

DIREKTORATET FOR E-HELSE
Postboks 221 Skøyen
0213 OSLO

V/ E-helse standardisering

Vår referanse:
20/00701-4
Saksbehandler:

Deres referanse:
20/375

Dato:
15.10.2020

Yvonne Garshol, t.90519909

Innspillsrunde: Veileder for bruk av FAIR-prinsippene for helsedatakilder

Innspill til veilederen generelt

Helse Sør-Øst RHF takker Direktoratet for e-helse for invitasjon og mulighet til å bidra i innspillsrunden knyttet til veileder for bruk av FAIR-prinsippene for helsedatakilder.

Etter gjennomgang av prinsippene er Helse Sør-Øst av den oppfatning at disse er innholdsmessig faglig fornuftig. Helse Sør-Øst er imidlertid ett stykke fra å kunne innføre et slikt rammeverk mtp. kompetanse, kapasitet og finansiering.

Veilederen beskriver arkitekturprinsipper for helseregistre. Dette går inn på et område hvor Helse Sør-Øst har flere registre. Disse registrene skal i neste omgang harmoniseres med EPJ-løsningene. Slik vi leser invitasjonen er målgruppen i første omgang helseregistrene. Dersom målet er automatisert innmelding fra primære datakilder til helsedataplattformen vil dette medføre en betydelig arbeidsmengde, som det vil være urealistisk å kartlegge for dagens portefølje i Helse Sør-Øst.

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.

Direktoratet for e-helse anbefaler at forvaltere av helsedatakilder bruker evalueringsverktøyet som et virkemiddel for innføring av FAIR-prinsippene, og for å identifisere forbedringsområder.

Helse Sør-Øst ser det som viktig å få klarhet i hva Direktoratet for e-helse sikter mot; Bruk av prinsippene mot registermiljøene eller bruk av prinsippene også inn mot primære datakilder.

Kort oppsummert:

- **Omfanget av tiltakene** for å kunne applisere FAIR prinsippene på datasett som rapporteres til nasjonale helseregistre er antageligvis stor, men ikke kartlagt.
- Realisme av å applisere FAIR prinsippene på Helse Sør-Østs IKT portefølje er ikke kartlagt fordi RHF **ikke har oversikt over alle dataelementer som rapporteres til helseregistrene og hvilke datakilder** de kommer fra.
- Dersom man skal ha FAIR prinsipper som en realistisk nasjonal føring, vil det ha **konsekvenser for Helse Sør-Østs oppdrag til Sykehuspartner**, dvs. Sykehuspartner må få i oppdrag å etablere mekanismer som sikrer sporing på variabelnivå og forvaltningen av FAIR prinsippene på datasettet som er berørt.
- Konsekvenser av FAIR prinsippene når det gjelder **finansieringen av relevante tiltak** for å applisere disse på relevante deler av IKT porteføljen er ikke estimert.

Innspill knyttet til prinsippene og evalueringsskjemaet

Ved gjennomgang av innholdet i FAIR prinsippene vil vi innledningsvis si at disse fremstår som gode kriterier og innholdsmessig faglig fornuftige. En kartlegging av prinsippene vil kunne gi en god oversikt, men dette vil kreve at miljøene med den rette innsikten styres slik at kartlegging gjennomføres og gir et godt datagrunnlag.

Gruppering av spørsmålene

Det virker til å være en viss overlapp mellom arkene i evalueringsskjemaet. Det bør vurderes å systematisere spørsmålene på en mer pedagogisk måte, hvor det blir lettere å ta for seg hver datamengde.

- Det kan vurderes om det skal være færre spørsmål i ett ark, der hvert spørsmål kan være klassifisert som en kombinasjon av F A I R. Flere spørsmål og underpunkter går igjen i flere av F A I R arkene.
- Det kan vurderes om det vil være bedre å presentere spørsmålene i ett ark, gjerne systematisert etter et kriterium/gruppering, som for eksempel standarder, API, metadata, formater, forvaltning.
- For hvert spørsmål kan man vurdere å inkludere en beregning av hvordan de bidrar til hhv. F A I R kriteriene. Dette vil kunne bidra til en mindre mengde spørsmål, presentert i en pedagogisk rekkefølge og gruppering.

Forslag til systematisering av spørsmålene:

Gruppe	spørsmål	kommentar	rating	Faste vekter				Beregnes
				F	A	I	R	Vektet sum
Metadata (deklarasjon)	Standard format?	(karakter settes)	8	1	3	2	5	7.35
	Åpen API	Noen spørsmål gjentas i flere seksjoner						
Metadata (forvaltning)								
Aggregert pr metadata (pr system)								
(ikke sensitive) Data	Standard format?							
	Åpen API							
Sensitive data	Beskrivelse av tilganger, prosess for tildeling, prosess for endring.							
Aggregerte (fortsatt sensitive) data								
Godt nok anonymisert (feks v aggregering)								
Kodeverk								
Begrep/terminologi								

Metadata, aggregerte data og sensitive data

Inndelingen i metadata, aggregerte data og sensitive data – men ikke ordinære data – er noe uklar.

