

Høringssvar

Navn: Universitetet i Innlandet

Hva beskriver best deg/dere som avsender:

- ☒ Offentlig virksomhet (statlig, kommunal eller fylkeskommunal)
- ☐ Næringsliv (privat virksomhet eller bransjeorganisasjon)
- ☐ Forskning eller utdanning
- ☐ Presse eller medier
- ☐ Sivilsamfunn eller interesseorganisasjon
- ☐ Privatperson
- ☐ Annen – Beskriv gjerne her:

1. Veilederens relevans og nytteverdi

På hvilken måte kan veilederen være nyttig for offentlige virksomheter som tilgjengeliggjør data?

Veilederen har potensial til å bli et viktig verktøy for offentlige virksomheter som forvalter og deler data. Den samler juridiske, tekniske og praktiske krav og anbefalinger på ett sted, og gir en strukturert framstilling av både gjeldende og kommende regelverk, samt FAIR-prinsippenes ambisjonsnivåer. Denne tabellstrukturen, med tydelig skille mellom krav og anbefaling, gjør det enklere for virksomheter å forstå hvor de står, og å prioritere hvilke tiltak som må komme først.

For en kunnskapsinstitusjon som Universitetet i Innlandet, med både forsknings- og utdanningsvirksomhet, er det særlig verdifullt at veilederen løfter fram dokumentasjon, metadata og datakvalitet, ikke bare selve publiseringen av data. Henvisningene til nasjonale kataloger (som data.norge.no, Helsedata.no og Geonorge) og til standardiserte metadatabeskrivelser er konkrete tiltak som kan bidra til at data faktisk blir mulig å finne, forstå og vurdere.

Samtidig framstår veilederen i dagens form som relativt abstrakt. Den skiller mellom åpne og beskyttede data, men illustrerer i liten grad med konkrete datatyper, og mangler gjennomgående eksempler som viser veien fra dokumentasjon til tilgjengeliggjøring for et faktisk datasett. For virksomheter som skal plassere egne data i rammeverket, er det krevende å gå fra prinsipper til praktisk håndtering.

Vi anbefaler derfor at veilederen suppleres med:

- typiske eksempler på datasett i ulike kategorier,
- ett-to gjennomgående eksempler som følger et datasett gjennom de aktuelle punktene, gjerne ett for åpne data og ett for beskyttede/sensitive data.

For studieforvaltningen, som håndterer student-, studie- og eksamensdata og benytter nasjonale systemer, er veilederen særlig relevant. Her blir det tydelig hvor viktig begrepsklarhet er. I dag er «datasett» og «data» definert relativt generelt, og det er vanskelig å vite om veilederen sikter til for eksempel:

- studentdata,
- studiedata og emnehistorikk (bl.a. til nasjonalt utdanningsregister),
- eksamensdata og timeplaner,
- undervisningsopplegg i læringsplattformer,
- nemndssaker mv.

Vi anbefaler en tydeligere og mer praktisk definisjon av «datasett», og en klargjøring av ansvarsfordelingen mellom virksomhet og systemleverandør/nasjonale tjenesteeiere (f.eks. FS, opptakssystemer). Konkret bør det framgå:

- hvem som har ansvar for tilgjengeliggjøring,
- om leverandører har plikt til å tilrettelegge teknisk (for eksempel via API-er),
- og om virksomheter kan avvise krav dersom leverandører ikke tilbyr nødvendige løsninger.

Et sentralt mål er at data skal være enkle å finne, forstå og ta i bruk. Rådet ønsker særlig innspill fra virksomheter og personer som bruker data fra offentlig sektor.

Dekker veilederen de behovene dere har som brukere av offentlige data? Er det viktige brukerbehov som ikke er omtalt?

Som forskningsmiljø innen helse-, sosial- og utdanningsvitenskap, og som bruker av data fra offentlig sektor (blant annet helseregistre, registerdata fra SSB, data via Helsedata.no, kommunale helse- og velferdsdata og ulike utdanningsregistre), opplever vi at veilederen dekker flere sentrale behov knyttet til åpne data. Den gir klare føringer for hvordan datasett kan gjøres mer gjenfinnbare, tilgjengelige og gjenbrukbare.

