

Utkast til innspill – versjon 0.9

Veileder for bruk av FAIR- prinsippene for helsedatakilder



Tittel:

Veileder for bruk av FAIR-prinsippene for
helsedatakilder

Rapportnummer

HITR 1236 utkast 2020

Utgitt:

06/2020

Dokumenttype

Veileder

Utgitt av:

Direktoratet for e-helse

Kontakt:

postmottak@ehelse.no

Publikasjonen kan lastes ned på:

www.ehelse.no

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 dette området.

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 innføring av FAIR-prinsippene, og for å identifisere forbedringsområder.

Dette dokumentet inneholder en norsk versjon av FAIR-prinsippene og er utarbeidet av Helsedataprogrammet¹ i regi av Direktoratet for e-helse. Arbeidet legger de internasjonale prinsippene til grunn, men har i tillegg til å etablere en norsk oversettelse, utarbeidet detaljerte spørsmål og et tilhørende evalueringsskjema. Dette muliggjør at dataforvaltere i sektoren enkelt kan utføre en evaluering for å vurdere sin oppfyllelse av FAIR-prinsippene. Formålet med arbeidet er å bidra til et felles vokabular, og tilrettelegge for mer effektiv deling og gjenbruk av helsedata.

Denne versjon av de norske FAIR-prinsippene er i hovedsak rettet mot norske helseregistre. Prinsippene vil også kunne brukes til evaluering av andre helsedatakilder som f.eks. helseundersøkelser, biobanker og forskerdatasett.

¹ <https://ehelse.no/programmer/helsedataprogrammet>

Hva er FAIR-prinsipper?

FAIR-prinsippene [4] 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 gitt i Figur 1.

De norske FAIR-prinsippene er beskrevet utførlig f.o.m. side 8 i denne veilederen.

Figur 1: FAIR principles [5]

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.

Hva er 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 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 evaluering og forklaringer til hvorfor dette er viktig.

FAIR-scoren som resulterer fra evalueringen skal gi helseregistrene økt innsikt i egen oppfyllelse av FAIR-prinsippene. Videre vil en sammenstilling av FAIR-scorene fra ulike helseregistre kunne øke forståelsen for den generelle graden av ivaretagelse av de ulike prinsippene i Norge. På nasjonalt nivå vil evalueringene kunne gi et godt bilde av hvilke områder som bør ha fokus for å i større grad tilrettelegge for deling og gjenbruk av data, og

være grunnlag for prioriteringer.

Selvevalueringsmodellen kan lastes ned på:
www.ehelse.no

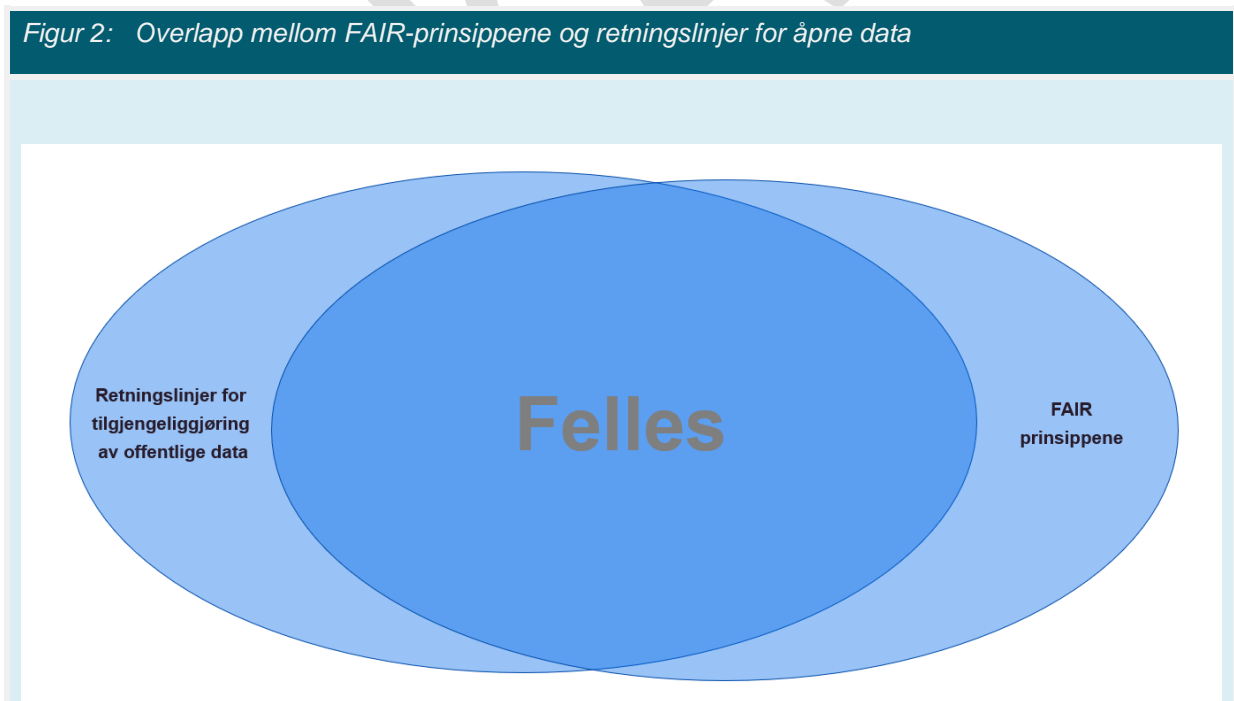
Hvorfor anbefales bruk av FAIR-prinsippene?

