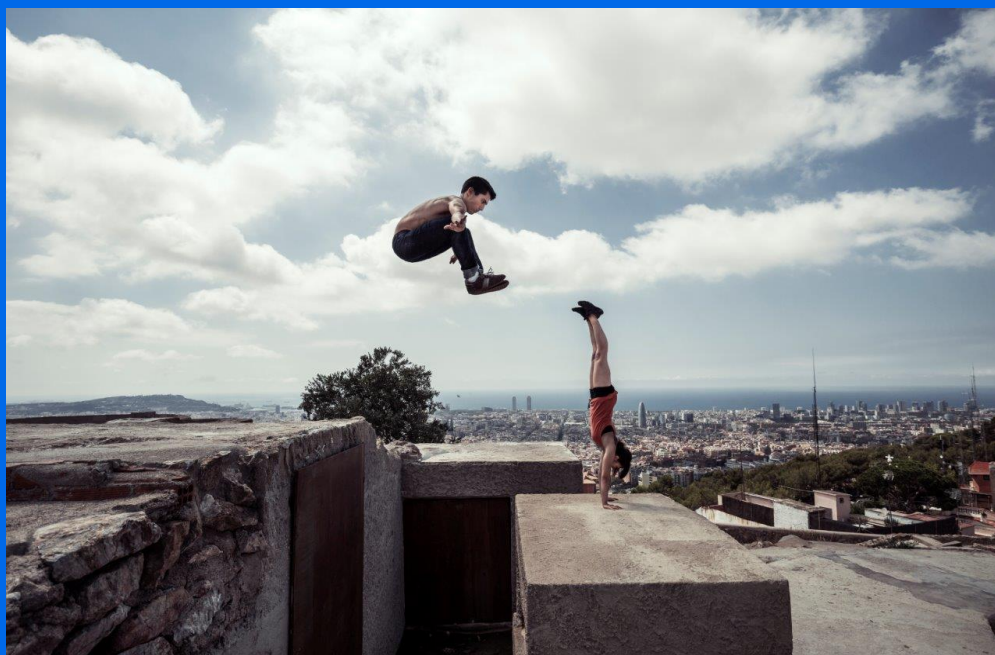


Nasjonal spesifikasjon for metadata om helsedata

Versjon 1.1



HITS 1230:2021

Publikasjonens tittel:

Nasjonal spesifikasjon for metadata om
helsedata

Rapportnummer

HITS 1230:2021

Utgitt:

11/2021

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 helsedataavdelingen i Direktoratet for e-helse 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 blir beskrevet på en lik måte og kan utveksles maskinelt. Metadatamodellen i denne spesifikasjonen vil også danne grunnlaget for å kunne spesifisere dataprodukt som skal overføres til Helseanalyseplattformen.

Spesifikasjonen bygger på arbeidet som ble igangsatt gjennom infrastrukturprosjektet Health Registries for Research der de sentrale helseregistrene etablerte en felles minimumsstandard for dokumentering av registervariabler [1]. Gjennom arbeidet i Helsedataprogrammet i Direktoratet for e-helse har minimumsstandarden blitt videreutviklet til å inkludere dokumentering av metadata om både datakildene selv, tilhørende datasamlinger, versjoner og kodeverkene som anvendes. Spesifikasjonen er i denne versjonen også utvidet til å kunne dokumentere metadata om medisinske kvalitetsregistre og befolkningsbaserte helseundersøkelser, samt fylle behov relatert til funksjonalitet i den nasjonale Helseanalyseplattformen.

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 Helseanalyseplattformen. Spesifikasjonen vil etter hvert revideres for å favne ytterligere behov relatert til utvikling av den Nasjonale variabelkatalogen og Helseanalyseplattformen, og for å kunne favne 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 samarbeid med registerforvaltere og relevante fagavdelinger i Direktoratet for E-helse.

Spesifikasjonen utvikles i tråd med FAIR-prinsippene og de internasjonale standardene DCAT og SKOS, og gjennom anbefalinger innhentet gjennom samarbeide i Europeiske og Nordiske prosjekter og arbeidsgrupper.

Versjonslogg

Versjon	Dato	Endringsbeskrivelse
1.0	01.01.2020	Første versjon
1.1	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 (valgfri). Denne erstatter TemaLokalt, som utgår som egen property i variabelmetadata. Importeres som egen fil med tilhørende mapset. • Ny klasse for Instrument (valgfri). Importeres som egen fil med tilhørende mapset ved behov. • Ny klasse for Filter (valgfri). Importeres 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 eller anbefalte <p>Importfiler:</p>

Versjon	Dato	Endringsbeskrivelse
		<ul style="list-style-type: none"> • Kolonnen «Required» er endret til «Status» med verdiene «Obligatorisk», «Anbefalt» og «Valgfri». • 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 • Vedlegg 3: Obligatoriske og frivillige klasser i den logiske informasjonsmodellen

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	14
4 Innrapportering av metadata til Nasjonal variabelkatalog	16
4.1 Prosess for innrapportering av metadata.....	16
4.2 Importfiler for metadata i Nasjonal variabelkatalog	19
Begrepsliste	56
Referanser	62
Vedlegg 1: Mapping av properties i KILDEMETADATA-filen til DCAT-AP-NO og DCAT-EN 63	
Vedlegg 2: Mapping av properties i VARIABELMETADATA-filen til SKOS-AP-NO og SKOS-EN	72
Vedlegg 3: Obligatoriske og frivillige klasser i den logiske informasjonsmodellen....	80

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 i Direktoratet for e-helse videreførte dette arbeidet. Gjennom 2020 har spesifikasjonen blitt utvidet til å inkludere metadata fra nasjonale medisinske kvalitetsregistre og befolkningsbaserte helseundersøkelser.

Nasjonal variabelkatalog ble etablert i løsningen HealthTerm i 2019. HealthTerm ble levert av CareCom AS og er anskaffet av Direktoratet for e-helse,

Etableringen av variabelkatalogen er avgjørende for utviklingen av en god variabelutforsker på helsedata.no og inngår som en viktig komponent i utviklingen av en effektiv søknadstjeneste og Helseanalyseplattformen.

Gjennom etableringen av data- og analysetjenester på Helseanalyseplattformen, vil det å overføre dataprodukter fra utvalgte datakilder til Helseanalyseplattformen stå sentralt. Gode metadata vil være avgjørende for at Helsedataservice skal kunne forvalte dataprodukter på Helseanalyseplattformen. Videre vil datamodellen for metadata danne grunnlaget for å spesifisere dataprodukter 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 mellom aktører og systemer, også på tvers av landegrenser.

Spesifikasjonen er tiltenkt å dekke informasjonsbehovet for bruk av data til sekundære formål som forskning, kvalitetsforbedring og annen analyse. Spesifikasjonen legger derfor føringer på hvilken informasjon som det er obligatorisk å innrapportere, hvilken som er anbefalt og hvilken som er valgfri.

Metadata vil utveksles med helsedata.no for å understøtte søknadsprosessen i et felles søknadsskjema. Innrapportering av metadata vil også 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.

Metadata vil også utveksles med andre metadatakataloger som f.eks. [Felles datakatalog](#) og [data.europa.eu](#), noe som krever en gradvis tilpasning til internasjonale standarder for utveksling og deling av metadata.

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 utover 2021 videreutvikles og etableres som en normerende retningslinje på linje med *Nasjonal spesifikasjon for metadata om helsedata*.

Det er gjort flere utvidelser av metadatamodellen fra versjon 1.0 til versjon 1.1 av spesifikasjonen. Dette er gjort for å kunne beskrive datakilder som medisinske kvalitetsregistre og befolkningsbaserte helseundersøkelser på en mer presis måte, samt å gi mer relevant informasjon til brukerne av dataene. Enkelte av klassene i modellen er frivillig å ta i bruk. Videre har flere *properties* som var satt som valgfrie i forrige versjon blitt gjort om til obligatoriske eller anbefalte i denne. Dette er gjort ut fra en vurdering av hvilken informasjon som vil være mest nyttig for brukere av helsedataene, samt etter tilbakemeldinger fra forvaltere av de aktuelle helsedatakildene.

1.3 Omfang og avgrensninger

Dette dokumentet er utformet som en spesifikasjon. Etter hvert som spesifikasjonen blir implementert, metadata fra nye datakilder blir importert og nye erfaringer gjøres, er målet at spesifikasjonen får status som en anbefalt standard. Dette er en lengre prosess og avhenger av god forankring i sektoren og utviklingen av rammeverk for tverrsektorielt og internasjonalt samarbeide.

Spesifikasjonen omfatter foreløpig sentrale helseregistre, nasjonale medisinske kvalitetsregistre og befolkningsbaserte helseundersøkelser. Fremtidige versjoner av spesifikasjonen vil også kunne stille krav til innrapportering av metadata fra andre helsedatakilder som f.eks. biodata og gendata. Spesifikasjonen må i så fall utvides til også å stille krav om hvordan denne typen data skal dokumenteres.

Etter hvert vil sannsynligvis en god del metadata kunne bli generert og hentet fra primærkilder som f.eks. PAS/EPJ, labsystemer og radiologisystemer, medisinsk teknisk utstyr og personlige applikasjoner. Hvor og hvordan dette skal spesifiseres, og hvordan dette vil påvirke sektoren og systemleverandører, vil bli vurdert når den tid kommer.

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

Denne versjonen av spesifikasjonen støtter ikke automatisert overføring av metadata til den nasjonale variabelkatalogen, og det legges derfor opp til at metadata skal overføres via e-post, se 4.1.1. Dette er ikke en skalerbar løsning, og Direktoratet for e-helse har derfor til hensikt å sende på høring en oppdatert spesifikasjon i løpet av 2022 der det legges til rette for å automatisere overføring av metadata til nasjonal variabelkatalog ved bruk av API.

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 spesifikasjon for metadata om helsedata (dette dokumentet) og de nevnte standardene finnes i Vedlegg 1 og Vedlegg 2.

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

2 Logisk informasjonsmodell for metadata om helsedata

Dette kapittelet 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 dagens tekniske løsning.

Informasjonsmodellen i Figur 1 er utvidet med et flere nye klasser fra versjon 1.0 av denne spesifikasjonen. Ikke alle klassene er obligatoriske, og enkelte vil kun være relevant for kun enkelte typer av datakilder. I Vedlegg 3: Obligatoriske og frivillige klasser i den logiske informasjonsmodellen, redegjøres det nærmere for hvilke klasser som er obligatoriske og frivillige, samt hvilke som ikke skal benyttes. Dette gjøres for å synliggjøre hva som minimum forventes av datakildene som skal rapportere inn metadata, samt hva det vil være mulig å kunne rapportere inn.




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 en datakilde 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*.
- **Datasett** er ikke en klasse i informasjonsmodellen for metadata, men beskriver i stedet de faktiske helse- og personopplysningene som en versjon av en datasamling representerer. I *Spesifikasjon for dataoverføring til Helseanalyseplattformen* inngår **Datasett** i den logiske informasjonsmodellen for relasjon mellom data og metadata og betegner ett gitt sett med data som inngår i et dataprodukt som skal overføres 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 datakilder 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.

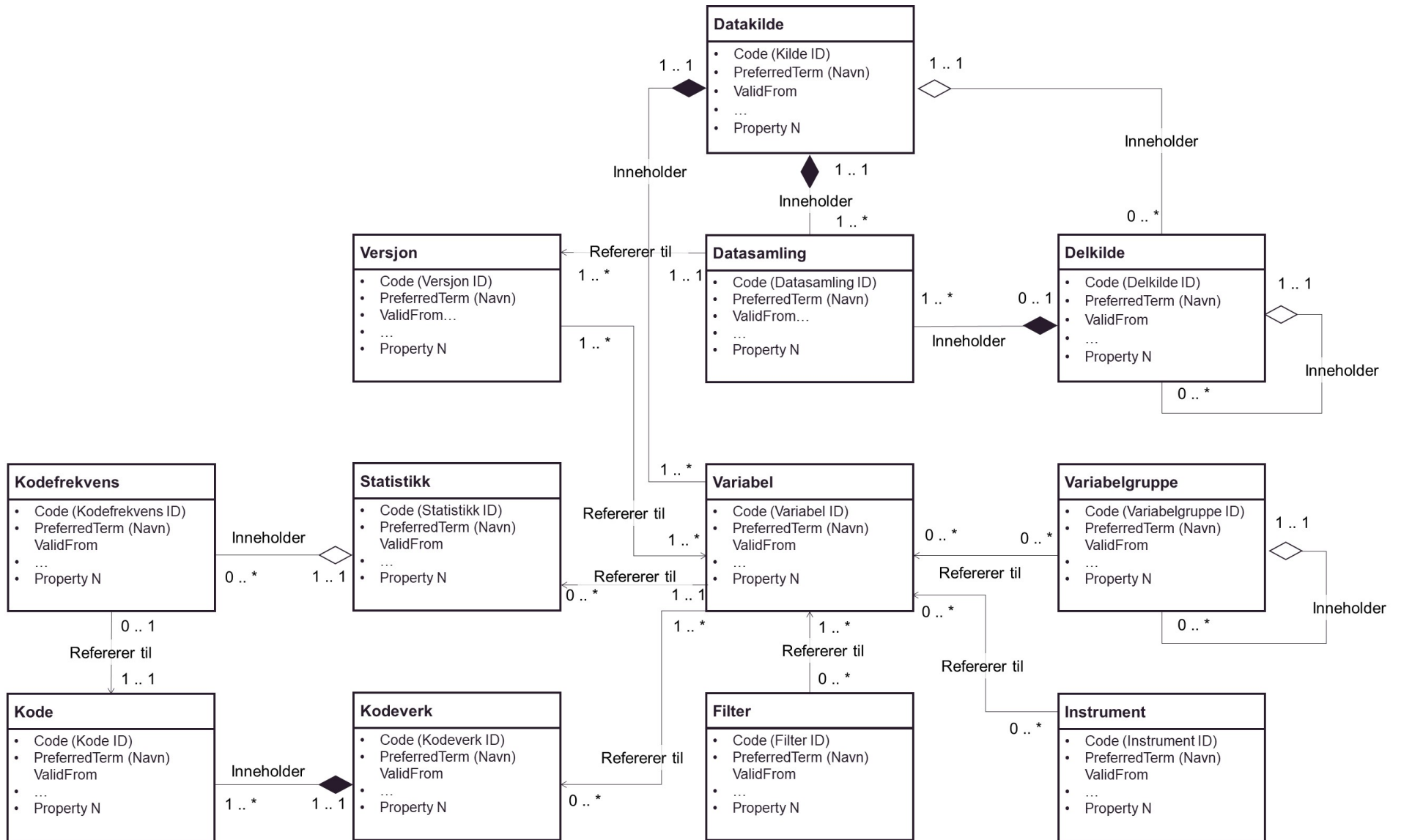
- **Versjon** er et teknisk hjelpemiddel som brukes til å definere hvilke variabler som inngår i ulike årganger/versjoner av de ulike datasamlingene/dataproduktene i en datakilde.
- **Variabler** er de informasjonsbærende elementene i datakilden og definerer hvilken informasjon som finnes i de nasjonale helsedatakildene.
- **Variabelgruppe** er en kategorisering av variabler. Strukturen kan bygges opp hierarkisk. Variabelgruppe erstatter den nå utgåtte *property* LokaltTema, men vil også kunne brukes til å gruppere variabler på tvers av helsedatakilder.
- **Kodeverk** representerer i denne sammenhengen et verdsett. Et kodeverk kan for eksempel være et administrativt kodeverk, et helsefaglig kodeverk eller egendefinerte svaralternativer for en variabel.
- **Kode** er i denne sammenheng det samme som et svaralternativ i et kodeverk (verdsett). En kode består av en kodeverdi og en kodetekst.
- **Statistikk** er definert som de statistiske egenskapene ved en årgang/versjon av en variabel. Statistikk som metadata har som formål å si noe om registrerings-/kodingskvaliteten og gi en pekepinn om antall forekomster. Statistikken egner seg ikke for analyse som sådan.
- **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 mer komplett oversikt over relevante fagbegrep er gjengitt i begrepslisten.