Samtidig er det nettopp beskyttede data som i praksis er mest sentrale for forskningen. De er ikke fritt tilgjengelige, men gjøres tilgjengelige gjennom søknad, rettslig grunnlag og sikker infrastruktur. Den brukerreisen veilederen legger opp til for beskyttede data er lite omtalt. Det gjelder særlig:

- hvordan brukere oppdager at beskyttede datasett eksisterer,
- hvilke vilkår og tilgangsprosesser som gjelder,
- og hvordan tilgangen gis i sikre miljøer.

Vi anbefaler at veilederen tydeliggjør at metadata skal være åpne selv når selve dataene er skjermet. Dette gjør det mulig for forskere og andre profesjonelle databrukere å finne og vurdere beskyttede datasett uten å ha tilgang til innholdet. Dette følger både av prinsippet «data skal åpnes når de kan, og skjermes når de må», og av FAIR-prinsippene om gjenfinnbarhet og autorisasjonsstyrt tilgang.

Videre vil varige, unike identifikatorer (for eksempel DOI) for datasett være avgjørende for sitering og etterprøvbarhet i forskning. Vi anbefaler også en tydeligere kobling til datahåndteringsplaner, som allerede er et krav fra forskningsfinansiører som Forskningsrådet og EU.

For studieforvaltningen oppleves veilederen som lite konkret når den møter den faktiske praksisen med komplekse datasett. Vi savner særlig:

- avklaringer av hvilket nivå studentdata kan deles på,
- tydeligere skille mellom ulike datatyper (persondata, fagdata, administrative data, saksdata),
- beskrivelse av data som ligger i systemer virksomheten ikke eier, men bruker (som FS og opptakssystemer),
- og en samlet framstilling av de juridiske rammene.

Forholdet mellom innsyn etter offentlighetsloven og krav til tilgjengeliggjøring etter dataforvaltningsloven er uklart. Det bør tydeliggjøres:

- om allment tilgjengelige datasett kan være grunnlag for å avvise innsynskrav,
- hvordan klageadgang ivaretas,
- og om veilederen innebærer realitetsendringer i praktisering av eksisterende regelverk.

Hva kan man gjøre i offentlig sektor for å sikre at veilederen blir tatt i bruk og har en positiv effekt?

En veileder alene endrer ikke praksis i en sektor med knappe ressurser. Skal veilederen få reell effekt, må den støttes av:

- felles infrastruktur for deling og katalogisering,
- kompetansetiltak,
- og tydelige insentiver.

Dette gjelder særlig for mindre virksomheter og fagmiljøer uten egne datateam.

Konkret anbefaler vi:

- gode maler og verktøy for metadata,
- tett tilknytning til eksisterende kataloger,
- at etterlevelse av FAIR og god metadatapraksis anerkjennes som del av kjernevirksomheten.

I forskningssektoren bør god datahåndtering telle på linje med publisering når kvalitet og innsats vurderes. Tydelig ledelsesforankring og klart definerte minimumsnivåer vil gjøre det enklere å prioritere arbeidet.

Vi anbefaler også at veilederen i større grad omtaler hvordan offentlige data kan brukes i forbindelse med kunstig intelligens. For forskningssektoren er spørsmålet ikke bare om data er åpne, men:

- under hvilke vilkår data kan brukes i eksterne eller kommersielle KI-tjenester,
- hvordan ansvarlig gjenbruk og datakontroll skal ivaretas.

Det er behov for tydeligere signaler om intensjonen med åpne data i en KI-kontekst, og om hvordan man skal forstå gjenbruk og opphavsrett der KI genererer nye datasammenstillinger og analyser.

For studieforvaltningen mener vi at faktisk bruk og positiv effekt forutsetter at veilederen:

- presiserer og harmoniserer begrepsbruken, særlig «datasett», «tilgjengeliggjøring», «innsyn» og «metadata», også i lys av etablerte begreper i universitets- og høyskolesektoren,
- inneholder flere konkrete, sektorrettede eksempler fra vår sektor, som viser hvordan datasett identifiseres, beskrives og deles,
- tydeliggjør ansvarsdelingen mellom virksomhet og systemleverandør/nasjonale tjenesteeiere, inkludert eventuell plikt til å tilby API-er og finansiering av slike løsninger,
- operasjonaliserer krav og anbefalinger, for eksempel når det gjelder løpende publisering av data, publisering av metadata i nasjonale portaler og adgang til å ta betalt,
- og supplerer omfanget av lenker og støttedokumenter med kortfattede oversikter og sektorrettede sammendrag.

