

Innspill til rapport om rettighets- og lisensspørsmål ved tilgjengeliggjøring av forskningsdata

Uttalelse - Universitetsbiblioteket i Tromsø

Status	Innsendt til Norges forskningsråd Innsendt og bekreftet av instansen via: philipp.conzett@uit.no
Innsendt av	Philipp Conzett
Innsenders e-post:	philipp.conzett@uit.no
Innsendt dato	01.09.2021
Hvilket fakultet eller institutt?:	Universitetsbiblioteket i Tromsø Fakultet / Institutt / Senter ved universitet eller høyskole (Offentlig)
Stilling	Førstebibliotekar

✓ Jeg bekrefter at uttalelsen er på vegne av hele fakultetet eller instituttet.

Innspill til rapport om rettighets- og lisensspørsmål ved tilgjengeliggjøring av forskningsdata

Avsender: Norges forskningsråd

1. 1. INNLEDNING

På vegne av Kunnskapsdepartementet har Forskningsrådet og UNIT nedsatt et utvalg som har fått i oppdrag å utrede rettighets- og lisensspørsmål knyttet til deling av datasett fra forskning. Forskningsrådet er sekretariat for utvalget. I fjor høst leverte utvalget et notat til Kunnskapsdepartementet som beskriver prinsipielle spørsmål de mener bør behandles videre, hvilke strukturelle endringer som bør vurderes for at Norge skal oppnå mer deling og gjenbruk av forskningsdata, og hvilke lovverk som berører disse problemstillingene. Basert på notatet ba utvalget om innspill fra relevante aktører i løpet av vinteren og våren 2020/2021.

Innspillene har kommet inn i flere former, blant annet gjennom skriftlige innspill, møter med ulike aktører og en åpen høringskonferanse i mars 2021. Nå har utvalget gjennomgått innspillene fra første fase og innarbeidet dette i et utkast til rapport. Det er denne utkastet til rapport du nå er i ferd med å gi innspill til. Endelig rapport fra utvalget leveres til Kunnskapsdepartementet til høsten.

Fristen for innspill er 1. september.

2. 2. HVA ØNSKER VI INNSPILL PÅ?

Vi ønsker generelt innspill på innholdet i utkastet til rapporten, om det er dekkende og om det er nyansert nok. Vi setter pris på konkrete tilbakemeldinger og forslag. Vi ønsker spesielt innspill på foreløpige anbefalinger (som ennå ikke er diskutert ferdig i utvalget).

3. 3. HVEM KAN SPILLE INN?

Vi ønsker innspill fra bredden av interessenter i kunnskapssektoren. Vi oppfordrer til ett felles innspill pr organisasjon, men det er også mulig å sende inn individuelle svar eller svar på vegne store og små grupper. Vi ber om at det informeres om hvem man representerer i svaret.

4. 4. OM HØRINGSVERKTØYET

- Det er mulig å videresende skjemaet og gi både lese- og skrivetilgang til andre.
- For å sende inn det endelige innspillet må du huske å trykke "lagre" og "send inn". Du vil motta en kvittering med svarene du har sendt inn.

5. 5. SPØRSMÅL OM HØRINGEN

For evt. spørsmål som gjelder revisjonsarbeidet, vennligst kontakt: Siri Lader Bruhn slb@forskningsradet.no eller Nenitha Charlotte Dagslott ncd@forskningsradet.no

6. 6. MED DITT SAMTYKKE

Ved å sende innspillene, samtykker du/dere til at vi kan behandle og samle innspillene for spesifikke formål som for eksempel i rapporter og for publisering på nettsiden.

Vi ber om dine kontaktopplysninger (navn, e-post og evt. institusjon). Vårt behandlingsgrunnlag er GDPR art. 6 (1) (e). Høringsinnspillene regnes som saksdokumenter og vil bli arkivert. [Les vår personvernerklæring](#) for mer informasjon om hvordan vi behandler dine personopplysninger.

? Til del 1: Er beskrivelsene av forskningsdata dekkende? Begrunn gjerne svaret.

Delvis enig

Innleiinga og del 1 summerer opp og problematiserer på ein god måte viktige aspekt og utfordringar ved deling og gjenbruk av forskingsdata. Vi har derimot nokre innspel til enkelte punkt.

