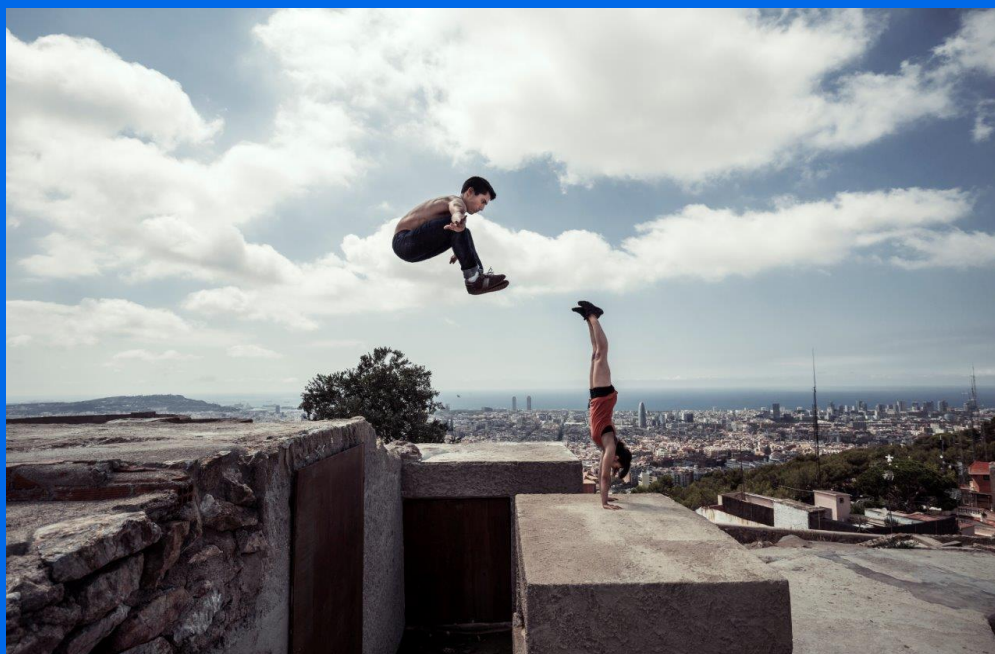


Nasjonal spesifikasjon for metadata om helsedata



HITS 1230:2020

Publikasjonens tittel:

Nasjonal spesifikasjon for metadata om
helsedata

Rapportnummer

HIST 1230:2020

Utgitt:

02/2020

Utgitt av:

Direktoratet for e-helse

Kontakt:

postmottak@ehelse.no

Postadresse:

Postboks 6737 St. Olavs plass,
0130 OSLO

Besøksadresse:

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

Publikasjonen kan lastes ned på:

www.ehelse.no

Forord

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

Målgruppen for dokumentet er dataforvaltere og tekniske utviklere i de nasjonale helsedatakildene. Spesifikasjonen setter føringer for hvordan helsedata til sekundære formål skal dokumenteres. Videre legger spesifikasjonen til rette for etableringen av en nasjonal Helseanalyseplattform ved å sikre at data som skal forvaltes på plattformen er beskrevet på en lik måte og kan utveksles maskinelt.

Spesifikasjonen bygger videre på arbeidet som ble igangsatt gjennom infrastrukturprosjektet Health Registries for Research (HRR) der de sentrale helseregistrene etablerte en felles minimumsstandard for dokumentering av registervariabler. [1] Gjennom arbeidet i Helsedataprogrammet er minimumsstandarden blitt utvidet til også å inkludere dokumentering av metadata om datakildene og tilhørende datasamlinger. Man har også adressert behovene som oppstår som følge av etableringen av en nasjonal Helseanalyseplattform. Arbeidet er blitt forankret i Helsedataprogrammets temagruppe for metadata, arbeidsgrupper tilknyttet helsedataprogrammet og andre interessenter i Direktoratet for e-helse.

Spesifikasjonen vil etter hvert revideres for å favne ytterligere behov relatert til Helseanalyseplattformen. Det er også ønskelig å utvide spesifikasjonen til å kunne adressere andre helsedatakilder som helseundersøkelser og biobanker.

Innhold

1	Innledning	6
1.1	Bakgrunn.....	6
1.2	Formål.....	6
1.3	Omfang og avgrensninger	6
2	Logisk informasjonsmodell for metadata om helsedata.....	8
3	Objektstruktur for helsedatakilder i Nasjonal variabelkatalog	10
4	Innrapportering av metadata til Nasjonal variabelkatalog	12
4.1	Prosess for innrapportering av metadata.....	12
4.2	Importfiler for metadata i Nasjonal variabelkatalog	14
	Begrepsliste.....	32
	Referanser	34

1 Innledning

1.1 Bakgrunn

Arbeidet med å definere og innrapportere standardiserte metadata fra helsedatakilder ble initiert gjennom infrastrukturprosjektet Health Registries for Research (HRR). [1] HRR ble avsluttet i 1. kvartal 2019, og Helsedataprogrammet viderefører dette arbeidet.

Nasjonal variabelkatalog er etablert i løsningen HealthTerm, levert av CareCom A/S, anskaffet av Direktoratet for e-helse. Løsningen skal forvaltes av Helsedataservice. Inntil Helsedataservice er etablert som permanent organisasjon vil løsningen bli forvaltet gjennom Helsedataprogrammet.

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

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

1.2 Formål

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

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

1.3 Omfang og avgrensninger

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

Denne spesifiseringen omfatter sentrale helseregistre og nasjonale medisinske kvalitetsregistre. Fremtidige versjoner av spesifiseringen vil også kunne stille krav til innrapportering av metadata fra andre helsedatakilder som f.eks. befolkningsbaserte helseundersøkelser og biobanker. Det legges videre til grunn at det må gjøres en mindre revisjon av spesifiseringen for å legge til rette for eksport av data til Helseanalyseplattformen.

Spesifikasjonen gir ikke føringer for hvordan metadata skal struktureres og forvaltes i lokale løsninger, kun om hvordan dette skal innrapporteres til Nasjonal variabelkatalog.

