

Veileder for bruk av FAIR- prinsippene for helsedatakilder



HITR 1236:2020

Tittel:

Veileder for bruk av FAIR-prinsippene for
helsedatakilder

Rapportnummer

HITR 1236:2020

Utgitt:

12/2020

Dokumenttype

Veileder

Utgitt av:

Direktoratet for e-helse

Kontakt:

postmottak@ehelse.no

Publikasjonen kan lastes ned på:

www.ehelse.no

Innholdsfortegnelse

Innledning	4
Anbefaling	4
Formål	4
Målgruppe	4
Bakgrunn	5
Kort om FAIR-prinsippene	6
FAIR evalueringsskjema	7
FAIR i kontekst av annet arbeid	7
Retningslinjer for tilgjengeliggjøring av offentlige data	7
Forankring i overordnede arkitekturprinsipper	9
Internasjonal standardisering	10
De norske FAIR-prinsippene	12
Søkbarhet (Findable)	12
Tilgjengelighet (Accessible)	16
Interoperabilitet (Interoperable)	20
Gjenbrukbarhet (Reusable)	24
Appendiks	28
Begrepsdefinisjoner	28
Referanser	35

Innledning

Anbefaling

Direktoratet for e-helse anbefaler bruk av FAIR-prinsippene for å sikre en systematisk tilrettelegging for deling og gjenbruk av helseregisterdata, og felles forståelse for hva som kreves for å lykkes med dette.

FAIR-prinsippene skal bidra til økt innsikt i hvilke områder som i dag er godt tilrettelagt for datadeling.

Det er utarbeidet et evalueringsverktøy for å understøtte FAIR-prinsippene. Direktoratet for e-helse anbefaler at forvaltere av helsedatakilder bruker evalueringsverktøyet som et virkemiddel for å kunne ta i bruk FAIR-prinsippene, og for å identifisere forbedringsområder.

Formål

Dette dokumentet inneholder en norsk versjon av FAIR-prinsippene og er utarbeidet av Helsedataprogrammet [1] i regi av Direktoratet for e-helse. Arbeidet legger de internasjonale prinsippene til grunn, men har i tillegg til å etablere en norsk oversettelse, innarbeidet veiledende spørsmål, samt et tilhørende evalueringsskjema. Dette er ment å gjøre det enklere for dataforvaltere i sektoren å utføre en evaluering for å vurdere egen oppfyllelse av FAIR-prinsippene.

Formålet med arbeidet er å bidra til et felles språk og forståelse, og tilrettelegge for mer effektiv deling og gjenbruk av helsedata både for "mennesker og maskiner".

Målgruppe

Denne versjonen av de norske FAIR-prinsippene er i hovedsak rettet mot norske helseregistre. Innholdet er av forholdsmessig teknisk natur og kan dermed være litt vanskelig å forstå for f.eks. registerforvaltere med en ikke-teknisk bakgrunn. Vi anbefaler derfor at registerforvaltere og systemforvaltere samarbeider rundt tolkningen av prinsippene og kriteriene. Prinsippene vil også kunne brukes til evaluering av andre helsedatakilder som f.eks. helseundersøkelser, biobanker og forskerdatasett.

Prinsippene gjelder for metadata om helsedatakilder, tilhørende datasamlinger (datasett) og variablene som datasamlingene består av.

Prinsippene må sees i sammenheng med oppfylling av kravene i "Nasjonal spesifisering for metadata om helsedata" [2] (Direktoratet for e-helse) og Digitaliseringsdirektoratet sin "Veileder for beskrivelse av datasett" [3].

Bakgrunn

Prinsippene ble første gang anbefalt av G20 toppledere på Hangzhou Summit China 4-5 September 2016 [4]. De begynner å få stor utbredelse i europeiske forskningsmiljøer og benyttes blant annet av European Data Portal [5] og EOSC (European Open Science Cloud) [6]. Sistnevnte er et omfattende EU-initiativ med målsetning om å understøtte datadrevet forskning og innovasjon og i "The European Code of Conduct for Research Integrity" [7] fra ALLEA som er en europeisk sammenslutning av mer enn 50 forskningsinstitutter fra 40 land i Europa.

FAIR-prinsippene benyttes også av det nordiske programmet for helse og velferd [8] som blant annet skal bidra til styrket nordisk samarbeid om forskningsinfrastrukturer og harmonisering på nordisk nivå.

Konseptvalgutredningen for Helseanalyseplattformen [9] har vurdert at FAIR-prinsippene kan være hensiktsmessige både for å understøtte forvaltning av data og metadata, og for å støtte opp under prosesser rundt tilrettelegging, deling og gjenbruk av helsedata.

Helsedataprogrammet vil derfor jobbe videre med innføring av FAIR-prinsippene for å bidra til enklere tilgang til helsedata, bedre datakvalitet og mer effektiv registerforvaltning.

Ønske om en norsk versjon av FAIR-prinsippene ble identifisert gjennom deltakelse fra Helsedataprogrammet i "Nordic health metadata working group" som er en del av NordForsk programmet «Nordic Programme on Health and Welfare» [10, 8]. FAIR-prinsippene ble benyttet for å kartlegge i hvilken grad metadata for helseregistre og andre helsedatakilder i Norden er søkbare, tilgjengelige/nedlastbare, og tilrettelagt for samhandling og gjenbruk.

Arbeidet bygger blant annet på rapporten "Kriterier för FAIR Forskningsdata" [11] utarbeidet av det svenske Vetenskapsrådet, og RDA sin analyse av eksisterende FAIR-scoringverktøy - "Results of an Analysis of Existing FAIR assessment tools" [12].

Kort om FAIR-prinsippene

FAIR-prinsippene [13] [14] er et sett med veiledende arkitekturprinsipper som skal tilrettelegge for deling og gjenbruk av data gjennom at dataene er søkbare (**F**indable), tilgjengelige (**A**ccessible), understøtter interoperabilitet (**I**nteroperable) og er gjenbrukbare (**R**eusable). En oversikt over den internasjonale versjonen av FAIR-prinsippene er presentert i [15] og gjengitt under.

De norske FAIR-prinsippene med underliggende kriterier tilsvarer den internasjonale versjonen på kriterienivå. Disse er beskrevet utførlig i denne veilederen. I tillegg er det formulert noen veiledende spørsmål, som skal støtte forståelsen av de ulike kriteriene.

To be Findable:

- F1.** (meta)data are assigned a globally unique and eternally persistent identifier.
- F2.** data are described with rich metadata.
- F3.** (meta)data are registered or indexed in a searchable resource.
- F4.** metadata specify the data identifier.

To be accessible:

- A1** (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1** the protocol is open, free, and universally implementable.
 - A1.2** the protocol allows for an authentication and authorization procedure, where necessary.
- A2** metadata are accessible, even when the data are no longer available.

To be interoperable:

- I1.** (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2.** (meta)data use vocabularies that follow FAIR principles.
- I3.** (meta)data include qualified references to other (meta)data.

To be re-usable:

- R1.** meta(data) have a plurality of accurate and relevant attributes.
 - R1.1.** (meta)data are released with a clear and accessible data usage license.
 - R1.2.** (meta)data are associated with their provenance.
 - R1.3.** (meta)data meet domain-relevant community standards.

Figur 1 De internasjonale FAIR prinsippene [15]

FAIR evalueringsskjema

Helsedataprogrammet har utarbeidet et evalueringsskjema i Excel for å kunne evaluere helseregistre opp mot FAIR-prinsippene. Evalueringsskjemaet er bygget opp slik at helseregistre kan vurdere sin egen oppfyllelse av FAIR-prinsippene basert på en skala fra 1 til 5. I tillegg er det utarbeidet et sett med veiledende spørsmål som støtte til evalueringen med forklaringer for hvorfor dette er viktig.

FAIR-scoren fra evalueringen er ment å gi helseregistrene økt innsikt i egen oppfyllelse av FAIR-prinsippene. En sammenstilling av FAIR-score fra ulike helseregistre kan også øke forståelsen for den generelle graden av ivaretagelse av de ulike prinsippene i Norge. På nasjonalt nivå kan evalueringene dermed gi et godt bilde av hvilke områder som bør ha fokus for i større grad å kunne tilrettelegge for deling og gjenbruk av data.

Evalueringsskjemaet kan lastes ned på:

www.ehelse.no

FAIR i kontekst av annet arbeid

Retningslinjer for tilgjengeliggjøring av offentlige data

Statens IKT-politikk innebærer et mål om økt gjenbruk av offentlige data. Regjeringens føringer for deling av åpne offentlige data er regulert i regjeringens «Retningslinjer ved tilgjengeliggjøring av offentlige data» [16]. Ifølge disse retningslinjer må data tilgjengeliggjøres på en måte som gjør det mulig for brukere å realisere verdien av dem. Dette innebærer både juridisk beskrivelse av hvordan data kan brukes og teknisk beskrivelse av hvordan data er tilgjengeliggjort.

I "Digitaliseringsrundskrivet" [17], understrekes det at "Den enkelte virksomhet skal ha oversikt over hvilke data den håndterer, hva dataene betyr, hva de kan brukes til, hvilke prosesser de inngår i, og hvem som kan bruke dem (informasjonsforvaltning). I samsvar med bruksbestemmelsene i offentleglova skal virksomheten gjøre egnet informasjon tilgjengelig i maskinlesbare formater, fortrinnsvis gjennom API'er."