Informasjonsmodellen benytter seg videre av UML-baserte ikoner og begrep. Disse er nærmere forklart i Tabell 1.

Tabell 1: Forklaring av UML-syntaks benyttet i informasjonsmodellen		
Ikon	Definisjon	Beskrivelse
	Klasse	En <i>klasse</i> er en logisk gruppering av entiteter som beskriver en gitt del av modellen. Klasser kjennetegnes ved at de brukes til å beskrive en gitt type objekter. Eksempler på plasser er Variabel som brukes til å definere alle variablene i gitt datakilde.
	Assosiativ relasjon	En <i>assosiativ relasjon</i> er den mest generelle formen for relasjon innenfor UML. I modellen under brukes assosiative relasjoner der det ikke er noen form for hierarkisk relasjon mellom <i>klassene</i> i form av en <i>referer til</i> -relasjon.
	Aggregerende relasjon	En <i>aggregerende relasjon</i> uttrykker at deler settes sammen til en større helhet, i invers forstand ofte kalt en «part of» relasjon. I modellen under

		brukes aggregerende relasjoner der en <i>klasse</i> kan <i>inneholde</i> av 0, 1 eller flere instanser av en <i>sub-klasse</i> .
	Komposisjon-relasjon	En <i>komposisjon-relasjon</i> uttrykker er en spesiell form for en <i>aggregerende relasjon</i> i de tilfellene der det er et strengt krav at en klasse må <i>inneholde</i> minst én instans av en gitt sub-klasse.



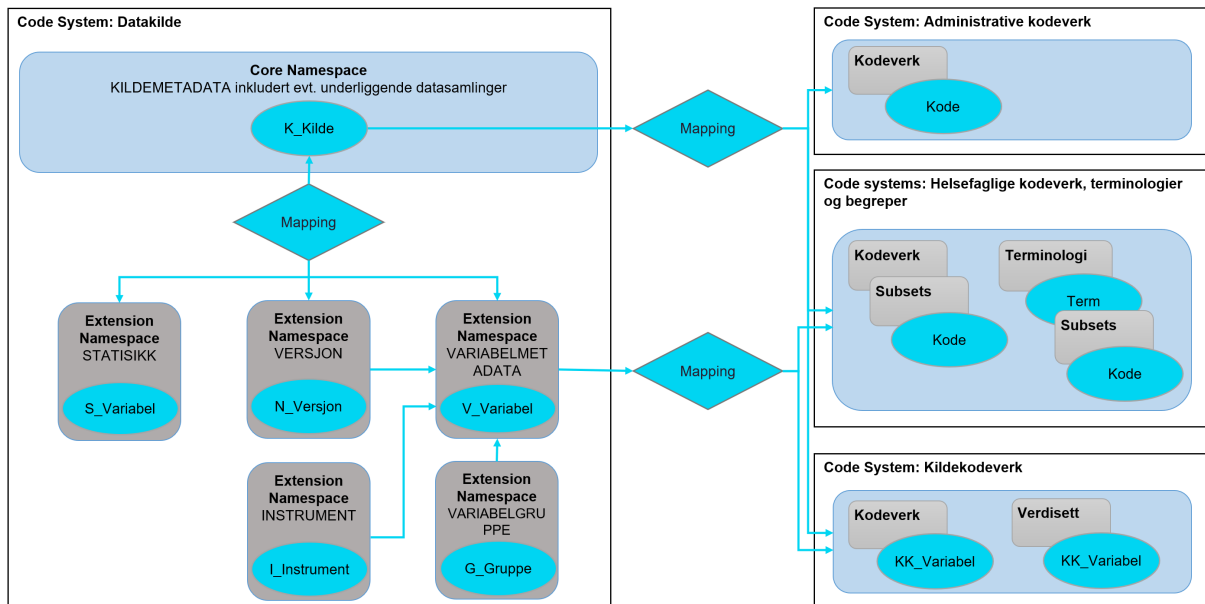
Punktlisten under forklarer informasjonslisten illustrert i Figur 1.

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 helsedatakilder som leverer dataprodukt til Helseanalyseplattformen, skal klassen datasamling benyttes til å dokumentere metadata om dataprodukt. Klassen delkilde kan brukes til å gruppere dataprodukt og dokumentere metadata om logiske grupperinger av 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 datakilder som skal levere dataprodukt til Helseanalyseplattformen.
8. Hver variabel kan *referere til* null, én eller flere variabelgrupper.
9. Variabler kan *referere* til kodeverk. For kategoriske 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 kategoriske variabler kan statistikk *inneholde* kodefrekvens, det vil si antall utfall av de ulike mulige verdiene i variabelens standardiserte vokabular.
11. Hver variabel kan bli *referert til* fra null, ett eller flere instrument, og flere ulike variabler kan bli *referert til* fra samme 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 punktlisen under. Løsningsspesifikke begrep som *Code System*, *concept* og *property* er forklart nærmere i vedlagt begrepsliste. Alle HealthTerm-spesifikke begrep er benyttet på originalspråket (engelsk) for å slippe å innføre nye begrep med tilhørende risiko for å innføre nye misforståelser.



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.
3. 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.

4. 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.
5. STATISTIKK importeres som en *Extension* til *Core Namespace*. STATISTIKK inneholder statistisk informasjon om utvalgte variabler definert i VARIABELMETADATA fordelt per datasamling og VERSJON som variabelen inngår i.
6. 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.
7. VARIABELGRUPPE importeres som en *Extension* til *Core Namespace*. VARIABELGRUPPE brukes for å ordne variabler etter tema. VARIABELGRUPPER kan bygges opp hierarkisk.
8. 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.
9. 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 en liste 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.
10. 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.
11. Datasamlinger mappes til respektive VERSJONER som igjen mappes til de variablene som inngår i gjeldende versjon av datasamlingen.
12. Variablene i VARIABELMETADATA mappes til respektive KILDEKODEVERK, respektiv STATISTIKK*, og evt. andre referanser som f.eks. begrepskataloger
13. 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

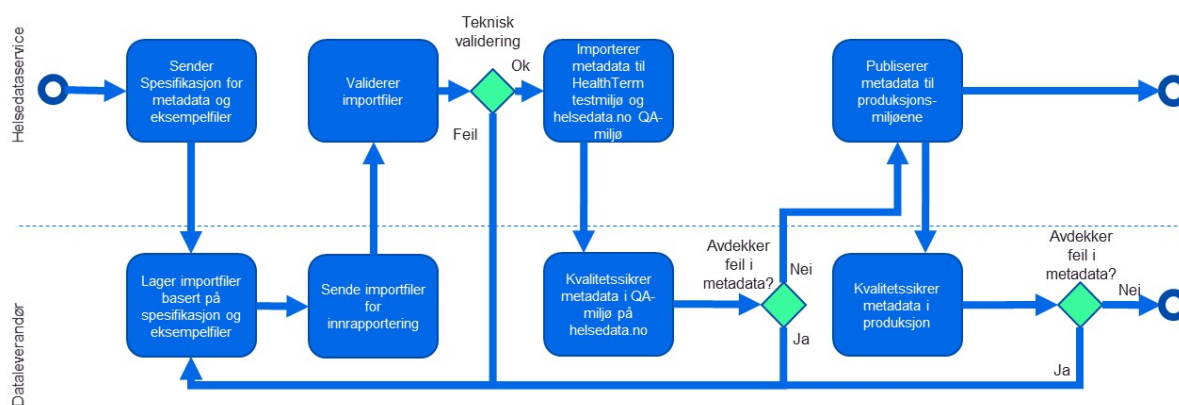
4.1 Prosess for innrapportering av metadata

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. Det vil også være mulig å gjøre oppdateringer hyppigere i de tilfellene der dette fremstår som hensiktsmessig. For helsedatakilder som skal overføre dataprodukt til Helseanalyseplattformen, vil behovet for oppdatering av metadata bli styrt av oppdateringsfrekvensen for data som eksporteres til Helseanalyseplattformen.

I fremtiden er det ønskelig at innrapporteringsprosessen i størst mulig grad skal være automatisert. Ettersom dagens tekniske løsning for den nasjonale variabelkatalogen foreløpig ikke støtter automatisert innlesning av metadata, vil man i en overgangsfase legge opp til en prosess der de nasjonale helsedatakildene lager et sett med eksportfiler som sendes per e-post til metadata@helsedata.no.

Det utredes nærmere løsninger for automatisert mottak av metadata til Nasjonal variabelkatalog, og man tar sikte på å støtte mottak av oppdaterte metadatafiler via API i løpet av 2022.

Figur 3 viser hvordan dagens prosess for oppdatering av metadata i Nasjonal variabelkatalog foregår. Det forutsettes at man har kontaktet Helsedataservice og planlagt en prosess for å inngå i økosystemet rundt Helseanalyseplattformen.



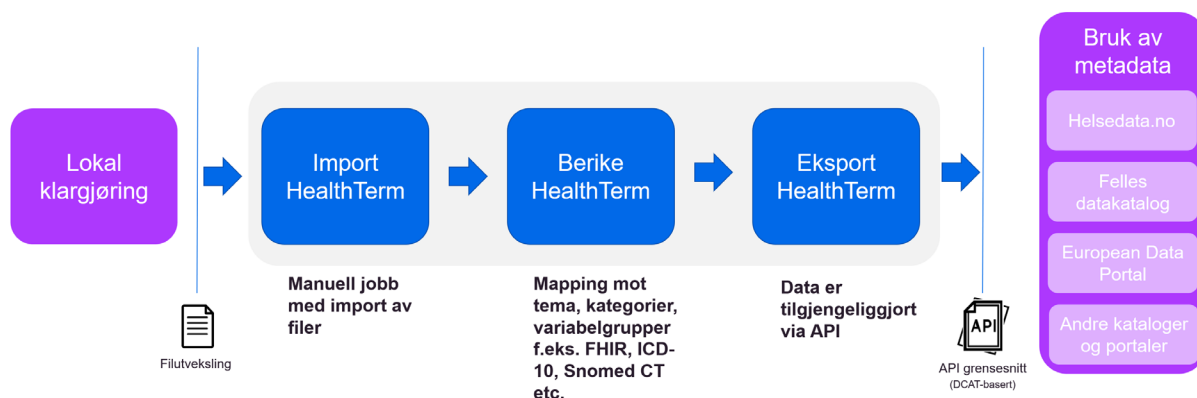
Figur 3 Stegene for manuell innrapportering av metadata til HealthTerm og publisering til helsedata.no

[Felles datakatalog](#) og data.europa.eu. Arbeidet med å forvalte og tilgjengeliggjøre metadata fra den nasjonal variabelkatalogen vil gjøres i nært samarbeid med forvalterne av de aktuelle datakildene og metadata skal ikke eksponeres på helsedata.no eller andre kilder før forvalterne av den

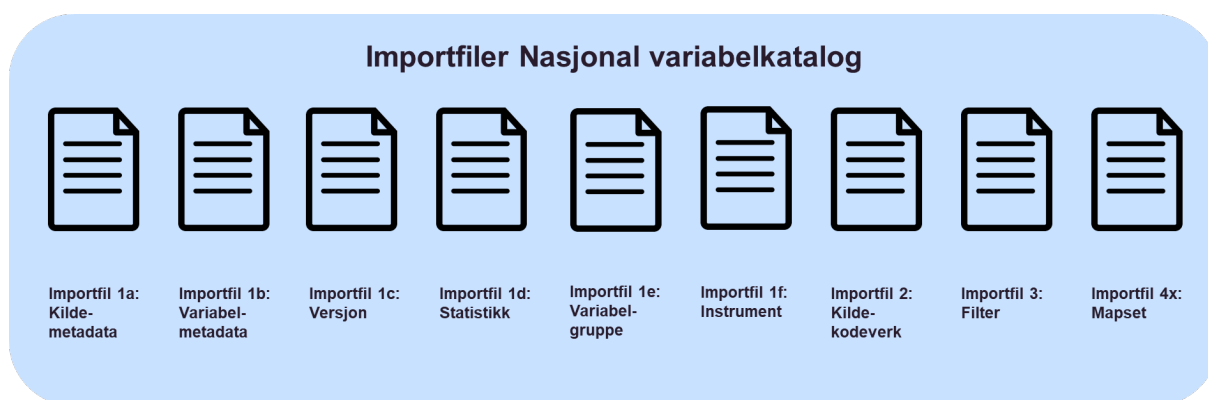
enkelte datakilde har fått anledning til å kvalitetssikre og godkjenne innholdet. Det legges derfor opp til et kontrollregime der dataforvalterne gis anledning til å kontrollere egne variabler og metadata før disse publiseres åpent på helsedata.no.

Ved førstegangs import av metadata er det mulig å bygge opp en endringshistorikk tilbake i tid ved å anvende *properties* 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 *properties* Beskrivelse eller Kommentar, og at signifikante kvalitetsbrister i datagrunnlaget framgår av *property* Kvalitetsnote. Det redegjøres nærmere for bruken av ValidFrom og ValidTo i kapittel 4.1.2.

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



For å kunne realisere den logiske informasjonsmodellen i Kapittel 2, vil de nasjonale helsedatakildene 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 de nasjonale helsedatakildene skal rapportere *Importfil 3: Filter* selv.



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 som en del av et komplett sett av importfiler. Dette gjøres for å sikre at metadataene til enhver tid er konsistente. 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.

4.1.2 Bruk av systemproperties ValidFrom og ValidTo

Systempropertyene ValidFrom og ValidTo har i etterkant av at versjon 1.0 av denne spesifikasjonen ble publisert, vært gjenstand for mye forvirring. I dette avsnittet redegjøres det derfor nærmere for betydningen av disse propertyene.