Følgjande overordna innvending går på heile rapporten:

Dersom det berre er forskaren sjølv og arbeidsgjevaren til forskaren som har eventuelle rettigheter til dataa til forskaren og begge ønskjer at dataa skal delast, så er deling heilt uproblematisk. Dette er sjølvsagt eit opplagt poeng, men det kan lett hamna i skuggen for mange til dels kompliserte juridiske vurderingar i rapporten. Dei uproblematiske sidene ved datadeling bør derfor koma tydelegare fram i rapporten.

For meir konkrete innspel til innleiinga og del 1 sjå nedanfor.

? Har du andre kommentarer til del 1 av rapporten? Henvis gjerne til kapittelnummer ved kommentarer til spesifikt innhold.

INNLEDNING

Bakgrunn

«Innenfor enkelte fagområder, som for eksempel humaniora og samfunnsvitenskap, kan det i tillegg være metodiske aspekter ved datainnsamlingen som gjør det vanskelig å skille "rene" data ut fra selve forskningsaktiviteten.»

>> Denne setninga er litt uklar. Utfordringa består meir i at det i enkelte fag ikkje er beint fram å skilja data frå forskingsresultat og/eller at det er brukt andre omgrep enn «data» om grunnlagsmateriale som er brukt for å nå fram til forskingsresultat. Jamfør det som står i innleiinga til kap. 1 i del 1.

DEL 1

1.3. Kildedata vs. data generert underveis i forskningsprosessen

«Det kan også være snakk om data fra annen forskning.»

>> Det som er sagt i dette avsnittet om kjildedata, gjeld også for deriverte/prosesserte data: Om ein gjenbruker slike data av ein annan forskar, så må ein ta omsyn til kva lisens(ar), avtaler og lovverk som gjeld for desse dataa. Dette kan med fordel presiserast i im omtalen av kjildedata, t.d. slik: «Det kan også være snakk om genererte data fra annen forskning.».

1.4 Kvalitative vs. kvantitative data

«Kvantitativ metode brukes gjerne for å teste ideer og hypoteser som er framlagt i starten av et prosjekt og innebærer at man innhenter store mengder data ut fra forhåndsbestemte parametere.»

>> Det finst mange og gode døme på kvantitativ forskning som er basert på små eller mellomstore datamengder. Den typen forskning er i utgangspunktet ikkje mindre viktig enn forskning basert på stordata. Vi føreslår derfor å sløyfa «store mengder» i denne setninga.

1.5 Persondata vs. data om verden for øvrig

Det finst også andre typar data som treng eit særleg vern, men av andre grunnar enn personvern, t.d. data som er relaterte til artsmangfald. Dette bør kort nemnast i rapporten.

«Men med et eksplisitt samtykke om [til?] deling, så vil det være mulig etter de vilkår som er beskrevet. Da vil det være aktuelt å bruke en lisens eller en frigivelseserklæring.»

>> I og med at eit samtykke kan trekkjast tilbake, herskar det ein del tvil og forvirring om deling, og då særleg open deling, av persondata. Dette bør tematiserast i rapporten.

2.1 Åpne data

Opne data blir definert slik i dette avsnittet: «At data er åpne, betyr at de deles uten begrensinger på hvem som kan bruke dataene (tilgangsbegrensninger) eller hva dataene skal brukes til (formålsbegrensinger).»

>> Etter vårt syn er dette ein for snever definisjon av opne data. Det finst ulike gradar av openheit. Eit datasett som er lisensiert med CC0 er meir open enn eit datasett med CC BY-NC, men det blir rart å kalla sistnemnde for lukka data. Med prinsippet «så ope så mogleg, så lukka som nødvendig» opnar nasjonale retningsliner for ei meir kontinuerleg fortolking av opne data enn den heller snevre tilnærminga som aktørar som Open Knowledge Foundation har. Ei for snever tolking av kva type data som «får lov» til å få stempelen «opne data», bidreg etter vår vurdering ikkje til kulturendringa som er nødvendig for å få til ein overgang mot meir open deling av data.

5. Forskningsdata som er kvalifisert for deling og gjenbruk

«Et annet eksempel er persondata som må anonymiseres i så stor grad at det i praksis forringer datakvaliteten og i noen tilfeller ikke lenger kan defineres som gjenbrukbare forskningsdata.»

