

Forskningsdata i et datadrevet samfunn

Et notat fra utredningsutvalget for rettighets- og lisensspørsmål
i forbindelse med deling av datasett i forskningen

Oktober 2020

Innhold

Bakgrunn	3
Om dette notatet	4
1. Hva er det som er så annerledes med forskningsdata?	4
Bruk av begrepet forskningsdata i dette notatet.....	5
Åpne data versus lukkede data.....	5
Åpne data versus FAIR data	6
Forskningsdataenes livssyklus og egenskaper	7
2. Forskningsdata i et ouversiktlig juridisk landskap	8
3. Om lisenser for deling og gjenbruk av forskningsdata.....	10
Akademisk frihet, institusjonenes styringsrett og forskningsetikk	11
4. Utfordringer knyttet til deling av forskningsdata spesielt	12
Åpne forskningsdata i et resultatstyrt og merittbasert system.....	12
Hva er formålet – deling eller gjenbruk?	13
Infrastrukturer for deling og gjenbruk av forskningsdata og bærekraftige finansieringsmodeller	13
Komplekse forskningsdata, ressursbehov og krav til kompetanse.....	14
Kommersialisering basert på forskningsdata.....	15
Om lovverk og regulering for deling og gjenbruk av forskningsdata	16
Oppsummerende sluttkommentar fra utvalget	16

Bakgrunn

Åpenhet og kunnskapsdeling står i sentrum for forskning og innovasjon. Samtidig står samfunnet overfor store utfordringer som må løses blant annet for å sikre global bærekraft og utvikling av kunnskapsbasert næringsliv i Norge. Den teknologiske utviklingen de siste tiårene har ført til store endringer i hvordan forskning utføres, hvordan resultatene deles og hvordan de kan brukes på nye måter. Forskning og innovasjon blir i stadig større grad drevet framover av tilgangen til nye og store datamengder, og effektive strukturer for åpenhet og deling rundt forskningsdata er nødvendig for å både sikre datakvalitet, økt utnyttelse av forskningsfunn i samfunnet og effektiv utnyttelse av ressursene i forskningsmiljøene.

Data som er generert gjennom offentlig finansiert forskning bør i prinsippet forventes å bli delt på lik linje med andre offentlige data, i tråd med nasjonale og internasjonale føringer. Resultatene fra slik forskning er et fellesgode som har verdi både for vitenskapen og for samfunnet. Bedre tilgang til forskningsdata vil styrke kvaliteten og tilliten til forskningen ved at resultater kan valideres og etterprøves på en bedre måte. Forskningssystemet har tradisjonelt delt kunnskap i form av publisering av vitenskapelige artikler, mens data som artiklene er basert på i mindre grad er blitt offentliggjort. Det er i mange forskningsfelt urovekkende mange forskningsfunn som tilbakevises som følge av svak datakvalitet (lav reproduserbarhet). Samtidig er det en utfordring at det utarbeides mye forskningsdata som aldri blir brukt i artikler av ulike årsaker, og dermed aldri blir tilgjengelig for samfunnet. Mer systematisk deling av disse dataene vil gjøre at offentligheten og andre forskere kan etterprøve kvaliteten og holdbarheten i dataene som forskningen og publikasjonene er bygget på.

Samtidig vil deling av data representere en samfunnsverdi ved at både andre forskere og samfunnet ellers kan utnytte dataene på nye måter og i kombinasjon med andre datasett. Blant annet vil gode strukturer for datadeling kunne legge til rette for langt mer effektive oppsummeringer av det samlede kunnskapsgrunnlaget om ulike felt. Det er også et stort potensial for å øke samfunnets utnyttelse av forskningsdata til innovasjon og kommersialisering. Mer deling av forskningsdata vil legge til rette for utvidet tverrfaglig forskning og nyskapning i samfunns- og næringsutvikling, samtidig som det vil bidra til en mer nyansert offentlig debatt og bedre demokratiske prosesser, ved at befolkningen får økt innsikt i grunnlaget for forskningen.