Når man oppretter et nytt *concept* i HealthTerm, får dette *conceptet* og alle tilhørende *properties* en *history*. En *history* vil være en komplett kronologisk hendelseslogg over alle endringer som skjer med et *concept* eller en *property*. Den første hendelsen i enhver *history* er en *Create*. Tidspunktet for denne, settes av ValidFrom fra importfil, eller fra en ValidFrom dato man velger i importgrensesnittet. Dette tidspunktet kan aldri endres, og man kan ikke legge til hendelser før dette tidspunktet i ettertid. Samme tidspunkt gjelder også for alle *properties* som tilhører *conceptet*. Da ValidFrom er en dato og hendelsen logges med et tidspunkt, defineres klokkeslett for hendelsen til 00:00:00.

Det er mulig å legge inn en ValidTo dato i importfilen, for å indikere at et *concept* ikke lenger skal være gyldig, eller at det på et tidspunkt tilbake i tid har opphørt å være gyldig. Bruk av ValidTo fører til en *Retire* hendelse, som er en avslutning på en *history*. NB! det er viktig å være klar over at det ikke tillates noen form for oppdateringer når denne datoen først er satt.

Forsøk på dette fører til en feilhendelse i importgrensesnittet. Derfor anbefales at utgåtte *concepts* fjernes fra importfilene etter at denne oppdateringen har skjedd.

Ved oppdatering av eksisterende *concepts*, har ValidFrom en endret betydning. ValidFrom vil da være hendelsestidspunkt for eventuelle endringer. Derfor skal ikke ValidFrom brukes i importfiler for endring. Endringer skal logges på et tidspunkt som styres i importgrensesnittet. NB! Hvis en endringsfil inneholder nye *concepts*, er det mulig å styre ValidFrom for *enkelt-concepts*. Dette vil kreve at man har full kontroll på hva som er levert til HealthTerm tidligere. Eller det må avtales spesielt med Helsedataservice.

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). Mapping-fil skal være i CSV-format (.csv) med komma (',') som separasjonstegn, ikke semikolon (;). For csv-filen skal tegnsett følge UTF-8. Disse kravene følger av systemkrav fra HealthTerm.

4.2.2 Navnstandard for importfiler

For å sikre sporbarhet skal importfilene definert i Figur 5 gis et standardisert navn. Navnene kan kun inneholde store og små bokstaver (A-Z og a-z) og tall (0-9). Bindestrek (-) og understrek (_) brukes som vist i definisjonene under og skal ikke inngå i de delene av filnavnet som defineres fritt av datakilden.

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

**DATAKILDE_Importfiltype-Importfilnavn_Årssett*_Versjon*_Dato sendt
(YYYYMMDD)**

*Årssett er siste komplette årgang som er inkludert i datagrunnlaget

*Versjon kan brukes for kilder som oppdateres og/eller endres flere ganger i året eller for å angi datakildens interne versjonsangivelse.

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 (definert i Tabell 14)
- F for filter
- For andre mappings brukes en egendefinert logisk forkortelse

Eksempler:

- KREG_3a_MAPSET_K-N_2018_V34d_20190402
(Filen mapper KILDEMETADATA 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 Navngivning av nøkkelfeltet **Code** i importfilene

Propertien Code er nøkkel for de forskjellige metadataobjektene. For at den skal kunne fungere som en unik og persistent ID skal den navngis på følgende måte:

- Det er kun tillatt med store bokstaver fra (A-Z), heltall (0-9) og samt bindestrek. Underscore «_» og punktum «.» benyttes bare som skille tegn for nøklene som vist i eksemplene under. Underscore er tillatt i variabel.Kortnavn
- Første tegn i nøklene kan ikke være ett siffer.
- Nøklene prefixes med hva slags type objekt som gjelder, K for Kildemetadata, V for variabelmetadata osv. slik som beskrevet for importfiltypen mapset. Deretter «_» og kortnavn på datakilden. For objekttype datakilde er dette nok, som f.eks K_HUNT, K_DAR m.m.

De andre objekttypene skal i tillegg til type og kilde ha med egne navn og hierarki. Tabell 2 viser hvordan man kan navngi de forskjellige nøkler.

Tabell 2: Navnstandard for bruk av <i>code</i>	
Object code	Navngivning av kode
Datakilde	Eksempel: K_HUNT Prefix + «_» + datakilde.KortNavn
Delkilde	Eksempel: K_HUNT.HUNT1 Prefix + «_» + datakilde.code + «.» + delkilde.KortNavn
Datasamling	Eksempel: K_HUNT.HUNT1_QUESTIONAIRE Prefix + «_» + datakilde.code + «.» + datasamling.KortNavn
Variabel	Eksempel: V_HUNT.AGE_YEAR Prefix + «_» + datakilde.kortnavn + «.» + variabel.KortNavn

Tabell 2: Navnstandard for bruk av <i>code</i>	
Object code	Navngivning av kode
Versjon	Eksempel: N_HUNT.HUNT1_QUESTIONAIRE_2018 Prefix + «_» + datakilde.kortnavn + «.» + datasamling.KortNavn + «_» + versjonsnummer/id
Statistikk	Eksempel: S_HUNT_2017.HUNT1_QUESTIONAIRE.GENDER.1 Prefix + «_» + datakilde.kortnavn + «_» + årstall + «.» + datasamling.KortNavn + «.» + variabel.KortNavn + «.» + KodeverLokalid
Variabelgruppe	Eksempel: G_HUNT.EXPOSURE Prefix + «_» + datakilde.KortNavn + «.» + variabelgruppe.KortNavn
Kildekodeverk	Eksempel: KK_HUNT.GENDER.1 Prefix + «_» + datakilde.kortnavn + «.» + variabel.KortNavn + «.» + KodeverLokalid
Instrument	Eksempel: I_HUNT.AbbottAXSymx Prefix + «_» + datakilde.code + «.» + instrumentkilde.KortNavn
Filter	Som instrument

4.2.4 Format for importfiler

Importfilene for metadata i HealthTerm er spesifisert i Tabell 5 til Tabell 13. Følgende begrep brukes for å definere importspesifikasjonen.

- **Concept** er det grunnleggende dataobjektet i HealthTerm. Variabler er derav et eksempel på et *concept* i modellen.
- **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 *properties* er definert i samarbeid med forvalterne av de aktuelle datakildene og med utgangspunkt i aktuelle standarder. Informasjonsinnholdet importeres som tekst, og har en maksimal lengde på 4000 tegn for hver *property*.
- **Status** spesifiserer om informasjonsinnholdet er obligatorisk, anbefalt eller valgfritt.
- **Beskrivelse** inneholder beskrivelser av informasjonen i kolonnene. Innholdet kan være definert som et **Standardisert Vokabular**. Et standardisert vokabular er en liste med lovlige svaralternativ og skal kun representeres ved kodeverdien i innrapporteringsfilen.

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: Kildemetadata. 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 **Topppnodene** i hver av klassene i kodesystemet. Klassene er Kildemetadata, Versjon, Variabelmetadata, Statistikk, Variabelgruppe og

Instrument. Disse toppnodene vil defineres ved hjelp av 1. linje i de enkelte importfilene, med unntak av kildemetadata, der toppnoden defineres av 2. linje i importfil 1a. Toppnodene skal defineres som beskrevet i tabell 3. *Code* – feltet i toppnoden vil være ParentCode til alle de etterfølgende radene.

Tabell 3: Hierarkisk oppbygging av importfilene					
Filtype	Code	Parent Code	Hierar chy	PreferredTerm	ConceptType
1a: Kildemetadata	DAR			Dødsårsaksregisteret	GROUPING
1a: Kildemetadata	K_DAR	DAR		KILDEMETADATA	REGULAR
1b: Variabelmetadata	V_DAR	DAR		VARIABLEMETADATA	GROUPING
1c: Versjon	N_DAR	DAR		VERSJON	GROUPING
1d: Statistikk	S_DAR	DAR		STATISTIKK	GROUPING
1d: Statistikk	S_DAR_2017. DAR	S_DAR		DØDSÅRSAKSREGISTER ET 2017	GROUPING
1e: Variabelgruppe	G_DAR	DAR		VARIABLEGRUPPE	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 4: Polyhierarkisk oppbygging av Importfil 2: Kildekodeverk				
Code	ParentCode	Hierarchy	PreferredTerm	ConceptType
KK_KODEVERK*			KILDEKODEVERK	GROUPING
KK_ALLE*	KK_KODEVERK		ALLE	GROUPING
KK_EIER*	KK_KODEVERK		EIER	GROUPING
KK_KILDE*	KK_KODEVERK		KILDE	GROUPING
KK_<Eier>**	KK_EIER		<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		FHI	GROUPING
KK_DAR	KK_FHI		DØDSÅRSAKSREGISTERET	GROUPING

KK_DAR	KK_KILDE		DØDSÅRSAKSREGISTERET	
KK_DAR.KJONN_K	KK_ALLE		Kjønn	
KK_DAR.KJONN_K	KK_DAR			
<p>*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.</p>				

4.2.5 Formatering av tekst

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

Formatet støtter bare et utvalg formatteringskoder:

Tabell 5: Bruk av markdown i tekst			
Atributt	Syntaks	Eksempel på bruk	Eksempel utseende *
Linjeskift	 	Første linje tekst. Andre linje tekst	Første linje tekst Andre linje tekst
Avsnitt	 	Første linje tekst. Nytt avsnitt	Første linje tekst Nytt Avsnitt
Bold tekst	**<tekst>**	**Uthevet tekst**	Uthevet tekst
Link	[Tekst](url)	Se https://helsedata.no	Se https://helsedata.no
Ordnet Punktliste	n.	1. Punkt 1 2. Punkt 2 3. Punkt 3	1. Punkt 1 2. Punkt 2 3. Punkt 3
Uordnet punktliste	*	* Første * Andre * Tredje	• Første • Andre • Tredje

Utseende vil ikke være eksakt som i eksemplet over, men styres av stildefinisjonen for nettsiden helsedata.no

4.2.6 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 datakilder som skal levere dataprodukt til Helseanalyseplattformen vil importfilen også brukes til å definere dataprodukt i form av datasamlinger og evt. logiske grupperinger av dataprodukt i form av delkilder.

Iht Prop. 63 (2019-2020) skal dataprodukt inneholde informasjon om dataproduktets bruksområder (Anbefalte bruksområder og Rettsgrunnlag). Dette representeres som metadata om datasamlinger i importfilen definert under. Anbefalte bruksområder er definert i *property* 16, mens Rettsgrunnlag er definert i *properties* 30-37.

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. Disse *propertyene* har også navn som stammer fra HealthTerm.

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 kildemetadatatypene.

Tabell 6: Format for importfil 1a-Kildemetaddata.

Gruppering	Nr	Property	Status	Beskrivelse
<i>Administrative formål</i>	1	Code	Obligatorisk	Unik navnestreng. Tillatte tegn er store bokstaver og tall, uten mellomrom og æ, ø, å. Datakilde: Prefiks K for kilde, underscore, kortnavn på datakilden. F.eks. K_HUNT eller K_NPR. Delkilde: Prefiks K for kilde, underscore, kortnavn på datakilden, punktum, og navn på delkilde. F.eks. K_HUNT.HUNT1, eller K_NPR.SOMATIKK. Datasamling: Prefiks K for kilde, underscore, kortnavn på datakilden, punktum, og navn på datasamling. F.eks. K_HUNT.HUNT1_QUESTIONAIRE1, eller K_NPR.SOMATIKK_EPISODE.

Tabell 6: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Status	Beskrivelse
	2	ParentCode	Obligatorisk	<p>Datakilde: Skal være lik kortnavnet på datakilden, f.eks. HUNT.</p> <p>Delkilde: Skal være lik Code for tilhørende datakilde, f.eks. K_HUNT.</p> <p>Datasamling: Skal være lik Code for tilhørende datakilde. Hvis datasamlingen tilhører en delkilde, skal delkildens Code benyttes, f.eks. K_HUNT.HUNT1.</p>
HealthTerm	3	Hierarchy	Valgfri	Skal være tom.
Tittel og beskrivelse	4	PreferredTerm	Obligatorisk	<p>Kort beskrivende navn på datakilden, delkilden eller datasamlingen.</p> <p>F.eks. «Den norske mor, far og barn undersøkelsen», «MOBA Hovedskjema» eller «Spørreskjema1» for henholdsvis datakilde, delkilde og datasamling.</p>
HealthTerm	5	ConceptType	Valgfri	Skal være tom
Administrative formål	6	PresentationOrder	Valgfri	<p>Datakilde: Skal være tom.</p> <p>Delkilde, datasamling: Brukes dersom man ønsker å presentere delkildene eller datasamlingene i en egendefinert rekkefølge som avviker fra alfabetisk. Gyldige verdier er positive heltall og delkildene eller datasamlingene presenteres i stigende rekkefølge basert på disse tallene.</p>
	7	ValidFrom	Obligatorisk	<p>Brukes for å angi fra når en datakilde, delkilde eller datasamling er gyldig. Settes ved første import.</p> <p>Format: YYYYMMDD</p>
	8	ValidTo	Valgfri	Skal være tom

Tabell 6: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Status	Beskrivelse
<i>Tittel og beskrivelse</i>	9	Synonym	Valgfri	Skal være tom (inntil videre)
	10	PatientFriendlyTerm	Valgfri	Skal være tom (inntil videre)
	11	ClinicalFriendlyTerm	Valgfri	Skal være tom (inntil videre)
	12	KortNavn	Anbefales	Forkortelse av navnet på datakilden, delkilden eller datasamlingen. F.eks. MOBA eller NPR.
	13	NavnEngelsk	Anbefales	Engelsk navn på datakilden, delkilden eller datasamlingen.
	14	Beskrivelse	Obligatorisk	En utførlig beskrivelse. Skal være forståelig også for innbyggere. Se eksempler på helsedata.no
	15	Beskrivelse Engelsk	Anbefales	En utførlig beskrivelse på engelsk. Skal være forståelig også for innbyggere.
	16	AnbefalteBruksområder	Obligatorisk	En beskrivelse av hvilke typer formål (eksempelvis forskning, analyse, styringsdata, beredskap osv) og faglige temaer (eksempelvis kreft, hjerte og kar, kirurgi osv), en datakilde, delkilde og/eller datasamlinger er egent for.
	17	Emneord	Valgfri	Beskrivende emneord (tag). Skilles ved å bruke «;». F.eks. Emneord;Nytt emneord;
<i>Opphav og ferskhet</i>	18	Opprettet	Obligatorisk	Året datakilden, delkilden eller datasamlingen ble opprettet og begynte å samle inn data.