Beskrivelsen er i stor grad basert på Difi's standard for begrepsbeskrivelse [3] og Standard for beskrivelse av datasett og datakataloger (DCAT-AP-NO) [4]. Spesifiseringen støtter opp

Nasjonal spesifisering for metadata om helsedata

under FAIR-prinsippene om at metadata og data skal være "Findable", "Accesible", "Interoperable" og "Reusable". [5]

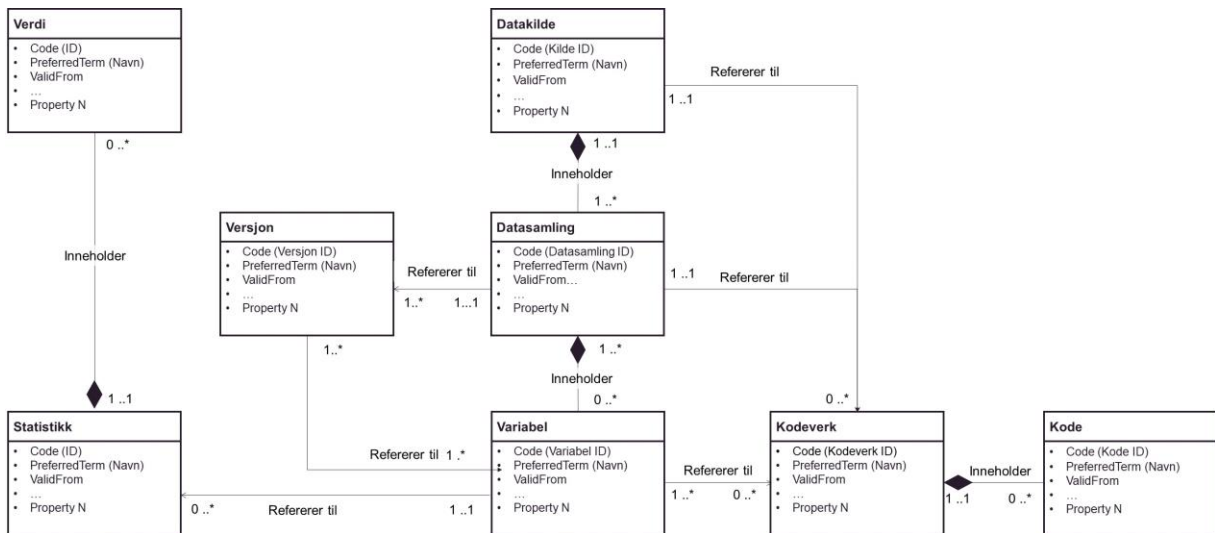
2 Logisk informasjonsmodell for metadata om helsedata

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 den valgte tekniske løsningen.

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

- **Datakilde** er definert som kildene til dataene. I denne versjonen av spesifikasjonen er en datakilde ekvivalent med et helseregister.
- **Datasamling** er definert som et sub-register innenfor en datakilde, altså en delmengde av variablene i den aktuelle datakilden. Eksempler på datasamlinger kan være sektorer i NPR eller skjema i medisinske kvalitetsregistre.
- **Versjon** er et teknisk hjelpemiddel som brukes til å definere hvilke variabler som inngår i ulike årganger/versjoner av de ulike datasamlingene i en datakilde.
- **Variabler** er de informasjonsbærende elementene i datakilden og definerer hvilken informasjon som finnes i registrene.
- **Kodeverk** representerer i denne sammenhengen et verdsett. Et kodeverk kan for eksempel representere en variabels verdsett. Det kan også representere en samling verdier som brukes til å beskrive en datakilde.
- **Kode** er i denne sammenheng det samme som et svaralternativ i et verdsett.
- **Statistikk** er definert som de statistiske egenskapene ved en årgang/versjon av en variabel.
- **Verdi** er definert som antall gyldige tilfeller av et svaralternativ/utfall av en variabel.

En komplett oversikt over relevante fagbegrep er gjengitt i begrepslisten.



Figur 1 Logisk informasjonsmodell for metadata om helsedata til sekundærbruk

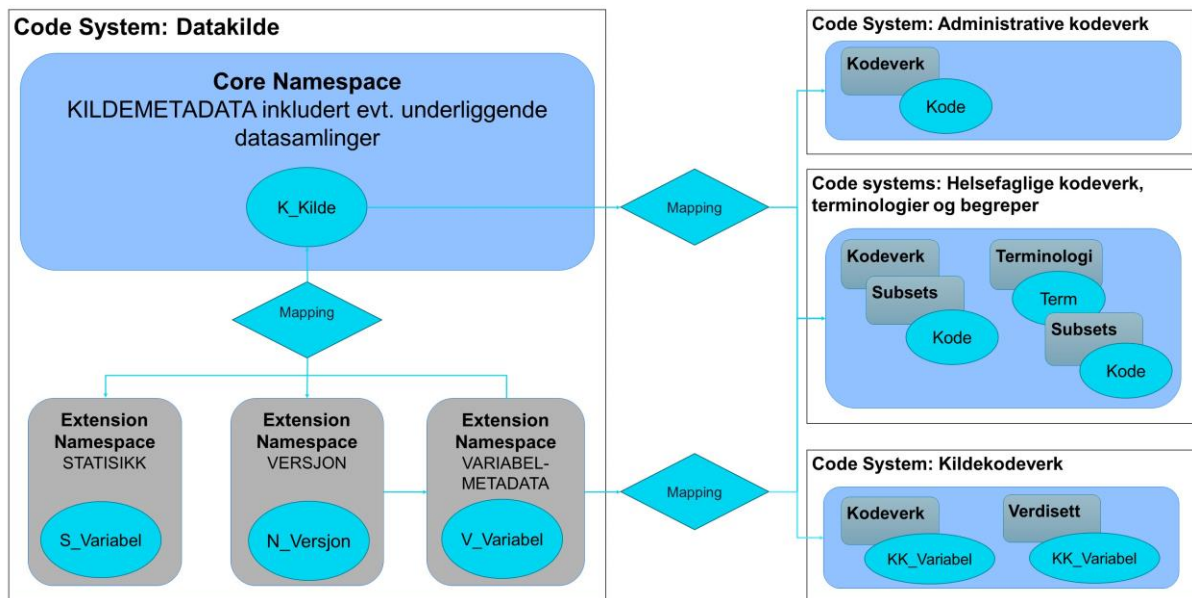
Punktlisten under forklarer informasjonslisten illustrert i Figur 1.

1. Hver enkelt datakilde er definert ut fra et sett *properties* om datakilden.
2. En datakilde *inneholder* én eller flere datasamlinger.
3. Hver datakilde og datasamling er definert ut fra et sett med egne *properties*.
4. Hver datasamling *inneholder* variabler som er definert ut fra et sett *properties* om variabelen. Samme variabel kan inngå i flere datasamlinger.
5. 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.
6. Hver versjon *refererer til* en eller flere variabler.
7. Variabler kan *referere* til kodeverk. For diskrete variabler (variabler med svaralternativ) utgjør dette kodeverket variabelens verdsett der kodeverket *inneholder* koder som representerer alle mulige verdier (svaralternativ) en variabel kan ha.
8. Hver variabel kan også *referere* til statistikk om seg selv, som gjennomsnitt, standardavvik og årstall statistikken er gyldig for. For diskrete variabler kan statistikk *inneholde* antall utfall av de ulike mulige verdiene i variabelens verdsett.
9. Datakilder og datasamlinger kan også *referere* til kodeverk for å forenkle filtrering og søk på tvers av de ulike kildene/datasamlingene. For eksempel kan en datakilde *referere* til de ulike hovedkapitlene i ICD-10 som kilden adresserer.