Rådet foreslår at veilederen erstatter Retningslinjer ved tilgjengeliggjøring av offentlige data (lenke til regjeringen.no). Rådet mener at en veileder er mer fleksibel og enklere å holde oppdatert. Samtidig kan en veileder oppfattes som mindre forpliktende enn et rundskriv og/eller retningslinjer.

Hvilke fordeler og ulemper ser dere med at det er en veileder fremfor retningslinjer?

Vi ser klare fordeler ved å bruke en veileder framfor statiske retningslinjer. En veileder:

- er mer fleksibel, og enklere å oppdatere i takt med både regelverksutvikling (som dataforvaltningsloven) og teknologiske endringer,
- gir større rom for å beskrive «beste praksis» og å bygge inn erfaringer over tid.

Ulempen er at en veileder ofte oppfattes som mindre forpliktende enn et rundskriv eller retningslinjer. I en sektor med knappe ressurser vil det som ikke er eksplisitt krav lett bli nedprioritert. Dersom rådet mener at FAIR og god metadatapraksis er avgjørende, er det viktig at veilederen:

- tydelig angir hvilke minimumsnivåer som forventes,
- og skiller klart mellom rettslig bindende krav og veiledende anbefalinger.

Vi anbefaler at veilederen eksplisitt angir hvilke punkter som bygger på konkrete rettslige krav, med henvisning til:

- offentlighetsloven,
- dataforvaltningsloven,

- personvernregelverket mv., og hvilke punkter som er anbefalinger eller uttrykk for beste praksis. Et mulig kompromiss er å kombinere veilederens fleksibilitet med et kort, forpliktende minstekrav.

For vår del er det avgjørende at dette skillet er tydelig, slik at vi kan:

- planlegge ressursbruk og organisering,
- vurdere behovet for egne funksjoner eller enheter for dataforvaltning og datadeling,
- og sikre etterlevelse av regelverk (inkludert GDPR og universell utforming).

2. Skillet mellom krav og anbefalinger

Veilederen skiller mellom gjeldende og kommende rettslige krav på den ene siden, og anbefalinger for god praksis på den andre siden. Rådet ønsker innspill på om denne balansen er hensiktsmessig og forståelig.

Er forskjellen på krav og anbefalinger hensiktsmessig og forståelig? Er det anbefalinger som i realiteten er krav, og i så fall, etter hvilket regelverk?

Den overordnede balansen mellom krav og anbefaling, og mellom gjeldende og kommende regelverk, framstår som hensiktsmessig og pedagogisk. Tabellen med kolonnene «krav/anbefaling» og «gjeldende/kommende regelverk» gjør det mulig å se både dagens plikter og planlagte endringer.

Vi har likevel noen forbehold. Flere sentrale FAIR-elementer står som anbefaling uten tilhørende regelverk, men er i praksis forutsetninger for at datadeling skal ha verdi. Eksempler er:

- faste, unike identifikatorer (punkt 12),
- at metadata følger dataene (punkt 13),
- og dokumentert datakvalitet.

Slike punkter kan i dag formelt plasseres som anbefaling, men realiteten er at de allerede er innarbeidede forventninger, og i deler av sektoren på vei til å bli eksplisitte krav.

Helsedataforordningen stiller for eksempel krav om «Quality and Utility»-merking for offentlig finansierte helsedata. Å plassere datakvalitet som ren anbefaling kan undervurdere hvor bindende og sentralt dette er i ferd med å bli.

Vi ser ellers ikke mange anbefalinger som i realiteten er rettslige krav utover det veilederen selv markerer, men enkelte anbefalinger om metadata og lisensiering ligger nært rettslig regulerte områder, blant annet gjennom offentleglova og åndsverkloven. Det er viktig at veilederen tydeliggjør hva som anses som nødvendig minimum for god og ansvarlig datadeling, selv om ikke alt er direkte hjemlet i lov.