Understøtter retningslinjer ved tilgjengeliggjøring av offentlige data

Statens IKT-politikk innebærer et mål om økt gjenbruk og viderebruk av åpne offentlige data. Regjeringens føringer for deling av åpne offentlige data er regulert i regjeringens «Retningslinjer ved tilgjengeliggjøring av offentlige data» [12]. 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. Se appendix for alle 15 retningslinjene. Ved å gi tilgang til data som brukere selv kan gjøre nytte av, bidrar retningslinjene til innovasjon, næringsutvikling og transparens.

De norske FAIR-prinsippene overlapper i stor grad med retningslinjer ved tilgjengeliggjøring av offentlige data som illustrert i Figur 2 men gir en mer konkret veiledning i hav som skal til for å oppfylle retningslinjene. I tillegg favner FAIR-prinsippene enda bredere og gir veiledning i flere aspekter.

Figur 2: Overlapp mellom FAIR-prinsippene og retningslinjer for åpne data



Utbredelse

Det er stor interesse for og økende antall implementasjoner av FAIR-prinsippene både nasjonalt og internasjonalt.

Prinsippene ble anbefalt av G20 toppledere Hangzhou Summit China 4-5 September 2016 [6]. De begynner å få stor utbredelse i europeiske forskningsmiljøer og benyttes blant annet i EOSC (European Open Science Cloud) [7] som er et omfattende EU initiativ med målsetning å understøtte datadrevet forskning og innovasjon og i "The European Code of Conduct for Research Integrity" [8] fra ALLEA² som er en europeisk sammenslutning av vitenskapsselskap og består av 59 akademi fra 40 land fra Europarådets region.

FAIR-prinsippene benyttes av det nordiske programmet for helse og velferd³ 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» [13]. 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" [14] utarbeidet av det svenske Vetenskapsrådet, og RDA sin analyse av eksisterende FAIR-scoringverktøy - "Results of an Analysis of Existing FAIR assessment tools" [15].

Bygger på arkitekturprinsippene

Arbeidet med FAIR-prinsippene bygger på arbeid med arkitekturprinsipper i Helsedataprogrammet og den norske versjonen er lagt tett opp til de internasjonale FAIR-prinsippene. Det kan være behov for sektorielle, virksomhetsspesifikke og domenespesifikke tilpasninger, men dette bør minimeres for å oppnå enklere vedlikehold og tilrettelegging for samarbeid. En oversikt over de ulike nivåene er nærmere illustrert i Figur 3.

² [Allea, All European Academics](#)

³ [Nordic Programme on Health and Welfare, Nordforsk](#)

Figur 3: Ulike nivåer for arkitekturprinsipper



De Norske FAIR-prinsippene

Dette kapitlet 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.