3 Objektstruktur for helsedatakilder i Nasjonal variabelkatalog

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

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



Figur 2 Objektstruktur for helsedatakilder i HealthTerm

1. Hver enkelt datakilde importeres som et eget *Code System* som kan betraktes som en mappe der all informasjon om kilden med tilhørende variabler ligger lagret.
2. KILDEMETADATA importeres som *Code System* og blir dermed *Core Namespace* under *Code System*. KILDEMETADATA består av et *concept* med tilhørende *properties* om datakilden. Dersom datakilden inneholder flere underliggende datasamlinger, inneholder KILDEMETADATA også et *concept* for hver av datasamlingene. Datasamlingene defineres som *children* av datakilden og 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. VARIABELMETADATA importeres som en *Extension* til *Core Namespace* (KILDEMETADATA). VARIABELMETADATA inneholder en oversikt over alle

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

4. STATISTIKK importeres som en *Extension* til *Core Namespace*. STATISIKK inneholder statistisk informasjon om utvalgte variabler definert i VARIABELMETADATA fordelt per datasamling og VERSJON som variabelen inngår i.
5. VERSJON importeres som en *Extension* til *Core Namespace*. VERSJON brukes for å koble hvilke variabler som inngår i hvilke versjoner av en datasamling, og for å vite når datagrunnlaget er oppdatert.
6. 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 som kodeverkene anvendes i, hvem som eier disse, samt et samlehierarki der alle kodeverkene med tilhørende koder (verdier/svaralternativ) er sortert i en flat alfabetisk liste. Sistnevnte for enklere å kunne identifisere kodeverk og/eller koder som kan/bør harmoniseres på tvers av kildene. Strukturen er lik den som brukes for administrative kodeverk.
7. 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.
8. Datasamlinger mappes til respektive VERSJONER som igjen mappes til de variablene som inngår i gjeldende versjon av datasamlingen.
9. Variablene i VARIABELMETADATA mappes til respektive KILDEKODEVERK, respektiv STATISTIKK, og evt. andre referanser som f.eks. begrepskataloger
10. For koordinerings-, harmoniserings- og berikningsformål kan variabler mappes mot flere relevante kodeverk/koder, klassifiseringer, terminologier/termer, og etter hvert f.eks. internasjonale standarder, begreps- og informasjonsmodeller som HL7 FIHR, SnomedCT, Contsys etc.

4 Innrapportering av metadata til Nasjonal variabelkatalog

4.1 Prosess for innrapportering av metadata

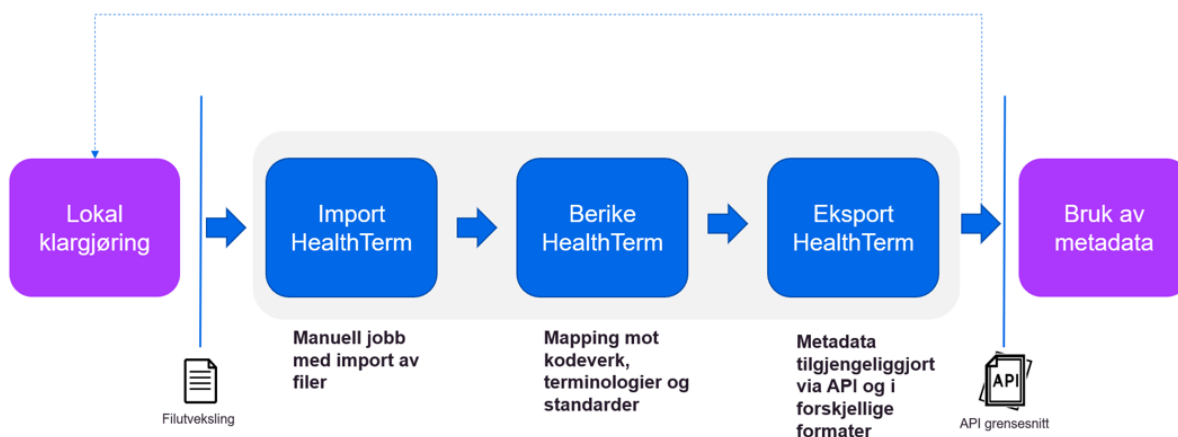
Metadata om datakilder, datasamlinger, versjoner, variabler og kodeverk må i størst mulig grad speile datagrunnlaget i kildene, både i nåtid og historisk. Filstrukturen er utviklet blant annet for å kunne knytte versjoner av datasamlinger og variabler opp mot metadata og statistikk, på et gitt tidspunkt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

I et fremtidig mål bilde, er det ønskelig at innrapporteringsprosessen i størst mulig grad skal være automatisert. Etersom dagens tekniske løsning for Nasjonal variabelkatalog foreløpig ikke støtter inngående APIer, vil man nå, i en overgangsfase, legge opp til en prosess der registrene genererer et sett med eksportfiler som oversendes til Helsedataservice. Det må utredes nærmere hvordan en mer automatisert prosess kan se ut.

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

Helsedataservice vil ha i oppgave å importere metadataene i Nasjonal variabelkatalog, oppdatere dem, og tilgjengeliggjøre dem via et API-grensesnitt til eksterne kilder som helsedata.no og evt. andre løsninger som ønsker å benytte seg av informasjonen i Nasjonal variabelkatalog. Arbeidet med å forvalte og tilgjengeliggjøre metadata fra Nasjonal variabelkatalog vil gjøres i nært samarbeid med registermiljøene, og metadata skal ikke eksponeres på helsedata.no før registerforvalterne har fått anledning til å kvalitetssikre innholdet. Det legges derfor opp til et kontrollregime der registermiljøene gis anledning til å kontrollere egne variabler og metadata før disse publiseres åpent på helsedata.no.

Forvaltning av Nasjonal variabelkatalog omfatter også tilrettelegging for harmoniseringsarbeid for realisering av *Felles språk i helse- og omsorgssektoren*, [6] samt videreutvikling av mappinger og relasjoner mot andre kodeverk, terminologier og standarder. Dette er illustrert i Figur 3.



Figur 3: Prosess for import, forvaltning og eksport av variabler og metadata

For å kunne realisere den logiske informasjonsmodellen i Kapittel 2, vil registrene bli bedt om å generere seks ulike filtyper. Disse er illustrert i Figur 4. Merk at en datakilde i Nasjonal variabelkatalog vil ha behov for flere mapset, derav Importfil 3x.



Figur 4 Importfiler til Nasjonal variabelkatalog per datakilde

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

- **11+N-formatet:** En fil med 11 obligatoriske *properties*, definert i HealthTerm, og med et ubegrenset antall tillegg*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 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 3x: Mapset.

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

4.1.1 Prosess for oppdatering av innrapporterte metadata

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

4.2 Importfiler for metadata i Nasjonal variabelkatalog

Dette kapittelet beskriver i detalj kravene til de ulike importfilene definert i Figur 4 og relaterer de ulike importfilene til den logiske informasjonsmodellen illustrert i Figur 1.

