabelgruppe. PreferredTerm på toppnoden skal ha navn lik VARIABELGRUPPER med store bokstaver.
5	ConceptType	Nei	Skal være tom for VARIABELGRUPPE
6	PresentationOrder	Nei	Brukes dersom man ønsker å presentere variabelgruppene i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk.
7	ValidFrom	Ja	Brukes for å angi fra når en variabelgruppe er gyldig. Format: YYYYMMDD
8	ValidTo	Nei	Skal ikke ha verdi
9	Synonym	Nei	Brukes dersom man ønsker å legge til synonymer til en variabelgruppe. Krever en ny rad i importfilen dersom mer enn ett synonym.

Tabell 7: Format for importfil 1e-Variabelgruppe			
Nr	Property	Obligatorisk	Beskrivelse
10	PatientFriendlyTerm	Nei	Pasientvennlig navn på variabelgruppen
11	ClinicalFriendlyTerm	Nei	Helsefaglig begrep på variabelgruppen
12	Beskrivelse	Ja	Utførlig beskrivelse av variabelgruppen.
13	NavnEngelsk	Nei	Engelsk oversettelse av PreferredTerm
14	BeskrivelseEngelsk	Nei	Engelsk oversettelse av beskrivelsen av variabelgruppen.
15	Filter	Nei	Definerer om variabelgruppen skal brukes til filtrering i Helsedata.no <u>Standardisert vokabular:</u> <ul style="list-style-type: none"> • 1 = Ja • 2 = Nei

4.2.10 Importfil 1f-Instrument

Importfil 1f-Instrument brukes til å beskrive instrument med utgangspunkt i et sett standardiserte *properties*. Filen er basert på HealthTerm 11+N importformat.

Tabell 8: Format for importfil 1f-Instrument			
Nr	Property	Obligatorisk	Beskrivelse
1	Code	Ja	Unik navnestreng for et instrument. Store bokstaver, uten mellomrom og æ, ø, å. Prefiks I for Instrument, underscore, kortnavn på datakilden, punktum, og samme ID som i kilden, f.eks. I_HUNT.AbbottAXSymx
2	ParentCode	Ja	Forkortelse for kildenaavn med prefiks I for Instrument og underscore, f.eks. I_HUNT "ParentCode" for toppnoden skal være lik kortnavnet på datakilden, f.eks. HUNT.
3	Hierarchy	Nei	Brukes ikke for INSTRUMENT
4	PreferredTerm	Ja	Kort beskrivende navn på det enkelte Instrumentet. PreferredTerm på toppnoden skal ha navn lik INSTRUMENT med store bokstaver.
5	ConceptType	Nei	Skal være tom for INSTRUMENT
6	PresentationOrder	Nei	Brukes dersom man ønsker å presentere instrumentene i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk.
7	ValidFrom	Ja	Brukes for å angi fra når et instrument er gyldig. Settes ved første import. Format: YYYYMMDD
8	ValidTo	Nei	Skal ikke ha verdi.
9	Synonym	Nei	Brukes dersom man ønsker å legge til synonymer til et instrumentnavn. Krever en ny rad i importfilen dersom mer enn ett synonym.
10	PatientFriendlyTerm	Nei	Brukes ikke for INSTRUMENT
11	ClinicalFriendlyTerm	Nei	Brukes ikke for INSTRUMENT
12	Beskrivelse	Ja	Utførlig beskrivelse av instrumentet.

Tabell 8: Format for importfil 1f-Instrument			
Nr	Property	Obligatorisk	Beskrivelse
13	Filter	Nei	Definerer om instrumentet skal brukes til filtrering i Helsedata.no <u>Standardisert vokabular:</u> <ul style="list-style-type: none"> • 1 = Ja • 2 = Nei
14	Dokumentasjon	Nei	Lenke(r) til opprinnelig dokumentasjon/referanser for instrumentet.

4.2.11 Importfil 2-Kildekodeverk

Importfil 2-Kildekodeverk brukes til å beskrive kodeverk og koder som variablene er basert på. Filen skal inneholde alle kodeverk som benyttet av den aktuelle datakilden. Endringer og historikk i nasjonale kodeverk og terminologier forutsettes ivaretatt av respektive "kodeverkseiere". Filen er basert på HealthTerm 11+N importformat og har spesifikasjoner som definert i tabellen under.