Digitaliseringsdirektoratet sin "Veileder for orden i eget hus" [18], jf. Figur 2, følger opp med å si at "den enkelte virksomhet skal ha oversikt over hvilke data den håndterer, hva dataene betyr, hva de brukes til, hvilke prosesser de inngår i, og hvem som kan bruke dem (informasjonsforvaltning). Dette innebærer også å ta stilling til hvilke data som kan gjøres tilgjengelig for gjenbruk i offentlig sektor, og viderebruk av privat sektor."



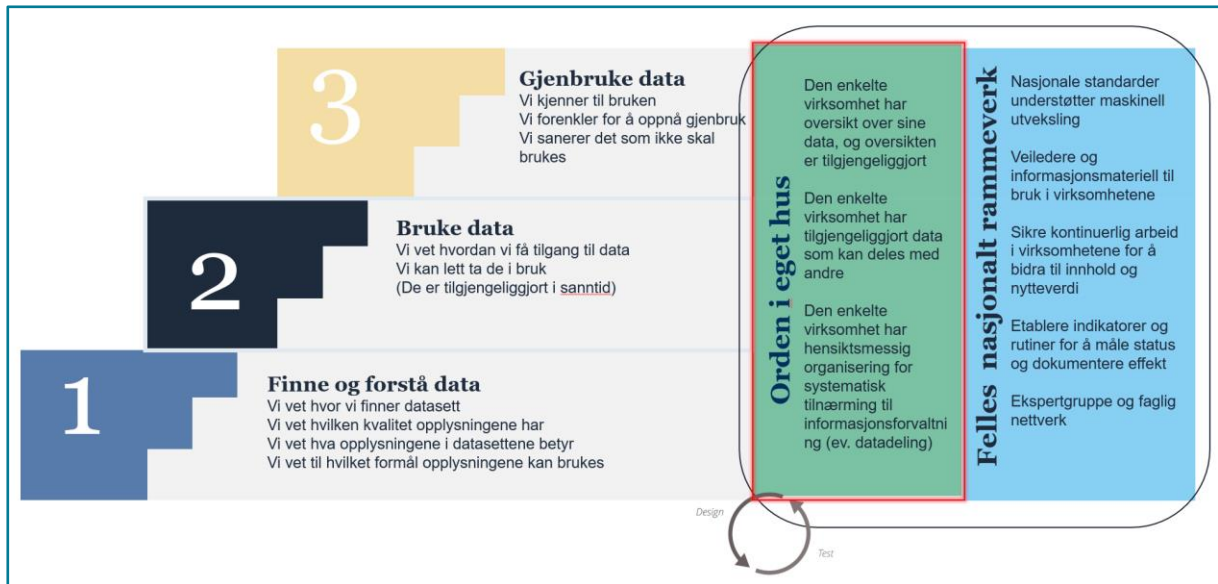
Figur 2 Hovedsteg for å skape "Orden i eget hus" (Digitaliseringsdirektoratet)

Rapporten "Informasjonsforvaltning i offentlig sektor" [19] konkretiserer dette i en stegvis måloppnåelse i fem punkter med økende ambisjonsnivå:

1. Datasettene i virksomheten er beskrevet i en oversikt
2. Oversikten fra punkt 1 er publisert
3. Tilgang til data er vurdert og beskrevet
4. Dataelementene er beskrevet
5. Beskrivelsene er strukturerte og maskinlesbare

For å understøtte dette har Digitaliseringsdirektoratet etablert et "Rammeverk for informasjonsforvaltning" [20] og en Felles datakatalog [21] for deling av datasett, API'er, begreper og informasjonsmodeller.

FAIR-prinsippene med sine tilhørende kriterier bidrar til å konkretisere hva som faktisk må på plass operasjonelt for å realisere myndighetskravene. Figur 3 viser hvordan Digitaliseringsdirektoratet tenker seg en stegvis tilnærming til hvordan en virksomhet kan og bør arbeide for å få "orden i eget hus" innenfor rammene av et felles internasjonalt rammeverk. Trinnene samsvarer godt med de 5 stegene gjengitt ovenfor.



Figur 3 Stegvis realisering av Digitaliseringsdirektoratets "Felles rammeverk for informasjonsforvaltning" og "Orden i eget hus".

Forankring i overordnede arkitekturprinsipper

I det nye nasjonale arkitekturrammeverket for samhandling [22], innledes det med å si at "skal vi få til økt samhandling i offentlig sektor, så må IT-systemene snakke sammen, vi må bruke like begreper, vi må ha god oversikt over juridiske forhold og vi må ha vilje og evne til å samhandle. Rammeverket bygger på "European Interoperability Framework" [23], jf. Figur 4.

FAIR-prinsippene understøtter spesielt operasjonaliseringen av den semantiske og tekniske samhandlingsevnen, og arkitekturprinsipp 4: "Del og gjenbruk data" [24], men favner også de juridiske aspektene med f.eks. referanser til formål, lovverk, tilgangsnivå, lisenser etc.



Figur 4 European Interoperability Framework, Norsk versjon oversatt av Digitaliseringsdirektoratet.

Internasjonal standardisering

"Nasjonal spesifisering for metadata om helsedata" [2] er basert på Digitaliseringsdirektoratet sin "Veileder for beskrivelse av datasett" [3] og "Standard for beskrivelse av datasett og datakataloger (DCAT-AP-NO)" [25], som igjen er basert på EUs DCAT-AP [26]. DCAT-AP-NO skal sikre at beskrivelser av offentlige data utføres på en felles, strukturert måte og i en maskinlesbar form. Standarden stiller krav til hva som skal, bør og kan være med i beskrivelsene, med spesifisering av dataformat.

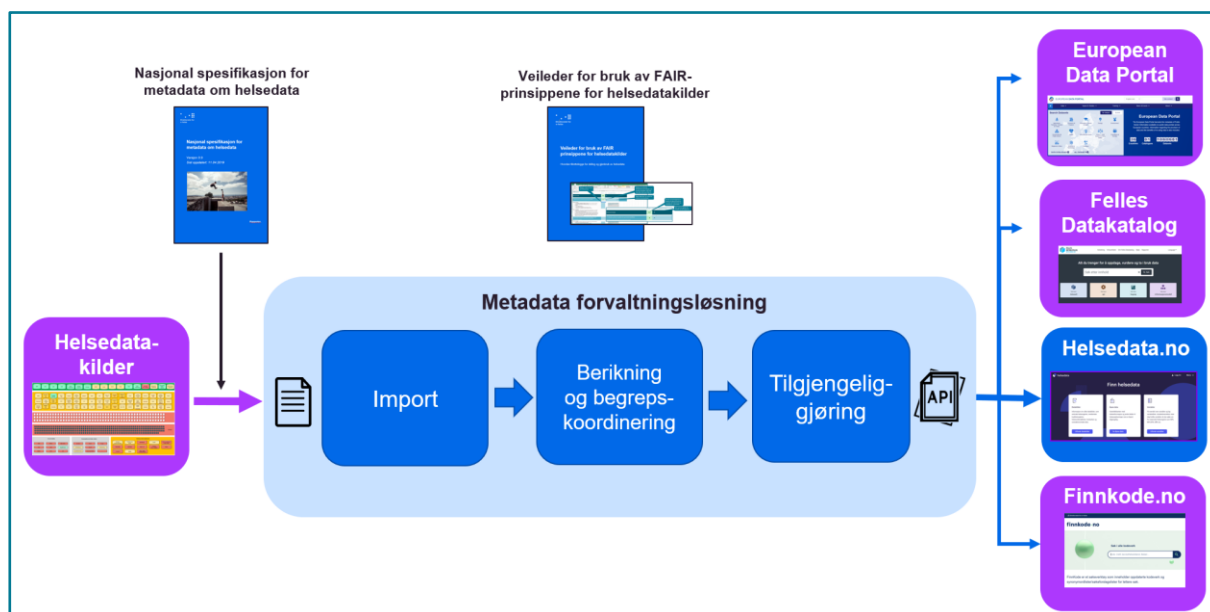
Ved å basere metadatabeskrivelsene på samme internasjonale standard muliggjøres maskinell utveksling av metadata mellom sektorielle og internasjonale metadata og datakataloger. I praksis betyr det at f.eks. "Den europeiske dataportalen" [5] kan høste metadata om informasjon som offentlig sektor forvalter (PSI) og som er publisert på dataportaler i europeiske land.

Både i den "Den europeiske dataportalen" [5] og i "Felles datakatalog" [21] utvikles det nå automatiserte valideringstjenester som scorer kvaliteten på metadata med utgangspunkt i FAIR-prinsippene og relevante DCAT-properties. "Nasjonal spesifisering for metadata om helsedata" [2], som ble utarbeidet og publisert av Direktoratet for e-helse i februar 2020, er i stor grad basert på DCAT-AP-NO [25], og det er en målsetting at kommende revisjoner skal bidra til en ytterligere samordning.

"Veileder for bruk av FAIR-prinsippene for helsedatakilder" og "FAIR evalueringsskjema" er i denne versjonen tilrettelagt for å kunne bli forstått og brukt av "mennesker" og ikke "maskiner", men de veiledende spørsmålene refererer i stor grad til properties i DCAT-AP-

NO [25] uten at selve referansen er oppgitt (i denne versjonen). Direktoratet for E-helse vil følge nært med på utviklingen av den FAIR- og DCAT-baserte valideringstjenesten som er under utvikling i Digitaliseringsdirektoratet.

Figur 5 illustrer hvordan metadataene "flyter" fra helsedatakildene, via den sektorielle forvaltningsløsningen til de nevnte kataloger og portaler.



Figur 5 Verdikjeden for metadata fra helsedatakildene, via sektoriell metadataforvaltning til helsedata.no, Felles datakatalog og European Data Portal

De norske FAIR-prinsippene

Dette kapittelet inneholder en norsk oversettelse av FAIR-prinsippene samt veiledende spørsmål som kan benyttes i vurderingen av om registrene etterlever disse og bidra til å identifisere områder som bør adresseres.