4.2.1 Navnstandard for importfiler

For å sikre sporbarhet skal importfilene definert i Figur 4 gis et standardisert navn.

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

DATAKILDE_Importfiltype-Importfilnavn_Årssett*_Versjon*_Dato sendt (YYYYMMDD)

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

*Årssett er siste komplette årgang som er inkludert i datagrunnlaget

Eksempler:

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

Filtype 3x 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 og tilhørende datasamlinger
- N for versjon
- S for statistikk
- KK for kildekodeverk
- AK for administrative kodeverk
- For helsefaglige kodeverk brukes
 - ICD for ICD-10
 - SNO for SNOMED CT

- For andre mappings brukes en egendefinert logisk forkortelse

Eksempler

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

4.2.2 Format for importfiler

Importfilene for metadata i HealthTerm er spesifisert i Tabell 4 til Tabell 8. Følgende begrep brukes for å definere importspesifikasjonen.

- **Concept** er det grunnleggende dataobjektet i HealthTerm. Variabler er derav definert som et *concept*.
- **Property** inneholder egenskaper ved, og metadata om, et *concept* – en variabel. De første 11 standard *propertyene* er systemspesifikke og begrepene tilsvarer de som vises i HealthTerm. Begrepene som anvendes på de resterende *propertyene* er definert i samarbeid med registerforvalterne og med utgangspunkt i aktuelle standarder.
- **DataType** spesifiserer formatet på elementene under *Column Name* (definert av HealthTerm) Skriver man f.eks. inn en tekst (*String*) i en celle som har format "*Number*" vil dette før til en feil (*Error*) når man skal importere filen.
- **Length** spesifiserer maksimum lengde på informasjonsinnholdet.
- **Required** spesifiserer om informasjonsinnholdet er obligatorisk eller valgfritt.
- **Editable** spesifiserer om informasjonen i en *property* kan endres etter første gangs import eller registrering. F.eks. kan man ikke endre *Code* eller *ValidFrom* på et *concept* etter at det er publisert.
- **Description** inneholder beskrivelser av informasjonen i kolonnene.

Propertyene **Code** og **ParentCode** beskriver en hierarkisk struktur i importfilene.

Den videre beskrivelsen refererer til den 1. linjen i filen som **Topnode**.

Toppnoden skal bare ha verdi i *Code*, *PreferredTerm* og *ConceptType*.

Code – feltet i toppnoden vil være *ParentCode* til alle de etterfølgende radene.

Tabell 1: Hierarkisk oppbygging av importfilene					
Filtype	Code	Parent Code	Hierar chy	PreferredTerm	ConceptType
1a: Kildemetadata	DAR			Dødsårsaksregisteret	GROUPING
1a: Kildemetadata	K_DAR	DAR		KILDEMETADATA	GROUPING
1b: Variabelmetadata	V_DAR	DAR		VARIABELMETADATA	GROUPING
1c: Versjon	N_DAR	DAR		VERSJON	GROUPING
1d: Statistikk	S_DAR	DAR		STATISTIKK	GROUPING
1d: Statistikk	S_DAR_2017. DAR	S_DAR		DØDSÅRSAKSREGISTERET 2017	GROUPING

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

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

Tabell 2: Polyhierarkisk oppbygging av Importfil 2: Kildekodeverk				
Code	ParentCode	Hierarchy	PreferredTerm	ConceptType
KK_KODEVERK*		KILDEKODEVERK	KILDEKODEVERK	GROUPING
KK_ALLE*	KK_KODEVERK	KILDEKODEVERK	ALLE	GROUPING
KK_EIER*	KK_KODEVERK	KILDEKODEVERK	EIER	GROUPING
KK_KILDE*	KK_KODEVERK	KILDEKODEVERK	KILDE	GROUPING
KK_<Eier>**	KK_EIER	KILDEKODEVERK	<Kortnavn>	GROUPING
Eksempel: Variabelen KJONN_K Tilhører datakilden KK_DAR (Dødsårsaksregisteret) har eieren KK_FHI og inngår også i strukturen «Alle»				
KK_FHI	KK_EIER	KILDEKODEVERK	FHI	GROUPING
KK_DAR	KK_FHI	KILDEKODEVERK	DØDSÅRSAKSREGISTERET	GROUPING
KK_DAR	KK_KILDE	KILDEKODEVERK	DØDSÅRSAKSREGISTERET	
KK_DAR.KJONN_K	KK_ALLE	KILDEKODEVERK	Kjønn	
KK_DAR.KJONN_K	KK_DAR	KILDEKODEVERK	Kjønn	
*Grunnstruktur som KILDEKODEVERK importeres inn i defineres som en del av systemadministrasjon før import.				
**Eiere defineres som en del av grunnstrukturen. Eksempel KK_FHI, KK_HDIR, KK_KREFT etc.				

4.2.3 Importfil 1a-Kildemetadata

Importfil 1a-Kildemetadata brukes til å beskrive datakilden og underliggende datasamlinger, på et overordnet nivå ved hjelp av et sett standardiserte *properties*. Filen er basert på HealthTerm 11+N importformat og har følgende spesifikasjoner.

Tabell 3: Format for importfil 1a-Kildemetadata. Dette filformatet benyttes også for import av metadata om datasamlinger.						
Nr	Property	Data Type	Length	Required	Editable	Description
1	Code	String	4000	Ja	Nei	Unik navnestreng for en kilde og evt. underliggende datasamlinger. Store bokstaver, uten mellomrom og æ, ø, å. Prefiks K for kilde, underscore, kortnavn på datakilden, punktum, og navn på evt. datasamling, f.eks. K_NPR.SOMATIKK
2	ParentCode	String	4000	Ja	Nei	Forkortelse av datakildens navn med prefiks K for kilde, f.eks. K_NPR. "ParentCode" for toppnoden skal være lik kortnavnet på datakilden, f.eks. DAR.
3	Hierarchy	String	298	Nei	Ja	Skal være tom for KILDEMETADATA
4	PreferredTerm	String	3700	Ja	Ja	Kort beskrivende navn på kilden og evt. underliggende datasamlinger, f.eks. "Norsk pasientregister". "PreferredTerm" på toppnoden skal ha navn lik KILDEMETADATA med store bokstaver.
5	ConceptType	ENUM	N/A	Nei	Ja	Skal være tom for KILDEMETADATA
6	PresentationOrder	Number	264 – 1 (Long)	Nei	Ja	Brukes dersom man ønsker å presentere datasamlingene i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk.
7	ValidFrom	String (YYYYMMDD)	4000	Ja	Nei	Brukes for å angi fra når en datakilde eller datasamling er gyldig. Settes ved første import. Forutsetning for å kunne tilrettelegge for søk på gyldige datakilder og datasamlinger på en gitt dato eller endringer i en valgt tidsperiode.