Importfil 2-Kildekodeverk for den enkelte kilde importeres og oppdateres til et felles *Code System* og *Core Namespace* som etter hvert vil inneholde samtlige kodeverk som anvendes i kildene. Kodeverkene er strukturert i polyhierarkier som viser "Eiere", "Kilde" og "Alle". Under "Alle" ligger samtlige kodeverk, for alle registrene, sortert alfabetisk i en flat liste.

Variabler som er basert på administrative kodeverk, helsefaglige kodeverk eller andre masterkodeverk som definert i Tabell 11 skal ikke inngå i Importfil 2-Kildekodeverk.

Tabell 9: Format for importfil 2-Kildekodeverk			
Nr	Property	Obligatorisk	Beskrivelse
1	Code	Ja	Standardisert og unik navnestreng for et kodeverk med tilhørende koder. Store bokstaver, uten mellomrom og æ, ø, å. Prefiks for KK for kildekodeverk, kortnavn på datakilden, punktum, variabelnavn, punktum og KodeverkLokalld (som i kilden), f.eks. KK_DAR. KJONN_K.1

Tabell 9: Format for importfil 2-Kildekodeverk			
Nr	Property	Obligatorisk	Beskrivelse
2	ParentCode	Ja	Kildekodeverkene importeres i en forhåndsdefinert struktur som viser hvem som eier de, hvilken datasamling (kilde) de inngår i, samt en samlegruppe der alle kildekodeverkene presenteres alfabetisk i en flat liste. Dette kalles polyhierarkier. En kilde vil derfor ha flere "ParentCodes", henholdsvis for eier, f.eks. KK_FHI, for samling, f.eks. KK_MFR, og for den flate samlelisten, KK_ALLE.
3	Hierarchy	Nei	Brukes for å presentere kodeverkene i flere kontekster, også kalt polyhierarkier.
4	PreferredTerm	Ja	Kort beskrivende navn på det enkelte kodeverk, f.eks. Kjønn, eller Kode, f.eks. Kvinne. "PreferredTerm" på toppnoden skal ha navn lik KILDEKODEVERK med store bokstaver
5	ConceptType	Nei	Skal være tom for KILDEKODEVERK
6	PresentationOrder	Nei	Brukes dersom man ønsker å presentere kodeverkene og/eller kodene i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk.
7	ValidFrom	Ja	Brukes for å angi fra når et kodeverk eller en kode er gyldig. Settes ved første import. Forutsetning for å kunne tilrettelegge for søk på gyldige kodeverk og/eller koder på en gitt dato eller endringer på et kodeverk og/eller på en kode i en tidsperiode. Format: YYYYMMDD
8	ValidTo	Ja	Brukes for å angi at et kodeverk eller en kode ble, eller vil bli, ugyldig fra en gitt dato. Forutsetning for å kunne tilrettelegge for søk på gyldige kodeverk og/eller koder på en gitt dato eller endringer på et kodeverk og/eller på en kode i en tidsperiode Format: YYYYMMDD
9	Synonym	Nei	Skal være tom for KILDEKODEVERK

Tabell 9: Format for importfil 2-Kildekodeverk			
Nr	Property	Obligatorisk	Beskrivelse
10	PatientFriendly Term	Nei	Skal være tom for KILDEKODEVERK
11	ClinicalFriendly Term	Nei	Skal være tom for KILDEKODEVERK
12	Beskrivelse	Nei	Utførlig beskrivelse/definisjon av en kode
13	Kommentar	Nei	Generelle kommentarer f.eks. til endringer
14	Eier	Nei	Hvilken institusjon som er ansvarlig for å forvalte kodeverket
16	MasterKodeverk	Nei	Hvilket masterkodeverk som dette kodeverket er basert på, evt. er et subset av, f.eks. ICD-10 (ValueSet)
17	ReferanseTerm	Nei	Dersom begrepet som brukes i navnet på kodeverket og/eller koden kan forankres/defineres av en standardisert term.
18	KodeverkLokalID	Ja	Lokal ID for koden, f.eks. 1 eller 2 for hhv Mann og Kvinne.
19	NavnEngelsk	Nei	PreferredTerm oversatt til Engelsk. Dvs. Kode- eller kodeverdi –beskrivelse.
20-N	Mulige tilleggspproperties	Nei	Gir mulighet for å legge til ytterligere properties ved behov

4.2.12 Importfil 3-Filter

Beskrivelse av filter detaljeres i en senere versjon.