Figur 4: Norsk versjon av FAIR-prinsippene

F	Søkbarehet (Findable)	A	Tilgjengelighet (Accessible)
	<ul style="list-style-type: none"> F1: Metadata og data har en globalt unik og konsistent 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 		<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 åpne, tilgjengelige 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
I	Interoperabilitet (Interoperable)	R	Gjenbrukbarhet (Reusable)
	<ul style="list-style-type: none"> I1: Metadata og data er beskrevet med utgangspunkt i 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. 		<ul style="list-style-type: none"> R1: Datakilden, evt. datasamlinger, datasett og dataprodukt 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 bearbejdet, 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

Søkbarehet (Findable)

Figur 5: Søkbarehet (Findable):

Kortversjon	Metadata og data er godt beskrevet, de har en unik og persistent ID, og er åpent tilgjengelig via en utforsker på en website.
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 (persistent)?
- 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.
- For eksempel: En 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](#)

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

Internasjonal versjon

Data are described with rich metadata

Veiledende spørsmål

- Er kilder, evt. Datasamlinger og datasett beskrevet med strukturerte og maskinlesbare metadata slik at de enkelt (maskinelt) kan tilrettelegges for søk og filtreringer?
- Er variablene med tilhørende metadata strukturert og definert slik at de enkelt (maskinelt) kan tilrettelegges for søk og filtreringer?
- Inneholder metadataene aggregert statistikk på svaralternativnivå? For eksempel fordeling av antall døde per år fordelt på kjønn.

- Er beskrivelsen av metadata dokumentert og tilgjengeliggjort?

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

[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 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?
- Er metadata om datakilden, evt. datasamlinger, datasett, og variabler tematisert slik at søk kan tilrettelegges på tvers av helsedatakilder?
- 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, så vil dette kunne hjelpe. Dersom søk med utgangspunkt i metadata kan kombineres med f.eks. aggregerte data vil dette gi en 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](#)

Tilgjengelighet (Accessible)

Figur 6: Tilgjengelighet (Accessible)

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 veldefinerte grensesnitt og formater A1.1: Protokollene som brukes er åpne, tilgjengelige 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 og formater

Internasjonal versjon

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

Veiledende spørsmål

- Kan *metadata* om kilden, evt. datasamlinger, datasett og variabler deles og/eller utveksles i kjente formater som f.eks. JSON, XML, CSV og Excel?
- 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 *aggregerte (anonymiserte) data* deles og/eller utveksles i kjente formater som f.eks. JSON, XML, CSV og Excel?
- Er *aggregerte (anonymiserte) data* representert på en slik måte at de kan aksesseres maskinelt gjennom et programmeringsgrensesnitt (API) og er egnet for automatisert, maskinell prosessering?
- Kan *sensitive data* deles og/eller utveksles i kjente formater som f.eks. JSON, XML, CSV og Excel?
- Er *sensitive data* representert på en slik måte at de kan aksesseres maskinelt gjennom et programmeringsgrensesnitt (API) og er egnet for automatisert, maskinell prosessering?
- Er tilgangen til *aggregerte (anonymiserte) data* gratis?
- Er tilgangen til *sensitive data* gratis?

- Tilbys tilgjengeliggjøring av komplette predefinerte datasett og/eller dataprodukter fra en eller flere datakilder? F.eks. et utvalg av de mest etterspurte variablene for et forskningsformål med tilhørende metadata og data

Hvorfor er dette viktig?

- 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.
- 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 betalingsssatsene i elektronisk form. Alle opplysninger om grunnlaget for utregning av betalingsssatsene skal også publiseres elektronisk, slik at de er lette å finne for potensielle brukere.
- 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 analystejeneste, 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](#)

[Skatteetatens begrepskatalog - API-beskrivelse](#)

[HL7 FHIR](#)

Spesifisering for dataoverføring til Helseanalyseplattformen

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 og serialiseringsformater som f.eks. XML eller JSON?
- Er *data* tilgjengelige gjennom åpne og gratis protokoller og serialiseringsformater?

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 tilgjengeliggjøres 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 dataen (metadata) og den faktiske dataen muliggjør bedre søkbarhet og forståelse av innholdet uten å måtte eksponere sensitiv data. Forvaltning og lagring av metadata er også generelt mindre ressurskrevende enn lagring av den faktiske dataen, samt også mindre sensitivt i forhold til lagring over tid. Metadata gir verdifull informasjon, for eksempel ved planlegging av replika studier – selv om de opprinnelige dataene ikke lenger er tilgjengelige.