Tabell 6: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Status	Beskrivelse
				Format: YYYY.
	19	SistOppdatert	Obligatorisk	Dato for siste oppdatering av KILDEMETADATA Format: YYYYMMDD
	20	Frekvens	Valgfri	<p>Datakilde, delkilde: Skal være tom.</p> <p>Datasamling: Oppdateringsfrekvens brukes for å beskrive hvor ofte datagrunnlaget oppdateres på Helseanalyseplattformen.</p> <p>Standardisert vokabular:</p> <ul style="list-style-type: none"> 1 = Sjeldnere enn årlig 2 = Årlig 3 = Halvårlig 4 = Tertialvis 5 = Kvartalsvis 6 = Månedlig 7 = Ukentlig 8 = Daglig 9 = Hyppigere enn daglig
	21	Innsamlingsmetode	Valgfri	Kort beskrivelse av metode for innsamling av data.

Tabell 6: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Status	Beskrivelse
<i>Innhold og kvalitet</i>	22	Kildetype	Obligatorisk	Kategorisering av datakilden. Standardisert vokabular: 1 = Sentralt helseregister ² 2 = Nasjonalt medisinsk kvalitetsregister 3 = Annet medisinsk kvalitetsregister 4 = Befolkningsbasert helseundersøkelse 5 = Biobank 6 = Annen datakilde
	23	InklusjonsOgEksklusjonskriterier	Obligatorisk	En faglig overordnet beskrivelse av hvilke kriterier som stilles for at en person skal være registrert i datakilden, delkilden eller datasamlingen. Dette tilsvarer en beskrivelse av populasjonen som inngår i datakilden, delkilden eller datasamlingen.
	24	Geografisk Avgrensning	Valgfri	Beskrivelse av opptaksområdet for innbyggerne som inngår i datakilden, delkilden eller datasamlingen.
	25	Tilslutningsgrad	Valgfri	Andel enheter av totalt antall relevante enheter som rapporterer inn til datakilden, delkilden eller datasamlingen. Med enheter menes foretak, organisasjoner og virksomheter. Oppgis som desimaltall på formatet «0,875». (Kompletthet i form av dekningsgrad på institusjons- eller enhetsnivå)

² Inkluderer lovbestemte helseregistre og andre sentrale helseregistre (<https://www.fhi.no/div/datatilgang/om-sentrale-helseregistre/>)

Tabell 6: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Status	Beskrivelse
	26	BeregningTilslutningsgrad	Valgfri	Hvordan tilslutningsgraden er beregnet. F.eks. hva er nevneren i brøken?
	27	Dekningsgrad	Obligatorisk	<p>NB! Kun for Nasjonale kvalitetsregistre inntil videre.</p> <p>Dekningsgraden angir i hvilken grad de pasienter eller hendelser som skal registreres faktisk blir registrert. Dekningsgraden for et enkeltsykehus er et mål på hvor mange av de aktuelle pasientene som behandles ved sykehuset som blir registrert i kvalitetsregisteret, mens den samlede dekningsgraden for registeret er den aggregerte dekningsgraden for alle sykehusene som behandler pasienter i registerets målgruppe. Man kan for eksempel si at registeret har en samlet dekningsgrad på 80% og at spesifikt sykehus har en dekningsgrad på 85%.</p> <p>Andre helsedatkilder bruker proprietæren «Kvalitetsnote» for å beskrive kvaliteten på datagrunnlaget.</p> <p>Oppgis som desimaltall på formatet «0,875».</p>
	28	BeregningDekningsgrad	Valgfri	Hvordan dekningsgrad er beregnet. F.eks. sammenstilt med data fra NPR, og evt. formel.
	29	Kvalitetsnote	Anbefales	Overordnet tekstlig beskrivelse av kvaliteten på datagrunnlaget.
Rettsgrunnlag	30	Formaal	Obligatorisk	Kort beskrivelse av det juridiske formålet datakilden er opprettet for. Dersom mulig, en kopi av formålet slik det står skrevet i forskrift.

Tabell 6: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Status	Beskrivelse
	31	Lovverk	Obligatorisk	En eller flere URL-er til lovverk på lovdata.no. F.eks. URL til Lov om helseregistre og behandling av helseopplysninger. Flere URL-er skiller ved å bruke «;».
	32	Forskrift	Valgfri	En eller flere URL-er til forskrift på lovdata.no. F.eks. URL til Forskrift om innsamling og behandling av helseopplysninger i Medisinsk fødselsregister. Flere URL-er skiller ved å bruke «;».
	33	HjemmelTil gjengeliggj oring	Valgfri	URL til paragraf med hjemmel for tilgjengeliggjøring.
	34	Dataansvar lig	Obligatorisk	Navn på virksomheten eller enheten som eier og har det juridiske dataansvaret for datakilden, delkilden eller datasamlingen.
	35	Databehan dler	Obligatorisk	Navn på virksomheten eller enheten som faktisk behandler datakilden. Er det samme som Dataansvarlig med mindre denne har satt ut databehandlingen til ekstern tredjepart.
	36	GradAvPer sonidentifik asjon	Valgfri	Høyeste grad av identifisering det er mulig å søke om fra denne datakilden, delkilden eller datasamlingen. Standardisert vokabular: 1 = Direkte personidentifiserbare opplysninger 2 = Indirekte personidentifiserbare opplysninger 3 = Anonyme opplysninger

Tabell 6: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Status	Beskrivelse
	37	JuridiskNot e	Valgfri	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>	38	Hjemmesid e	Obligatorisk	URL til hjemmeside.
	39	Epost	Obligatorisk	E-postadresse for henvendelser om datakilden, delkilden eller datasamlingen.
	40	Telefonnum mer	Valgfri	Telefonnummer for henvendelser om datakilden, delkilden eller datasamlingen.
	41	OrgnrDataa nsvarlig	Valgfri	Organisasjonsnummer til dataansvarlig.
	42	HERidData kilde	Valgfri	Datakilde: Unik identifikator som tildeles av NHN, Brukes for identifisering av datakilden innenfor helsenettet. Kan brukes til spørringer.
	43	HERidData ansvarlig	Valgfri	Unik identifikator som tildeles av NHN, Brukes for identifisering av datakilden innenfor helsenettet. Kan brukes til spørringer.
<i>Administrative formål</i>	44	Kildemetad atatype	Obligatorisk	Angir hvilken type kildemetadata som beskrives. Standardisert vokabular: 1 = Datakilde 2 = Delkilde 3 = Datasamling

Tabell 6: Format for importfil 1a-Kildemetadata.				
Gruppering	Nr	Property	Status	Beskrivelse
	45	EriSamsvar Med	Valgfri	Referanse til evt. standard(er), spesifikasjon(er) etc. som datakilden, delkilden eller datasamlingen er i henhold til, f.eks. DCAT-AP-NO. Skilles ved å bruke «;». F.eks. Standard 1;Standard 2;
	46	StatistikkType	Valgfri	Hvorvidt medfølgende statistikk er akkumulert eller årsbasert. Akkumulert vil eksempelvis kunne være over flere år slik som for helseundersøkelser, mens årsbasert vil være innenfor et definert år slik som for DÅR. Standardisert vokabular: 1 = Akkumulert 2 = Årsbasert
	47	TelleEnhet	Valgfri	Datasamling: Telleenhet for datasamlingen angitt ved en tekstlig beskrivelse, eks. pasient, episode eller sykehusopphold.
Tillegg	48-N	Mulige tilleggsproperties	Valgfri	Gir mulighet for å legge til ytterligere properties som f.eks. registerspesifikke tilleggsproperties, properties for kategoriseringsformål etc. Disse kan defineres etter hvert som ønsker og behov oppstår.

4.2.7 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 Tabell 7 er det en kolonne for «Gruppering». Den skal ikke inkluderes i importfilen og er lagt til i spesifikasjonen for å øke lesbarheten.

Tabell 7: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Status	Beskrivelse
Administrative formål	1	Code	Obligatorisk	Unik navnestreng for en variabel som også benyttes for å koble metadata med data. Tillatte tegn er store bokstaver og tall, 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	Obligatorisk	Forkortelse for kildenavn 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	Valgfri	Brukes dersom man ønsker å presentere variablene i flere kontekster, også kalt polyhierarkier.
Tittel og beskrivelse	4	PreferredTerm	Obligatorisk	Kort beskrivende navn på den enkelte variabel, f.eks. Kjønn. PreferredTerm på toppnoden skal ha navn lik «VARIABLEMETADATA» med store bokstaver.
HealthTerm	5	ConceptType	Valgfri	Skal være tom.
Administrative formål	6	PresentationOrder	Valgfri	Brukes dersom man ønsker å presentere variablene i en egendefinert rekkefølge som avviker fra en alfabetisk. Gyldige verdier er positive heltall og variablene presenteres i stigende rekkefølge basert på disse tallene.
	7	ValidFrom	Obligatorisk	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. Med «gyldig» menes at det finnes data tilgjengelig for variabelen. Nye avledede variabler kan være gyldig tilbake i tid. Format: YYYYMMDD

Tabell 7: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Status	Beskrivelse
	8	ValidTo	Valgfri	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	Valgfri	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	Valgfri	Pasientvennlig navn på variabelen.
	11	ClinicalFriendlyTerm	Valgfri	Helsefaglig begrep på variabelen.
	12	NavnEngelsk	Anbefalt	Engelsk navn på variabelen.
	13	Beskrivelse	Obligatorisk	Utførlig beskrivelse av variabelen. Se eksempler på helsedata.no.
	14	BeskrivelseEngelsk	Anbefalt	Utførlig beskrivelse av variabelen på engelsk.
Opphav	15	Opprinnelse	Obligatorisk	Om variabelen er hentet direkte fra helsetjenesten, er beregnet, avledet eller kodet. Dersom en variabel tilhører flere av verdiene i det standardiserte vokabularet så skal det siste steget i prosessen benyttes. Standardisert vokabular: 1 = Beregnet (F.eks. utregnet basert på en formel; Ventetid, liggetid etc.)

Tabell 7: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Status	Beskrivelse
				<p>2 = Avledet (F.eks. en kategorisering; Aldersgrupper, fylke, etc.)</p> <p>3 = Kodet (F.eks. strukturering av tekst basert på et kodeverk/terminologi)</p> <p>4 = Direkte fra fagsystem/EPJ (F.eks. hentet/gjenbrukt direkte fra NPR-melding, «urørt»)</p> <p>5 = Direkte fra skjema (F.eks. hentet/gjenbrukt direkte fra papirskjema eller elektronisk skjema-løsning, «urørt»)</p> <p>6 = Annet (F.eks. etablert i helsedatakilden for administrative formål.)</p>
	16	OpprinnelseUtledning	Valgfri	Hvordan beregnede, avledede og kodete variabler er utledet. Ikke aktuell for variabler hentet direkte fra helsetjenesten eller en elektronisk skjema-løsning
<i>Innhold og kvalitet</i>	17	Kommentar	Valgfri	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	Valgfri	Engelsk oversettelse av Kommentar.
	19	Kvalitetsnote	Anbefalt	Overordnet tekstlig beskrivelse av kvaliteten på datagrunnlaget som variabelen representerer, f.eks. komplettethet i form av fullstendighet (kodingskvalitet) og/eller kodekvalitet.
	20	KvalitetsnoteEngelsk	Anbefalt	Engelsk oversettelse av Kvalitetsnote.
	21	MaaleType	Valgfri	Hva variabelen måler. F.eks. tid, alder, vekt og puls.
	22	MaaleEnhet	Valgfri	Hvilken måleenhet variabelen er basert på. F.eks. år, mnd., kg, gram, slag/minutt.

Tabell 7: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Status	Beskrivelse
	23	DataType	Obligatorisk	<p>Datatype som spesifisert i FHIR Primitive Types (https://www.hl7.org/fhir/datatypes.html).</p> <p>Standardisert vokabular:</p> <ul style="list-style-type: none"> 1 = String 2 = Integer 3 = Decimal 4 = Boolean 5 = Time 6 = Date 7 = Datetime 8 = Uri 9 = Base64binary 10 = Fødselsnummer (11 siffer)
	24	Lengde	Valgfri	Maksimal lengde på variabelen i antall tegn/siffer Kan brukes for variabler med datatype 1,2 og 3 (String, Integer og Decimal)
	25	Presisjon	Valgfri	Maksimalt antall desimaler.
	26	GrenseLav	Valgfri	Nedre grenseverdi (minimumsverdi) Kan brukes for variabler med datatype 2 og 3 (Integer og Decimal) Indikerer at lavere verdier med stor sannsynlighet er feil, og bør utelates fra analyse.

Tabell 7: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Status	Beskrivelse
	27	GrenseHoy	Valgfri	Øvre grenseverdi (maksimumsverdi) Kan brukes for variabler med datatype 2 og 3 (Integer og Decimal) Indikerer at høyere verdier med stor sannsynlighet er feil, og bør utelates fra analyse.
<i>Juridisk informasjon</i>	28	Identifiseringsgrad	Valgfri	Dataforvalters klassifisering av bidrag til risiko for identifisering av individ. Nyttig informasjon for forsker når prosjekt må vurdere tiltak for dataminimering. Standardisert vokabular: 1 = Ikke vurdert 2 = Ingen 3 = Lav 4 = Middels 5 = Høy
<i>Skjema</i>	29	Ledetekst	Valgfri	Dette er teksten som står i spørreskjema eller innrapporteringsskjema ved hver variabel. Ledetekst er primært aktuell for variabler som er registrert via et skjema/en skjemaløsning.
	30	LedetekstEngelsk	Valgfri	Engelsk ledetekst for variabelen.
	31	SpoersmaalNummer	Valgfri	Spørsmålsnummer i spørreskjema eller innrapporteringsskjema.
	32	SkjemaVersjon	Valgfri	Vise hvilke(n) versjon(er) av spørreskjema eller innrapporteringsskjema denne variabelen finnes i.

Tabell 7: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Status	Beskrivelse
Administrative formål	33	AnbefaltForDigitalInnsyn	Obligatorisk*	<p>Anbefaling fra forvalter av datakilden til HDS om hvorvidt variabelen bør inngå i et digitalt, automatisert innsyn til den registrerte via helsenorge.no.</p> <p>Er det anbefalt at variabelen, gitt at denne inngår i et dataprodukt som skal overføres til Helseanalyseplattformen, inngår i et digitalt, automatisert innsyn?</p> <p>Standardisert vokabular:</p> <p>1 = Ja</p> <p>2 = Nei</p> <p>*Gjelder kun for variabler som skal inngå i dataprodukt som skal overføres til Helseanalyseplattformen.</p>
	34	TabellvennligNavn	Valgfri	Alternativ kolonneoverskrift for eksport eller utskrift av data.
	35	TabellvennligNavnEngelsk	Valgfri	Alternativ kolonneoverskrift for eksport eller utskrift av data for engelskspråklige.
	36	DatabaseReferanse	Obligatorisk	API-streng/URL/SourceID/TechName/unikt navn på klasse og/eller attributt i datakildens egen database. Denne referansen vil bli benyttet ved forespørsel om henting av variabelen fra kilden.
	37	Erstatter	Anbefalt	Code som er erstattet av denne variabelen.
	38	ErstattesAv	Valgfri	Code til variabelen som evt. erstatter denne variabelen.
	39	StartdatoHistorikk	Valgfri	Dato for start reell historikk. Hvis angitt er dette den første datoen man har historikk på metadata.