4.2.13 Importfil 4x-Mapset

Mapset er et eget importfilformat i HealthTerm. Dette brukes for å mappe mellom *concepts* i to forskjellig *Code Systems*, eller mellom to *Namespaces* internt i et *Code System*.

For å realisere vår informasjonsmodell i HealthTerm, er følgende mapset nødvendige:

1. Fra DATASAMLINGER til respektive VERSJONER av datasamlingene
2. Fra VERSJONER til VARIABELMETADATA og variabler som inngår i respektive versjoner
3. Fra VARIABELMETADATA til LOKALE KODEVERK
4. Fra VARIABELMETADATA til ADMINISTRATIVE KODEVERK
5. Fra VARIABELMETADATA til HELSEFAGLIGE KODEVERK
6. Fra VARIABELMETADATA til STATISTIKK
7. Fra VARIABELGRUPPE til VARIABELMETADATA
8. Fra INSTRUMENT til VARIABELMETADATA

I tillegg kan man bruke *Mapsets* for å mappe *concepts* til f.eks. klassifikasjoner og standarder som er importert, og som man har tilgang til i HealthTerm. Dette kan være nyttig for f.eks. tilrettelegging av filtrering og søk på helsedata.no,

Mappingfiler inneholder et løpenummer, en *source* og en *target*. *Source* er konseptet det mappes fra, f.eks. en variabel. *Target* er konseptet det mappes til, f.eks. kodeverket som variabelen er basert på.

Tabell 10: Format for mappingfil 4x-MAPSETT		
Nr	Source	Target
1	CODE eller annen property	CODE eller annen property
Løpenummer	String	String
Løpenummer tildeles sekvensielt. Dette har ingen praktisk betydning.	Source_Code henviser til felt 1, Code, for det codesystem'et som skal være kilde for mappingen. Dette vil være Variabel-ID for mapset for statistikk, kodeverk og versjonen. Eksempel V_DAR.KJONN_K.	Target_Code henviser til felt 1, Code, for det codesystem'et som skal være mål for mappingen. For statistikk-mapset vil dette være Statistikk-ID. <ul style="list-style-type: none"> • Eksempel S_DAR.KJONN_K. For kodeverk vil det være Kodeverk-ID. <ul style="list-style-type: none"> • Eksempel for lokale kodeverk er KK_DAR.KJONN_K. • Eksempel for administrative kodeverk er 3101. • Eksempel for helsefaglige kodeverk er ICD-10. For Versjon vil det være Versjon-ID. <ul style="list-style-type: none"> • Eksempel V_DAR.DAR.2017Q3

4.2.14 Administrative og helsefaglige kodeverk

Administrative kodeverk omfatter Volven-samlingen. Her inngår NPR-kodeverk og kodeverk fra SSB.

Disse kodeverkene er tilgjengelig i Nasjonal variabelkatalog, og man refererer til dem ved egen mappingfil. Target code i mappingfilen vil være samme OID som brukes i Volven. Eksempler på dette er 3101 for Kjønn og 3402 for kommunenummer.

I de tilfellene der et register benytter seg av et av de helsefaglige kodeverkene i Tabell 11, så skal det mappes til respektive "Kortnavn" som definert i tabellen under, og kodeverket skal ikke følge med i Importfil 2-Kildekodeverk.

Tabell 11: Helsefaglige kodeverk	
Kortnavn	Beskrivelse
ICD-8	Den internasjonale statistiske klassifikasjonen av sykdommer og beslektede helseproblemer. Versjon 8
ICD-9	Den internasjonale statistiske klassifikasjonen av sykdommer og beslektede helseproblemer. Versjon 9
ICD-10	Den internasjonale statistiske klassifikasjonen av sykdommer og beslektede helseproblemer. Versjon 10
ICD-11	Den internasjonale statistiske klassifikasjonen av sykdommer og beslektede helseproblemer. Versjon 11
NCSP-NCNP-NCRP	Norsk klinisk prosedyrekodeverk
ICPC-2	Den internasjonale klassifikasjonen for primærhelsetjenesten
ATC	Anatomisk Terapeutisk Kjemisk legemiddelregister
DRG	Diagnoserelaterte grupper, pasientklassifiseringssystem
ICF	Internasjonal klassifikasjon av funksjon, funksjonshemming og helse
NLK	Norsk Laboratoriekodeverk
LOINC	Logical Observation Identifiers Names and Codes
NORPAT	Norsk patologikodeverk
PHBU	Multiaksial klassifikasjon i psykisk helsevern for barn og unge
SNOMED CT	Systematized Nomenclature of Medicine

Begrepsliste

Dette dokumentet benytter en del fag- og løsnings-spesifikke begrep. Disse er forklart nærmere i tabellen under.