For virksomheter som vurderer organisering og ressursbruk, er det avgjørende at:

- det går klart fram hvilke punkter som bygger på konkret regelverk,
- hvilke som er nødvendige minimum for ansvarlig datadeling,

- og hvilke som er rene ambisjonsnivåer.

3. FAIR-prinsippene

Veilederen viser til de internasjonale FAIR-prinsippene. FAIR står for Findable, Accessible, Interoperable, Reusable. Dette er en større endring sammenlignet med de nåværende retningslinjene ved tilgjengeliggjøring av offentlige data. Rådet har selv definert ambisjonsnivåene, som en hjelp for offentlige virksomheter når de skal vurdere hvilke prinsipper de bør etterleve.

På hvilken måte har det en verdi for databrukere om offentlig sektor i større grad etterlever FAIR-prinsippene? Eventuelt hvis det ikke har verdi, hvorfor ikke? Er FAIR-prinsippene gjennomførbare og mulig å etterleve for offentlige virksomheter som deler data?

For forskningssektoren har det stor verdi at offentlig sektor etterlever FAIR-prinsippene. FAIR er allerede det etablerte rammeverket for forskningsdata, Forskningsrådet og EU/Horisont Europa krever både datahåndteringsplaner og FAIR-tilpasning. Når offentlige datasett utformes etter samme prinsipper, reduseres friksjon og det blir enklere å:

- finne relevante datasett,
- forstå opphav og historikk,
- sammenstille data på tvers av kilder,
- og etterprøve analyser og resultater.

Særlig dokumentasjon av opphav, endringer og historikk (R1.2) er en forutsetning for reproduserbar forskning.

Prinsippene er gjennomførbare, men ressurskrevende. Ambisjonsnivåene som veilederen legger opp til, gjør arbeidet håndterbart ved å tilby en trinnvis tilnærming. Samtidig er det viktig at ambisjonsnivåene knyttes til datasettets samfunnsverdi og etterspørsel, slik at ressursinnsatsen står i rimelig forhold til nytten.

FAIR har sitt opphav i forskningssektoren (Wilkinson et al., 2016), og praksis er mest modent utviklet der. Vi anbefaler at veilederen:

- viser eksplisitt til hvordan FAIR praktiseres i forskningssektoren,
- bruker universitets- og høyskolesektoren som referanse for god praksis,
- og legger vekt på å unngå ulike tolkninger av FAIR mellom offentlig forvaltning og forskning, særlig fordi universiteter og høyskoler selv er offentlige virksomheter som deler data.

I en KI-kontekst blir FAIR særlig viktig. Her blir:

- avklart forhold til metadata,
- tydelige lisens- og opphavsrettsvilkår,

- og krav til ansvarlig viderebruk, avgjørende for om data kan gjenbrukes på en forsvarlig måte. Samtidig oppstår nye utfordringer: selv om data er åpne, kan KI-genererte sammenstillinger og analyser ha forskningsmessig og kommersiell verdi som må forvaltes etter regler for opphavsrett og datakontroll. Veilederen kan bidra ved å tydeliggjøre intensjonen med åpne data og FAIR i en KI-kontekst.

For studie- og utdanningsdata er FAIR-prinsippene også relevante, men krevende å fullt ut etterleve for komplekse og delvis sensitive datasett. For at ambisjonsnivået skal være gjennomførbart, anbefaler vi at veilederen:

- skiller tydeligere mellom åpne/aggregerte data og persondata,
- foreslår minimumsnivåer for FAIR-etterlevelse for ulike kategorier datasett,
- og tydeligere drøfter kost-nytte for virksomhetene.

Er rådets ambisjonsnivåer for FAIR-prinsippene forståelig og hensiktsmessig? Bør ambisjonsnivået justeres for noen av prinsippene, og i så fall hvilke og hvordan?

Ambisjonsnivåene; lavt, middels og høyt, framstår som forståelige og gir en god ramme for en gradvis tilnærming. Det er positivt at veilederen åpner for å starte med FAIR for metadata, og deretter arbeide videre med selve datasett.