Tabell 7: Format for importfil 1b-Variabelmetadata				
Gruppering	Nr	Property	Status	Beskrivelse
				<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>
	40	Synlig	Obligatorisk	<p>Definerer om variabelen skal presenteres på helsedata.no. Variabler som ikke er synlige vil heller ikke kunne inngå i en variabelliste som legges ved søknad om tilgang til helsedata via felles søknadsskjema på helsedata.no.</p> <p>Standardisert vokabular:</p> <p>1 = Ja</p> <p>2 = Nei</p>
Tillegg	41-N	Mulige tilleggsproperties	Valgfri	<p>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.</p> <p>Skal også gi mulighet for generelle eller registerspesifikke filteringsmekanismer.</p>

4.2.8 Importfil 1c-Versjon

Importfil 1c-Versjon brukes til å identifisere hvilke variabler som inngår i en datasamling i en gitt periode og evt. revisjon. Den skal også inneholde opplysninger om når datagrunnlaget er registrert. En variabel vil kunne inngå i flere datasamlinger i forskjellige perioder. Hver versjon gis en unik *code*, og en angivelse av start på gyldighetsperiode. En variabels tilhørighet til en versjon defineres ved hjelp av *Mapsets*. Ved import av en ny versjon av en eksisterende datasamling, gis forrige (gjeldene) versjon en sluttdato av metadataforvaltningsansvarlige slik at det kun er en gjeldende versjon av hver datasamling.

Dataprodukter identifiseres ved hjelp av Versjon. Derfor må det registreres en ny versjon i forkant av leveranse av et dataprodukt for å identifisere hvilke variabler som inngår i dataproduktet, hvilken tidsperiode datagrunnlaget er registrert, og når denne versjonen er gyldig.

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

Tabell 8: Format for importfil 1c-Versjon			
Nr	Property	Status	Beskrivelse
1	Code	Obligatorisk	Unik navnestreng for en versjon som også brukes til å koble en versjon av metadatafiler til et datasett ved overføring av data til Helseanalyseplattformen. 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	Obligatorisk	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	Valgfri	Skal være tom for VERSJON
4	PreferredTerm	Obligatorisk	Kort beskrivende navn på Versjonen, f.eks. Somatikk. Toppnoden skal ha navn lik VERSJON.
5	ConceptType	Valgfri	Skal være tom for VERSJON
6	Presentation Order	Valgfri	Skal være tom for VERSJON
7	ValidFrom	Valgfri	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. ValidFrom for nye versjoner settes av metadatforvalter i Helsedataservice
8	ValidTo	Valgfri	Brukes for å angi at en versjon ble, eller vil bli, ugyldig fra en gitt dato. Versjoner skal ikke ha overlappende gyldighetsdatoer mellom revisjoner. ValidTo for nye versjoner settes av metadataforvalter i Helsedataservice

Tabell 8: Format for importfil 1c-Versjon			
Nr	Property	Status	Beskrivelse
9	Synonym	Valgfri	Skal være tom for VERSJON
10	PatientFriendlyTerm	Valgfri	Skal være tom for VERSJON
11	ClinicalFriendlyTerm	Valgfri	Skal være tom for VERSJON
12	Datasamling	Obligatorisk	Kortnavn på datasamlingen, som angitt i Code f.eks. SOMATIKK_EPISODE
13	Versjon	Obligatorisk	Versjon, som angitt i Code. Dette kan være årgang, frys-versjon, skjemaversjon eller liknende
14	Forvaltningstype	Valgfri	<p>Hvorvidt dette datasettet benyttes til utlevering av data, eller om det er en beskrivelse av de innrapporterte dataene, som for eksempel et spørreskjema. Standardverdi er 1.</p> <p>Standardisert vokabular:</p> <ul style="list-style-type: none"> 1 = Utlevering 2 = Innrapportering 3 = Innrapportering og utlevering
15	DatabaseReferanse	Valgfri	<p>API-streng, URL eller unikt navn på klasse, tabell eller view i datakildens database.</p> <p>Sammen med databasereferanse på variabelmetadata skal dette utgjøre eksakt referanse til dataelement i databasen.</p>
16	DataFra	Obligatorisk	<p>Dato for tidligste registrerte hendelse i datagrunnlaget.</p> <p>Som dato for hendelse regnes også dato for utfylling av spørreskjema, dato for prøvetaking, dato for intervju etc.</p> <p>Format YYYYMMDD.</p>

Tabell 8: Format for importfil 1c-Versjon			
Nr	Property	Status	Beskrivelse
17	DataTil	Anbefalt	Dato for siste registrerte hendelse i datagrunnlaget. Hvis data er overført til HAP, eller innsamling av data er avsluttet, skal denne være utfylt. Format: YYYYMMDD
18	IdentKolonne	Valgfri	Kolonner for «Primærnøkkel». En eller flere Code-verdier for variabler som identifiserer en forekomst i datagrunnlaget. Flere kolonner skilles med semikolon. Primærnøkkel skal kunne brukes for å endre/slette enkeltelementer i datagrunnlaget. Ved feil eller ved reservasjon/inndratt samtykke.
19	IdentType	Valgfri	Nøkkeltype. Standardisert vokabular: 1 = Unik 2 = Duplikat
20	FinnesPaaHAP	Obligatorisk	Overføres det data fra den aktuelle VERSJONEN av DATASAMLINGEN til Helseanalyseplattformen: Standardisert vokabular: 1 = Ja 2 = Nei
21-N	Mulige tilleggsperties	Valgfri	Gir mulighet for å legge til ytterligere properties. Disse kan defineres etter hvert som evt. ønsker og behov oppstår.

4.2.9 Importfil 1d-Statistikk

Importfil 1d-Statistikk brukes til å presentere enkel statistikk for et utvalg av datakildens variabler.

Statistikk som metadata har som formål å si noe om registerings-/ kodingskvaliteten og gi en pekepinn om antall forekomster. Statistikken egner seg ikke for analyse som sådan..

En egen *property*, *StatistikkType*, på Kildemetadata angir om statistikken er årsbasert eller akkumulert. Årsbasert statistikk rapporteres på grunnlag av hele avsluttede årganger. Alle akkumuleringer og beregninger er basert på ett kalenderår. Akkumulert statistikk brukes der grunnlagsdataene ikke har naturlig inndeling i år, for eksempel helseundersøkelser. Alle akkumuleringer og beregninger er da basert på alle innsamlede data totalt. Legg merke til at statistikk rapporteres pr. datasamling.

Statistikk rapporteres ulikt for kategoriske variabler og kontinuerlige variabler. For kategoriske variabler, gjøres det kun en opptelling innenfor hver kategori/kodeverdi samt antall variabler som mangler verdi. For kontinuerlige variabler, gjøres opptelling av antall rapporterte og antall manglende forekomster, samt enkle standardiserte statistiske beregninger. For funksjonene median og standardavvik skal bare gyldige verdier inngå i grunnlaget, og beregningen skjer ut fra ett år eller akkumulert i henhold til *StatistikkType* som nevnt over.

Det må gjøres en vurdering av hvilke variabler det er naturlig å rapportere statistikk for. Kontinuerlige variable må inneholde data som representerer én rad i datasamlingen.

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

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

Tabell 9: Format for importfil 1d-Statistikk

Nr	Property	Status	Beskrivelse
1	Code	Obligatorisk	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	Obligatorisk	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. For kilder med akkumulert statistikk, brukes årstall for det året statistikken er produsert.
3	Hierarchy	Valgfri	Skal være tom for STATISTIKK

Tabell 9: Format for importfil 1d-Statistikk

Nr	Property	Status	Beskrivelse
4	PreferredTerm	Obligatorisk	Kort beskrivende navn på den enkelte variabel, f.eks. Kjønn. «PreferredTerm» på toppnoden skal ha navn lik STATISTIKK med store bokstaver.
5	ConceptType	Valgfri	Skal være tom for STATISTIKK
6	PresentationOrder	Valgfri	Brukes dersom man ønsker å presentere statistikken på verdier/koder i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk. Gyldige verdier er positive heltall og statistikken presenteres i stigende rekkefølge basert på disse tallene.
7	ValidFrom	Obligatorisk	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	Valgfri	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. Årsbasert statistikk skal ikke ha overlappende gyldighetsdatoer. ValidTo kan brukes ved innlesing av statistikk for tidligere årganger av årsbasert statistikk. Denne skal ellers være tom. Settes av metadataforvalter hos Helsedataservice. Format: YYYYMMDD
9	Synonym	Valgfri	Skal være tom for STATISTIKK
10	PatientFriendlyTerm	Valgfri	Skal være tom for STATISTIKK
11	ClinicalFriendlyTerm	Valgfri	Skal være tom for STATISTIKK

Tabell 9: Format for importfil 1d-Statistikk			
Nr	Property	Status	Beskrivelse
12	Kommentar	Valgfri	En mer utførlig beskrivelse av statistikken ved behov.
13	SisteOppdaterte Aarssett	Obligatorisk	Året statistikken gjelder for. Brukes for enklere å kunne filtrere på årssett. Kompletterer «ValidFrom» og «ValidTo».
14	GyldigeTilfeller	Obligatorisk	Antall rapportert inn totalt
15	ManglendeTilfeller	Anbefalt	Antall som mangler innrapportering («NULL-forekomster i datagrunnlaget»)
16	Desimaler	Valgfri	Antall desimaler avrundet til
17	Prosentandel	Valgfri	Dersom beregnet i kilden
18	MIN	Valgfri	Minimum
19	MAX	Valgfri	Maksimum
20	AVG	Valgfri	Gjennomsnitt
21	STD	Valgfri	Standardavvik. Det er registerets populasjon/innsamlende data som avgjør om dette er standard avvik for populasjon eller innsamlede data.
22	MED	Valgfri	Median
23	KodeverkLokalID	Valgfri	Lokal ID for kodeverket og koden, f.eks. 1 eller 2 for hhv Mann og Kvinne.
24-N	Mulige tilleggsproperties	Valgfri	Gir mulighet for å legge til ytterligere properties ved behov

4.2.10 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.

Variabelgruppe erstatter den nå utgåtte proprieteten LokaltTema, men vil også kunne brukes til å gruppere variabler på tvers av helsedatakilder.

Tabell 10: Format for importfil 1e-Variabelgruppe			
Nr	Property	Status	Beskrivelse
1	Code	Obligatorisk	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	Obligatorisk	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	Valgfri	Brukes ikke for variabelgrupper.
4	PreferredTerm	Obligatorisk	Kort beskrivende navn på den enkelte variabelgruppe. PreferredTerm på toppnoden skal ha navn lik VARIABELGRUPPER med store bokstaver.
5	ConceptType	Valgfri	Skal være tom for VARIABELGRUPPE
6	PresentationOrder	Valgfri	Brukes dersom man ønsker å presentere variabelgruppene i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk. Gyldige verdier er positive heltall og variabelgruppene presenteres i stigende rekkefølge basert på disse tallene.
7	ValidFrom	Obligatorisk	Brukes for å angi fra når en variabelgruppe er gyldig. Format: YYYYMMDD
8	ValidTo	Valgfri	Skal være tom for VARIABELGRUPPE

Tabell 10: Format for importfil 1e-Variabelgruppe			
Nr	Property	Status	Beskrivelse
9	Synonym	Valgfri	Brukes dersom man ønsker å legge til synonymer til en variabelgruppe. Krever en ny rad i importfilen dersom mer enn ett synonym.
10	PatientFriendlyTerm	Valgfri	Pasientvennlig navn på variabelgruppen
11	ClinicalFriendlyTerm	Valgfri	Helsefaglig begrep på variabelgruppen
12	Beskrivelse	Obligatorisk	Utførlig beskrivelse av variabelgruppen.
13	NavnEngelsk	Anbefalt	Engelsk oversettelse av PreferredTerm
14	BeskrivelseEngelsk	Anbefalt	Engelsk oversettelse av beskrivelsen av variabelgruppen.
15	Filter	Valgfri	Definerer om variabelgruppen skal brukes til filtrering i Helsedata.no. Variabelgruppe er tenkt som en egen filtertype, og er ikke en del av klassen «Filter» Standardisert vokabular: 1 = Obligatorisk 2 = Valgfri
16-N	Mulige tilleggsperties	Valgfri	Gir mulighet for å legge til ytterligere properties ved behov

4.2.11 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 11: Format for importfil 1f-Instrument			
Nr	Property	Status	Beskrivelse
1	Code	Obligatorisk	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	Obligatorisk	Forkortelse for kildenavn 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	Valgfri	Skal være tom for INSTRUMENT
4	PreferredTerm	Obligatorisk	Kort beskrivende navn på det enkelte Instrumentet. PreferredTerm på toppnoden skal ha navn lik INSTRUMENT med store bokstaver.
5	ConceptType	Valgfri	Skal være tom for INSTRUMENT
6	PresentationOrder	Valgfri	Brukes dersom man ønsker å presentere instrumentene i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk. Gyldige verdier er positive heltall og instrumentene presenteres i stigende rekkefølge basert på disse tallene.
7	ValidFrom	Obligatorisk	Brukes for å angi fra når et instrument er gyldig. Settes ved første import. Format: YYYYMMDD
8	ValidTo	Valgfri	Skal være tom for INSTRUMENT
9	Synonym	Valgfri	Brukes dersom man ønsker å legge til synonymer til et instrumentnavn. Krever en ny rad i importfilen dersom mer enn ett synonym.

Tabell 11: Format for importfil 1f-Instrument			
Nr	Property	Status	Beskrivelse
10	PatientFriendlyTerm	Valgfri	Skal være tom for INSTRUMENT
11	ClinicalFriendlyTerm	Valgfri	Skal være tom for INSTRUMENT
12	Beskrivelse	Obligatorisk	Utførlig beskrivelse av instrumentet.
13	NavnEngelsk	Anbefalt	Engelsk oversettelse av PreferredTerm
14	BeskrivelseEngelsk	Anbefalt	Engelsk oversettelse av beskrivelsen av Instrumentet.
15	Filter	Valgfri	Definerer om instrumentet skal brukes til filtrering i Helsedata.no. Instrument er tenkt som en egen filtertype, og er ikke en del av klassen «Filter» Standardisert vokabular: 1 = Obligatorisk 2 = Valgfri
16	Dokumentasjon	Anbefalt	Lenke®®® til opprinnelig dokumentasjon/referanser for instrumentet.
17	UnikIdentifikator	Valgfri	Unik identifikator (DOI) til et instrument
18-N	Mulige tilleggsperties	Valgfri	Gir mulighet for å legge til ytterligere properties ved behov

4.2.12 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 Tabell 12.