Interoperabilitet

Figur 7: Interoperabilitet (Interoperable)

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

- Er *metadata* basert på en nasjonal og/eller internasjonal standard?
- Er begrepene som anvendes for å beskrive *variablene* basert på nasjonal og/eller internasjonal standard (terminologi)?
- Er begrepene som anvendes for å beskrive *variablene* oversatt til engelsk?
- Er kodeverkene (verdisettene) som anvendes basert på et nasjonalt og/eller et internasjonalt kodeverk (klassifikasjon)?
- 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, landegrensener og teknologier

Aktuelle linker

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

[Generic Statistical Information Model \(GSIM\), High-Level Group for the Modernisation of Official Statistics](#)

[Standard for begrepsbeskrivelser, Digitaliseringsdirektoratet](#)

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

[Felles datakatalog, Digitaliseringsdirektoratet](#)

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, 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 metadataen 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 metadataen også skal kunne maskinelt tolkes, må begreper og relasjonen mellom begreper beskrives på en formalisert og strukturert måte. Her vil bruk av OID'er, standarder for kodeverk og terminologi være sentralt. Maskinlesbare metadata vil også tilrettelegge for bedre, raskere og mer kostnadseffektiv funksjonell kobling av data fra ulike kilder.

Aktuelle linker

[Volven](#)

[FinnKode](#)

[Snomed CT browser](#)

[Felles begrepskatalog \(DigDir\)](#)

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 begrepsmodell og/eller informasjonsmodell som viser sammenhengen de inngår i?
- Er *variablene* dokumentert i en begrepsmodell og/eller informasjonsmodell som viser sammenhengen de inngår i?
- Er begreps- og/eller informasjonsmodellen som *metadataene* er basert på dokumentert og tilgjengelig?
- Er begreps- og/eller informasjonsmodellen som *variablene* er basert på dokumentert og tilgjengelig?
- 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 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 den aktuelle metadataen 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

[Generic Statistical Information Model \(GSIM\), High-Level Group for the Modernisation of Official Statistics](#)

[Contsys](#)

[Eksempel Nationell informationsstruktur, Sosialstyrelsen](#)

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

[HL7 FHIR](#)

[Snomed CT browser](#)

Gjenbrukbarhet (Reusable)

Figur 8: Gjenbrukbarhet (Reusable)

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, datasett og dataprodukter 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, 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 om datakvalitet, dekningsrader, kodingskvalitet, etc.?
- Finnes det *metadata* med opplysninger om historikk og endringer i f.eks. navn, beskrivelser, kodeverk, svaralternativ, etc.?

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

[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 dataen kommer med klare retningslinjer og betingelser for hvordan dataen 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. re-koding, 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, vil det være et behov for å vite hvor dataen kommer fra. Helt fra opprinnelsen og om den er blitt bearbeidet underveis, og i så fall av hvem. Dette gjelder både for å kunne dokumentere historikk for datakilden, men også for å kunne ha kontroll på referanser, og eksempelvis anerkjennelse ved publisering. For å kunne ha denne oversikten er det viktig at hele arbeidsprosessen og historikk for bearbeidelse av dataen beskrives. Dette gjelder både i forhold til; Hvem som har genererte eller samlet dataen? Hvordan har den blitt behandlet? Har den blitt publisert før? Inneholder den data fra andre kilder som grunnlag? Ideelt sett bør denne 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, dataen er organisert på en standardisert måte, dataen 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 \(DigDir\)](#)

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

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

Appendiks

Regjeringens retningslinjer ved tilgjengeliggjøring av offentlige data

Figur 9: Utklipp av Retningslinjer ved tilgjengeliggjøring av offentlige data [2]

1. Bruk åpne standardlisenser

Data som tilbys skal ha klare vilkår for hvordan de kan brukes. Vilkårene skal åpne for så mange bruksområder som mulig. Dersom data som tilgjengeliggjøres er beskyttet etter åndsverkloven, bør virksomheten si ifra seg sine egne eksklusive økonomiske rettigheter til bruk av datasettet. Dette kan enkelt gjøres ved å bruke åpne standardlisenser som Creative Commons 4.0 eller Norsk lisens for offentlige data (NLOD). Disse standardlisensene sikrer en helhetlig praksis for hvilke rettigheter brukerne har, og fritar samtidig utgiver for juridisk ansvar knyttet til datakvalitet og hva data blir brukt til.