Tabell 3: Format for importfil 1a-Kildemetadata. Dette filformatet benyttes også for import av metadata om datasamlinger.

Nr	Property	Data Type	Length	Required	Editable	Description
8	ValidTo	String (YYY YMM DD)	4000	Nei	Nei	Brukes for å angi at en kilde/datasamling ble, eller vil bli, ugyldig fra en gitt dato. Brukes kun ved initial import når man ønsker å få med historikk på utgåtte kilder/ datasamlinger. Kilder/ datasamlinger med ValidTo-dato må fjærens fra importfilen ved neste gangs import for å unngå feil i HealthTerm
9	Synonym	String	4000	Nei	Ja	Skal være tom for KILDEMETADATA
10	PatientFriendlyTerm	String	4000	Nei	Ja	Skal være tom for KILDEMETADATA
11	ClinicalFriendlyTerm	String	4000	Nei	Ja	Skal være tom for KILDEMETADATA
12	Oppdatert	String	4000	Ja	Ja	Dato for når KILDEMETADATA sist ble oppdatert (YYYYMMDD)
13	Beskrivelse	String	4000	Ja	Ja	En mer utførlig beskrivelse av kilden og tilhørende datasamlinger
14	Opprettet	String	4000	Ja	Ja	Året datakilden/datasamlingen(e) ble opprettet historisk.
15	KortNavn	String	4000	Ja	Ja	Forkortelse av datakildens/datasamlingens navn.
16	NavnEngelsk	String	4000	Ja	Ja	Det engelske navnet på datakilden/datasamlingen.
17	Beskrivelse Engelsk	String	4000	Ja	Ja	En mer utførlig beskrivelse av datakilden på engelsk
18	KildeType	String	4000	Ja	Ja	Om datakilden er et sentralt helseregister eller et nasjonalt kvalitetsregister. I fremtidige versjoner av denne standarden vil også helseundersøkelse, biobanker og mer inngå som typer. <u>Verdisett:</u> <ul style="list-style-type: none"> • Sentralt helseregister • Nasjonalt kvalitetsregister • Annet kvalitetsregister

Tabell 3: Format for importfil 1a-Kildemetadata. Dette filformatet benyttes også for import av metadata om datasamlinger.

Nr	Property	Data Type	Length	Required	Editable	Description
						<ul style="list-style-type: none"> • Befolkningsbasert helseundersøkelse • Biobank • Annen datakilde
19	Formaal	String	4000	Ja	Ja	Kort beskrivelse av formålet registrerte er opprettet for.
20	Inklusjonskriterier	String	4000	Ja	Ja	En faglig overordnet beskrivelse av hvilke kriterier som stilles for at en person skal være registrert i datakilden.
21	Tilslutningsgrad	String	4000	Nei	Ja	Andel enheter av totalt antall relevante enheter som rapporterer inn til datakilden. Med enheter menes virksomheter.
22	BeregningTilslutningsgrad	String	4000	Nei	Ja	Hvordan tilslutningsgraden er beregnet. Hva er "nevneren"?
23	Dekningsgrad	String	4000	Nei	Ja	Andel pasienter som er rapportert inn sett opp mot totalt antall pasienter som er blitt behandlet for samme sykdom i en periode (individnivå)
24	BeregningDekningsgrad	String	4000	Nei	Ja	Hvordan dekningsgrad er beregnet, eks. sammenstilt med data fra NPR.
25	DataAnsvarlig	String	4000	Ja	Ja	Den som eier datakilden og har det juridiske dataansvaret.
26	DataBehandler	String	4000	Ja	Ja	Den som faktisk behandler datakilden. Er det samme som DataAnsvarlig med mindre denne har satt ut databehandlingen til ekstern tredjepart.
27	Telefonnummer	String	4000	Ja	Ja	Telefonnummer for å kunne komme i kontakt med forvalter av datakilden.
28	Adresse	String	4000	Ja	Ja	Adresse for å kunne komme i kontakt med forvalter av datakilden.
29	E-post	String	4000	Ja	Ja	E-post for å kunne komme i kontakt med forvalter av datakilden.
30	Hjemmeside	String	4000	Ja	Ja	Lenke til datakildens hjemmeside.

Tabell 3: Format for importfil 1a-Kildemetadata. Dette filformatet benyttes også for import av metadata om datasamlinger.

Nr	Property	Data Type	Length	Required	Editable	Description
31	RettsligGrunnlag	String	4000	Ja	Ja	Om oppføringer i datakilden er samtykkebasert, reservasjonsbasert eller lovpålagte. <u>Verdisett:</u> <ul style="list-style-type: none"> • Samtykke • Reservasjon • Lovpålagt
32	Lovverk	String	4000	Ja	Ja	Hvilket lovverk opprettelsen av datakilden er forankret i.
33	Forskrift	String	4000	Nei	Ja	URL til forskrift
34	Sensitivitetsnivå	String	4000	Ja	Ja	Hvilken type opplysninger som ligger i datakilden og hvor sensitive disse er. Om datakilden inneholder direkte personidentifiserbare opplysninger indirekte personidentifiserbare opplysninger, pseudonymiserte opplysninger eller anonyme opplysninger. <u>Verdisett:</u> <ul style="list-style-type: none"> • Direkte personidentifiserbart • Indirekte personidentifiserbart • Pseudonymt • Anonymt
35	Arkiveringsinformasjon	String	4000	Nei	Ja	Hvor lang tid datakilden har adgang til å lagre innsamlet data før denne må slettes.
36	Koblingsmulighet	String	4000	Nei	Ja	En oversikt over hvilke andre datakilder det er mulig å koble den aktuelle kilden mot i.h.t. lover, forskrifter og samtykkeerklæringer.
37	HERID_Datakilde	String	4000	Nei	Ja	Unik identifikator som tildeles av NHN, Brukes for identifisering av datakilden innenfor helsenetten. Kan brukes til spørringer.
38	HERID_Dataansvarlig	String	4000	Nei	Ja	Unik identifikator som tildeles av NHN, Brukes for identifisering av datakilden innenfor helsenetten. Kan brukes til spørringer.

Tabell 3: Format for importfil 1a-Kildemetadata. Dette filformatet benyttes også for import av metadata om datasamlinger.

Nr	Property	Data Type	Length	Required	Editable	Description
39	OppdateringsFrekvens	String	4000	Nei	Ja	Oppdateringsfrekvens brukes for å beskrive hvor ofte datasamlinger oppdateres på Helseanalyseplattformen. Informasjonselementet er ikke relevant for datakilder og datasamlinger som ikke ligger på Helseanalyseplattformen.
40-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 ønsket og behov oppstår.

4.2.4 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.