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 samtlige datakilder, sortert alfabetisk i en flat liste.

Variabler som er basert på administrative kodeverk, helsefaglige kodeverk eller andre masterkodeverk som definert i Tabell 14 skal ikke inngå i Importfil 2-Kildekodeverk.

Tabell 12: Format for importfil 2-Kildekodeverk

Nr	Property	Status	Beskrivelse
1	Code	Obligatorisk	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_KILDE.KJONN_K.1
2	ParentCode	Obligatorisk	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	Valgfri	Brukes for å presentere kodeverkene i flere kontekster, også kalt polyhierarkier.
4	PreferredTerm	Obligatorisk	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. For koder som ikke har en tekst, for eksempel en skala-verdi fra 1 – 10, kan kodeverdien legges her.
5	ConceptType	Valgfri	Skal være tom for KILDEKODEVERK

Tabell 12: Format for importfil 2-Kildekodeverk

Nr	Property	Status	Beskrivelse
6	PresentationOrder	Valgfri	Brukes dersom man ønsker å presentere kodeverkene og/eller kodene i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk. Gyldige verdier er positive heltall og kildekodeverkene og kodene presenteres i stigende rekkefølge basert på disse tallene.
7	ValidFrom	Obligatorisk	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	Valgfri	Brukes for å angi at et kodeverk eller en kode ble, eller vil bli, ugyldig fra en gitt dato. For aktive kodeverk skal denne være tom. Format: YYYYMMDD
9	Synonym	Valgfri	Skal være tom for KILDEKODEVERK
10	PatientFriendlyTerm	Valgfri	Skal være tom for KILDEKODEVERK
11	ClinicalFriendlyTerm	Valgfri	Skal være tom for KILDEKODEVERK
12	Beskrivelse	Valgfri	Utførlig beskrivelse/definisjon av en kode
13	Kommentar	Anbefalt	Generelle kommentarer f.eks. til endringer

Tabell 12: Format for importfil 2-Kildekodeverk			
Nr	Property	Status	Beskrivelse
14	Eier	Valgfri	Hvilken institusjon som er ansvarlig for å forvalte kodeverket
15	MasterKodeverk	Anbefalt	Hvilket masterkodeverk som dette kodeverket er basert på, evt. er et subset av, f.eks. ICD-10 (ValueSet)
16	ReferanseTerm	Valgfri	Dersom termen som brukes i navnet på kodeverket og/eller koden kan forankres/defineres av en standardisert term (begrep)
17	KodeverkLokalID	Obligatorisk	Lokal kodeverdi for koden, f.eks. 1 eller 2 for hhv Mann og Kvinne.
18	NavnEngelsk	Anbefalt	PreferredTerm oversatt til Engelsk. Dvs. kodeteksten oversatt til engelsk
19	BeskrivelseEngelsk	Anbefalt	Beskrivelse oversatt til Engelsk
20	KommentarEngelsk	Anbefalt	Evt. Kommentar oversatt til Engelsk
21-N	Mulige tilleggsproperties	Valgfri	Gir mulighet for å legge til ytterligere properties ved behov

4.2.13 Importfil 3-Filter

Beskrivelse av filter detaljeres i en senere versjon.

4.2.14 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. felles klassifikasjoner, standarder og informasjonsmodeller som er importert, og som man har tilgang til i HealthTerm. Dette kan være nyttig for f.eks. tilrettelegging og videreutvikling av filtrering og søk på helsedata.no og visualisering av hvilken kontekst variablene er samlet inn i/nngår i.

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 13: 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

4.2.15 Administrative og helsefaglige kodeverk

Med administrative kodeverk menes her kodeverkene som per i dag finnes på volven.no. Disse omfatter blant annet kodeverkene som anvendes i NPR og KPR, og er tildelt en unik og persistent identifikator (OID).

Disse kodeverkene er også tilgjengelige i Nasjonal variabelkatalog, og skal refereres til ved egne mappingfiler. *Target code* i mappingfilen vil være samme OID som brukes i Volven. Eksempler på OID'er er 3101 for Kjønn og 3402 for kommunenummer.

I de tilfellene der en datakilde benytter seg av et av de helsefaglige kodeverkene i Tabell 11, så skal det mappes til respektive «Kortnavn» som definert i Tabell 14, og kodeverket skal ikke følge med i Importfil 2-Kildekodeverk.

Tabell 14: 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
NCMP-NCSP-NCRP	Norsk klinisk prosedyrekodeverk
ICPC-2	Den internasjonale klassifikasjonen for primærhelsetjenesten. Inkluderer ICPC-2B.
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 løsningsspesifikke begrep. Disse er forklart nærmere i tabellen under.

Tabell 15: Begrepsliste	
Begrep	Forklaring
CareCom	CareCom er teknisk leverandør av HealthTerm – løsningen Nasjonal variabelkatalog er realisert i.
Code System	Code System er et løsningsspesifikt begrep som anvendes i HealthTerm. Et 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.
Concept	Concept er et løsningsspesifikt begrep som anvendes i HealthTerm. Concept er de meningsbærende dataobjektene i HealthTerm og brukes til å modellere variabler, datakilder, datasamlinger osv.
Concept type	Concept type er et løsningsspesifikt begrep som anvendes i HealthTerm. 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.
Core Namespace	Core namespace er et løsningsspesifikt begrep som anvendes i HealthTerm. Core Namespace er hovedmappen i et Code System. For de nasjonale helsedatakildene lagres kildemetadata om variablene som Core Namespace.
Dataansvarlig	Den som bestemmer formålet med behandlingen av helseopplysningene og hvilke hjelpemidler som skal brukes, og den som i eller i medhold av lov er pålagt et databehandlingsansvar. Kilde: Helseregisterloven § 2 bokstav e.

Tabell 15: Begrepsliste	
Begrep	Forklaring
Databehandler	Den som behandler personopplysninger på vegne av den [data]behandlingsansvarlige. Kilde: Personopplysningsloven § § 2 nr.5
Datakilde	Datakilde er definert som kildene til dataene. I denne versjonen av spesifikasjonen er en datakilde ekvivalent med et helseregister, et nasjonalt medisinsk kvalitetsregister eller en befolkningsbasert helseundersøkelse.
Dataprodukt	Dataprodukt er en konseptualisering av måten data skal overføres fra nasjonale helsedatakilder til Helseanalyseplattformen. Konseptet dataprodukt består av et datasett, samt metadata om variabler, versjon og datasamling. Sammenhengen mellom metadata og datasett som overføres til Helseanalyseplattformen er forklart nærmere i <i>Spesifikasjon for dataoverføring til Helseanalyseplattformen</i> .
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.
Dekningsgrad	Dekningsgraden angir i hvilken grad de pasienter eller hendelser som skal registreres faktisk blir registrert. Dekningsgraden for et enkeltsykehus er et mål på hvor mange av de aktuelle pasientene som behandles ved sykehuset som blir registrert i kvalitetsregisteret, mens den samlede dekningsgraden for registeret er den aggregerte dekningsgraden for alle sykehusene som behandler pasienter i registerets målgruppe. Man kan for eksempel si at registeret har en samlet dekningsgrad på 80% og at spesifikt sykehus har en dekningsgrad på 85%. I vår sammenheng er det kun den aggregerte dekningsgraden for de nasjonale kvalitetsregistrene som skal rapportere inn (DekningsgradAggregert).
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. Denne kan ha flere nivåer. For datakilder som skal levere dataprodukt til Helseanalyseplattformen, kan klassen Delkilde også brukes for å lage logiske grupperinger av flere dataprodukt.

Tabell 15: Begrepsliste	
Begrep	Forklaring
Extension Namespace	Extension Namespace er et løsningsspesifikt begrep som anvendes i HealthTerm. Extension Namespace (eller bare Extension) er tilleggsmapper i HealthTerm der annen informasjon lagres. Variabelmetadata, statistikk, versjon, instrument og variabelgruppe lagres i Extension Namespaces under respektive Core Namespaces
Filter	Filter er en kategorisering av 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.
Informasjonsmodell	En informasjonsmodell er en formalisert beskrivelse av et interesseområde f.eks. i form av modellelementer (klasserer) med tilhørende egenskaper (properties), og relasjoner mellom disse. Den formelle beskrivelsen av en informasjonsmodell er «en beskrivelse av informasjonen en virksomhet trenger å motta eller selv produsere for å utføre sitt daglige virke». Kilde: Skatteetaten gjennom Difi: https://www.difi.no/artikkel/2016/09/begreperfellesinformasjonsmodeller
Instrument	Instrument 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.
Kategorisk variabel	En kategorisk variabel er en variabel som kan ta på seg en av et begrenset, og vanligvis fast, antall mulige verdier, f.eks. en kodeverdi fra et kodeverk (Diagnose, kjønn, kommune etc.). Den består altså ikke av tallverdier, men av f.eks. kodeverdier eller kategoriseringer (klassifiseringer).
Kontinuerlig variabel	En kontinuerlig variabel kan anta alle verdier, ofte innenfor et bestemt intervall, f.eks. høyde, vekt, inntekt, osv.
Klasse	En klasse brukes til å klassifisere og beskrive et sett av objekter f.eks. variabler med tilhørende egenskaper (properties)
Kode	Kode er i denne sammenheng det samme som svaralternativene i et verdsett. Består av en kodeverdi (KodeverkLokalID) og en kodetekst (Kodebeskrivelsetekst), for eksempel «2» og «Kvinne».
Kodefrekvens	Kodefrekvens er definert som antall gyldige tilfeller av et svaralternativ/utfall av en variabel.

Tabell 15: Begrepsliste	
Begrep	Forklaring
Kodeverdi	Kodeverdien er koden som står foran en kodetekst, f.eks. 1 for Mann. Kodeverdier kan være alfanumeriske og inneholde både mellomrom og forskjellige tegn. Dersom man skal koble en kodeverdi med kodetekst er det derfor viktig at formatene er «en-til-en».
Kodeverk	Kodeverk favner i denne sammenheng bredere enn det man i helsevesenet normalt definerer som et kodeverk. Et kodeverk kan i denne sammenheng både representere en variabels verdsett/svaralternativ eller et standardisert (kontrollert) vokabular for å beskrive en properties egenskaper. Den formelle definisjonen av et kodeverk er «en samling av begreper eller en klassifikasjon hvor hver tilhørende begrepsdefinisjon er tilknyttet en unik kode innenfor samlingen eller klassifikasjonen».
Kodetekst	Kodeteksten er teksten som etterfølger en kodeverdi, f.eks. «Mann» for kodeverdien 1.
Mapping	Mapping brukes for å håndtere ikke-hierarkiske relasjoner i HealthTerm, eksempelvis referere til kodeverk. Mer formelt kan man si at mapping er en funksjon som knytter dataobjekter (concepts) i et gitt sett av unike dataobjekter (source) til et annet sett av unike dataobjekter (target).
Lovbestemte helseregistre	Lovbestemte helseregistre (§11) behandler direkte personidentifiserbare helseopplysninger uten krav til samtykke. De er etablert av den sentrale helseforvaltningen for å ivareta landsomfattende funksjoner og er meldepliktige. 10 av de sentrale helseregistrene er lovbestemte. De resterende 8 registrene er samtykkebaserte eller uten direkte personidentifiserende kjennetegn (§9a og 9b)
Nasjonal variabelkatalog	Den nasjonale variabelkatalogen er navnet på løsningen der innrapporterte metadata fra helsedatakildene forvaltes og tilgjengeliggjøres for helsedata.no og andre metadatakataloger i inn og utland.
Objektstruktur	En objekstruktur er en representasjon av hvordan klassene i en informasjonsmodell er modellert i henhold til hverandre og hvilke relasjoner som f.eks. finnes mellom klassene i informasjonsmodellen.

Tabell 15: Begrepsliste	
Begrep	Forklaring
Polyhierarki	Polyhierarki betyr at et begrep kan plasseres i flere hierarkier, f.eks. kan et helseregister både plasseres i et organisasjonshierarki, et hierarki for type registre og et hierarki der registeret er kategorisert etter hvilken pasientgruppe det favner.
Property	Properties er det samme som attributter som brukes til å beskrive egenskapene til dataobjekt (concept), f.eks. en variabel.
Retired	Retired brukes, i denne sammenheng, til å beskrive at et concept, eller en property, ikke lenger er aktive i løsningen. En gammel versjon av en property settes eksempelvis til retired når man redigerer navnet på den.
Rike metadata	Rike metadata er ekvivalent med utfyllende metadata – metadata som er tilstrekkelige for å utfylle et definert formål. Rike metadata må si noe om blant annet hvor man kan henvendes seg for å søke om å få tilgang til data, tilgangen til data i seg selv, restriksjoner som er lagt på datagrunnlaget, datakvaliteten, kodeverk og terminologi som er anvendt, kontekst f.eks. i form av en begreps-/informasjonsmodell etc.
Sentrale helseregistre	<p>De sentrale helseregistrene er etablert for å ivareta landsomfattende oppgaver. De brukes først og fremst til helseanalyser, statistikk, kvalitetsforbedring av helsetjenester, forskning, administrasjon, styring og beredskap.</p> <p>Registrene er opprettet med hjemmel i helseregisterloven og forskrifter. Registre med direkte personidentifiserbare opplysninger som ikke er basert på samtykke, er opprettet etter behandling i Stortinget. Det finnes også sentrale helseregistre som ikke inneholder personidentifiserbare opplysninger, eller bare indirekte personidentifiserbare opplysninger. Registrene forvaltes av ulike virksomheter i den sentrale helseforvaltningen.</p> <p>https://www.fhi.no/div/datatilgang/om-sentrale-helseregistre/</p>
Standardisert vokabular	<p>Et standardisert (kontrollert) vokabular er en liste med forhåndsdefinerte svaralternativ. Standardiserte vokabularer bidrar blant annet til enklere validering av datakvalitet, bedre semantisk interoperabilitet og enklere tilrettelegging av data for analyse, og består vanligvis av en kodeverdi og kodetekst.</p> <p>Et synonymt begrep er «kontrollert vokabular». Ref. https://data.norge.no/specification/dcat-ap-no/#Kontrollerte-vokabularer</p>

Tabell 15: Begrepsliste	
Begrep	Forklaring
Statistikk	Statistikk i denne sammenheng, er å betrakte som metadata som sier noe om volum, kodingskvalitet og datakvalitet generelt på variabelnivå. Statistikken som innrapporteres som metadata egner seg ikke for, og er heller ikke ment for, analyseformål.
Variabel	Variabler er de informasjonsbærende elementene i datakilden og definerer hvilken informasjon som finnes i den aktuelle helsedatakilden.
Variabelgruppe	Variabelgruppe er en kategorisering av variabler. Strukturen kan bygges opp hierarkisk. Inntil videre begrenset til 2 nivåer.
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.