Tabell 12: Begrepsliste	
Begrep	Forklaring
Dataprodukt	Dataprodukt er en konseptualisering av måten data skal overføres fra nasjonale helsedatakilder til Helseanalyseplattformen. Innholdet i dataprodukt spesifiseres gjennom metadata, og informasjon om dataproduktet som sådan spesifiseres i form av metadata i objektet datasamling.
Datakilde	Datakilde er definert som kildene til dataene. I denne versjonen av spesifikasjonen er en datakilde ekvivalent med et helseregister.
Delkilde	Delkilde kan brukes dersom datakilden har ulike deler som må beskrives ut ifra en logisk, kronologisk, juridisk eller populasjonsbasert betraktning. HUNT1, 2, 3 og 4 er eksempel på delkilder. Kan ha flere nivåer.
Datasamling	Datasamling er definert som et sub-register innenfor en datakilde, altså en delmengde av variablene i den aktuelle datakilden. Eksempler på datasamlinger kan være sektorer i NPR eller skjema i medisinske kvalitetsregistre.
Versjon	Versjon er et teknisk hjelpemiddel som brukes til å definere hvilke variabler som inngår i ulike årganger/versjoner av de ulike datasamlingene i en datakilde.
Variabel	Variabler er de informasjonsbærende elementene i datakilden og definerer hvilken informasjon som finnes i registrene.
Variabelgruppe	Variabelgruppe er en kategorisering av variabler. Strukturen kan bygges opp hierarkisk. Inntil videre begrenset til 2 nivåer.
Kodeverk	Kodeverk representerer i denne sammenhengen et verdisett. Et kodeverk kan for eksempel representere en variabels verdisett. Det kan også representere en samling verdier som brukes til å beskrive en datakilde.
Kode	Kode er i denne sammenheng det samme som svaralternativene i et verdisett. Består av en ID (KodeverkLokalID) og en beskrivelse (Kodebeskrivelse), for eksempel «2» og «Kvinne».
Statistikk	Statistikk er definert som de statistiske egenskapene ved en årgang/versjon av en variabel.
Kodfrekvens	Verdi er definert som antall gyldige tilfeller av et svaralternativ/utfall av en variabel.
Code System	Code System kan betraktes som den øverste mappen i en mappestruktur i HealthTerm. All informasjon lagres innenfor et Code System. Hvert register modelleres i løsningen som et separat Code System.

Tabell 12: Begrepsliste	
Begrep	Forklaring
Core Namespace	Core Namespace er hovedmappen i et Code System. For registre lagres metadata om variablene som Core Namespace.
Extension Namespace	Extension Namespace (eller bare Extension) er tilleggsmapper i HealthTerm der annen informasjon lagres. Kildemetadata, statistikk og versjon lagres i Extension Namespaces.
Concept	Concept er de meningsbærende dataobjektene i HealthTerm og brukes til å modellere variabler, datakilder, datasamlinger osv.
Property	Properties er det samme som metadataelement eller attributter og brukes til å beskrive concepts.
Mapping	Mapping brukes og å håndtere ikke-hierarkiske relasjoner i HealthTerm, eksempelvis referere til kodeverk.
Retired	Retired brukes om concepts og properties som ikke lengre er aktive i løsningen. En gammel versjon av en property settes eksempelvis til retired når man redigerer den og får ut en ny versjon.
CareCom	CareCom er teknisk leverandør av HealthTerm – løsningen Nasjonal variabelkatalog er realisert i.

Referanser

- [1] «Health Registries for Research (HRR), WP 4. National standard for registry data documentation,» [Internett]. Available: <https://hrr.w.uib.no/wp-4-national-standard-for-registry-data-documentation/>.
- [2] «Nasjonal e-helsestrategi 2017-2022,» Direktoratet for e-helse, 2017.
- [3] «Data Catalog Vocabulary (DCAT) - Version 2,» [Internett]. Available: <https://www.w3.org/TR/vocab-dcat-2/>.
- [4] «SKOS Simple Knowledge Organization System,» [Internett]. Available: <https://www.w3.org/2004/02/skos/>.
- [5] «Standard for beskrivelse av datasett, datatjenester og datakataloger (DCAT-AP-NO),» [Internett]. Available: <https://data.norge.no/specification/dcat-ap-no/>.
- [6] «SKOS-AP-NO-Begrep – Forvaltningsstandard for tilgjengeliggjøring av begrepsbeskrivelser basert på SKOS,» [Internett]. Available: <https://data.norge.no/specification/skos-ap-no-begrep/>.
- [7] «FAIR-prinsippene,» [Internett]. Available: <https://www.force11.org/group/fairgroup/fairprinciples>.
- [8] «Felles språk i helse- og omsorgssektoren,» Direktoratet for e-helse, 2019.