>> Nye teknologiar held på å bli utvikla som skal gjera det mogleg å analysa persondata utan tilgang til dataa i ikkje-anonymisert form. Eit døme på slik teknologi er såkalla «Differential Privacy» (jf. <https://opendp.org/>). Setninga ovanfor kan derfor med fordel presiserast noko, t.d. slik: «Et annet eksempel er persondata som må anonymiseres i så stor grad at det i praksis forringer datakvaliteten og i noen tilfeller - gitt dagens teknologiske muligheter - ikke lenger kan defineres som gjenbrukbare forskningsdata.»

? Til del 2: Er innholdet i kapittelet 6. "Juridisk rammeverk for deling og gjenbruk av forskningsdata" dekkende? Begrunn gjerne svaret.

Delvis enig

Vi er samde i konklusjonane og tilrådingane, men har ein del kommentarar på dei enkelte punkta.

6.1 Internasjonalt lovverk

Med tanke på kor viktig det er at nasjonalt lovverk i størst mogleg grad er harmonisert med internasjonalt lovverk for å fremja deling og gjenbruk av forskingsdata hadde det vore nyttig å få kommentert del 2 i rapporten basert på ein gjennomgang av internasjonalt lovverk og forholdet mellom nasjonalt og internasjonalt lovverk. I mange tilfelle vil forskarar ha behov for å gjenbruka data frå andre land, og derfor treng vi eit godt og harmonisert juridisk grunnlag som gjer det enkelt å gjenbruka data og gjera tilgjengeleg data som byggjer på gjenbrukte data.

6.2.1 Immaterialrettslig regelverk

Vern for samling av forskningsdata (databasevern)

«Det er verdt å merke seg at investeringene skal knytte seg til innsamlingen, kontrollen eller presentasjonen, altså arbeidet med struktureringen og sammensetningen av databasen.

Investeringer i forbindelse med produksjon av data gir ikke vern.»

>> I mange tilfelle blir data genererte gjennom innsamling (t.d. gjennom registrering, måling, intervju, spørjeskjema). Det er derfor uklart korleis setninga ovanfor skal tolkast. På den eine sida blir det her slege fast at «vesentlig investering» i innsamling gjev/kan gje databasevern, medan det på den andre sida blir slege fast at produksjon av data ikkje gjev vern. Det bør gå tydelegare fram kva det er som må ha kravd ei vesentleg investering? Er det 1) innsamling av data + strukturering av databasen, eller 2) berre struktureringa (= oppsettet/designet?) av databasen? Det bør presiserast kva ein meiner med «produksjon av data» i denne samanhengen.

6.3.1 Begrensninger som skal beskytte datasubjektet

Personopplysningsloven

«Personvernforordningen har til formål å verne fysiske personer i forbindelse med behandling av personopplysninger [...]».

>> Her hadde det vore fint med ei presisering av at dette vernet i norsk lovgjeving er avgrensa

til levande personar. (Andre land kan ha unntak frå dette, t.d. Danmark og Frankrike; jf. <https://aanoip.org/life-after-death-data-protection-rights-of-deceased-persons/>).

6.5 Om eierskap til forskningsdata – kan data eies?

«For data som ikke er vernet av noen regler, er rådende oppfatning at det er institusjonen som har råderett over dataene.»

>> Kva er denne rådande oppfatninga basert på? Finst det noko rettsleg grunnlag som stør opp om denne oppfatninga? Eller er oppfatninga mest basert på «institusjonens oppfatning om arbeidsgivers rettigheter til ansattes arbeidsresultater»?

«I den grad man bruker begrepet "eie", videreføres det en viss usikkerhet til hva denne posisjonen juridisk sett innebærer.»

>> Kva alternativ omgrep i staden for «eie»/«eierskap» tilrår utvalet at institusjonane skal bruka?

6.6. Utvalgets vurdering av det juridiske rammeverket

Her bør det også løftast fram at det er viktig med nasjonal koordinasjon og samarbeid for å sikra at institusjonane har/får utvikla den nødvendige juridiske kompetansen på dette området.

«Utvalget oppfatter derfor at det antagelig ikke er hensiktsmessig å lovregulere forskningsdata særskilt i dag.»

>> Derimot er det behov for å få etablert gode nasjonale praktiske rettleiingar som gjev eit samla oversyn over råd til forskarar om korleis dei på enklast mogleg måte kan gjenbruka andre sine forskingsdata og gjera sine egne data tilgjengelege.

«Etter utvalgets syn bør spørsmål rundt "eierskap" til data søkes avklart i størst mulig grad, gjennom utredning av juridiske spørsmål, gjennom retningslinjer ved institusjonene og i klare rammeverk for avtaler med tredjeparter.»