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 *propertyene* i denne spesifikasjonen og DCAT.

Vedlegg 1a: Mapping KILDEMETADATA til DCAT-AP-NO						
Gruppering	Nr	Property	Status	Beskrivelse	DCAT-AP-NO	DCAT-EN (DQV-EN)
Administrative formål	1	Code	Obligatorisk	<p>Unik navnestreng. Tillatte tegn er store bokstaver og tall, uten mellomrom og æ, ø, å.</p> <p>Datakilde: Prefiks K for kilde, underscore, kortnavn på datakilden. F.eks. K_HUNT.</p> <p>Delkilde: Prefiks K for kilde, underscore, kortnavn på datakilden, punktum, og navn på delkilde. F.eks. K_HUNT.HUNT1</p> <p>Datasamling: Prefiks K for kilde, underscore, kortnavn på datakilden, punktum, og navn på datasamling. F.eks. K_HUNT.HUNT1_QUESTIONAIRE1</p>	Datasett: identifikator *	dct:identifiser

Nasjonal spesifikasjon for metadata om helsedata

	2	ParentCode	Obligatorisk	<p>Datakilde: Skal være lik kortnavnet på datakilden, f.eks. HUNT.</p> <p>Delkilde: Skal være lik Code for tilhørende datakilde, f.eks. K_HUNT.</p> <p>Datasamling: Skal være lik Code for tilhørende datakilde. Hvis datasamlingen tilhører en delkilde, skal delkildens Code benyttes, f.eks. K_HUNT.HUNT1.</p>	Datasett: er del av	dct:isPartOf
HealthTerm	3	Hierarchy	Valgfri	Skal være tom for KILDEMETADATA	E-helse:Extension	
Tittel og beskrivelse	4	PreferredTerm	Obligatorisk	<p>Kort beskrivende navn på datakilden, delkilden eller datasamlingen.</p> <p>F.eks. «««Den norske mor, far og barn undersøkelsen»»», «««MOBA Hovedskjema»»» eller «««Spørreskjema1»»» for henholdsvis datakilde, delkilde og datasamling.</p>	Datasett: tittel (NB) *	dct:title (NB)
HealthTerm	5	ConceptType	Valgfri	Skal være tom For KILDEMETADATA	E-helse:Extension	
Administrative formål	6	PresentationOrder	Valgfri	<p>Datakilde: Skal være tom.</p> <p>Delkilde, datasamling: Brukes dersom man ønsker å presentere delkildene eller datasamlingene i en egendefinert rekkefølge som avviker fra alfabetisk. Gyldige verdier er positive heltall og delkildene eller datasamlingene presenteres i stigende rekkefølge basert på disse tallene.</p>	E-helse:Extension	

Nasjonal spesifikasjon for metadata om helsedata

	7	ValidFrom	Obligatorisk	Brukes for å angi fra når en datakilde, delkilde eller datasamling er gyldig. Settes ved første import. Format: YYYYMMDD	E-helse:Extension	
	8	ValidTo	Valgfri	Skal være tom for KILDEMETADATA	E-helse:Extension	
Tittel og beskrivelse	9	Synonym	Valgfri	Skal være tom for KILDEMETADATA (inntil videre)	E-helse:Extension	
	10	PatientFriendlyTerm	Valgfri	Skal være tom for KILDEMETADATA (inntil videre)	E-helse:Extension	
	11	ClinicalFriendlyTerm	Valgfri	Skal være tom for KILDEMETADATA (inntil videre)	E-helse:Extension	
	12	KortNavn	Valgfri	Forkortelse av navnet på datakilden, delkilden eller datasamlingen. F.eks. MOBA eller NPR.	(Datasett: tittel)	(dct:title)
	13	NavnEngelsk	Anbefales	Engelsk navn på datakilden, delkilden eller datasamlingen.	Datasett: tittel (EN) *	dct:title (EN) *
	14	Beskrivelse	Obligatorisk	En utførlig beskrivelse. Skal være forståelig også for innbyggere. Se eksempler på helsedata.no	Datasett: beskrivelse (NB) *	dct:description (NB) *
	15	BeskrivelseEngelsk	Anbefales	En utførlig beskrivelse på engelsk. Skal være forståelig også for innbyggere.	Datasett: beskrivelse (EN) *	dct:description (EN) *

Nasjonal spesifikasjon for metadata om helsedata

	16	AnbefalteBruk somraader	Obligatorisk	En beskrivelse av hvilke typer formål (eksempelvis forskning, analyse, styringsdata, beredskap osv) og faglige temaer (eksempelvis kreft, hjerte og kar, kirurgi osv), en datakilde, delkilde og/eller datasamling er egent for.	E-helse:Extension	
	17	Emneord	Valgfri	Beskrivende emneord (tag). Skilles ved å bruke «;». F.eks. Emneord;Nytt emneord;	Datasett: emneord	dcat:keyword
Opphav og ferskhet	18	Opprettet	Obligatorisk	Året datakilden, delkilden eller datasamlingen ble opprettet og begynte å samle inn data. Format: YYYY.	E-helse:Extension	
	19	SistOppdatert	Obligatorisk	Dato for siste oppdatering av KILDEMETADATA Format: YYYYMMDD	Datasett: endringsdato	dct:modified
	20	Frekvens	Valgfri	Datakilde, delkilde: Skal være tom. Datasamling: Oppdateringsfrekvens brukes for å beskrive hvor ofte datagrunnlaget oppdateres på Helseanalyseplattformen. Standardisert vokabular: 1 = Sjeldnere enn årlig 2 = Årlig 3 = Halvårlig 4 = Tertialvis 5 = Kvartalsvis 6 = Månedlig 7 = Ukentlig	Datasett: frekvens *	dct:accrualPeriodicity *

				8 = Daglig 9 = Hyppigere enn daglig		
	21	Innsamlingsmetode	Valgfri	Kort beskrivelse av metode for innsamling av data.	Datsett: ble generert ved	prov:wasGeneratedBy
Innhold og kvalitet	22	Kildetype	Obligatorisk	Kategorisering av datakilden. Standardisert vokabular: 1 = Sentralt helseregister ³ 2 = Nasjonalt medisinsk kvalitetsregister 3 = Annet medisinsk kvalitetsregister 4 = Befolkningsbasert helseundersøkelse 5 = Biobank 6 = Annen datakilde	Aktør: utgivertype	foaf:name
	23	InklusjonsOgEksklusjonskriterier	Obligatorisk	En faglig overordnet beskrivelse av hvilke kriterier som stilles for at en person skal være registrert i datakilden, delkilden eller datasamlingen. Dette tilsvarer en beskrivelse av populasjonen som inngår i datakilden, delkilden eller datasamlingen.	E-helse:Extension	
	24	GeografiskAvgrensning	Valgfri	Beskrivelse av opptaksområdet for innbyggerne som inngår i datakilden, delkilden eller datasamlingen.	Datsett: dekningsområde *	dct:spatial *
	25	Tilslutningsgrad	Valgfri	Andel enheter av totalt antall relevante enheter som rapporterer inn til datakilden, delkilden eller datasamlingen. Med enheter menes foretak, organisasjoner og virksomheter. Oppgis som desimaltall	Kompletthet (completeness)?	dqv:dimension

³ Inkluderer lovbestemte helseregistre og andre sentrale helseregistre (<https://www.fhi.no/div/datatilgang/om-sentrale-helseregistre/>)

				på formatet «0,875». (Kompletthet i form av dekningsgrad på institusjons- eller enhetsnivå)		
26	BeregningTilslutningsgrad	Valgfri		Hvordan tilslutningsgraden er beregnet. F.eks. hva er nevneren i brøken?	Kompletthet (completeness)?	dqv:dimension
27	Dekningsgrad	Obligatorisk		<p>NB! Kun for Nasjonale kvalitetsregistre inntil videre.</p> <p>Dekningsgraden angir i hvilken grad de pasienter eller hendelser som skal registreres faktisk blir registrert. Dekningsgraden for et enkeltsykehus er et mål på hvor mange av de aktuelle pasientene som behandles ved sykehuset som blir registrert i kvalitetsregisteret, mens den samlede dekningsgraden for registeret er den aggregerte dekningsgraden for alle sykehusene som behandler pasienter i registerets målgruppe.</p> <p>Man kan for eksempel si at registeret har en samlet dekningsgrad på 80% og at spesifikt sykehus har en dekningsgrad på 85%.</p> <p>Andre helsedatkilder bruker proprietæren «Kvalitetsnote» for å beskrive kvaliteten på datagrunnlaget.</p> <p>Oppgis som desimaltall på formatet «0,875».</p>	Kompletthet (completeness)?	dqv:dimension
28	BeregningDekningsgrad	Valgfri		Hvordan dekningsgrad er beregnet. F.eks. sammenstilt med data fra NPR, og evt. fomel.	Kompletthet (completeness)?	dqv:dimension
29	Kvalitetsnote	Anbefales		Overordnet tekstlig beskrivelse av kvaliteten på datagrunnlaget.	Datasett: Kvalitetsnote (*)	dqv:qualityAnnotation

Nasjonal spesifikasjon for metadata om helsedata

Rettsgrunnlag	30	Formaal	Obligatorisk	Kort beskrivelse av formålet datakilden er opprettet for. Dersom mulig, en kopi av formålet slik det står skrevet i forskrift.	Datasett: Følger (dcatno:objective? Finnes kun i veileder)	cpsv:follows
	31	Lovverk	Obligatorisk	En eller flere URL-er til lovverk på lovdata.no. F.eks. URL til Lov om helseregistre og behandling av helseopplysninger. Flere URL-er skiller ved å bruke «;».	Datasett: Følger	cpsv:follows
	32	Forskrift	Valgfri	En eller flere URL-er til forskrift på lovdata.no. F.eks. URL til Forskrift om innsamling og behandling av helseopplysninger i Medisinsk fødselsregister. Flere URL-er skiller ved å bruke «;».	Datasett: Følger	cpsv:follows
	33	HjemmelTilgjengeliggjøring	Valgfri	URL til paragraf med hjemmel for tilgjengeliggjøring.	Datasett: Følger	cpsv:follows
	34	Dataansvarlig	Obligatorisk	Navn på virksomheten eller enheten som eier og har det juridiske dataansvaret for datakilden, delkilden eller datasamlingen.	Datasett: kontaktpunkt	dcat:contactPoint
	35	Databehandler	Obligatorisk	Navn på virksomheten eller enheten som faktisk behandler datakilden. Er det samme som Dataansvarlig med mindre denne har satt ut databehandlingen til ekstern tredjepart.	Datasett: kontaktpunkt	dcat:contactPoint
	36	GradAvPersonidentifikasjon	Valgfri	Høyeste grad av identifisering det er mulig å søke om fra denne datakilden, delkilden eller datasamlingen. Standardisert vokabular: 1 = Direkte personidentifiserbare opplysninger 2 = Indirekte personidentifiserbare opplysninger	E-helse: Extension	

Nasjonal spesifikasjon for metadata om helsedata

				3 = Anonyme opplysninger		
	37	JuridiskNote	Valgfri	Relevant juridisk informasjon som ikke dekkes av andre properties. F.eks. informasjon om reservasjonsrett eller hvilke andre kilder data kan sammenstilles med.	Datasett: Følger	cpsv:follows
Kontaktinformasjon	38	Hjemmeside	Obligatorisk	URL til hjemmeside.	Datasett: kontaktpunkt	dcat:contactPoint
	39	Epost	Obligatorisk	E-postadresse for henvendelser om datakilden, delkilden eller datasamlingen.	Datasett: kontaktpunkt *	dcat:contactPoint *
	40	Telefonnummer	Valgfri	Telefonnummer for henvendelser om datakilden, delkilden eller datasamlingen.	Datasett: kontaktpunkt *	dcat:contactPoint *
	41	OrgnrDataansvarlig	Valgfri	Organisasjonsnummer til dataansvarlig.	Datasett: kontaktpunkt	dcat:contactPoint
	42	HERidDatakilde	Valgfri	Datakilde: Unik identifikator som tildeles av NHN, Brukes for identifisering av datakilden innenfor helsenettet. Kan brukes til spørringer. Delkilde, datasamling: Skal være tom.	E-helse: Extension	dcat:contactPoint
	43	HERidDataansvarlig	Valgfri	Unik identifikator som tildeles av NHN, Brukes for identifisering av datakilden innenfor helsenettet. Kan brukes til spørringer.	E-helse: Extension	
Administrative formål	44	Kildemetadatatype	Obligatorisk	Angir hvilken type kildemetadata som beskrives. Standardisert vokabular: 1 = Datakilde 2 = Delkilde	E-helse: Extension	

Nasjonal spesifikasjon for metadata om helsedata

				3 = Datasamling		
	45	EriSamsvarMed	Valgfri	Referanse til evt. standard(er), spesifikasjon(er) etc. som datakilden, delkilden eller datasamlingen er i henhold til, f.eks. DCAT-AP-NO. Skilles ved å bruke «;». F.eks. Standard 1;Standard 2;	Datsett: i samsvar med	dqv:conformsTo
	46	StatistikkType	Valgfri	Hvorvidt medfølgende statistikk er akkumulert eller årsbasert. Akkumulert vil eksempelvis kunne være over flere år slik som for helseundersøkelser, mens årsbasert vil være innenfor et definert år slik som for DÅR. Standardisert vokabular: 1 = Akkumulert 2 = Årsbasert	E-helse: Extension	
	47	TelleEnhet	Valgfri	Datasamling: Telleenhet for datasamlingen angitt ved en tekstlig beskrivelse, eks. pasient, opphold. Kilde, Delkilde: Skal være tom	E-helse: Extension	
Tillegg	48-N	Mulige tilleggsproperties		Gir mulighet for å legge til ytterligere properties som f.eks. registerspesifikke tilleggsproperties, properties for kategoriseringsformål etc. Disse kan defineres etter hvert som ønsker og behov oppstår.		

*Overføres til [Felles datakatalog, https://data.norge.no/](https://data.norge.no/)

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 *propertyene* i denne spesifikasjonen og SKOS.