2. Tilby data gratis

Hovedregelen er at data skal være gratis, og at det ikke er anledning til å ta betalt for kostnader til innsamling og produksjon av data for viderebruk. 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.

3. Tilby data uten brukerregistrering

Data skal i utgangspunktet gjøres tilgjengelig uten at brukeren må søke om tillatelse eller registrere seg. I tråd med bestemmelsene i offentlighetsloven skal data kunne benyttes hvor som helst, av hvem som helst, og til ethvert formål. Brukere bør samtidig oppmuntres til å oppgi kontaktinformasjon, slik at utgiveren kan varsle brukerne for eksempel ved endringer eller driftsavbrudd.

4. Dokumenter datasettene

Virksomhetene bør dokumentere datasettene slik at det blir enkelt å ta datasettene i bruk både for mennesker og maskiner. Med dokumentasjon mener vi beskrivelser som gjør det mulig for andre å oppdage, forstå og bruke dine data.

5. Tilby informasjon om datakvalitet

Kvaliteten på virksomhetens data påvirker hvor egnet de er til andre formål enn de først ble skapt for. Dokumentering av datakvalitet er til stor hjelp i prosessen med å vurdere om virksomhetens datasett er egnet til andre formål, og øker sjansen for bruk. Datakvaliteten bør derfor være dokumentert, og kjente utfordringer bør eksplisitt omtales i beskrivelsene.

6. Tilby oppdaterte data

Virksomheten bør tilby oppdaterte data og være tydelig på hvilken oppdateringsfrekvens som gjelder. Tjenesten som leverer data bør ha et tjenestenivå tilpasset brukernes behov.

7. Gjør data synlige

For at potensielle brukere av offentlige data enkelt skal kunne finne data, bør beskrivelser av datasett være tilgjengelig på data.norge.no, som er en katalog med beskrivelser av åpne datasett fra det offentlige. Difi gir anbefalinger om formater for dette formålet i dokumentet Standard for beskrivelser av datasett og datakataloger. Virksomheten bør vurdere å tilby beskrivelser på engelsk i tillegg til norsk.

8. Bruk maskinlesbare og standardiserte formater

Data skal være tilgjengelig i maskinlesbare formater. I tillegg bør formatene være standardiserte. Dette sikrer god samhandlingsevne med annen informasjon (interoperabilitet) og legger ikke unødvendige begrensninger på hva informasjonen kan brukes til i fremtiden. Eksempler på maskinlesbare og standardiserte formater er CSV, XML, JSON og RDF-serialiseringer som RDF/XML, JSON-LD og Turtle.

9. Tilby data gjennom et programmeringsgrensesnitt

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.

10. Tilby komplett nedlasting

Selv om et programmeringsgrensesnitt er en svært fleksibel måte å tilby data på, kan brukerne også ha behov for å laste ned komplette datasett for å etablere lokale kopier. Fullstendig nedlasting kan tilbys som funksjonalitet i et programmeringsgrensesnitt, eller som en maskinlesbar fil publisert på internett.

11. Bruk faste adresser og unike identifikatorer

Data bør ha unike, permanente og hensiktsmessige adresser på internett, slik at det er mulig å lenke data sammen. Dette gjelder selve datasettet og versjoner og serier av dette, samt datasettets elementer. Data får økt verdi om de refererer til andres data om samme ting, samme sted, samme hendelse, samme person osv. Det betyr at alle bør bruke samme identifikatorer på tvers av datasett, og legge til rette for at egne identifikatorer kan refereres til av andre.

12. Publisert oversikt over virksomhetens data

Hvilke data som tilgjengeliggjøres bør være drevet av brukernes behov. For at brukerne skal kunne finne og bruke dine data, må de vite hvilke data du har. Virksomheten bør derfor vedlikeholde og publisere en oversikt over hvilke data de forvalter. Dette gjelder også beskrivelser av datasett som av forskjellige grunner ikke er tilgjengeliggjort.

13. Tilpass data til brukernes behov

Virksomhetene bør tilpasse data slik at brukere enkelt kan ta dem i bruk, for eksempel ved å lage systemer for spørringer som er tilpasset spesifikke formål. Slike tilpasninger bør likevel ikke være til hinder for at datasett også blir gjort tilgjengelig i sin opprinnelige form, dersom dette blir etterspurt. Her er det viktig at virksomhetene er åpne for innspill fra de som ønsker å bruke datasettene.