Vedlegg 1: Mapping av properties i KILDEMETADATA-filen til DCAT-AP-NO og DCAT-EN

Arbeidet med å mappe kildemetadata mot henholdsvis den internasjonale versjonen av DCAT og Digitaliseringsdirektoratets norske versjon DCAT-AP-NO er et pågående arbeid.

Mappingen er viktig for å kunne utveksle og dele metadata om datasett på tvers av sektorer og landegrenser, samt for å opparbeide en felles forståelse av innholdet.

Mappingen er også en forutsetning for å kunne etablere maskinell validering av kvaliteten på metadataene.

Der det i tabellen er brukt "E-helse:Extension , "?" eller (...) har vi foreløpig ikke funnet en god match, eller er usikre på mappingen, mellom propertiene i denne spesifikasjonen og DCAT.

PresentationOrder_V.1.1	KILDEMETADATA_Properties	DCAT-AP-NO	DCAT-EN (DQV-EN)
1	Code	Datasett: identifikator	dct:identifier
2	ParentCode	Datasett: er del av	dct:isPartOf
3	Hierarchy	E-helse:Extension	?
4	PreferredTerm	Datasett: tittel	dct:title
5	ConceptType	E-helse:Extension	?
6	PresentationOrder	E-helse:Extension	?
7	ValidFrom	E-helse:Extension	?
8	ValidTo	E-helse:Extension	?
9	Synonym	E-helse:Extension	(dcat:keyword)
10	PatientFriendlyTerm	E-helse:Extension	(dcat:keyword)
11	ClinicalFriendlyTerm	E-helse:Extension	(dcat:keyword)
12	KortNavn	Datasett: tittel (kortnavn)	dct:title
13	NavnEngelsk	Datasett: tittel (EN)	dct:title
14	Beskrivelse	Datasett: beskrivelse (NB)	dct:description
15	BeskrivelseEngelsk	Datasett: beskrivelse (EN)	dct:description
16	Emneord	Datasett: emneord	dcat:keyword
17	Opprettet	Ehelse:Extension	?
18	SistOppdatert	Datasett: endringsdato	dct:modified

Nasjonal spesifikasjon for metadata om helsedata

19	Frekvens (Oppdateringsfrekvens)	Datsett: frekvens	dct:accrualPeriodicity
20	Innsamlingsmetode	Datsett: Ble generert ved	prov:wasGeneratedBy
21	KildeType	Aktør: utgivertype	foaf:name
22	InklusjonsOgEksklusjonskriterier (Inklusjonskriterier)	E-helse:Extension	?
23	GeografiskAvgrensning	Datsett: dekningsområde	dct:spatial
24	Tilslutningsgrad	Kompletthet (completeness)	dqv:Dimension
25	BeregningTilslutningsgrad	Kompletthet (completeness)	dqv:Dimension
26	Dekningsgrad	Kompletthet (completeness)	dqv:Dimension
27	BeregningDekningsgrad	Kompletthet (completeness)	dqv:Dimension
28	Kvalitetsnote (KjenteProblemer)	Datsett:Kvalitetsnote	dqv:QualityAnnotation
29	Formaal	dcatno:objective (Kun i veileder)	?
30	Loverk	(Datsett: følger)	cspv:follows
31	Forskrift	(Datsett: følger)	cspv:follows
32	HjemmelTilgjengeliggjøring	(tom)	(tom)
33	DataAnsvarlig	Datsett: kontaktpunkt	dcat:contactPoint
34	DataBehandler	Datsett: kontaktpunkt	dcat:contactPoint
35	GradAvPersonidentifikasjon	(tom)	(tom)
36	JuridiskNote	(tom)	(tom)
37	Hjemmeside	Datsett: kontaktpunkt (Datsett: dokumentasjon?)	dcat:contactPoint
38	E-post	Datsett: kontaktpunkt	dcat:contactPoint
39	Telefonnummer	Datsett: kontaktpunkt	dcat:contactPoint
40	Adresse	Datsett: kontaktpunkt	dcat:contactPoint
41	OrgnrDataansvarlig	Datsett: kontaktpunkt	dcat:contactPoint
42	HERidDatakilde	(tom)	(tom)
43	HERidDataansvarlig	Aktør: organisasjonsidentifikator	dct:identifier
44	Kildemetadatatype	E-helse:Extension	?