>> Og då er det viktig at desse rammeverka og avtalane er utforma, slik at det vil vera enkelt å eventuelt dela slike data i tråd med rammeverka og avtalane.

? Er innholdet i kapittel 7. "Om lisenser og frigivelseserklæringer for deling og gjenbruk av forskningsdata" dekkende? Begrunn gjerne svaret.

Kapittelet dekkjer i stor grad utfordringane knytte til temaet. Men vi ønskjer nedanfor å kommentera enkelte punkt i utgreiinga.

7.1 Om lisenser og deres funksjon

I avsnitt 2 (jf. «En lisens er en kontrakt...») bør det gå tydelegare fram at også lisensar og ev. andre bruksvilkår knytte til data som er blitt gjenbrukte, må ivaretakast i ein lisens. Kanskje dette inngår i «tredjeparters rettigheter til dataene», men det er såpass viktig, og til dels også krevjande å etterleva, at det bør nemnast eksplisitt.

7.2.1 Lisensgivers ansvar – Avklaring av rettigheter og delingsrestriksjoner

«En lisensgiver er i dette tilfellet en forsker eller institusjon hvor forskningsdataene er produsert

av ansatte gjennom sitt ansettelsesforhold.»

>> Her hadde det vore nyttig med ei avklaring på kven det er som har rett til å lisensiera? Når er det forskaren, når er det institusjonen, når er det begge, når er det ingen av dei?

7.3.1 Opne standardlisenser

«Både internasjonale og nasjonale styrende dokumenter, policy og retningslinjer anbefaler hovedsakelig bruk av åpne standardlisenser på offentlig finansierte forskningsdata som åpner for videst mulig tilgang, gjenbruk og videredistribusjon.»

>> Her hadde det vore nyttig å nemna at det ikkje er semje om kva lisens som bør klassifisera som open lisens. Til dømes opnar OpenAIRE (jf. <https://www.openaire.eu/research-data-how-to-license/>) for at også SA og ND kan vera i tråd med «the Open Access definition».

«Disse lisensene er mye brukt blant forskere på grunn av vilkåret om attribusjon som er til stede i alle seks CC-lisenser, men vilkåret er juridisk sett bare gjeldende der det lisensierte materialet er gjenstand for opphavsrett eller tilgrensande vern. Dette framkommer av lisensteksten under Del 8 – Tolkning: "a. For å unngå tvil, denne Offentlige Lisensen begrenser ikke, og skal ikke bli tolket slik at den reduserer, begrenser eller setter vilkår for, bruken av det Lisensierte Materialet dersom bruken er lovlig uten at det er gitt tillatelse under denne Offentlige Lisensen." Bruk av CC-lisenser på data uten opphavsrettslig vern, som et virkemiddel for å oppnå kildehenvisning og sitering, bidrar til forvirring rundt hva som er vernet av loven og hva lisensgiver ønsker å oppnå med lisensen og bør dermed unngås.»

>> I dette avsnittet diskuterer rapporten bruk av CC-lisensar på data som er gjenstand for opphavsrett (eller tilgrensande vern), og data utan opphavsrettsleg vern. Rapporten peiker på at vilkåra i CC-lisensane er juridisk bindande berre for data MED opphavsrettsleg vern og refererer til Del 8 pkt. a av den juridiske teksten for CC-lisensane (sjå t.d. her:

<https://creativecommons.org/licenses/by/4.0/legalcode.no>).

Men del 8 pkt. a handlar etter vår tolking ikkje om spørsmålet om materiale er opphavsrettsleg verna eller ikkje. Del 8 pkt. a handlar om bruk av lisensierte materiale (altså materiale som ER gjenstand for opphavsrett eller tilgrensande vern) som ikkje treng løyve frå rettighetshavar. Slik bruk kan vera det som fell under avgrensing av opphavsretten (exceptions and limitations to copyright), og som i norsk åndsverklov er avgrensingar av opphavsretten beskrivne i kapittel 3. Når det gjeld databasar, kan slik bruk (altså bruk som ikkje treng løyve frå rettighetshavar) i tillegg vere uttrekk eller gjenbruk av uvesentlege delar av ein database – vel å merka berre dersom uttrekket/gjenbruken ikkje er gjenteken/systematisk (åvl, kap. 2, § 24).