14. Oppmuntre til bruk

Å gjøre data åpent tilgjengelig er ikke nok for at data skal bli brukt. Utgivere bør samhandle med brukerne og aktivt oppmuntre til bruk av deres data.

15. Legg til rette for tilbakemeldinger

Gjennom å tilby brukerne muligheten for å gi tilbakemeldinger, vil virksomheten ha bedre forutsetninger for å forstå behovene til brukerne. Dette vil bidra til å forbedre kvaliteten på publiserte data og til å bygge tillit mellom utgiveren og brukerne.

UTKAST

Begrepsdefinisjoner

Begrep	Definisjon	Eksempler	Kilde	Link til 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. Kilde: Wikipedia</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. Kilde: helsedata.no</p>		Wikipedia Helsedata.no	https://helsedata.no/soknadshjelp-aggregerte-opplysninger/
Analyserom	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.	TSD, Analyserom på HAP		
API	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.	https://fellesdatakatalog.brreg.no/apis	Standardiserte tjenestegrensesnitt (API) for helseregistre	https://ehelse.no/publikasjoner/standardiserte-tjenestegrensenitt-api-for-helseregistre
Autentisering	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.	HelseID, ID-porten, BankID, Commfides	Standardiserte tjenestegrensesnitt (API) for helseregistre	https://ehelse.no/publikasjoner/standardiserte-tjenestegrensenitt-api-for-helseregistre
Autorisering	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.		Standardiserte tjenestegrensesnitt (API) for helseregistre	https://ehelse.no/publikasjoner/standardiserte-tjenestegrensenitt-api-for-helseregistre

Begrep	Definisjon	Eksempler	Kilde	Link til Kilde
Data	"Data" brukes i Helsedataprogrammet i praksis som synonymt med informasjon i helseregistrene.		Standardiserte tjenestegrensesnitt (API) for helseregistre	https://ehelse.no/publikasjoner/standardiserte-tjenestegrensesnitt-api-for-helseregistre
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		Konseptvalgutredningen Helseanalyseplattformen	https://ehelse.no/publikasjoner/konseptvalgutredning-for-helseanalyseplattformen
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		Konseptvalgutredningen Helseanalyseplattformen	https://ehelse.no/publikasjoner/konseptvalgutredning-for-helseanalyseplattformen
Datakilde	Datakilde er definert som kildene til dataene. I denne sammenheng er et helseregister ekvivalent med en datakilde.		Nasjonal spesifisering for metadata om helsedata	
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:		Konseptvalgutredningen Helseanalyseplattformen	https://ehelse.no/publikasjoner/konseptvalgutredning-for-helseanalyseplattformen

Begrep	Definisjon	Eksempler	Kilde	Link til Kilde
	<ul style="list-style-type: none"> 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 o Sammenlignbarhet: I hvilken grad data lar seg sammenligne over tid, mellom ulike institusjoner og på tvers av geografi o Reliabilitet: Hvor pålitelig / nøyaktig en målemetode er (hvor reproduserbare data er) o Relevans: I hvor stor grad data oppfyller nåværende og fremtidige behov hos brukerne (*) <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.		Nasjonal spesifisering for metadata om helsedata	
Datasett	Datasett er et teknisk hjelpemiddel som brukes til å definere hvilke variabler som inngår i ulike årganger/versjoner av de ulike datasamlingene i en datakilde.		Nasjonal spesifisering for metadata om helsedata	
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			https://www.kvalitetsregistre.no/malpopulasjon-og-dekningsgrad#idx-1