<p>F Søklarhet (Findable)</p> <ul style="list-style-type: none"> • F1: Metadata og data har en globalt unik og persistent identifikator • F2: Dataobjekter er beskrevet med utfyllende, strukturerte og maskinlesbare metadata • F3: Metadata inkluderer en identifikator til selve dataobjektet i kilden • F4: Metadata og data er indekserte og søkbare gjennom en åpen tjeneste 	<p>A Tilgjengelighet (Accessible)</p> <ul style="list-style-type: none"> • A1: Metadata og data er tilgjengelige og kan utveksles gjennom veldefinerte grensesnitt og formater • A1.1: Protokollene som brukes er åpent tilgjengelige, gratis og lesbare med standard IT-verktøy • A1.2: Protokollene støtter autentisering og autorisasjon der dette er aktuelt • A2: Metadata er tilgjengelige selv om selve data i kilden ikke er tilgjengelig
<p>I Interoperabilitet (Interoperable)</p> <ul style="list-style-type: none"> • I1: Metadata og data er beskrevet ved bruk av definerte begreper og/eller semantiske standarder. • I2: Begreper, kodeverk, terminologier og ontologier som anvendes er søkbare, tilgjengelige, interoperable og har en veletablert forvaltningsprosess. • I3: Både metadata og data er beskrevet slik at sammenhengen de inngår i er forståelig og presis, og slik at sammenlikning på tvers av kilder muliggjøres. 	<p>R Gjenbrukbarhet (Reusable)</p> <ul style="list-style-type: none"> • R1: Datakilden, evt. datasamlinger og datasett er beskrevet med utfyllende metadata • R1.1: Det finnes metadata som beskriver vilkårene for hvordan datagrunnlaget kan benyttes • R1.2: Det finnes metadata som beskriver hvordan datagrunnlaget er bearbeidet, f.eks. avledet, beregnet, anonymisert etc. på veien fra kilden til slik dataene foreligger nå • R1.3: Beskrivelsen av metadata er basert på internasjonale og domenebaserte begreper og standarder

Figur 6 Oversikt over de norske FAIR-prinsippene

Søklarhet (Findable)

Handler om at det finnes informasjon i form av metadata som gjør det enkelt for både mennesker og maskiner å finne datasett samt tilhørende informasjon.

Søklarhet	
Kortversjon	Metadata og data er godt beskrevet, de har en unik og persistent ID og er åpent tilgjengelig via en utforsker på en webside.
Kriterier	<p>F1: Metadata og data har en globalt unik og persistent identifikator</p> <p>F2: Dataobjekter er beskrevet med utfyllende, strukturerte og maskinlesbare metadata</p> <p>F3: Metadata inkluderer en identifikator til selve dataobjektet i kilden</p> <p>F4: Metadata og data er indekserte og søkbare gjennom en åpen tjeneste</p>

F1: Metadata og data har en globalt unik og persistent identifikator

Internasjonal versjon

(Meta)data are assigned a globally unique and eternally persistent identifier

Veiledende spørsmål

- Har metadata om kilden, evt. underliggende datasamlinger, datasett, og variabler en globalt unik ID?
- Har data en tilsvarende globalt unik ID?
- Har kodeverkene og/eller verdsettene som anvendes en globalt unik ID?
- Er IDene konsistente over tid (persistente)?
- Er IDen basert på en internasjonalt anerkjent standard med en tilhørende tekstuell beskrivelse som gjør den maskinlesbar og søkbar?
- Forvaltes IDen av en organisasjon som tildeler og sikrer at IDen forblir unik over tid?

Hvorfor er dette viktig?

- Dataobjekter i en kilde skal ikke kunne forveksles/mikses med dataelementer i en annen kilde, hverken av mennesker eller datamaskiner. Endringer på et dataobjekt skal kunne spores og være entydig over tid, på tvers av datasett og systemer.
- Identifikatoren kan bli brukt som del av en URI, men det er fortsatt verdifullt å ha identifikatoren representert eksplisitt.
- Ved bruk av en unik OID vil man også kunne få ekstra informasjon om elementet som en del av et hierarki. OID er bygd opp hierarkisk, som en streng, gjerne med en landskode innledningsvis, deretter bransje, organisasjon, kilde etc., og helt ned til den enkelte property (attributt) på variabelen.
- OIDs for helse- og omsorgssektoren tildeles og forvaltes av Direktoratet for e-helse.

Aktuelle linker

[Pekere til offentlige ressurser på nett, Digitaliseringsdirektoratet](#)

[OID identifikatorserier i helse- og omsorgstjenesten, Direktoratet for e-helse](#)

[Datasett-identifikator](#)

[DCAT:dct:identifiser](#)

[Nasjonal spesifisering for metadata om helsedata, Direktoratet for e-helse](#)

F2: Dataobjekter er beskrevet med utfyllende, strukturerte og maskinlesbare metadata

Internasjonal versjon

Data are described with rich metadata

Veiledende spørsmål

- Er alle obligatoriske properties i Nasjonal spesifisering for metadata om helsedata [2] utfylt?
- Inneholder metadata tittel og en fritekstbeskrivelse av datasettet?
- Er dataobjektene tematisert slik at de kan sorteres under gitte kategorier?
- Er dataobjektene tilordnet tidsbestemt informasjon som gjør det mulig å tilrettelegge for tidsbaserte søk?
- Er det tilknyttet en eller flere emneord til dataobjektene som kan bidra til å utvikle effektive filtrerings- og søketjenester?
- Er datakilder og datasamlinger geografisk avgrenset slik at det er mulig å tilrettelegge for geografisk baserte søk?
- Finnes metadata på engelsk?

Hvorfor er dette viktig?

- For at data skal være lett å finne og tas i bruk, er det behov for å etablere gode beskrivelser som muliggjør søk og maskinell prosessering
- Det er lettere å velge ut og gjenbruke data hvis de har tydelige og rike beskrivelser som sier noe om formål, kvaliteten på datagrunnlaget, endringer i koding etc. Se mer detaljerte kriterier under R (Gjenbruk)

Aktuelle linker

[Nasjonal spesifisering for metadata om helsedata, Direktoratet for e-helse](#)

[Veileder for beskrivelse av datasett, Digitaliseringsdirektoratet](#)

[Standard for begrepsbeskrivelser, Digitaliseringsdirektoratet](#)

[Standard for beskrivelse av datasett og datakataloger \(DCAT-AP-NO\), Digitaliseringsdirektoratet](#)

F3: Metadata inkluderer en identifikator til selve dataobjektet i kilden

Internasjonal versjon

Metadata clearly and explicitly include the identifier of the data they describe

Veiledende spørsmål

- Inkluderer variabelens metadata en entydig identifikator som muliggjør maskinell kobling med tilhørende data?
- Inkluderer variabelens metadata en URL til hvor dataen ligger slik at den kan hentes ut f.eks. via et API?

Hvorfor er dette viktig?

- Entydige identifikatorer er viktige for å kunne koble metadata med data og vice versa.
- Entydige identifikatorer er avgjørende for å kunne hente ut og dele data fra en datakilde via f.eks. API'er.
- Entydige identifikatorer er også nødvendig for å kunne gjenskape et datagrunnlag i ettertid.

Aktuelle linker

[Pekere til offentlige ressurser på nett, Digitaliseringsdirektoratet](#)

[Standardiserte tjenestegrensesnitt \(API\) for helseregistre, Direktoratet for e-helse](#)

F4: Metadata og data er indekserte og søkbare gjennom en åpen tjeneste

Internasjonal versjon

(Meta)data are registered or indexed in a searchable resource

Veiledende spørsmål

- Er metadata om datakilden, evt. datasamlinger, datasett og variabler søkbare og åpent tilgjengelig på en webside?
- Muliggjør tjenesten temabaserte søk på tvers av flere helsedatakilder?
- Muliggjør tjenesten tidsbaserte søk?
- Muliggjør tjenesten geografisk baserte søk?
- Er det mulig å opprette et sett av variabler i en variabelliste slik at man kan bestille, dele og/eller gjenfinne et egendefinert utvalg av variabler?

- Er ikke-sensitive data (åpne data) som f.eks. aggregert statistikk tilgjengelig via en åpen webtjeneste?
- Muliggjør en evt. analysetjeneste søk på data fra flere helsedatakilder samtidig?
- Er tjenesten tilgjengelig på engelsk, herunder navn på kilder, evt. underliggende datasamlinger, variabler, og tilhørende metadata?

Hvorfor er dette viktig?

- Oversikt over hvilke datakilder som finnes og god informasjon om innholdet i disse, er en grunnleggende forutsetning for effektiv forskning.
- Verdien av en slik oversikt øker jo flere datakilder den inneholder informasjon om.
- Er informasjonen standardisert og maskinlesbar, vil brukeren kunne søke på informasjon fra flere datakilder samtidig og definere sine søk mer presist. Finnes det statistikk på variabelnivå vil dette kunne hjelpe. Dersom søk med utgangspunkt i metadata kan kombineres med f.eks. aggregerte data vil dette gi et langt bedre grunnlag for vurderinger av datagrunnlaget enn metadata alene.

Aktuelle linker

[Helsedata.no](https://helsedata.no)

[Linker til nasjonale nordiske variabelkataloger, Nordic Health Data](#)

[Felles datakatalog, Digitaliseringsdirektoratet](#)

[European Data Portal](#)

Tilgjengelighet (Accessible)

Handler om at data og metadata er tilrettelagt på en måte som gjør det enkelt for mennesker og maskiner å få tak i datasett med tilhørende informasjon.