Det at vilkåra i CC-lisensar er juridisk sett gjeldande berre for materiale med opphavsrettsleg vern går fram av del 1, pkt. f og g, som beskriv kva «lisensiert materiale» og «lisensierte rettigheter» kan vera under denne lisensen:

- f. Lisensiert Materiale betyr det kunstneriske eller litterære verk, database, eller annet materiale som Lisensgiver har underlagt denne Offentlige Lisensen.

- g. Lisensierte Rettigheter betyr rettigheter som gis Deg i henhold til vilkårene og betingelsene i denne Offentlige Lisensen, som er begrenset til Opphavsrett og Lignende Rettigheter som gjelder for Din bruk av det Lisensierte Materialet og som Lisensgiveren har rett til å lisensiere. (<https://creativecommons.org/licenses/by/4.0/legalcode.no>)

7.3.2 CC0 og Public Domain Mark – data i det fri

«Det innebærer at opphaveren til en database som er vernet som åndsverk, eller rettighetshaver til enkelte verk som inngår i en database, alltid kan kreve å bli navngitt eller å påtale bruk av verket som er i strid med respektretten, uansett om datasamlingen er gjort tilgjengelig med en CC0 frigivelseserklæring eller annen rettighetsfraskrivelse.»

>> CC0-vilkåra tek eksplisitt høgde for tilfelle der CC0 ev. er i strid med lovgjevinga, og opphavaren stadfestar gjennom bruk av CC0 at han/ho ikkje kjem til å forfølgja gjenbruk av dataa sine dersom det skulle visa seg at det ifølgje lov ikkje var mogleg å gje avkall på opphavsretten eller andre rettigheter han/ho måtte ha til datasettet. Dette skriv CC:

«Should any part of the License for any reason be judged legally invalid or ineffective under applicable law, such partial invalidity or ineffectiveness shall not invalidate the remainder of the License, and in such case Affirmer hereby affirms that he or she will not (i) exercise any of his or her remaining Copyright and Related Rights in the Work or (ii) assert any associated claims and causes of action with respect to the Work, in either case contrary to Affirmer's express Statement of Purpose.»

Det er med andre ord heller ikkje i Noreg problematisk å bruka CC0 på forskingsdata, heller ikkje for data som i utgangspunktet er verna av opphavsretten, dersom det berre er forskaren sjølv som har laga datasettet og det ikkje er basert på andre sine data eller andre produkt. Grunnen til det er at dersom CC0 skulle vera i strid med norsk opphavsrett fordi ein ikkje kan gje avkall på opphavsrettigheter, så vil jo ein misbruk av denne opphavsretten berre vera til ulempe for sjølve opphavaren, dvs. forskaren som sjølv har valt å fråskriva seg denne retten. Ved å velja CC0 har forskaren stadfesta at "he or she will not (i) exercise any of his or her remaining Copyright and Related Rights in the Work or (ii) assert any associated claims and causes of action with respect to the Work, in either case contrary to Affirmer's express Statement of Purpose".

Det uproblematisk ved bruk av CC0 i slike tilfelle bør gå fram av rapporten. Dette for å unngå (vidarespreiing av) ei ganske utbreidd feilaktig oppfatning av at det i Noreg «ikkje er lov» å bruka CC0 på opphavsrettsleg verna materiale.

7.3.4 God henvisningsskikk for forskningsdata

Dei etiske prinsippa for tilvising/kreditering det er referert til, gjeld forskning/vitskap. Men kva med gjenbruk av data til andre formål enn forskning / utanfor akademia, t.d. kommersialisering, journalistikk, osv.? Korleis sikrar vi tilstrekkeleg kreditering ved slik gjenbruk? Er det då nok med forskningsetiske normer, eller treng vi meir generelle etiske retningslinjer?

? Har du andre kommentarer til del 2 av rapporten? Henvis gjerne til kapittelnummer ved kommentarer til spesifikt innhold.

8.1 Mer lukket enn nødvendig

«Siden det kan være vanskelig å forutse hva data skal brukes til når de deles åpent, gir informert samtykke bare mulighet til å dele data som inneholder personopplysninger, der datasubjektet har blitt informert eksplisitt om potensiell bruk og har samtykket til dette.»

>> Vil det seia at det å spørja om lov til å gjera dataa ope tilgjengeleg for ein generell gjenbruk av typen «forskning» ikkje oppfyller krava til informert og eksplisitt samtykke som definert i GDPR?