- Ark A: Spørsmål 12, 13, 14 repeteres for inndelingen metadata, aggregerte data og sensitive data, men ikke for ordinære data.
- Ark I: Spørsmål repeteres for inndelingen metadata og variabler.

Masterdata, referansedata/kodeverk kommer senere i andre ark. Det er ingen kategori for metadata hvis man tolker dette i en videre forstand. Vi kan anta at dette er fordi man ser for seg at den slags data er en del av den ordinære datamodellen. Språkbruken er konsistent i dokumentet på at metadata er «data som beskriver data». Mange miljøer bruker imidlertid metadata om instansdata som typisk beskriver mediedata:

- Ikke selve telefonsamtalen, men tidspunkt, varighet, deltagende telefonnummer, oppkoblingskarakteristika
- Ikke selve bildet, men størrelsen, encoding, sted/tid/utstyr/opphavsperson
- Ikke selve epost-innholdet, men fra/til, tidspunkter, routing

Vi kan anta at dette ikke er skilt ut som «metadata» i FAIR prinsippene, men er en del av den ordinære datamodellen, slik at man bør betrakte størrelse, encoding,

sted/tid/utstyr/opphavsperson, anatomisk lokasjon etc. som felter i klassen som representerer for eksempel et røntgenbilde.

Aggregerte og anonymiserte data

Det er en kobling mellom aggregerte og anonymiserte data i arkene som er noe uklar, i A1 underpunkter L15, L16, L19.

- Hvis man aggregerer slik at person-id blir borte, har man (kanskje) anonymisert?
- Hvis man aggregerer over en annen dimensjon, har du ikke anonymisert?

Det bør vurderes om veilederen skal inkludere spørsmål om anonymisering. Aggregerte data er ikke nødvendigvis anonymiserte.

- Hvis man aggregerer medisinforbruk ved å aggregere pr person pr medisin, så har man et sensitivt og ikke-anonymisert resultat.
- Hvis man aggregerer pr distrikt pr medisin pr år, så har man antagelig et mangelfullt anonymisert datasett. Man kan potensielt utlede for mye informasjon om den eneste personen i distriktet som bruker denne typen medisin.
- Hvis du aggregerer pr medisin pr år, kan vi anta at ingen kan få ut noe sensitivt.

Anonymisering og «faking» av realistiske testdata er også et tema. Det er for eksempel ikke akseptabelt å bygge et testsett ved å ta en produksjonsdatabase og bytte ut personnummer, navn og adresse med tilfeldige «randomiserte» strenger. Man må strekke det langt for at det ikke skal være mulig å trekke ut sensitive data

Det kan finnes viktigere spørsmål rundt sensitive data enn A1 L17, L18, L20. Man kan vurdere å spørre om definisjoner av sperringer, blokkeringer, forvaltning av disse, på instansnivå, gruppenivå etc.

Spørsmålsinndelingen i ark A, der man har samme spørsmål over forskjellige typer informasjon: Er det et interessant spørsmål om sensitive data er gratis, eller burde det heller være spørsmål knyttet til prosesser for å få tilgang til data?

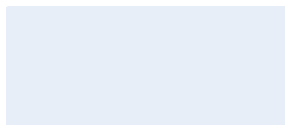
Oppsummert

Basert på våre ovenforstående innspill oppfatter vi ikke veilederen som tydelig. Helse Sør-Øst oppfatter omfanget av kartleggingen som uklar. Det er uklart hva Direktoratet for e-helse inkluderer i kartleggingen når det henvises til helseregisterdata og helsedatakilder.

Helse Sør-Øst oppfatter at prinsippene og evalueringsskjemaet vil gi svar på praktisk nyttige og relevante spørsmål, som vil gi god verdi om kartleggingen gjøres av miljøene med den rette innsikten.

Vi håper at våre innspill oppleves av Direktoratet for e-helse som konstruktive og berikende på innhold og anbefalingen i rapporten. Vi ønsker med dette å også oppfordre Direktoratet for e-helse til å ta kontakt med oss, dersom dere har spørsmål til innholdet.

Med vennlig hilsen
Helse Sør-Øst RHF



Rune Simensen
direktør teknologi og e-helse

Ulf Sigurdsen
leder for eHelse