Tilgjengelighet	
Kortversjon	Metadata og data kan deles og utveksles maskinelt via åpne, veldefinerte grensesnitt og formater som ivaretar personvernet.
Kriterier	A1: Metadata og data er tilgjengelige og kan utveksles gjennom standardiserte grensesnitt A1.1: Protokollene som brukes er åpent tilgjengelige, gratis og lesbare med standard IT-verktøy A1.2: Protokollene støtter autentisering og autorisasjon der dette er aktuelt A2: Metadata er tilgjengelige selv om selve data i kilden ikke er tilgjengelig

A1: Metadata og data er tilgjengelige og kan utveksles gjennom standardiserte grensesnitt

Internasjonal versjon

(Meta)data are retrievable by their identifier using a standardized protocol

Veiledende spørsmål

- Er det mulig å laste ned metadata og åpne data via en webtjeneste?
- Inneholder metadata en tilgangs-URL (lenke) til webtjenesten?
- Inneholder metadata en nedlastings-URL (lenke) til selve datasettene?
- Er *metadata* representert på en slik måte at de kan aksesseres maskinelt gjennom et programmeringsgrensesnitt (API) og er egnet for automatisert, maskinell prosessering?
- Er tilgangen til *metadataene* gratis?
- Kan metadata og data deles og/eller utveksles i kjente formater som f.eks. JSON, XML, CSV og Excel?
- Er *data* representert på en slik måte at de kan aksesseres maskinelt gjennom et programmeringsgrensesnitt (API) og er egnet for automatisert, maskinell prosessering?
- Er tilgang til data gratis eller baserer prisen seg på marginalkostnadsprinsippet?

Hvorfor er dette viktig?

- Metadata bør i utgangspunktet gjøres tilgjengelig uten at brukeren må søke om tillatelse eller registrere seg. I tråd med bestemmelsene i offentlighetsloven skal metadata kunne benyttes hvor som helst, av hvem som helst, og til ethvert formål.
- Hovedregelen er at offentlig data skal være gratis, og at det ikke er anledning til å ta betalt for kostnader til innsamling og produksjon av data for videre bruk. Det finnes enkelte unntak i offentlighetsloven (§ 8) og -forskriften (§ 4) som gir anledning til å ta betalt for data. Virksomheter som krever betaling for informasjon, skal offentliggjøre betalingssatsene i elektronisk form. Alle opplysninger om grunnlaget for utregning av betalingssatsene skal også publiseres elektronisk, slik at de er lette å finne for potensielle brukere. De fleste sentrale helseregistre forholder seg til en egen forskrift for registeret som gir mulighet for å kreve betaling for behandling og tilrettelegging av opplysninger.
- URL står for Uniform Resource Locator, og er en enkelt forklart en lenke / adresse til en nettside. URL blir ofte kalt for en nettadresse.
- Et programmeringsgrensesnitt (API) er en måte å tilby data på som gjør det mulig for annen programvare å gjøre oppslag i hele eller spesifikke deler av virksomhetens data via internett. Det gjør det for eksempel mulig å bruke data i sanntid, filtrere på forespørsel, og å arbeide med data på dataelementnivå uten at brukerne må opprette lokale kopier av datasettene. Et programmeringsgrensesnitt er den beste måten å

gjøre data tilgjengelig på dersom datasettene er store, komplekse eller oppdateres ofte.

- Enkel tilgang til predefinerte datasett basert på f.eks. de erfaringsmessig mest etterspurte variablene vil være et viktig bidrag for å effektivisere tilgangen til data, spesielt dersom de kan inneholde en kombinasjon av variabler fra flere datakilder. Denne type datasett bør kunne tilbys som komplette dataprodukter, bestående av både data og metadata, enten via en analysetjeneste, et programmeringsgrensesnitt, eller som en maskinlesbar fil.

Aktuelle linker

[Retningslinjer ved tilgjengeliggjøring av offentlige data, Regjeringen.no](#)

[Nasjonal spesifisering for metadata om helsedata, Direktoratet for e-helse](#)

[HL7 FHIR](#)

A1.1: Protokollene som brukes er åpent tilgjengelige, gratis og lesbare med standard IT-verktøy

Internasjonal versjon

The protocol is open, free, and universally implementable

Veiledende spørsmål

- Er *metadata* tilgjengelige gjennom åpne og gratis protokoller som f.eks. HTTP eller FTP?
- Er *data* tilgjengelige gjennom åpne og gratis protokoller?

Hvorfor er dette viktig?

- For at metadata og data teknisk sett skal være enkelt tilgjengelig for brukerne, er det viktig at den tilgjengeliggjøres ved bruk av åpne protokoller som ikke har noen begrensninger i form av for eksempel kompleks implementasjon, eller begrensede lisenser. Eksempler på åpne kommunikasjonsprotokoller er blant annet: HTTP, FTP, SMTP
- Åpne protokoller bør anvendes slik at lukkede eller leverandørspesifikke protokoller unngås.
- Unngå å lage unødvendige begrensninger og kostnader som følge av begrensninger til enkelt leverandører ved lukkede og proprietære protokoller.

A1.2: Protokollene støtter autentisering og autorisasjon der dette er aktuelt

Internasjonal versjon

The protocol allows for an authentication and authorization procedure, where necessary

Veiledende spørsmål

- Brukes det protokoller for autentisering og autorisering i henhold til nasjonale føringer?
- Hvordan håndteres autentisering for tilgang til sensitive data?
- Hvordan håndteres autorisering for tilgang til sensitive data?
- Hvordan tilgjengeligjøres sensitive data?
- Finnes det prosedyrer/rutiner hos eier av datakilden for hvordan man følger opp at utleverte sensitive data blir forvaltet på en trygg og sikker måte og iht. regelverket?
- Finnes det et regelverk, en prosess og/eller en policy som beskriver hva man kan få tilgang til av data?

Hvorfor er dette viktig?

- Tilgangen til forskningsdata bør være så åpen som mulig, og så begrenset som nødvendig i de tilfeller det er snakk om sensitive data. For å kunne møte dette behovet er det behov for å benytte en protokoll som støtter verifisering av at brukeren faktisk er den som den gir seg ut for å være, dvs. autentisering, samt å kunne avgrense tilgangen til forskningsdata til kun den dataen brukeren skal ha tilgang til, dvs. autorisasjon.
- Eksempler på protokoller: OAuth, SAML, JWT token

Aktuelle linker

[General Data Protection Regulation, GDPR, EU](#)

[Norm for informasjonssikkerhet og personvern i helse og omsorgstjenesten, Normen](#)

[Rammeverk for autentisering og uavviselighet i elektronisk kommunikasjon med og i offentlig sektor, Regjeringen.no](#)

A2: Metadata er tilgjengelige selv om selve data i kilden ikke er tilgjengelig

Internasjonal versjon

Metadata are accessible, even when the data are no longer available

Veiledende spørsmål

- Er *metadata* tilgjengelig selv om dataene i kilden ikke er tilgjengelig?
- Er *metadata* lagret, og blir de forvaltet, i en driftssikker løsning?
- Finnes det en dokumentert langsiktig plan for forvaltning og tilgjengeliggjøring av *metadata*?

Hvorfor er dette viktig?

- Det å skille på beskrivelsen av data (metadata) og de faktiske data muliggjør bedre søkbarhet og forståelse av innholdet uten å måtte eksponere sensitive data. Forvaltning og lagring av metadata er også generelt mindre ressurskrevende enn lagring av de faktiske data, samt også mindre sensitivt i forhold til lagring over tid. Metadata gir verdifull informasjon, for eksempel ved planlegging av replikasjonsstudier – selv om de opprinnelige dataene ikke lenger er tilgjengelige.

Interoperabilitet (Interoperable)

Handler om at metadata og variabler er beskrevet i henhold til internasjonale standarder, terminologier, klassifikasjoner og kodeverk, som gjør det mulig å bruke datasett sammen med datasett fra andre kilder.

Interoperabilitet	
Kortversjon	Metadata og variabler er beskrevet og dokumentert med utgangspunkt i standarder, terminologier og kodeverk som forvaltes i henhold til FAIR-prinsippene.
Kriterier	I1: Metadata og data er beskrevet ved bruk av definerte begreper og/eller semantiske standarder. I2: Begreper, kodeverk, terminologier og ontologier som anvendes er søkbare, tilgjengelige, interoperable og har en veletablert forvaltningsprosess. I3: Både metadata og data er beskrevet slik at sammenhengen de inngår i er forståelig og presis, og slik at sammenlikning på tvers av kilder muliggjøres.

I1: Metadata og data er beskrevet ved bruk av definerte begreper og/eller semantiske standarder.

Internasjonal versjon

(Meta)data use a formal, accessible, shared and broadly applicable language for knowledge presentation

Veiledende spørsmål

- Samsvarer metadata med Nasjonal spesifisering for metadata om helsedata?
- Er variablene beskrevet ved bruk av en begrepskatalog og/eller en internasjonal terminologi?
- Er de viktigste begrepene som anvendes for å beskrive *variablene* oversatt til engelsk?
- Er kodeverkene (verdisettene) som anvendes basert på et nasjonalt og/eller et internasjonalt kodeverk (klassifisering)?
- Er kodeverkene (verdisettene) som anvendes oversatt til engelsk?

Hvorfor er dette viktig?