I kapittel 7.1 «Om lisenser og deres funksjon» står det at ein lisens er nødvendig når det ligg føre rettigheter til data. Men vi kan ikkje sjå at det går tydeleg fram ein stad i rapporten at det ikkje er høve til / noko poeng med å bruka restriktive lisensar (anna enn CC0 eller PD) på data som ikkje har nokon rettigheter knytte til seg. Ved gjenbruk av opne data utan rettigheter vil ein kunna sjå bort frå slike «uberettiga» lisenskrav. Dette aspektet kunne med fordel ha blitt drøfta i kap. 8.1.

8.2 Navngivelsesstabling

«Et juridisk argument for å unngå navngivelsesstabling vil være å anbefale bruk av frigivelseserklæringer, uten vilkår om navngivelse. Men som det er diskutert ovenfor er god henvisningsskikk ved bruk av andres forskningsdata et lovbestemt krav og forskningsetisk prinsipp og plikt i forskningen.»

>> Kva indikerer «men» i andre setninga? At det forskningsetiske prinsippet om god tilvisingsskikk strir imot det juridiske argumentet i den første setninga? Eller at det stør opp om det juridiske argumentet?

Det nemnde forskningsetiske prinsippet dikterer ikkje at ein rettighetshavar skal bruka ein lisens med namngjevingsvilkår på dataa sine. Det forskningsetiske kravet om kreditering er uavhengig av opphavsretten og ev. lisensar og andre juridiske verktøy, med eller utan namngjevingsvilkår. Når ein namngjev ut frå forskningsetiske prinsipp, gjer ein det i samsvar med god tilvisingsskikk på fagområdet sitt. Namngjevingsvilkåret i ein lisens (t.d. i ein CC-lisens) derimot er ein streng mekanisme som krev at bestemte element er med i namngjevinga, uavhengig av tilvisingsskikk på eit fagområde.

10.1 Tydelig ansvarsfordeling mellom den enkelte forsker og institusjon

Sjølv om dette til dels er drøfta i kap. 6.5 «Om eierskap til forskningsdata – kan data eies?», så hadde det her (kap. 10.1) vore nyttig med ei tydeleg oppsummering av det juridiske grunnlaget ein institusjon kan bruka for å etablere ei tydeleg ansvarsfordeling mellom forskar og institusjon. Kva slags plikter, rettigheter og ansvar har forskaren og institusjonen for ulike typar data: data

med opphavsrett, database med databasevern osv. Eit viktig punkt her er også retten til definera når data skal delast.

? Til del 3: Gir innholdet i del 3 en god beskrivelse av ikke-juridiske barrierer for deling og gjenbruk av forskningsdata? Begrunn gjerne svaret.

Delvis enig

Del 3 gjev ei god beskriving av ikkje-juridiske barrierar for deling og gjenbruk av forskingsdata. Vi har derimot nokre innspel til enkelte punkt; sjå nedanfor.

? Har du kommentarer til anbefalingene som står under overskriften "Regjeringen og departementene"?

«Kreve at etablering og utvikling av nasjonale systemer og tjenester som er relevant for deling i data, har en internasjonal orientering.»

>> Dette er eit veldig viktig poeng, som gjerne kan drøftast litt meir detaljert i brødteksten, og då gjerne med referanse til nokre sentrale initiativ/rammeverk på området, t.d. Invest in Open Infrastructure (<https://investinopen.org/>) og The Global Sustainability Coalition for Open Science Services (SCOSS; <https://scoss.org/>).

? Har du kommentarer til anbefalingene som står under overskriften "Finansierer og virkemiddelapparatet"?

«Forskningsrådet bør lyse ut egne midler til datainfrastrukturer som bidrar til mer deling og gjenbruk av forskingsdata.»

>> Her kan det gjerne leggjast til forskning på FAIR data.

? Har du kommentarer til anbefalingene som står under overskriften "Forskningsinstitusjonene"?

Nei.

? **Har du andre kommentarer til del 3 av rapporten? Henvis gjerne til kapittelnummer ved kommentarer til spesifikt innhold.**

12.2 Internasjonal orientering

«Mange norske forskere forholder seg ...»

>> Bruk gjerne ei meir inkluderande formulering, t.d. «Mange forskere i Norge forholder seg ...».