Tabell 4: Format for importfil 1b-Variabelmetadata

Nr	Property	Data Type	Length	Required	Editable	Description
1	Code	String	4000	Ja	Nei	Unik navnestreng for en variabel. Store bokstaver, uten mellomrom og æ, ø, å. Prefiks V for variabel, underscore, kortnavn på datakilden, punktum, og samme variabelnavn som i kilden, f.eks. V_DAR.KJONN_K.
2	ParentCode	String	4000	Ja	Nei	Forkortelse for kildenaavn med prefiks V for Variabel og underscore, f.eks. V_DAR. "ParentCode" for toppnoden skal være lik kortnavnet på datakilden, f.eks. DAR.
3	Hierarchy	String	298	Nei	Ja	Brukes dersom man ønsker å presentere variablene i flere kontekster, også kalt polyhierarkier. Aktuelt for registre med flere datasamlinger/delregistre/skjemaer.
4	PreferredTerm	String	3700	Ja	Ja	Kort beskrivende navn på den enkelte variabel, f.eks. Kjønn. PreferredTerm på toppnoden skal ha navn lik VARIABELMETADATA med store bokstaver.

Tabell 4: Format for importfil 1b-Variabelmetadata						
Nr	Property	Data Type	Length	Required	Editable	Description
5	ConceptType	ENUM	N/A	Nei	Ja	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.
6	PresentationOrder	Number	264 – 1 (Long)	Nei	Ja	Brukes dersom man ønsker å presentere variablene i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk.
7	ValidFrom	String (YYYYMMDD)	4000	Ja	Nei	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.
8	ValidTo	String (YYYYMMDD)	4000	Nei	Nei	Brukes for å angi at en variabel ble, eller vil bli, 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 fjærens fra importfilen ved neste gangs import for å unngå feil i HealthTerm
9	Synonym	String	4000	Nei	Ja	Brukes dersom man ønsker å legge til synonymer til et variabelnavn. Krever en ny rad i importfilen dersom mer enn ett synonym.
10	PatientFriendlyTerm	String	4000	Nei	Ja	Pasientvennlig navn på variabelen
11	ClinicalFriendlyTerm	String	4000	Nei	Ja	Helsefaglig begrep på variabelen
12	Beskrivelse	String	4000	Ja	Ja	Utførlig beskrivelse av variabelen.
13	Kommentar	String	4000	Nei	Ja	Kommentar til f.eks. endringer og beskrivelse av datakvalitet. Dette kan f.eks. være når variabelen gikk over fra å være basert på ICD-9 til ICD-10 eller andre endringer av betydning.

Tabell 4: Format for importfil 1b-Variabelmetadata						
Nr	Property	DataT ype	Length	Req uired	Edita ble	Description
14	Opprinnelse	String	4000	Nei	Ja	<p>Om variabelen er hentet direkte fra helsetjenesten, er beregnet, avledet eller kodet.</p> <p><u>Verdisett:</u></p> <ul style="list-style-type: none"> • Beregnet (F.eks. basert på en formel) • Avledet (F.eks. en kategorisering) • Kodet (F.eks. manuell koding basert på tekst) • Direkte fra fagsystem/EPJ (eks. hentet direkte fra NPR-melding) • Direkte fra skjema (eks. hentet direkte fra elektronisk skjemaløsning) • Annet
15	OpprinnelseUtleddning	String	4000	Nei	Ja	Hvordan beregnede, avlede og kodete variabler er utledet. Ikke aktuell for variabler hentet direkte fra helsetjenesten eller en elektronisk skjemaløsning
17	MaaleType	String	4000	Nei	Ja	Hva variabelen måler, f.eks. tid, alder, vekt, puls etc.
18	MaaleEnhet	String	4000	Nei	Ja	Hvilken måleenhet variabelen er basert på, f.eks. år, mnd, kg, gram, slag/minutt ect.
19	DataType	String	4000	Ja	Ja	<p>Datatype som spesifisert i FHIR Primitive Types.</p> <p><u>Verdisett:</u></p> <ul style="list-style-type: none"> • String • Integer • Decimal • Boolean • Time • Date • Datetime • Uri • Base64binary
20	DataKvalitet	String	4000	Nei	Nei	Datakildens egen vurdering om datakvaliteten på variabelen. Angis som en verdi fra 1-4 der 1 er meget dårlig og 4 er meget god.
21	TemaLokalt	String	4000	Ja	Nei	Identifiserende kortnavn f.eks. AVDØDE, DØDSFALL, PASIENT, OPPHOLD. Kan betraktes som en klassifikasjon i.h.h.t. Difis Standard for begrepsbeskrivelse.

Tabell 4: Format for importfil 1b-Variabelmetadata						
Nr	Property	Data Type	Length	Required	Editable	Description
22	DatabaseReferanse	String	4000	Ja	Ja	API-streng/URL/SourceID/TechName/unikt navn på klasse og/eller attributt i datakildens database. Denne referansen vil bli benyttet ved forespørsel om henting av variabelen fra kilden.
23	NavnEngelsk	String	4000	Nei	Ja	Engelsk oversettelse av variabelnavn
24	BeskrivelseEngelsk	String	4000	Nei	Ja	Engelsk oversettelse av beskrivelsen av variabelen.
25	Ledetekst	String	4000	Nei	Ja	Dette er teksten som står i registrets innregistreringsskjema ved hver variabel. Ledetekst er primært aktuell for variabler som er registrert via et skjema/en skjemaløsning.
26	GammellID	String	4000	Nei	Ja	VariabelID som er erstattet av denne variabelen
27	Synlig	String	4000	Nei	Ja	Definerer om variabelen skal presenteres på helsedata.no eller gjøres tilgjengelig via API-grensesnittet i HealthTerm <u>Verdisett:</u> <ul style="list-style-type: none"> • Ja • Nei
28	Utleverbar	String	4000	Nei	Ja	Definerer om variabelen er tilgjengelig via søknadstjenesten, eller kun ved spesielle behov <u>Verdisett:</u> <ul style="list-style-type: none"> • Ja • Kun ved spesielle behov • Nei
29-N	Mulige tilleggsperties	String	4000	Nei	Ja	Gir mulighet for å legge til ytterligere properties som f.eks. registerspesifikke tilleggsperties, properties for "harmoniseringsarbeid", properties for kategoriseringsformål, properties med referanser til begrepskataloger, terminologier, standarder etc.

4.2.5 Importfil 1c-Versjon

Importfil 1c-Versjon brukes til å identifisere hvilke variabler som inngår i f.eks. et årssett av en datakilde/datasamling. En variabel vil kunne inngå i flere datasamlinger i forskjellige

perioder. Hver versjon gis en unik *code* og ID, og en angivelse av gyldighetsperiode. En variabels tilhørighet til en versjon defineres ved hjelp av *Mapsets*.