Begrep	Definisjon	Eksempler	Kilde	Link til Kilde
	en nødvendig, men ikke tilstrekkelig betingelse for god datakvalitet.			
Domene	Domene betyr opprinnelig krongods (via fransk domaine fra latin dominus = herre). Felt som en rår over; interesseel. Spesialområde. I denne sammenheng brukes ordet for et virksomhets- og aktivitetsområde som forskningsfelt eller organisatorisk område.	Helsesektoren	Wikipedia, Kriterier for FAIR forskningsdata VR2018	https://www.vr.se/analys/rapporter/vara-rapporter/2018-12-07-kriterier-for-fair-forskningsdata.html
Filoverføringstjeneste	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.	Filoverføringstjenesten https://www.nhn.no/filoverfoeringstjenesten/		
Globalt unik ID	Globalt unik identifikator som kun kan referere til ett eneste digitalt objekt.	OID, URI, DOI, GUID	GoFAIR	https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/
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.		Wikipedia	
Informasjonsmodell	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. 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		Felles språk i helse- og omsorgssektoren Målbildeversjon 1.0 Konseptvalgutredningen Helseanalyseplattformen	https://ehelse.no/publikasjoner/felles-sprak-i-helse-og-omsorgssektoren-malbildeversjon-1.0 https://ehelse.no/publikasjoner/konseptvalgutredning-for-helseanalyseplattformen

Begrep	Definisjon	Eksempler	Kilde	Link til Kilde
	<p>applikasjonssegmenter</p> <p>Kilde: Skatteetaten gjennom DigDir: https://www.difi.no/artikkel/2016/09/begreper-fellesinformasjonsmodeller</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>		Standardiserte tjenestegrensesnitt (API) for helseregistre	https://ehelse.no/publikasjoner/standardiserte-tjenestegrensenitt-api-for-helseregistre
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. 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>		Felles språk i helse- og omsorgssektoren Målbildes versjon 1.0	https://ehelse.no/publikasjoner/felles-sprak-i-helse-og-omsorgssektoren-malbilde-versjon-1.0

Begrep	Definisjon	Eksempler	Kilde	Link til Kilde
Kode	Kode er i denne sammenheng det samme som svaralternativene i et verdisett.		Nasjonal spesifisering for metadata om helsedata	
Kodeverk	Et kodeverk er en samling unike begreper med tilhørende kode. Typiske eksempler vil være kodeverkene som per i dag publiseres på Volven.no.		Felles språk i helse- og omsorgssektoren Målbilde versjon 1.0	https://ehelse.no/publikasjoner/felles-sprak-i-helse-og-omsorgssektoren-malbilde-versjon-1.0
Maskinlesbar	Informasjon i et format som er strukturert slik at programvarer enkelt kan identifisere, ekstrahere og behandle informasjonens individuelle bestanddeler.		Kriterier for FAIR forskningsdata VR2018	https://www.vr.se/analys/rapporter/vara-rapporter/2018-12-07-kriterier-for-fair-forskningsdata.html
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 tjenestegrensensnitt (API) for helseregistre	https://ehelse.no/publikasjoner/standardiserte-tjenestegrensensnitt-api-for-helseregistre
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		Felles språk i helse- og omsorgssektoren Målbilde versjon 1.0	https://ehelse.no/publikasjoner/felles-sprak-i-helse-og-omsorgssektoren-malbilde-versjon-1.0

Begrep	Definisjon	Eksempler	Kilde	Link til Kilde
	forelder-begrep samtidig. En ontologi kan fremstilles på forskjellige måter, men strukturen bygges basert på regler, og de relasjoner som er angitt.			
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 Kriterier for FAIR forskningsdata VR 2018	https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/ https://www.vr.se/analys/rapporter/vara-rapporter/2018-12-07-kriterier-for-fair-forskningsdata.html
Protokoller	Kommunikasjonsprotokoll: Formelt regelverk som spesifiserer hvordan utveksling av data mellom to aktører skal foregå og hvilket format som anvendes.	HTTP, TCP/IP, FTP	Kriterier for FAIR forskningsdata VR2018	https://www.vr.se/analys/rapporter/vara-rapporter/2018-12-07-kriterier-for-fair-forskningsdata.html
Sensitive data	Med sensitive data menes i denne sammenheng sensitive personopplysninger: 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		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	https://www.regjeringen.no/no/dokumenter/prop.-56-ls-20172018/id2594627/?ch=7