Vedlegg 1b: Mapping VARIABELMETADATA til SKOS-AP-NO						
Gruppering	Nr	Property	Status	Beskrivelse	SKOS-AP-NO	SKOS Engelsk
Administrative formål	1	Code	Obligatorisk	Unik navnestreng for en variabel som også benyttes for å koble metadata med data. Tillatte tegn er store bokstaver og tall, 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.	Begrep: identifikator	dct:identifiser
	2	ParentCode	Obligatorisk	Forkortelse for kildenavn med prefiks V for Variabel og underscore, f.eks. V_DAR. «««ParentCode»»» for toppnoden skal være lik kortnavnet på datakilden, f.eks. DAR.	E-helse: Extension	

Vedlegg 1b: Mapping VARIABELMETADATA til SKOS-AP-NO						
Gruppering	Nr	Property	Status	Beskrivelse	SKOS-AP-NO	SKOS Engelsk
HealthTerm	3	Hierarchy	Valgfri	Brukes dersom man ønsker å presentere variablene i flere kontekster, også kalt polyhierarkier.	E-helse: Extension	
Tittel og beskrivelse	4	PreferredTerm	Obligatorisk	Kort beskrivende navn på den enkelte variabel, f.eks. Kjønn. PreferredTerm på toppnoden skal ha navn lik «««VARIABELMETADATA»»» med store bokstaver.	Begrep: anbefaltTerm (NB)	skosxl:prefLabel (NB)
HealthTerm	5	ConceptType	Valgfri	Skal være tom.	E-helse: Extension	
Administrative formål	6	PresentationOrder	Valgfri	Brukes dersom man ønsker å presentere variablene i en egendefinert rekkefølge som avviker fra en alfabetisk. Gyldige verdier er positive heltall og variablene presenteres i stigende rekkefølge basert på disse tallene.	E-helse: Extension	
	7	ValidFrom	Obligatorisk	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. Med «gyldig» menes at det finnes data tilgjengelig for variabelen. Nye avledede variabler kan være gyldig tilbake i tid. Format: YYYYMMDD	Begrep: gyldighetsperiode.gyldigFra OgMed	dct:temporal
	8	ValidTo	Valgfri	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	Begrep: gyldighetsperiode.gyldigTilOgMed	dct:temporal

Vedlegg 1b: Mapping VARIABELMETADATA til SKOS-AP-NO						
Gruppering	Nr	Property	Status	Beskrivelse	SKOS-AP-NO	SKOS Engelsk
Tittel og beskrivelse	9	Synonym	Valgfri	Brukes dersom man ønsker å legge til synonymer til et variabelnavn. Krever en ny rad i importfilen dersom mer enn ett synonym legges til.	Begrep: tillattTerm	skosxl:prefLabel
	10	PatientFriendlyTerm	Valgfri	Pasientvennlig navn på variabelen.	Begrep: tillattTerm	skosxl:prefLabel
	11	ClinicalFriendlyTerm	Valgfri	Helsefaglig begrep på variabelen.	Begrep: tillattTerm	skosxl:prefLabel
	12	NavnEngelsk	Anbefalt	Engelsk navn på variabelen.	Begrep: anbefaltTerm (EN)	skosxl:prefLabel (EN)
	13	Beskrivelse	Obligatorisk	Utførlig beskrivelse av variabelen. Se eksempler på helsedata.no.	Begrep: definisjon (NB)	skosno:definisjon (NB)
	14	BeskrivelseEngelsk	Anbefalt	Utførlig beskrivelse av variabelen på engelsk.	Begrep: definisjon (EN)	skosno:definisjon (EN)
Opphav	15	Opprinnelse	Obligatorisk	<p>Om variabelen er hentet direkte fra helsetjenesten, er beregnet, avledet eller kodet. Dersom en variabel tilhører flere av verdiene i det standardiserte vokabularet så skal det siste steget i prosessen benyttes.</p> <p>Standardisert vokabular:</p> <p>1 = Beregnet (F.eks. utregnet basert på en formel; Ventetid, liggetid etc)</p> <p>2 = Avledet (F.eks. en kategorisering; Aldersgrupper, fylke, etc.)</p> <p>3 = Kodet (F.eks. strukturering av tekst basert på et kodeverk/terminologi)</p>	E-helse: Extension	

Vedlegg 1b: Mapping VARIABELMETADATA til SKOS-AP-NO						
Gruppering	Nr	Property	Status	Beskrivelse	SKOS-AP-NO	SKOS Engelsk
				<p>4 = Direkte fra fagsystem/EPJ (F.eks. hentet/gjenbrukt direkte fra NPR-melding, «urørt»)</p> <p>5 = Direkte fra skjema (F.eks. hentet/gjenbrukt direkte fra papirskjema eller elektronisk skjema-løsning, «urørt»)</p> <p>6 = Annet (F.eks. etablert i helsedatakilden for administrative formål.)</p>		
	16	OpprinnelseUtl edning	Valgfri	Hvordan beregnede, avledede og kodete variabler er utledet. Ikke aktuell for variabler hentet direkte fra helsetjenesten eller en elektronisk skjema-løsning	E-helse: Extension	
Innhold og kvalitet	17	Kommentar	Valgfri	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.	E-helse: Extension (NB)	
	18	KommentarEn gelsk	Valgfri	Engelsk oversettelse av Kommentar.	E-helse: Extension (EN)	
	19	Kvalitetsnote	Anbefalt	Overordnet tekstlig beskrivelse av kvaliteten på datagrunnlaget som variabelen representerer, f.eks. kompletthet i form av fullstendighet (kodingskvalitet) og/eller kodekvalitet.	(Datasett: Kvalitetsnote (NB))	
	20	KvalitetsnoteE ngelsk	Anbefalt	Engelsk oversettelse av Kvalitetsnote.	(Datasett: Kvalitetsnote (EN))	
	21	MaaleType	Valgfri	Hva variabelen måler. F.eks. tid, alder, vekt og puls.	E-helse: Extension	

Vedlegg 1b: Mapping VARIABELMETADATA til SKOS-AP-NO						
Gruppering	Nr	Property	Status	Beskrivelse	SKOS-AP-NO	SKOS Engelsk
	22	MaaleEnhet	Valgfri	Hvilken måleenhet variabelen er basert på. F.eks. år, mnd., kg, gram, slag/minutt.	E-helse: Extension	
	23	DataType	Obligatorisk	Datatype som spesifisert i FHIR Primitive Types (https://www.hl7.org/fhir/datatypes.html). Standardisert vokabular: 1 = String 2 = Integer 3 = Decimal 4 = Boolean 5 = Time 6 = Date 7 = Datetime 8 = Uri 9 = Base64binary	E-helse: Extension	
	24	Lengde	Valgfri	Maksimal lengde på variabelen i antall tegn/siffer Kan brukes for variabler med datatype 1,2 og 3 (String, Integer og Decimal)	E-helse: Extension	
	25	Presisjon	Valgfri	Maksimalt antall desimaler.	E-helse: Extension	
	26	GrenseLav	Valgfri	Nedre grenseverdi (minimumsverdi) Kan brukes for variabler med datatype 2 og 3 (Integer og Decimal)	E-helse: Extension	

Vedlegg 1b: Mapping VARIABELMETADATA til SKOS-AP-NO						
Gruppering	Nr	Property	Status	Beskrivelse	SKOS-AP-NO	SKOS Engelsk
				Indikerer at lavere verdier med stor sannsynlighet er feil, og bør utelates fra analyse.		
	27	GrenseHoy	Valgfri	Øvre grenseverdi (maksimumsverdi) Kan brukes for variabler med datatype 2 og 3 (Integer og Decimal) Indikerer at høyere verdier med stor sannsynlighet er feil, og bør utelates fra analyse.	E-helse: Extension	
Juridisk informasjon	28	Identifiserings grad	Valgfri	Dataforvalters klassifisering av bidrag til risiko for identifisering av individ. Nyttig informasjon for forsker når prosjekt må vurdere tiltak for dataminimering. Standardisert vokabular: 1 = Ikke vurdert 2 = Ingen 3 = Lav 4 = Middels 5 = Høy	E-helse: Extension	
Skjema	29	Ledetekst	Valgfri	Dette er teksten som står i spørreskjema eller innrapporteringsskjema ved hver variabel. Ledetekst er primært aktuell for variabler som er registrert via et skjema/en skjema-løsning.	E-helse: Extension	
	30	LedetekstEngelsk	Valgfri	Engelsk ledetekst for variabelen.	E-helse: Extension	
	31	SpoersmaalNummer	Valgfri	Spørsmålsnummer i spørreskjema eller innrapporteringsskjema.	E-helse: Extension	

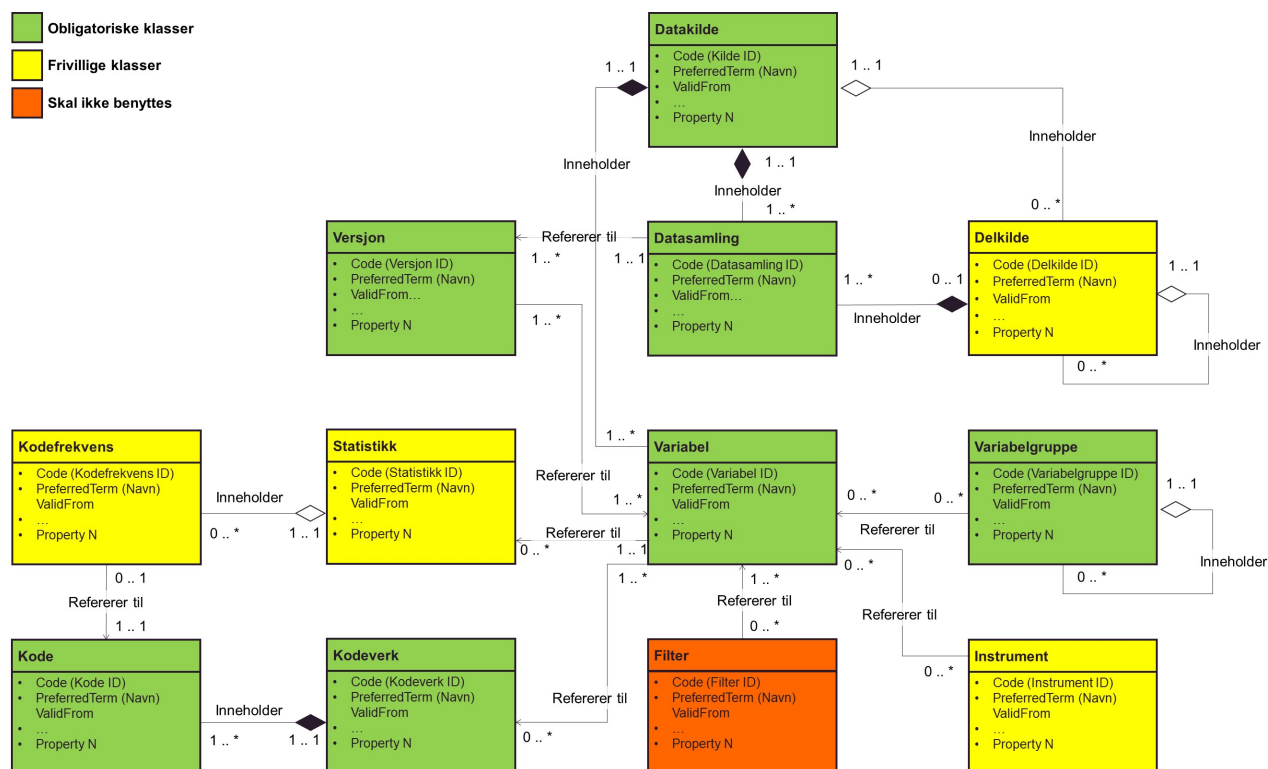
Vedlegg 1b: Mapping VARIABELMETADATA til SKOS-AP-NO						
Gruppering	Nr	Property	Status	Beskrivelse	SKOS-AP-NO	SKOS Engelsk
	32	SkjemaVersjon	Valgfri	Vise hvilke(n) versjon(er) av spørreskjema eller innrapporteringskjema denne variabelen finnes i.	E-helse: Extension	
Administrative formål	33	AnbefaltForDigitaltInnsyn	Obligatorisk*	Anbefaling fra forvalter av datakilden til HDS om hvorvidt variabelen bør inngå i et digitalt, automatisert innsyn til den registrerte via helsenorge.no. Er det anbefalt at variabelen, gitt at denne inngår i et dataprodukt som skal overføres til Helseanalyseplattformen, inngår i et digitalt, automatisert innsyn? Standardisert vokabular: 1 = Ja 2 = Valgfri *Gjelder kun for variabler som skal inngå i dataprodukt som skal overføres til Helseanalyseplattformen.	E-helse: Extension	
	34	TabellvennligNavn	Valgfri	Alternativ kolonneoverskrift for eksport eller utskrift av data.	E-helse: Extension	
	35	TabellvennligNavnEngelsk	Valgfri	Alternativ kolonneoverskrift for eksport eller utskrift av data for engelskspråklige.	E-helse: Extension	
	36	DatabaseReferanse	Obligatorisk	API-streng/URL/SourceID/TechName/unikt navn på klasse og/eller attributt i datakildens egen database. Denne referansen vil bli benyttet ved forespørsel om henting av variabelen fra kilden.	E-helse: Extension	
	37	Erstatter	Anbefalt	Code som er erstattet av denne variabelen.	Begrep: erstatter	dct:replaces

Vedlegg 1b: Mapping VARIABELMETADATA til SKOS-AP-NO						
Gruppering	Nr	Property	Status	Beskrivelse	SKOS-AP-NO	SKOS Engelsk
	38	ErstattesAv	Valgfri	Code til variabelen som evt. erstatter denne variabelen.	Begrep: erstattesAv	dct:replacedBy
	39	StartdatoHistorikk	Valgfri	Dato for start reell historikk. Hvis angitt er dette den første datoen man har historikk på metadata. Dette for å dokumentere at det kan være periode der variabelen er i bruk, men at man ikke kjenner hvilke endringer som er gjort. Hvis ikke angitt, regnes ValidFrom som start på historikk	E-helse: Extension	
	40	Synlig	Obligatorisk	Definerer om variabelen skal presenteres på helsedata.no. Variabler som ikke er synlige vil heller ikke kunne inngå i en variabeliste som legges ved søknad om tilgang til helsedata via felles søknadsskjema på helsedata.no. Standardisert vokabular: 1 = Ja 2 = Valgfri	E-helse: Extension	
Tillegg	41-N	Mulige tilleggsproperties	Valgfri	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.		

Vedlegg 3: Obligatoriske og frivillige klasser i den logiske informasjonsmodellen

Informasjonsmodellen som ligger til grunn for denne spesifikasjonen er utvidet med flere nye klasser siden versjon 1.0 av spesifikasjonen. Dette er gjort for å muliggjøre en mer presis beskrivelse av nye typer datakilder som nasjonale medisinske kvalitetsregistre og befolkningsbaserte helseundersøkelser.

De nye klassene er *Delkilde*, *Variabelgruppe*, *Instrument* og *Filter*. Utenom *Variabelgruppe*, som er løftet ut fra å være en obligatorisk egenskap ved klassen *Variabel*, er alle de nye klassene frivillige og skal kunne benyttes av aktører som selv opplever at de trenger disse klassene for å gå en presis beskrivelse av datagrunnlaget i egen datakilde. Det er dermed ikke forventet at alle aktører skal ta i bruk alle de frivillige klassene.



Figur 6 Obligatoriske og frivillige klasser i Logisk informasjonsmodell for metadata om helsedata til sekundærbruk