Nasjonal spesifikasjon for metadata om helsedata

45	EriSamsvarMed	Datasett:samsvar	dqv:conformsTo
46	StatistikkType	E-helse:Extension	?

Vedlegg 2: Mapping av properties i VARIABELMETADATA-filen til SKOS-AP-NO og SKOS-EN

Arbeidet med å mappe variabelmetadata mot henholdsvis den internasjonale versjonen av SKOS og Digitaliseringsdirektoratets norske versjon SKOS-AP-NO er et pågående arbeid.

Mappingen er viktig for å kunne utveksle og dele metadata om variabler på tvers av sektorer og landegrenser, samt for å opparbeide en felles forståelse av hva en variabel betyr.

Mappingen er også en forutsetning for å kunne etablere maskinell validering av kvaliteten på metadataene.

Der det i tabellen er brukt "E-helse:Extension" eller "?" har vi foreløpig ikke funnet en god match, eller er usikre på mappingen, mellom propertiene i denne spesifikasjonen og SKOS.

PresentationOrder_V.1.1	VARIABELMETADATA_Properties	SKOS-AP-NO	SKOS-EN
1	Code	Begrep.identifikator	dct:identifiser
2	ParentCode	E-helse:Extension	?
3	Hierarchy	E-helse:Extension	?
4	PreferredTerm	Begrep.anbefaltTerm	skosxl:prefLabel
5	ConceptType	E-helse:Extension	?
6	PresentationOrder	E-helse:Extension	?
7	ValidFrom	Begrep.gyldighetsperiode.gyldigFraOgMed	dct:temporal
8	ValidTo	Begrep.gyldighetsperiode.gyldigTilOgMed	dct:temporal
9	Synonym	Begrep.tillattTerm	skosxl:prefLabel
10	PatientFriendlyTerm	Begrep.tillattTerm	skosxl:prefLabel
11	ClinicalFriendlyTerm	Begrep.tillattTerm	skosxl:prefLabel
12	NavnEngelsk	Begrep.anbefaltTerm	skosxl:prefLabel
13	Beskrivelse	Begrep.definisjon	skosno:definisjon
14	BeskrivelseEngelsk	Begrep.definisjon	skosno:definisjon
15	Opprinnelse	E-helse:Extension	?

Nasjonal spesifikasjon for metadata om helsedata

16	OpprinnelseUtledning	E-helse:Extension	?
17	Kommentar	E-helse:Extension	?
18	KommentarEngelsk	E-helse:Extension	?
19	Kvalitetsnote (QAKommentar)	Datasett:Kvalitetsnote	dqv:QualityAnnotation
20	KvalitetsnoteEngelsk	Datasett:Kvalitetsnote	dqv:QualityAnnotation
21	MaaleType	E-helse:Extension	?
22	MaaleEnhet	E-helse:Extension	?
23	DataType	E-helse:Extension	?
24	Lengde	E-helse:Extension	?
25	Presisjon	E-helse:Extension	?
26	GrenseLav	E-helse:Extension	?
27	GrenseHoy	E-helse:Extension	?
28	Identifiseringsgrad	E-helse:Extension	?
29	Ledetekst	E-helse:Extension	?
30	SporsmaalNummer	E-helse:Extension	?
31	SkjemaVersjon	E-helse:Extension	?
32	AnbefaltForDigitaltInnsyn	E-helse:Extension	?
33	TabellvennligNavn	E-helse:Extension	?
34	TabellvennligNavnEngelsk	E-helse:Extension	?
35	DatabaseReferanse	E-helse:Extension	?
36	Erstatter (GammellID)	Begrep.erstatter	dct:replaces
37	ErstattesAv	Begrep.erstattesAv	dct:replacedBy
38	StartdatoHistorikk	Begrep.gyldighetsperiode.gyldigFraOgMed	dct:temporal
39	Synlig	E-helse:Extension	?