Begrep	Definisjon	Eksempler	Kilde	Link til Kilde
	ytterligere krav til behandling av slike opplysninger.		(EU) nr. 2016/679 (generell personvernforordning) i EØS-avtalen	
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 JSON	XML, JSON	Wikipedia	
Standardisert	Av standardiseringsorganisasjon fastslått eller innen domenen fastslått praksis.		Kriterier for FAIR forskningsdata VR2018	https://www.vr.se/analys/rapporter/vara-rapporter/2018-12-07-kriterier-for-fair-forskningsdata.html
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 norske basisprofiler for HL7 FHIR. https://ehelse.no/standarder/norske-basisprofiler-for-hl7-fhir		Standardiserte tjenestegrensesnitt (API) for helseregistre	https://ehelse.no/publikasjoner/standardiserte-tjenestegrensenitt-api-for-helseregistre
Statistikk	Statistikk er definert som de statistiske egenskapene ved en årgang/versjon av en variabel.		Nasjonal spesifisering for metadata om helsedata	
Terminologi	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. 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		Felles språk i helse- og omsorgssektoren Målbilde versjon 1.0	https://ehelse.no/publikasjoner/felles-sprak-i-helse-og-omsorgssektoren-malbilde-versjon-1.0

Begrep	Definisjon	Eksempler	Kilde	Link til Kilde
	sammenheng med hverandre.			
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	https://ehelse.no/publikasjoner/standardiserte-tjenestegrensesnitt-api-for-helseregistre
URL	URL er kort for Uniform Resource Locator og viser til «adresser» på Internett.		Wikipedia	
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	https://ehelse.no/publikasjoner/standardiserte-tjenestegrensesnitt-api-for-helseregistre
Variabel	Variabler er de informasjonsbærende elementene i datakilden og definerer hvilken informasjon som finnes i registrene.		Nasjonal spesifisering for metadata om helsedata	
Verdi	Verdi er definert som antall gyldige tilfeller av et svaralternativ/utfall av en variabel.		Nasjonal spesifisering for metadata om helsedata	

Referanser:

- [1] Direktoratet for e-helse (2016): Referansearkitektur og fellestjenester for helseregistre
<https://kilden.sykehusene.no/download/attachments/73043236/59b%20Sluttrapport%20-%20Referansearkitektur%20Helseregistre-%20v1.0.pdf?version=1&modificationDate=1560334868948&api=v2>
- [2] Forskningsrådet (2016): Enklere tilgang- mer forskning
<https://www.forskningsradet.no/om-forskningsradet/publikasjoner/2016/enklere-tilgang--mer-forskning/>
- [3] Et nytt system for enklere og sikrere tilgang til helsedata – Rapport fra Helsedatautvalget 2016-2017
https://www.regjeringen.no/contentassets/1fe9cf37e64344e1a3b3c62f950b100b/170630_helsedatalovutvalget.pdf
- [4] The FAIR data principles (Nature / Scientific Data March 2016):
<http://www.nature.com/articles/sdata201618.pdf>
- [5] The FAIR Data Principles (FORCE11)
<https://www.force11.org/group/fairgroup/fairprinciples>
- [6] G20 Leaders' Communique Hangzhou Summit, European Commission Press Release Database, 5th September 2016
http://europa.eu/rapid/press-release_STATEMENT-16-2967_en.htm
- [7] EOSC Declaration (European Open Science Cloud, Brussels 26 October 2017):
http://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf#view=fit&pagemo de=none
- [8] The European Code of Conduct for Research Integrity (March 2017)
https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf
- [9] Konseptvalgutredning for Helseanalyseplattformen (April 2018),
<https://ehelse.no/publikasjoner/konseptvalgutredning-for-helseanalyseplattformen>
- [10] Turning FAIR into reality European Commission November 2018
<https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283>
- [11] FAIR Guiding Principles for scientific data management and stewardship' (Go FAIR)
<https://www.go-fair.org/fair-principles/>
- [12] Kommunal- og moderniseringsdepartementet (2017): Retningslinjer ved tilgjengeliggjøring av offentlige data
<https://www.regjeringen.no/no/dokumenter/retningslinjer-ved-tilgjengeliggjoring-av-offentlige-data/id2536870/>
- [13] Nordforsk (2019): A vision of a Nordic secure digital infrastructure for health data: The Nordic Commons
<https://www.nordforsk.org/2020/vision-nordic-secure-digital-infrastructure-health-data-nordic-commons>
Vetenskapsrådet (2018): Kriterier för FAIR forskningsdata
<https://www.vr.se/analys/rapporter/vara-rapporter/2018-12-07-kriterier-for-fair-forskningsdata.html>

- [14] Research Data Alliance (2019): Results of an Analysis of Existing FAIR Assessment Tools
<https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/results-analysis-existing-fair-assessment-tools>

UTKAST