14.2 Investering i kompetanse om FAIR data

I arbeidet med å gjera forskingsdata så FAIR som mogleg er det i mange tilfelle nødvendig med disiplinifagleg forskarkompetanse. Dette arbeidet bør sjåast på som ein naturleg del av forskingsverksemda, og det er lite føremålstenleg å laga eit arbitrært skilje mellom forskning og administrasjon på dette området. Ein bør leggja til rette for attraktive vitskaplege karriereløp og akademisk meritering av arbeid som går på å utvikla og drifta infrastruktur og meir generelt god forvaltning av forskingsdata.

14.3 Prioriterte data for FAIR deling

«På grunn av store kostnader trengs det imidlertid en prioritering av hvilke forskningsdata som skal gjerest fullt ut FAIR, og det kan være nødvendig å sette kriterier for hvilke data som betraktes som prioriterte i en første fase.»

>> Ei slik prioritering som skissert i dette avsnittet kan tolkast som ei todeling av forskings-Noreg i éi gruppe som skaffar fram forskingsresultat som er samfunnsnyttige, og ei anna gruppe som produserer data og funn som ikkje utan vidare har nokon verdi for samfunnet. Erfaringa viser derimot at det ikkje utan vidare er mogleg å setja opp kriterium for kva data som har eller potensielt kan ha framtidig verdi for forskinga og samfunnet elles. Tvert imot har det vist seg at data frå eitt område etter mange år har verdi for forskning på eit anna område. Ved å leggja (for) stor vekt på ei siling av kva data som er «høgverdige» nok kan vi risikera at nyttige data ikkje blir tekne (godt nok) vare på for framtidig forskning. Ei prioritering som det er lagt opp til i dette avsnittet, kan også vera i strid med det Kunnskapsdepartementet skriv i sin Nasjonal strategi for tilgjengeliggjoring og deling av forskningsdata (s. 26): «Grunnprinsipp 3: Beslutninger om arkivering og tilrettelegging av forskningsdata må tas i forskerfellesskapene.» Dersom eit forskerfellesskap konkluderer med at eit datasett (eller meir generelt: ein type data) skal gjerest tilgjengeleg for vidare bruk, så bør sektoren ha dei nødvendige ressursane på plass for å gjera desse dataa så FAIR som mogleg.

«Det betyr bare at de [= data som ikkje blir prioriterte i første omgang] ikke nødvendigvis må tilfredsstillast FAIR-prinsippene (men likevel deles med tilstrekkelig informasjon for at dataene kan etterprøves og kvalitetssikres).»

>> Per i dag er det svært få tilfelle innanfor svært få fagområde som kan seiast å vera 100 % i tråd med FAIR-prinsippa. Som det tidlegare er nemnt i rapporten, vil det ofte vera snakk om at data er meir eller mindre FAIR. For mange forskarar, fagfelt og datatypar ligg det enno eit godt

stykke fram i tid før dei kan svara fullt ut på FAIR-prinsippa. Men det bør likevel ikkje bety at ein ikkje bør ha som ambisjon å gjera dataa sine så FAIR som mogleg med dei moglegheitene og ressursane som står til rådvelde per i dag. Den siterte setninga ovanfor kan lett tolkast som støtte for at det ikkje er så viktig å streva etter å gjera data så FAIR som mogleg. Setninga kan også tolkast slik at det å dokumentera data «med tilstrekkelig informasjon for at dataene kan etterprøves og kvalitetssikres» ikkje er ein naturleg del av FAIR-prinsippa. For å unngå ei slik mis-/vrangtolking bør det leggjast til «fullt ut» i setninga:

«Det betyr bare at de ikke nødvendigvis må tilfredsstille FAIR-prinsippene fullt ut, men de bør likevel deles med tilstrekkelig informasjon for at dataene kan etterprøves og kvalitetssikres. [---] På sikt vil det være nødvendig med en plan for hvordan de mer uensartede dataene, den såkalte lange halen av forskningsdata, skal bli et verdifullt bidrag til et datadrevet samfunn og en datadreven økonomi.»

>> Mykje av forskingsdata frå den lange halen av forskinga er allereie eit verdifullt bidrag til eit datadrive samfunn; sjå t.d. The e-IRG Task Force on the Long Tail of Data. 2016. 'Long Tail of Data'. E-IRG Task Force Document. The Hague: e-IRG. <http://e-irg.eu/documents/10920/238968/LongTailOfData2016.pdf>.

Med vennlig hilsen

Mari Sundli Tveit
Administrerende direktør i Norges
forskningsråd

Dokumentet er elektronisk godkjent