Samtidig ligger det en risiko i at «lavt nivå» blir en hvilepute. Dersom lavt nivå i praksis betyr at bare metadata, og ikke data, er FAIR-tilpasset, kan dataene fortsatt forbli vanskelig tilgjengelige for reell gjenbruk. Vi anbefaler derfor at veilederen presiserer at:

- lavt nivå er et startpunkt, ikke et mål,
- og at forventet ambisjonsnivå bør kobles til datasettets samfunnsverdi, bruksbehov og etterspørsel.

For virksomheter som forvalter komplekse personopplysningsdatasett, bør ambisjonsnivået justeres med realistiske krav. Det vil være hensiktsmessig å:

- prioritere høyt ambisjonsnivå for åpne og aggregerte data,
- definere tydelige minimumskrav for persondata,
- og gi eksempler som viser hva som anses som minimum for hvert prinsipp.

Er oversettelsene av prinsippene forståelig, og gjengir de meningsinnholdet fra originalspråket, engelsk? Hvilken norsk oversettelse av «Findable» er mest dekkende og presis – «gjenfinnbar» eller «søkbar»? Er det andre begreper som bør vurderes?

Oversettelsene framstår i all hovedsak som forståelige og dekkende. Når det gjelder «Findable», vurderer vi «gjenfinnbar» som mest presist. «Søkbar» sier bare at noe kan søkes etter, mens «Findable» i FAIR-sammenheng handler om at et datasett:

- kan finnes igjen entydig og varig over tid,
- har en varig og unik identifikator,

- har rike metadata,
- og er indeksert i relevante kataloger.

«Søkbar» ligger implisitt i «gjenfinnbar», men fanger ikke det systematiske forarbeidet som gjør gjenfinning mulig. Nettopp dette forarbeidet; identifikatorer, registrering og metadata, skiller F i FAIR fra en enkel søkefunksjon. Vi anbefaler derfor «gjenfinnbar» som begrep framfor «søkbar», eller en dobbel betegnelse der «søkbar» ikke blir overflødig.

Samtidig er det viktig med sektorovergripende begrepsharmonisering. Vi anbefaler at det igangsettes et arbeid der ulike sektorer involveres, og der det også vurderes mer brukervennlige beskrivelser, for eksempel «lett å finne/oppdage», dersom det gir bedre kommunikasjon til sluttbrukere uten å forskyve det faglige innholdet.

4. Andre innspill

Vi ønsker også innspill som ikke dekkes av spørsmålene over. Dette kan være synspunkter på enkeltpunkter i veilederen, forhold vi ikke har spurt om, eller andre perspektiver dere mener er relevante.

Er det noe i veilederen dere ønsker å kommentere spesifikt, bruker dere gjerne overskriftene og nummereringen fra veilederen.

Med utgangspunkt i veilederens egne overskrifter og nummerering vil vi særlig peke på følgende:

- **Punkt 7 – innebygd åpenhet og åpenhet som standardinnstilling**
For virksomheter som forvalter sensitive person-, helse- og sosialdata, vil skjerming av selve dataene nødvendigvis være standardinnstillingen; åpenheten må gjelde metadata og aggregerte eller anonymiserte data. Vi vil understreke at enkelte datatyper i praksis ikke kan anonymiseres fullt ut, genetiske data er et eksempel, og at gjenbruk også er bundet av samtykke og rettslig behandlingsgrunnlag. «Åpenhet som standard» må derfor avgrenses tydelig og forutsetter et gyldig rettslig grunnlag; formuleringen må ikke kunne leses slik at den senker terskelen for deling av sensitive data. Balansen mot personvernregelverket og taushetsplikt bør framgå like tydelig som i dagens retningslinjer.
- **Grenseflaten mot forskningsdata**
Veilederen bør omtale forholdet til den nasjonale strategien for tilgjengeliggjøring og deling av forskningsdata, og til eksisterende retningslinjer på forskningsområdet. Dette vil bidra til å unngå to parallelle FAIR-regimer, og til å klargjøre hvilke føringer som gjelder når offentlige virksomheter samtidig er forskningsutførende, slik universiteter og høgskoler er.
- **Fagspesifikke metadatastandarder (punkt 13–14)**
Veilederen bør tydelig anerkjenne at det vokser fram domenespesifikke metadatastandarder, og at den generelle veiledningen bør harmoniseres med disse. I helsesektoren er Health DCAT-AP i ferd med å bli obligatorisk metadatastandard under EU-regelverket (helsedataforordningen / European Health Data Space). Dette er i tråd med veilederens eget prinsipp R1.3 om domenebaserte og fagspesifikke standarder,

men bør sies eksplisitt, slik at virksomheter som er omfattet av sektorregelverk ikke må forholde seg til to ulike sett metadatakav.

