

Høringsvar

Veileder ved tilgjengeliggjøring av data fra offentlig sektor

Navn: Meteorologisk institutt

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?

Skriv inn svaret deres her:

Veilederen er enkel og forståelig, og den har med de viktigste prinsippene for god dataforvaltning. Den vil derfor være svært nyttig for de som tilgjengeliggjør data.

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?

Skriv inn svaret deres her:

MET er ikke en stor bruker av andre offentlige data, men vi ser at veilederen i nåværende form har høy fokus på søksmetadata og ikke så mye på bruksmetadata. Vi savner mer fokus også på bruksmetadata. Som bruker er det svært viktig at datatyper og definisjoner er oppgitt og tilgjengelig sammen med dataene, for eksempel hvilken enhet de er representert i og definisjonen av denne. Det kunne kanskje med fordel vært lagt til et punkt "Gjør dataene brukbare" der dette utdypes.

Ellers dekker veilederen de behovene vi har for øyeblikket.

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

Skriv inn svaret deres her:

- *Informere om veilederen gjennom publisering og omtale på møter, konferanser, m.m.*
- *Kontinuerlig videreutvikle veilederen og sikre at det ikke blir et statisk dokument som ikke følger med på datadelingsinfrastrukturens utvikling. Samtidig er det en utfordring innenfor datadeling å sikre balanse mellom anvendelse av ny teknologi og bevaring av funksjonalitet på tvers av systemene som bidrar. Fokus må være på effektiv deling av informasjonsinnhold og å sikre funksjonalitet og samvirke på tvers av systemer underveis i utviklingen.*
- *Det blir etterhvert viktig å demonstrere nytteverdien av datadeling i henhold til veilederen, og å bidra til at brukerne har mulighet til å utnytte dette.*

Rådet foreslår at veilederen erstatter Retningslinjer ved tilgjengeliggjøring av offentlige data ([lenke til regjeringen.no](https://www.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?

Skriv inn svaret deres her:

Så lenge veilederen skiller tydelig mellom krav og anbefalinger, mener vi at dette er dekket. Spesielt viktig er det som dere sier, at veilederen er et levende dokument som kontinuerlig videreutvikles. Det er nytteverdien og den hjelpen veilederen gir til de som skal dele dataene som på sikt gir god og nyttig datadeling i offentlig sektor.

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?

Skriv inn svaret deres her:

Ja, dette er en god måte å tydeliggjøre hva som er viktig i forhold til lovene. Vi er usikre på om unik identifikator kan være et krav, men den er både svært viktig og nyttig.

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?

Skriv inn svaret deres her:

Ja, FAIR-prinsippene er mulig å etterleve, men det vil også alltid være forbedringspotensiale. Derfor er en prioriteringsliste viktig og nyttig. Graden av FAIR etterlevelse gir ulik nytte for brukerne, fra å kunne finne data og til å kunne bruke de i enkle til avanserte analyser.

Etter tabellen med FAIR-prioritering er det gitt en liste over relevante punkter i veilederen. Det kan være nyttig å bygge denne informasjonen inn i tabellen fremfor å gjengi den etter tabellen.

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

Skriv inn svaret deres her:

Dette er bra, men det er kanskje lurt å øke ambisjonsnivået for FAIR-prinsippene. De er såpass generiske at det er viktig med veiledning på hvordan de skal tolkes i større grad enn det som er gjort.

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?

Skriv inn svaret deres her:

Vi har søkt etter ordet “gjenfinnbar” i norske ordbøker på nett. Det nærmeste vi kommer er “gjenfinne”, som defineres som å “finne igjen”. “Findable” gjelder etter vår forståelse mer enn dette, nemlig at et datasett også skal kunne “finnes”, “oppdages” eller “søkes opp” uten at man kjenner til det på forhånd. Derfor mener vi “søkbar” er bedre enn “gjenfinnbar”. Eventuelt er kanskje «finnbar» også dekkende.

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.

Skriv inn svaret deres her:

I bakgrunnsskrivet for veilederen påpekes det at «Dataprioriteringsrådet viderefører ikke punktet om å tilpasse data til brukernes behov» fordi det «er mer hensiktsmessig å standardisere heller enn å tilby skreddersøm». I den forbindelse kan det være nyttig å legge til et setningsledd i innledningen til veilederen: «Samtidig er det et stort potensial for økt deling og bedre tilrettelegging av dataene, slik at de er enkle å finne, vurdere og å ta i bruk for brukere utenfor den umiddelbare målgruppen», evt. «for brukere man ikke kjenner».

Kommentar til punkt 1, “Dokumenter datasettene”: Det er nevnt at man “som minimum [bør] dokumentere tittel, beskrivelse, tema, kontaktpunkt og utgiver”. Dette er viktige søksmetadata, men det kan med fordel påpekes at man også bør gjøre bruksmetadata tilgjengelig. Dette inkluderer spesifisering av enheter og definisjon av disse, ifølge en gitt standard. For eksempel er det viktig for brukerne å vite om temperatur er målt i Celsius, Fahrenheit eller Kelvin. CF-standarden løser dette ved hjelp av standardnavn og tilhørende definisjoner (f.eks., “air_temperature”).

Kommentar til punkt 4, «Bruk maskinlesbare og standardiserte formater»: Det nevnes eksempler på maskinlesbare og standardiserte formater (CSV, XML, JSON, ODT og RDF-serialiseringer som RDF/XML, JSON-LD og Turtle). Selv om dette er teknologiske standarder, så er de ikke standardisert i forhold til informasjonsinnhold og struktur. Dette gjør det utfordrende å følge opp punkt 13, «metadata følger dataene» og dermed enkelt gjenbruke data, fordi ulike dataprodusenter har full frihet i forhold til hvordan de registrerer, spesifiserer og formidler metadata. Det blir dermed utfordrende å sikre at metadata er forståelig, både ved menneskelig og maskinell bruk. Vi foreslår å endre anbefalingen til «Bruk maskinlesbare formater standardisert både på teknologisk grensesnitt og informasjonsinnhold», eller lignende. Ved Meteorologisk institutt bruker vi CF-NetCDF og ACDD (https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3) for å definere bruks- og søksmetadata i kildefilene. CF (bruksmetadata) og ACDD (søksmetadata) er ikke låst til NetCDF.

Punkt 4 påpeker også at data skal tilgjengeliggjøres i alle eksisterende formater og språkversjoner. For å unngå kostnadsdrivende duplisering, kan det påpekes at dette bør gjennomføres ved oversettelser fra et «master»-format.

Kommentar i forbindelse med punkt 12: For å begrense risiko med falske offentlige data, bør alle delte data registreres i en søkbar metadatakatalog, f.eks. data-norge, der unik identifikator og sjekksum for opprinnelsesfilen er registrert. På denne måten er det mulig å kontrollere at data er ekte. Det ser ut til at regelverket ikke legger opp til krav om ID, men dette bør være sterkt anbefalt.