- Bruk av utbredte og formelle språk herunder definerte begreper er en forutsetning for at metadata skal være automatisk søkbare og enkelt kunne utveksles på tvers av fagområder, sektorer, landegrensler og teknologier
- Nasjonal spesifisering for metadata om helsedata er i stor grad basert på den norske versjonen (DCAT-AP-NO) av den internasjonale metadata-standarder DCAT-AP

Aktuelle linker

[Nasjonal spesifisering for metadata om helsedata, Direktoratet for e-helse](#)

[Standard for begrepsbeskrivelser, Digitaliseringsdirektoratet](#)

[Veileder for beskrivelse av datasett, Digitaliseringsdirektoratet](#)

[Standard for beskrivelse av datasett og datakataloger \(DCAT-AP-NO\), Digitaliseringsdirektoratet](#)

[Felles datakatalog, Digitaliseringsdirektoratet](#)

[Snomed CT browser](#)

I2: Begreper, kodeverk, terminologier og ontologier som anvendes er søkbare, tilgjengelige, interoperable og har en veletablert forvaltningsprosess

Internasjonal versjon

(Meta)data use vocabularies that follow the FAIR principles

Veiledende spørsmål

- Er begreper, terminologier, ontologier og kodeverk (verdisett) som anvendes søkbare og åpent tilgjengelige?
- Er begreper, terminologier, ontologier og kodeverk (verdisett) som anvendes beskrevet med utfyllende metadata, herunder historikk over endringer, evt. lisenser for bruk m.m.?
- Kan begreper, terminologier, ontologier og kodeverk (verdisett) som anvendes deles og/eller utveksles i kjente formater?
- Er begreper, terminologier, ontologier og kodeverk (verdisett) representert på en slik måte at de kan aksesseres maskinelt gjennom et programmeringsgrensesnitt (API) og er egnet for automatisert, maskinell prosessering?
- Eksisterer det en veletablert forvaltningsprosess for begreper, terminologier, ontologier og kodeverk (verdisett) som anvendes?

Hvorfor er dette viktig?

- På samme måte som metadata og data i helsedatakildene, må offisielle begreper, terminologier, ontologier og kodeverk (verdisett) som anvendes være søkbare, tilgjengelige, interoperable og gjenbrukbare og tilfredsstillende FAIR-kriteriene. For å kunne utnytte metadata på en effektiv måte må beskrivelsen av innholdet være uttrykket på en slik måte at mottakeren av dataen kan forstå kontekst og koblingen mellom de ulike elementene. For at metadata også skal kunne maskinelt tolkes, må begreper og relasjonen mellom begreper beskrives på en formalisert og strukturert måte. Her vil bruk av OID'er og standarder for kodeverk og terminologi være sentralt. Maskinlesbare metadata vil også tilrettelegge for bedre, raskere og mer kostnadseffektiv kobling av data fra ulike kilder.

Aktuelle linker

[Volven](#)

[FinnKode](#)

[Snomed CT browser](#)

[Felles datakatalog, Digitaliseringsdirektoratet](#)

[Norske basisprofiler for HL7 FHIR](#)

I3: Både metadata og data er beskrevet slik at sammenhengen de inngår i er forståelig og presis, og slik at sammenlikning på tvers av kilder muliggjøres

Internasjonal versjon

(Meta)data include qualified references to other (meta)data

Veiledende spørsmål

- Er *metadata* dokumentert i en informasjonsmodell som viser sammenhengen de inngår i?
- Er informasjonsmodellen som *metadataene* er basert på dokumentert og tilgjengelig?
- Er informasjonsmodellen basert på, eller mappet til, en internasjonal standard (DCAT)?
- Er *variablene* dokumentert i en begrepsmodell som viser sammenhengen/konteksten de inngår i?
- Er begrepsmodellen som *variablene* er basert på tilgjengelig?
- Har variablene referanser til en begrepskatalog og/eller en internasjonal terminologi?
- Er *kilden*, evt. *datasamlinger* og *datasett* beskrevet med metadata, evt. mappet mot en internasjonal standard, som gjør det mulig å tematisere dem?
- Er *variablene* beskrevet med metadata, evt. mappet mot en internasjonal standard, som gjør det mulig å tematisere variabler med tilsvarende meningsinnhold på tvers av datakilder?
- Er evt. mappinger (mapsett) og/eller relasjoner tilgjengelige for deling og/eller utveksling gjennom et standardisert format?

Hvorfor er dette viktig?

- For å bedre kunne forstå innholdet i en samling av dataobjekter og deres relasjon til andre dataobjekt er det ikke nok å beskrive de digitale objektene isolert sett, men også relasjonen mellom disse, og konteksten for hva innholdet uttrykker. Sammenhengen beskrives ved å uttrykke relasjoner mellom ulike deler av de aktuelle metadataene både i forhold til ulike begreper, terminologier, ontologier og samlinger av dataobjekter. Dette er spesielt viktig når man skal vurdere og sammenlikne metadata og data på tvers av forskjellige helsedatakilder.

Aktuelle linker

[Contsys](#)

[Eksempel Nationell informationsstruktur, Sosialstyrelsen](#)

[HL7 FHIR](#)

[Snomed CT browser](#)

[Variabelbiblioteket](#)

Gjenbrukbarhet (Reusable)

Handler om at det finnes informasjon i form av metadata som beskriver vilkårene for gjenbruk av datasettet, hvordan datasettet er bearbeidet og kvaliteten på dataene.

Gjenbrukbarhet	
Kortversjon	Datakilden, evt. datasamlinger, datasett og variabler er beskrevet godt nok til at innholdet enkelt lar seg vurdere og evt. gjenskape
Kriterier	<p>R1: Datakilden, evt. datasamlinger og datasett er beskrevet med utfyllende metadata</p> <p>R1.1: Det finnes metadata som beskriver vilkårene for hvordan datagrunnlaget kan benyttes</p> <p>R1.2: Det finnes metadata som beskriver hvordan datagrunnlaget er bearbeidet, f.eks. avledet, beregnet, anonymisert etc. på veien fra kilden til slik dataene foreligger nå</p> <p>R1.3: Beskrivelsen av metadata er basert på internasjonale og domenebaserte begreper og standarder</p>

R1: Datakilden, evt. datasamlinger og datasett er beskrevet med utfyllende metadata

Internasjonal versjon

Meta(data) are richly described with a plurality of accurate and relevant attributes

Veiledende spørsmål

- Finnes det *metadata* med opplysninger om dataansvarlig, databehandler, utgiver, kontaktopplysninger, etc.?
- Finnes det *metadata* med opplysninger om formålet med å samle inn datagrunnlaget?
- Finnes det metadata med opplysninger om hvorfra og hvordan datagrunnlaget er innhentet?
- Finnes det metadata med opplysninger som sier noe om datakvaliteten, som f.eks. kompletthet (Dekningsgrad), korrekthet (kodingskvalitet), etc.?
- Finnes det metadata med opplysninger om historikk og endringer i f.eks. navn, beskrivelser, kodeverk, svaralternativ, etc.?
- Inneholder metadataene aggregert statistikk på svaralternativnivå, for eksempel fordeling av antall døde per år fordelt på kjønn?

Hvorfor er dette viktig?

- Det er mye lettere å finne og gjenbruke data hvis det er mange elementer knyttet til dataene. Prinsipp R1 er relatert til F2, men R1 fokuserer på brukerens evne (maskin eller menneske) til å bestemme om dataene egentlig er nyttige i en bestemt sammenheng. For å kunne avgjøre dette, skal datautgiveren ikke bare gi metadata som tillater utforsking av dataen, men også metadata som rikelig beskriver sammenhengen der dataene ble generert. Dette kan inkludere eksperimentelle protokoller, produsenten og merkevaren til maskinen eller sensoren som genererte dataene, artene som brukes, stoffregimet etc. Videre sier R1 at datautgiveren ikke bør forsøke å forutsi datakonsumentens behov, men heller være så generøs som mulig når det gjelder dokumentasjon av metadata. Selv om informasjon i enkelte tilfeller kan virke irrelevant, vil den i andre tilfeller være helt avgjørende.

Aktuelle linker

[Nasjonal spesifisering for metadata om helsedata, Direktoratet for e-helse](#)

[Spesifisering for beskrivelse av kvalitet på datasett, Digitaliseringsdirektoratet](#)

[Data Management Plan, Universitetet i Oslo](#)

[Tilgjengeliggjøring av forskningsdata Revidert 2017 \(PDF\), Norges forskningsråd](#)

[Retningslinjer ved tilgjengeliggjøring av offentlige data, Regjeringen.no](#)

R1.1: Det finnes metadata som beskriver vilkårene for hvordan datagrunnlaget kan benyttes

Internasjonal versjon

(Meta)data are released with a clear and accessible data usage license

Veiledende spørsmål

- Gir *metadata* gode og tydelige retningslinjer for hvordan dataene kan benyttes (juridisk)?
- Inneholder *metadata* all relevant informasjon om samtykker, regelverk, beskrivelse av formål og relaterte godkjenninger (lisenser) for tilgang til og bruk av datagrunnlag?

Hvorfor er dette viktig?

- Tilgjengeliggjøring av offentlige data er et viktig bidrag til innovasjon, næringsutvikling og åpenhet i samfunnet. Å gi tilgang til offentlige data betyr at næringsliv, forskere, sivilsamfunn og offentlig sektor selv kan gjøre nytte av informasjon offentlig sektor forvalter for verdiskaping, økt effektivitet og økt åpenhet og transparens. For å kunne tilrettelegge for gjenbruk av data er det viktig at dataene kommer med klare retningslinjer og betingelser for hvordan de kan gjenbrukes. Gjerne i form av en datalicens f. eks. Norsk lisens for offentlig data.