- **Datakvalitet (punkt 8)**

Datakvalitet omtales som anbefaling. I forskningssektoren er dokumentert datakvalitet allerede en innarbeidet forventning, og i deler av sektoren på vei til å bli eksplisitt krav. Helsedataforordningen (art. 78) stiller blant annet krav om merking av kvalitet og nytte for offentlig finansierte helsedata. Vi anbefaler at veilederen løfter datakvalitet nærmere minimumsnivået, og tydeliggjør at ansvarlig datadeling forutsetter at kvalitet er dokumentert.

- **Ressurser, insentiver og delt infrastruktur**

Etterlevelse av FAIR og god metadatapraksis krever delte nasjonale løsninger; kataloger, metadata- og analyseinfrastruktur, og kan ikke bæres fullt ut av hver enkelt virksomhet. Dersom byrden legges for tungt på den enkelte, kan etterlevelseskostnadene overstige nytten, og anbefalingene forbli intensjoner. Dette gjelder særlig små og spesialiserte fagmiljøer.

Vi vil også peke på at FAIR-data impliserer gjenbruk. I en KI-kontekst kan dette skape nye dilemmaer:

- man kan miste direkte kontroll over hvordan data blir brukt,
- nye datasammenstillinger og analyser kan få forskningsmessig og kommersiell verdi,
- samtidig som man ønsker å ivareta grunnleggende prinsipper for åpen vitenskap og offentlig nytte.

Her vil veilederen kunne bidra med tydeligere føringer for hvordan intensjonen med åpne data skal forstås i møte med KI-teknologi.

For studieforvaltningen har vi noen særskilte innspill:

- Ansvarsfordelingen mellom virksomheter og systemleverandører/nasjonale tjenesteeiere når det gjelder tilgjengeliggjøring og teknisk tilrettelegging (API-er mv.) bør beskrives klarere.
- Veilederen forutsetter i praksis en viss organisering av dataforvaltningen. Det vil være nyttig med eksempler på modeller for ansvar og koordinering i større virksomheter, og avklaringer av om det forventes egne enheter eller funksjoner.
- Omfanget av lenker og støttedokumenter gjør det krevende å få oversikt. Kortfattede, sektorrettede sammendrag, for eksempel for universitets- og høyskolesektoren, vil øke nytteverdien.
- Det bør tydeliggjøres om, og i hvilken grad, allment tilgjengelige datasett kan påvirke behandlingen av individuelle innsynskrav og tilhørende klageadgang.

Begrepet forskningsdata bør omtales eksplisitt i veilederen eller inkluderes i definisjonen av data. Slik veilederen er formulert nå, kan det være uklart om forskningsdata omfattes av begrepet «data fra offentlig sektor». En tydeligere definisjon vil bidra til å avklare hvilke datatyper

som inngår i veilederens virkeområde, og til å sikre en mer ensartet forståelse på tvers av sektorer og virksomheter.

I tillegg mener vi at veilederen i større grad bør anerkjenne at ikke alle data nødvendigvis kan eller bør tilgjengeliggjøres. Selv om åpenhet og videre bruk av data er viktige mål, finnes det også legitime hensyn som må ivaretas. Utover juridiske og etiske begrensninger kan enkelte datasett representere betydelige investeringer, unik kompetanse og langsiktig forsknings- og utviklingsinnsats. For universiteter, forskningsinstitusjoner og andre kunnskapsorganisasjoner kan slike data utgjøre en strategisk ressurs og et konkurransefortrinn i forskning, innovasjon og eksterne samarbeidsprosjekter. Vi anbefaler derfor at veilederen tydeliggjør at vurderinger av tilgjengeliggjøring må baseres på en balansert avveining mellom ønsket om åpenhet og hensynet til å beskytte legitime institusjonelle interesser.