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

Tabell 5: Format for importfil 1c-Versjon						
Nr	Property	Data Type	Length	Required	Editable	Description
1	Code	String	4000	Ja	Nei	Unik navnestreng for en versjon. Store bokstaver, uten mellomrom og æ, ø, å. Prefiks N, kortnavn på datakilden, punktum, kortnavn på datasamlingen, underscore og årstall. N_NPR.AMK_2018
2	ParentCode	String	4000	Ja	Nei	Forkortelse av datakildenavn med prefiks N for versjon, f.eks. N_NPR. "ParentCode" for toppnoden skal være lik kortnavnet på datakilden, f.eks. DAR.
3	Hierarchy	String	298	Nei	Ja	Skal være tom for VERSJON
4	PreferredTerm	String	3700	Ja	Ja	Kort beskrivende navn på Versjonen, f.eks. Somatikk. Toppnoden skal ha navn lik VERSJON.
5	ConceptType	ENUM	N/A	Ja	Ja	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.
6	PredefinedOrder	Number	264 – 1 (Long)	Nei	Ja	Skal være tom for VERSJON
7	ValidFrom	String (YYY YMM DD)	4000	Ja	Ja	Brukes for å angi fra når en versjon er gyldig. Den første versjonen av en datasamling vil ha ValidFrom lik den datoen datagrunnlaget er gyldig fra.
8	ValidTo	String (YYY YMM DD)	4000	Ja	Ja	Brukes for å angi at en versjon ble, eller vil bli, ugyldig fra en gitt dato. Versjoner skal ikke ha overlappende gyldighetsdatoer mellom revisjoner.
9	Synonym	String	4000	Nei	Ja	Skal være tom for VERSJON
10	PatientFriendlyTerm	String	4000	Nei	Ja	Skal være tom for VERSJON
11	ClinicalFriendlyTerm	String	4000	Nei	Ja	Skal være tom for VERSJON

Tabell 5: Format for importfil 1c-Versjon						
Nr	Property	DataT ype	Length	Req uired	Edita ble	Description
12	KortNavn	String	4000	Ja	Ja	Kortnavn på datakilden/datasamlingen, som angitt i code f.eks. DAR
13	Versjon	String	4000	Ja	Ja	Versjon, som angitt i code. Dette kan være årgang, frys-versjon, skjemaversjon eller liknende
15	Databaser eferanse	String	4000	Nei	Ja	API-streng, URL eller unikt navn på klasse, tabell eller view i datakildens database. Sammen med databasereferanse på variabelmetadata skal dette utgjøre eksakt referanse til dataelement i databasen.
16	DataFra	String	4000	Nei	Nei	Beskriver fra når det finnes data fra versjonen på Helseanalyseplattformen. Format YYYYMMDD.
17	DataTil	String	4000	Nei	Nei	Beskriver til når det finnes data fra versjonen på Helseanalyseplattformen. Format YYYYMMDD.
17- N	Mulige tilleggsprop erties	String	4000	Nei	Ja	Gir mulighet for å legge til ytterligere properties. Disse kan defineres etter hvert som evt. ønsker og behov oppstår.

4.2.6 Importfil 1d-Statistikk

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

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

I første versjon av denne standarden vil det ikke være obligatorisk å produsere statistikk.

Tabell 6: Format for importfil 1d-Statistikk						
Nr	Property	DataT ype	Length	Requi red	Edita ble	Description
1	Code	String	4000	Ja	Nei	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 kode, f.eks. S_NPR_2017.BUP.KJONN.1

Tabell 6: Format for importfil 1d-Statistikk						
Nr	Property	Data Type	Length	Required	Editable	Description
2	ParentCode	String	4000	Ja	Nei	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 årsett brukes i tillegg et årstall etterfulgt av kortnavnet på datasamlingen, f.eks. S_NPR_2017.BUP.
3	Hierarchy	String	298	Nei	Ja	Skal være tom for STATISTIKK
4	PreferredTerm	String	3700	Ja	Ja	Kort beskrivende navn på den enkelte variabel, f.eks. Kjønn. "PreferredTerm" på toppnoden skal ha navn lik STATISTIKK med store bokstaver.
5	ConceptType	ENUM	N/A	Nei	Ja	Skal være tom for STATISTIKK
6	PresentationOrder	Number	264 – 1 (Long)	Nei	Ja	Brukes dersom man ønsker å presentere statistikken på verdier/koder i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk.
7	ValidFrom	String (YYYYMMDD)	4000	Ja	Nei	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.
8	ValidTo	String (YYYYMMDD)	4000	Ja	Nei	Brukes for å angi til når statistikken er gyldig. Forutsetning for å kunne tilrettelegge for søk på statistikk på en gitt dato eller endringer i en valgt tidsperiode f.eks. utvikling over flere år. Statistikk skal ikke ha overlappende gyldighetsdatoer .
9	Synonym	String	4000	Nei	Ja	Skal være tom for STATISTIKK
10	PatientFriendlyTerm	String	4000	Nei	Ja	Skal være tom for STATISTIKK
11	ClinicalFriendlyTerm	String	4000	Nei	Ja	Skal være tom for STATISTIKK
12	Kommentar	String	4000	Nei	Ja	En mer utførlig beskrivelse av statistikken ved behov.
13	SisteOppdaterteÅrssett	String	4000	Ja	Ja	Året statistikken gjelder for. Brukes for enklere å kunne filtrere på årsett. Kompletterer "ValidFrom" og "ValidTo".
14	GyldigeTilfeller	String	4000	Ja	Ja	Antall rapportert inn totalt

Tabell 6: Format for importfil 1d-Statistikk						
Nr	Property	Data Type	Length	Required	Editable	Description
15	Manglende Tilfeller	String	4000	Nei	Ja	Antall som mangler innrapportering
16	Desimaler	String	4000	Nei	Ja	Antall desimaler avrundet til
17	Prosentandel	String	4000	Nei	Ja	Dersom beregnet i kilden
18	MIN	String	4000	Nei	Ja	Minimum
19	MAX	String	4000	Nei	Ja	Maksimum
20	AVG	String	4000	Nei	Ja	Gjennomsnitt
21	STD	String	4000	Nei	Ja	Standardavvik
22	MED	String	4000	Nei	Ja	Median
23	Kodeverk LokalID	String	4000	Nei	Ja	Lokal ID for kodeverket og koden, f.eks. 1 eller 2 for hhv Mann og Kvinne.
24-N	Mulige tilleggsperties	String	4000	Nei	Ja	Gir mulighet for å legge til ytterligere properties ved behov

4.2.7 Importfil 2-Kildekodeverk

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

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

Tabell 7: Format for importfil 2-Kildekodeverk						
Nr	Property	Data Type	Length	Required	Editable	Description
1	Code	String	4000	Ja	Nei	Unik navnestreng for et kodeverk med tilhørende koder (Verdisett og verdier). Store bokstaver, uten mellomrom og æ, ø, å. Prefiks for KK for kildekodeverk, kortnavn på datakilden, punktum, variabelnavn, underscore og kodeverdi (som i kilden) , f.eks. KK_DAR. KJONN_K.MANN