- Informasjon som er underlagt taushetsplikt, skal ikke tilgjengeliggjøres for videre bruk. Dette gjelder både personlige forhold og forretningshemmeligheter (forvaltningsloven §13). Fødested, fødselsdato, statsborgerskap, sivilstand, yrke, bopel eller arbeidssted regnes i denne sammenheng ikke som personlige forhold, med mindre de kan røpe et klientforhold eller lignende.

Linker

[Nasjonal spesifikasjon for metadata om helsedata, Direktoratet for e-helse](#)

[Data Management Plan, Universitetet i Oslo](#)

[Tilgjengeliggjøring av forskningsdata Revidert 2017, Norges forskningsråd](#)

[Retningslinjer ved tilgjengeliggjøring av offentlige data, Regjeringen.no](#)

[Oppsummering av vilkårene i Norsk lisens for offentlige data \(NLOD\)](#)

R1.2: Det finnes metadata som beskriver hvordan datagrunnlaget er bearbeidet, f.eks. avledet, beregnet, anonymisert etc. på veien fra kilden til slik dataene foreligger nå

Internasjonal versjon

(Meta)data are associated with detailed provenance

Veiledende spørsmål

- Finnes det *metadata* med opplysninger om hvordan datagrunnlaget er bearbeidet f.eks. rekoding, normalisering, avledninger, beregninger, sammenstilling og anonymisering?
- Blir spørringer, script, (evt. verktøy) og algoritmer for uthenting, behandling og sammenstilling av data lagret (hele veien fra opprinnelig kilde, delresultater frem til sluttresultat/utlevert datasett) – slik at eksakt samme resultat vil kunne gjenskapes ved rekjøring på et senere tidspunkt?
- Finnes det metadata med opplysninger om hvordan bruk av data fra datakilden skal siteres eller refereres til?

Hvorfor er dette viktig?

- Både for bruk og gjenbruk av data er det viktig å vite hvor data kommer fra opprinnelig og om data er bearbeidet underveis – i så fall av hvem og på hvilken måte. Dette gjelder både for å kunne dokumentere historikk for datakilden, og for å kunne ha kontroll på referanser, og eksempelvis akkreditering ved publisering. For å kunne ha denne oversikten er det viktig at hele arbeidsprosessen og historikk for bearbeidelse av dataen beskrives. Viktige spørsmål er: Hvem genererte eller samlet inn dataen? Hvordan har dataen blitt behandlet? Har dataen blitt publisert tidligere? Inneholder den data fra andre kilder som grunnlag? Ideelt sett bør arbeidsprosessen være beskrevet i et maskinlesbart format.

R1.3: Beskrivelsen av metadata er basert på internasjonale og domenebaserte begreper og standarder

Internasjonal versjon

(Meta)data meet domain-relevant community standards

Veiledende spørsmål

- Ivaretar metadata relevante standarder for domenet det beskriver?
- Forvaltes data og *metadata* iht. Digitaliseringsdirektoratets «Rammeverk for informasjonsforvaltning»?

Hvorfor er dette viktig?

- Det er lettere å gjenbruke, og koble, datasett hvis de er like: samme type data, data er organisert på en standardisert måte, data er basert på veletablerte filformater, dokumentasjon av metadata er basert på en standardisert mal og bruker standardisert begrepsbruk. Hvis domenestandarder eller beste praksis for dataarkivering og deling eksisterer, bør de følges. For eksempel har flere områder innen helsesektoren informasjonsstandarder som sier noe om hva som er minimum av hva som skal dokumenteres (eksempelvis MIAME og MIAPE for biodata). FAIR-data skal som et minimum oppfylle denne type standardene. Andre domenestandarder kan være mindre formelle, men fremmer likevel publisering av metadata og data på en måte som gjør dataen mer gjenbrukbar. Dette er det primære formålet med FAIRness.
- Med domene menes f.eks. helseregistre, helseundersøkelser, biobanker, labsvardata, datasett generert av/for forskere m.fl.
- Bruk av internasjonale og domenebaserte begreper og standarder for å beskrive metadata tilrettelegger for gjenbruk – også i andre sektorer og land.

Aktuelle linker

[Rammeverk for informasjonsforvaltning, Digitaliseringsdirektoratet](#)

[Minimum Information About a Microarray Experiment \(MIAME\)](#)

[Minimum Information About a Proteomics Experiment \(MIAPE\)](#)

Appendiks

Begrepsdefinisjoner

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

Begrep	Forklaring	Kilde
Aggregerte data	<p>Aggregering betyr innen statistikk å kombinere eller slå sammen data om enkeltindivider, grupper eller perioder til større grupper eller lengre tidsperioder. Aggregering innebærer at mengden informasjon reduseres.</p> <p>Aggregerte data brukes her om opplysninger som er slått sammen slik at det ikke er mulig å identifisere enkeltpersoner. Andre begreper som ofte brukes om aggregerte data er statistikk, tabelldata, aidentifiserte - eller anonymiserte opplysninger.</p>	Helsedata.no [27]
Analyserom	<p>En analyseinfrastruktur som tilbyr et spekter av analyseverktøy, samt funksjonalitet for å lagre data, validere data og sammenstille data. Ofte med høy fokus på personvern og sikkerhet.</p> <p>Eksempler: TSD, Analyserom på HAP</p>	
API	<p>Application programming interface: Grensesnitt i en programvare som åpner for å aktivere (kjøre) spesifikke deler av denne fra en annen programvare. I kontekst av denne utredningen har API to betydninger: (1) Et tjenestegrensesnitt som kan være implementert med ulike teknologier. Dette er i tråd med en moderne bruk av ordet med en bredere betydning, der hensikten er å støtte samhandling mellom virksomheter. (2) I kontekst av API-management betyr API et web-basert grensesnitt som tilbys av en komponent. Betydningen er her snevrere.</p>	<p>Standardiserte tjenestegrensesnitt (API) for helseregistre [28]</p> <p>Felles datakatalog [21]</p>
Autentisering	<p>Å etablere et visst nivå av tillit for at en gitt pålogget bruker er den vedkommende utgir seg for å være. For helseopplysninger kreves i Norge at man er autentisert på Nivå 4.</p> <p>Eksempler: HelseID, ID-porten, BankID, Commfides</p>	Standardiserte tjenestegrensesnitt (API) for helseregistre [28]
Autorisering	<p>Det å gi en bruker eller system adgang til data og/eller funksjoner i et IT-system. Utføres på basis av en autentisert identitet. Hvilken tilgang brukeren får bestemmes oftest ut fra rolle, eventuelt i kombinasjon med andre opplysninger.</p>	Standardiserte tjenestegrensesnitt (API) for helseregistre [28]
Begrepsmodell	<p>Begrepsmodeller er konseptuelle modeller som beskriver begrepenes sammenheng med hverandre uttrykt ved setninger som uttrykker kunnskap med enkel semantikk. Dette gir et meget godt utgangspunkt for å lage Strukturmodeller som er sammensatt av entiteter</p>	Folkeregisteret API dokumentasjon, Skatteetaten [29]

Begrep	Forklaring	Kilde
	og egenskaper på en logisk måte hvor entiteter og egenskaper er basert på begrepene i Begrepsmodellen.	
Data	Data brukes i Helsedataprogrammet i praksis som synonymt med informasjon i helseregistrene.	Standardiserte tjenestegrensesnitt (API) for helseregistre [28]
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 En fysisk eller juridisk person, en offentlig myndighet, en institusjon eller ethvert annet organ som alene eller sammen med andre bestemmer formålet med behandlingen av personopplysninger og hvilke midler som skal benyttes; når formålet med og midlene for behandlingen er fastsatt i unionsretten eller i medlemsstatenes nasjonale rett, kan den behandlingsansvarlige, eller de særlige kriteriene for utpeking av vedkommende, fastsettes i unionsretten eller i medlemsstatenes nasjonale rett. Kilde: EUs personvernforordning artikkel 4 nr.7	Konseptvalgutredning en Helseanalyseplattform en [9]
Databehandler	Den som behandler personopplysninger på vegne av den [data]behandlingsansvarlige Kilde: Personopplysningsloven § § 2 nr.5 En fysisk eller juridisk person, offentlig myndighet, institusjon eller ethvert annet organ som behandler personopplysninger på vegne av den [data]behandlingsansvarlige Kilde: EUs personvernforordning artikkel 4 nr.8	Konseptvalgutredning en Helseanalyseplattform en [9]
Datakilde	Datakilde er definert som kildene til dataene. I denne sammenheng er et helseregister ekvivalent med en datakilde.	Nasjonal spesifisering for metadata om helsedata [2]
Datakvalitet	ISO-definisjonen (8402-1986) av datakvalitet er: «The totality of features and characteristics of an entity that bears on in its ability to satisfy stated and implied needs» En enkel definisjon av datakvalitet er at det er et mål på hvorvidt data kan anvendes i henhold til intensjonen. Følgende seks kvalitetsdimensjoner benyttes for å beskrive datakvalitet: o Kompletthet: I hvilken grad alle tilfeller av en målpopulasjon er inkludert o Korrekthet / validitet: I hvilken grad data er gyldige og gir et riktig bilde av virkeligheten o Aktualitet: I hvilken grad data er oppdatert – hvor lang tid tar det fra en hendelse skjer, til data er oppdatert og kvalitetssikret	Konseptvalgutredning en Helseanalyseplattform en [9]

Begrep	Forklaring	Kilde
	<p>o Sammenlignbarhet: I hvilken grad data lar seg sammenligne over tid, mellom ulike institusjoner og på tvers av geografi</p> <p>o Reliabilitet: Hvor pålitelig / nøyaktig en målemetode er (hvor reproduserbare data er)</p> <p>o Relevans: I hvor stor grad data oppfyller nåværende og fremtidige behov hos brukerne (*)</p> <p>Kilde: Valideringshåndboken, Nasjonalt Servicemiljø for Medisinske Kvalitetsregistre: https://www.kvalitetsregistre.no/valideringshandboken</p> <p>(*) Relevans ligger ikke i Valideringshåndboken når denne rapporten skrives, men Helseanalyseplattformen anser det som en viktig kvalitetsdimensjon å ha med</p>	
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.	Nasjonalt spesifisering for metadata om helsedata [2]
Datasett	Et datasett er samling av data som er betraktet som en enhet. I denne veilederen brukes begrepet synonymt med hvilke metadata og/eller data som inngår i en definert versjon av en bestemt datakilde/datasamlingen,	Nasjonalt spesifisering for metadata om helsedata [2] Felles datakatalog [21]
Dekningsgrad	Et registers dekningsgrad er et mål på hvor stor andel av målpopulasjonen som finnes i registeret. Dekningsgraden påvirker hvordan resultater kan eller bør tolkes. Høy dekningsgrad er avgjørende for at registeret skal kunne tjene sitt formål. Dekningsgraden er kanskje den viktigste indikatoren på et registers datakvalitet. Men selv om høy dekningsgrad er en forutsetning for god datakvalitet, er det ikke slik at høy dekningsgrad nødvendigvis medfører god datakvalitet – høy dekningsgrad er en nødvendig, men ikke tilstrekkelig betingelse for god datakvalitet.	Målpopulasjon og dekningsgrad, Nasjonalt Servicemiljø for Medisinske Kvalitetsregistre [30]
Domene	Domene betyr opprinnelig krongods (via fransk domaine fra latin dominus = herre). Felt som en rår over; interesse- el. Spesialområde. I denne sammenheng brukes ordet for et virksomhets- og aktivitetsområde som forskningsfelt eller organisatorisk område. Eksempler: Sentrale helseregistre, nasjonale kvalitetsregistre, helseundersøkelser, biobanker	Kriterier for FAIR forskningsdata VR2018 [11] Felles datakatalog [21]
Filoverførings-tjeneste	Et webtjeneste for filoverføring, dvs. deling av filer som er for store eller sensitive til å legge ved en e-post ofte med høyt fokus på sikkerhet. Eksempel: Filoverføringstjenesten [31]	
Globalt unik ID	Globalt unik identifikator som kun kan referere til ett eneste digitalt objekt. Eksempler: URI, DOI, GUID	GoFAIR [14]

Begrep	Forklaring	Kilde
ID	En identifikator er en streng som gir navn til en entitet som et dataprogram, en variabel, en klasse, et objekt, en funksjon, en prosedyre el.l.	
Informasjonsmodell	<p>Informasjonsmodeller angir hva slags informasjon som skal brukes og hvordan informasjonselementer sorteres og lagres i forhold til hverandre, internt i et system eller ved utveksling mellom systemer.</p> <p>Formalisert beskrivelse av et begreps egenskaper og relasjoner mellom disse. Informasjonsmodellen beskriver den informasjonen en virksomhet trenger å motta eller selv produserer for å utføre sitt daglige virke. En felles informasjonsmodell er en modell til felles bruk på tvers av virksomheter forretningsområder og/eller applikasjonssegmenter</p> <p>Kilde: Skatteetaten gjennom Difi: https://www.difi.no/artikkel/2016/09/begreper-fellesinformasjonsmodeller</p>	<p>Felles språk i helse- og omsorgssektoren Målbilde versjon 1.0 [32]</p> <p>Konseptvalgutredning en Helseanalyseplattform en [9]</p> <p>Felles datakatalog [21]</p>
Interoperabilitet	<p>En egenskap ved et produkt eller et system som innebærer at dets grensesnitt er fullstendig forstått, slik at det kan arbeide sammen med andre produkter eller systemer, nåværende eller fremtidige, i en hvilken som helst implementasjon eller tilgang, uten noen restriksjoner.</p> <p>Semantisk interoperabilitet er muligheten for et datasystem å utveksle data med et annet system uten å være avhengig av at personer må tolke dataenes betydning («semantikk»).</p> <p>Kilde: Wikipedia</p> <p>interoperabilitet er evnen til ett produkt eller system, for hvilket alle grensesnitt er fullstendig oppgitt, å samhandle og fungere med andre produkter eller systemer, uten noen tilgang- og implementasjonsrestriksjoner.</p>	<p>Standardiserte tjenestegrensesnitt (API) for helseregistre [28]</p> <p>Felles datakatalog [21]</p>
Klassifikasjon	<p>En klassifikasjon er en samling unike begreper med tilhørende koder i meningsbærende hierarkier (17) (18) definert i ISO 17115 slik: "-- an exhaustive set of mutually exclusive categories to aggregate data at a pre- prescribed level of specialization for a specific purpose".</p> <p>I klassifikasjoner har kodene en strengt hierarkisk og strukturert inndeling, der alle begrep er plassert under ett, og kun ett, forelder-begrep. Kategoriseringen er basert på en eller flere logiske regler. Klassifikasjoner må i tillegg ha kodingsregler for bruk av kodene for å få konsistens i kodingen, og for å kunne sammenligne kodet data over tid og mellom ulike geografiske lokalisasjoner.</p> <p>Klassifikasjoner er laget og utformet for bruk til standardisert koding av informasjon, for statistiske formål.</p> <p>Et typisk eksempel på en klassifikasjon er WHO's International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10).</p>	<p>Felles språk i helse- og omsorgssektoren Målbilde versjon 1.0 [32]</p>

Begrep	Forklaring	Kilde
Kode	Kode er i denne sammenheng det samme som svaralternativene i et verdisett.	Nasjonal spesifisering for metadata om helsedata [2]
Kodeverk	Et kodeverk er en samling unike begreper med tilhørende kode. Typiske eksempler vil være kodeverkene som inngår i en rekke meldingsstandarder eller rapporteringer og som publiseres på Volven. En klassifisering er også et kodeverk i denne sammenheng.	Felles språk i helse- og omsorgssektoren Målbilde versjon 1.0 [32]
Maskinlesbar	Informasjon i et format som er strukturert slik at programvarer enkelt kan identifisere, ekstrahere og behandle informasjonens individuelle bestanddeler. Maskinlesbare data er data i et format som kan leses og bearbeides av en datamaskin og dermed lett kan deles på tvers av IT-systemer.	Kriterier for FAIR forskningsdata VR2018 [11] Felles datakatalog [21]
Metadata	Metadata er data om data. Et eksempel kan være dekningsgrad på en variabel, eller en liste over lovlige verdier for en variabel. (Fra gresk meta "«om"» og latin data "«opplysninger"».) Data som tjener til å definere eller beskrive andre data. Kilde: Wikipedia. Metadata, data om data, informasjon som beskriver annen informasjon. Metadata inneholder typisk et emneord, tittel og tidspunkt for opprettelse og endring av dokumentet. Bruk av metadata gjør det enklere å holde oversikt ved at metadata ordnes i en søkbar database sammen med metadata for andre filer. Kilde: Store Norske Leksikon, https://snl.no/metadata	Standardiserte tjenestegrensesnitt (API) for helseregistre [28] Felles datakatalog [21]
Ontologi	En ontologi er innen informasjonsvitenskap en formell representasjon av et sett begreper innenfor et kunnskapsområde. En ontologi definerer forhold mellom disse begrepene ved hjelp av relasjonene de har med hverandre. Disse relasjoner kan også representeres slik at de forstås av en datamaskin. Ontologier gir muligheten til å definere ethvert begrep ved sine relasjoner til flere omliggende begrep. Dette åpner også for muligheten til å definere et polyhierarki der begrep kan være plassert under flere forelderbegrep samtidig. En ontologi kan fremstilles på forskjellige måter, men strukturen bygges basert på regler, og de relasjoner som er angitt. Eksempler: SnomedCT, Consys OWL	Felles språk i helse- og omsorgssektoren Målbilde versjon 1.0 [32]
Persistent identifikator	Konsistent over tid. Identifikatoren er bestandig lenket til metadata som definerer den og til objektet den identifiserer. Dette garanteres av tjenesteleverandøren som tildeler identifikatoren.	GoFAIR [14] Kriterier for FAIR forskningsdata VR 2018 [11]
Protokoller	Kommunikasjonsprotokoll: Formelt regelverk som spesifiserer hvordan utveksling av data mellom to aktører skal foregå og hvilket format som anvendes. Eksempler: HTTP, TCP/IP, FTP	Kriterier for FAIR forskningsdata VR2018 [11]