Tabell 7: Format for importfil 2-Kildekodeverk						
Nr	Property	Data Type	Length	Required	Editable	Description
2	ParentCode	String	4000	Ja	Nei	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	String	298	Nei	Ja	Brukes for å presentere kodeverkene i flere kontekster, også kalt polyhierarkier.
4	PreferredTerm	String	3700	Ja	Ja	Kort beskrivende navn på det enkelte kodeverk, f.eks. Kjønn. "PreferredTerm" på toppnoden skal ha navn lik KILDEKODEVERKET med store bokstaver
5	ConceptType	ENUM	N/A	Ja	Ja	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.
6	PresentationOrder	Number	264 – 1 (Long)	Nei	Ja	Brukes dersom man ønsker å presentere kodeverkene og/eller kodene i en egendefinert rekkefølge som f.eks. avviker fra en alfabetisk.
7	ValidFrom	String (YYYYMMDD)	4000	Ja	Ja	Brukes for å angi fra når et kodeverk eller en kode er gyldig. Settes ved første import. Forutsetning for å kunne tilrettelegge for søk på gyldige kodeverk og/eller koder på en gitt dato eller endringer på i et kodeverk og/eller på en kode i en tidsperiode.
8	ValidTo	String (YYYY)	4000	Ja	Ja	Brukes for å angi at et kodeverk eller en kode ble, eller vil bli, ugyldig fra en gitt dato. Forutsetning for å kunne tilrettelegge for søk på gyldige kodeverk og/eller koder

Tabell 7: Format for importfil 2-Kildekodeverk						
Nr	Property	Data Type	Length	Required	Editable	Description
		MMDD)				på en gitt dato eller endringer på i et kodeverk og/eller på en kode i en tidsperiode
9	Synonym	String	4000	Nei	Ja	Skal være tom for KILDEKODEVERK
10	PatientFriendlyTerm	String	4000	Nei	Ja	Skal være tom for KILDEKODEVERK
11	ClinicalFriendlyTerm	String	4000	Nei	Ja	Skal være tom for KILDEKODEVERK
12	Beskrivelse	String	4000	Ja	Ja	Utførlig beskrivelse/definisjon av en kode
13	Kommentar	String	4000	Nei	Ja	Generelle kommentarer f.eks. til endringer
14	Eier	String	4000	Nei	Ja	Hvilken institusjon som er ansvarlig for å forvalte kodeverket
16	MasterKodeverk	String	4000	Nei	Ja	Hvilket masterkodeverk som dette kodeverket er basert på, evt. er et subset av, f.eks. ICD-10 (ValueSet)
17	ReferanseTerm	String	4000	Nei	Ja	Dersom begrepet som brukes i navnet på kodeverket og/eller koden kan forankres/defineres av en standardisert term.
18	KodeverkLokalID	String	4000	Ja	Ja	Lokal ID for koden, f.eks. 1 eller 2 for hhv Mann og Kvinne.
19-N	Mulige tilleggspproperties	String	4000	Nei	Ja	Gir mulighet for å legge til ytterligere properties ved behov

4.2.8 Importfil 3x: Mapset

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

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

1. Fra KILDEMETADATA, og evt. underliggende 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 STATISTIKK

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

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

Tabell 8: Format for mappingfil 3x-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. Eksempel S_DAR.KJONN_K. For kodeverk vil det være Kodeverk-ID, Eksempel for lokale kodeverk K_DAR.KJONN_K. Eksempel for Administrative kodeverk 3101. For Versjon vil det være Versjon-ID, Eksempel V_DAR.DAR.2017Q3

Begrepsliste

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

Tabell 9: Begrepsliste	
Begrep	Forklaring
Datakilde	Datakilde er definert som kildene til dataene. I denne versjonen av spesifikasjonen er en datakilde ekvivalent med et helseregister.
Datasamling	Datasamling er definert som et sub-register innenfor en datakilde, altså en delmengde av variablene i den aktuelle datakilden. Eksempler på datasamlinger kan være sektorer i NPR eller skjema i medisinske kvalitetsregistre.
Versjon	Versjon er et teknisk hjelpemiddel som brukes til å definere hvilke variabler som inngår i ulike årganger/versjoner av de ulike datasamlingene i en datakilde.
Variabel	Variabler er de informasjonsbærende elementene i datakilden og definerer hvilken informasjon som finnes i registrene.
Kodeverk	Kodeverk representerer i denne sammenhengen et verdisett. Et kodeverk kan for eksempel representere en variabels verdisett. Det kan også representere en samling verdier som brukes til å beskrive en datakilde.
Kode	Kode er i denne sammenheng det samme som svaralternativene i et verdisett.
Statistikk	Statistikk er definert som de statistiske egenskapene ved en årgang/versjon av en variabel.
Verdi	Verdi er definert som antall gyldige tilfeller av et svaralternativ/utfall av en variabel.
Code System	Code System kan betraktes som den øverste mappen i en mappestruktur i HealthTerm. All informasjon lagres innenfor et Code System. Hvert register modelleres i løsningen som et separat Code System.
Core Namespace	Core Namespace er hovedmappen i et Code System. For registre lagres metadata om variablene som Core Namespace.
Extension Namespace	Extension Namespace (eller bare Extension) er tilleggsmapper i HealthTerm der annen informasjon lagres. Kildemetadata, statistikk og versjon lagres i Extension Namespaces.
Concept	Concept er de meningsbærende dataobjektene i HealthTerm og brukes til å modellere variabler, datakilder, datasamlinger osv.
Property	Properties er det samme som metadataelement eller attributter og brukes til å beskrive concepts.

Tabell 9: Begrepsliste	
Begrep	Forklaring
Mapping	Mapping brukes og å håndtere ikke-hierarkiske relasjoner i HealthTerm, eksempelvis referere til kodeverk.
Retired	Retired brukes om concepts og properties som ikke lenger er aktive i løsningen. En gammel versjon av en property settes eksempelvis til retired når man redigerer den og får ut en ny versjon.
CareCom	CareCom er teknisk leverandør av HealthTerm – løsningen Nasjonal variabelkatalog er realisert i.

Referanser

- [1] «Health Registries for Research (HRR), WP 4. National standard for registry data documentation,» [Internett]. Available: <https://hrr.w.uib.no/wp-4-national-standard-for-registry-data-documentation/>.
- [2] «Nasjonal e-helsestrategi 2017-2022,» Direktoratet for e-helse, 2017.
- [3] «Standard for begrepsbeskrivelse,» Direktoratet for Forvaltning og IKT, 2012.
- [4] «Beskrivelse av datasett og datakataloger,» direktoratet for forvaltning og IKT, 2015.
- [5] «FAIR-prinsippene,» [Internett]. Available: <https://www.force11.org/group/fairgroup/fairprinciples>.
- [6] «Felles språk i helse- og omsorgssektoren,» Direktoratet for e-helse, 2019.