Begrep	Forklaring	Kilde
Sensitive data	<p>Med sensitive data menes i denne sammenheng sensitive personopplysninger:</p> <p>Sensitive personopplysninger er i personopplysningsloven § 2 nr. 8 definert som opplysninger om rasemessig eller etnisk bakgrunn, eller politisk, filosofisk eller religiøs oppfatning (bokstav a), at en person har vært mistenkt, siktet, tiltalt eller dømt for en straffbar handling (bokstav b), helseforhold (bokstav c), seksuelle forhold (bokstav d) og medlemskap i fagforeninger (bokstav e). Personopplysningsloven § 9 stiller ytterligere krav til behandling av slike opplysninger.</p>	Personopplysningsloven [33]
Serialiseringsformater	Serialisering er å transformere data i internminnet på en datamaskin til en sekvens av byter som kan lagres i en datafil eller overføres over et nettverk. Det finnes mange serialiseringsformater og noen av de mest benyttede er XML og JSOM	
Standardisert	Av standardiseringsorganisasjon fastsatt eller innen domenet fastslått praksis.	Kriterier for FAIR forskningsdata VR2018 [11]
Standardisert API	Ethvert API må ha en form for datakontrakt, og et standardisert API innebærer at innholdsformatet er basert på en nasjonal eller internasjonal standard. Et eksempel kan være FHIR grunnprofiler eller norske profiler.	Standardiserte tjenestegrensesnitt (API) for helseregistre [28]
Statistikk	<p>Tallfestede opplysninger om en gruppe eller et fenomen, som fremkommer ved sammenstilling og bearbeiding av opplysninger om de enkelte enhetene i gruppen eller et utvalg av disse enhetene, eller ved systematisk observasjon av fenomenet.</p> <p>I denne sammenheng er statistikk definert som de statistiske egenskapene ved en årgang/versjon av en variabel.</p>	<p>Felles datakatalog [21]</p> <p>Nasjonal spesifisering for metadata om helsedata [2]</p>
Terminologi	<p>En terminologi er en samling ord og uttrykk (termer) innenfor et fag eller emne. En terminologi er også et kodeverk når termen har en kode knyttet til seg. Terminologiens formål er å lette kommunikasjonen og samhandling mellom personer som arbeider innenfor det samme.</p> <p>I vår sammenheng benyttes begrepet terminologi for å angi at man har en samling begrep med tilhørende koder, og at begrepene er satt i en sammenheng med hverandre.</p>	Felles språk i helse- og omsorgssektoren Målbilde versjon 1.0 [32]
Tilgjengeliggjøring	Databehandler/databehandlingsansvarlig gjør data tilgjengelig for analyse ved utlevering (distribusjon av data) eller ved å gi bruker tilgang til data i et analysemiljø	Standardiserte tjenestegrensesnitt (API) for helseregistre [28]
URL	URL er kort for Uniform Resource Locator og viser til «adresser» på Internett.	
Utlevering	Med «Utlevering» forstås hele prosessen fra søknad er godkjent til data er mottaker (forsker) i hende. Denne prosessen omfatter både uttrekk av data, dialog om detaljer i dette, sammenstilling, og Forsendelse til mottaker.	Standardiserte tjenestegrensesnitt (API) for helseregistre [28]

Begrep	Forklaring	Kilde
Variabel	Variabler er de informasjonsbærende elementene i datakilden og definerer hvilken informasjon som finnes i registrene.	Nasjonal spesifisering for metadata om helsedata [2]
Verdi	Verdi er definert som antall gyldige tilfeller av et svaralternativ/utfall av en variabel.	Nasjonal spesifisering for metadata om helsedata [2]

Referanser

- [1] «Helsedataprogrammet,» Direktoratet for e-helse, [Internett]. Available: <https://ehelse.no/programmer/helsedataprogrammet>.
- [2] «Nasjonal spesifikasjon for metadata om helsedata,» Direktoratet for e-helse, 2020. [Internett]. Available: <https://ehelse.no/standarder/ikke-standarder/nasjonal-spesifikasjon-for-metadata-om-helsedata>.
- [3] «Veileder for beskrivelse av datasett,» Digitaliseringsdirektoratet, [Internett]. Available: <https://doc.difi.no/veiledere/veileder-for-beskrivelse-av-datasett/>.
- [4] G. L. C. H. Summit, «European Commission Press Release Database,» 5 September 2016. [Internett]. Available: https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967.
- [5] «Den europeiske dataportalen,» Publications Office of the European Union, [Internett]. Available: <https://www.europeandataportal.eu/no>.
- [6] «European Open Science Cloud, Brussels,» EOSC Summit, 2017. [Internett]. Available: http://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf#view=fit&pagemode=none.
- [7] «The European Code of Conduct for Research Integrity,» Allea, All European Academics, Mars 2017. [Internett]. Available: https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf.
- [8] «Nordisk program for helse og velferd,» NordForsk, [Internett]. Available: <https://www.nordforsk.org/sv/programs/nordisk-program-helse-og-velferd>.
- [9] «Konseptvalgutredning for Helseanalyseplattformen,» Direktoratet for e-helse, April 2018. [Internett]. Available: <https://ehelse.no/publikasjoner/konseptvalgutredning-for-helseanalyseplattformen>.
- [10] «A vision of a Nordic secure digital infrastructure for health data: The Nordic Commons,» NordForsk, 2019. [Internett]. Available: <https://www.nordforsk.org/2020/vision-nordic-secure-digital-infrastructure-health-data-nordic-commons>.
- [11] «Kriterier för FAIR forskningsdata,» Vetenskapsrådet, 2018. [Internett]. Available: <https://www.vr.se/analys/rapporter/vara-rapporter/2018-12-07-kriterier-for-fair-forskningsdata.html>.
- [12] «Results of an Analysis of Existing FAIR Assessment Tools,» Research Data Alliance, 2019. [Internett]. Available: <https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/results-analysis-existing-fair-assessment-tools>.

- [13 «The FAIR data principles,» *Nature/ Scientific Data*, March 2016.
]
- [14 «Go FAIR,» [Internett]. Available: <https://www.go-fair.org/fair-principles/>.
]
- [15 «The FAIR data principles,» FORCE11, [Internett]. Available:
] <https://www.force11.org/group/fairgroup/fairprinciples>.
- [16 «Retningslinjer ved tilgjengeliggjøring av offentlige data,» Kommunal- og moderniseringsdepartementet, 27 01 2017. [Internett]. Available: <https://www.regjeringen.no/no/dokumenter/retningslinjer-ved-tilgjengeliggjoring-av-offentlige-data/id2536870/>. [Funnet 16 03 2020].
- [17 «Digitaliseringsrundskrivet,» [Internett]. Available:
] <https://www.regjeringen.no/no/dokumenter/digitaliseringsrundskrivet/id2623277/>.
- [18 «Veileder for orden i eget hus,» Digitaliseringsdirektoratet, [Internett]. Available:
] <https://www.difi.no/fagomrader-og-tjenester/digitalisering-og-samordning/nasjonal-arkitektur/informasjonsforvaltning/veileder-orden-i-eget-hus>.
- [19 «Informasjonsforvaltning i offentlig sektor,» Digitaliseringsdirektoratet, [Internett].
] Available: <https://www.difi.no/sites/difino/files/rapport-informasjonsforvaltning-i-offentleg-sektor-2013-10-10.pdf>.
- [20 «Rammeverk for informasjonsforvaltning,» Digitaliseringsdirektoratet, [Internett].
] Available: <https://www.difi.no/fagomrader-og-tjenester/digitalisering-og-samordning/nasjonal-arkitektur/informasjonsforvaltning/rammeverk-informasjonsforvaltning>.
- [21 «Felles datakatalog,» Digitaliseringsdirektoratet, [Internett]. Available:
] <https://data.norge.no/>.
- [22 «Norsk arkitekturrammeverk for samhandling,» Digitaliseringsdirektoratet, [Internett].
] Available: <https://www.difi.no/fagomrader-og-tjenester/digitalisering-og-samordning/nasjonal-arkitektur/arkitekturrammeverk-samhandling>.
- [23 «European Interoperability Framework,» European Union, [Internett]. Available:
] https://ec.europa.eu/isa2/eif_en.
- [24 «Del og gjenbruk data,» Digitaliseringsdirektoratet, [Internett]. Available:
] <https://www.digdir.no/digitalisering-og-samordning/prinsipp-4-del-og-gjenbruk-data/1061>.
- [25 «Standard for beskrivelse av datasett og datakataloger (DCAT-AP-NO),»
] Digitaliseringsdirektoratet, [Internett]. Available: <https://data.norge.no/specification/dcat-ap-no/>.
- [26 «Data Catalog Vocabulary (DCAT) - Version 2,» W3C Recommendation, February 2020. [Internett]. Available: <https://www.w3.org/TR/vocab-dcat-2/>.

- [27 «Helsedata.no,» Direktoratet for e-helse, [Internett]. Available: <https://helsedata.no/>.
]
- [28 «Standardiserte tjenestegrensesnitt (API) for helseregistre,» Direktoratet for e-helse, 2018 03 26. [Internett]. Available: <https://ehelse.no/publikasjoner/standardiserte-tjenestegrensenitt-api-for-helseregistre>.
- [29 «Api dokumentasjon - Informasjonsmodell,» Folkeregisteret, [Internett]. Available: <https://skatteetaten.github.io/folkeregisteret-api-dokumentasjon/informasjonsmodell/>.
- [30 «Målpopulasjon og dekningsgrad,» Nasjonalt Servicemiljø for Medisinske Kvalitetsregistre, [Internett]. Available: <https://www.kvalitetsregistre.no/malpopulasjon-og-dekningsgrad#idx-1>.
- [31 N. Helsenett, «Filoverføringstjenesten,» Norsk Helsenett, [Internett]. Available: <https://www.nhn.no/filoverfoeringstjenesten/>.
- [32 «Felles språk i helse- og omsorgssektoren,» Direktoratet for e-helse, [Internett]. Available: <https://ehelse.no/publikasjoner/felles-sprak-i-helse-og-omsorgssektoren-malbilde-versjon-1.0>.
- [33 «Prop. 56 LS (2017-2018) Lov om behandling av personopplysninger (personopplysningsloven) og samtykke til deltakelse i en beslutning i EØS-komiteen om innlemmelse av forordning (EU) nr. 2016/679 (generell personvernforordning) i EØS-avtalen,» Regjeringen.no, [Internett]. Available: <https://www.regjeringen.no/no/dokumenter/prop.-56-ls-20172018/id2594627/?ch=7>.