Forskningssektoren står overfor mange av de samme utfordringene som andre sektorer når det gjelder deling av data. Etablerte standardformater, gode infrastrukturer, finansieringsmodeller, kompetanseheving og insentiver på både individ- og institusjonsnivå, vil være avgjørende for å oppnå mer deling og gjenbruk av data. Det er imidlertid noen ekstra utfordringer knyttet til deling av forskningsdata, både på grunn av egenskaper ved dataene, et uoversiktlig lovverk som regulerer rettigheter til dataene og manglende insentiver for deling i forskningssystemet. Lisenser brukes i økende grad for å sikre kreditering og regulere gjenbruk, og hensiktsmessig bruk av slike lisenser kan være et viktig virkemiddel for å oppnå mer deling og gjenbruk av forskningsdata. Bruken av lisenser er pr. i dag ikke konsistent, og det mangler gode retningslinjer. Til en viss grad brukes også lisenser med begrensninger for gjenbruk av dataene, og det er lite kunnskap om konsekvensene av bruksbegrensninger for den videre bruken av dataene i samfunnet. Det er behov for mer kunnskap og et

bevisst forhold til bruk av ulike lisenser i forskningsmiljøene for at forskningsdata skal bli et verdifullt bidrag i verdiskapingen i samfunnet.

For å oppnå mer deling og gjenbruk av forskningsdata er det behov for en systematisk gjennomgang av hvilke typer data som genereres gjennom forskning, hvilke delingsrestriksjoner som de ulike datatypene eventuelt er underlagt og hvilke lisenser som kan brukes for ulike datatyper for å oppnå mer deling og gjenbruk, uten at delingsrestriksjonene kompromitteres. Det er videre behov for å diskutere overordnede konsekvenser av lisensbruk for blant annet forskningskvalitet, publisering, kommersialisering og åpenhet i forskningen. Det er også behov for overordnede diskusjoner knyttet til hvordan forskningssystemet bør innrettes, herunder hvilke kompetanser som må styrkes, for å tilrettelegge for mer deling og gjenbruk av data.

Om dette notatet

For å finne gode løsninger på utfordringene som er skissert over, har Kunnskapsdepartementet gitt Forskningsrådet, i samarbeid med UNIT, i oppdrag å nedsette et utvalg som skal utrede blant annet rettighets- og lisensspørsmål i forbindelse med deling av datasett fra forskning. Utvalget har holdt tre møter og utarbeidet dette dokumentet, som blant annet kan tjene som underlag til Kunnskapsdepartementets innspill til stortingsmeldingen om datadrevet økonomi og innovasjon.

Utvalget har gjennomgått strukturer for rettigheter til forskningsdata og diskutert mekanismer som påvirker deling av data og utnyttelse av disse i samfunnet, og foreslår konkrete endringer som kan bidra til å stimulere til mer deling og gjenbruk av data. Forskningssystemet er klart preget av at insentivsystemene for institusjoner og enkeltforskere stimulerer til høyt publikasjonsvolum av artikler, men i liten grad til å dele forskningsdata. Personvern, immaterialrettigheter og kommersielle hensyn er andre forhold som må balanseres mot samfunnets behov for deling av forskningsdata. Utvalget har sett på rettigheter og strukturer for deling i lys av disse forholdene, og vurdert hvilke strukturer som kan legge til rette for effektiv deling av forskningsdata og økt kvalitetssikring og utnyttelse av slike data i samfunnet, inkludert i offentlig forvaltning, i næringslivet og innad i forskningssystemet.

I notatet gis det først en gjennomgang av hvilke egenskaper som kjennetegner forskningsdata og som det er nødvendig å være spesielt oppmerksom på i denne sammenhengen. Deretter gis det en oppsummering av hvilket juridisk rammeverk som enten regulerer eller kan påvirke hvordan forskningsdata deles i dagens system. I det etterfølgende diskuteres ulike typer lisenser og hvilke tilsiktede og utilsiktede konsekvenser valg av lisenser kan ha. Avslutningsvis presenteres ulike dilemmaer og problemstillinger som knytter seg til deling av forskningsdata spesielt. Under hvert punkt foreslår utvalget mulige tiltak som kan/bør vurderes for at forskningsdata skal bli et så bra bidrag som mulig til et datadrevet samfunn.

1. Hva er det som er så annerledes med forskningsdata?

Mange av løsningene som nå utredes og utvikles for deling av offentlige data generelt, vil også kunne overføres til forskningsdata. For å utnytte potensialet som ligger i forskningsdata, er det imidlertid noen særskilte egenskaper og trekk som bør tas i betraktning når det skal avgjøres hvilke retningslinjer som skal gjelde. Utvalget forsøker i dette notatet å skissere noen hovedlinjer som vi mener er viktige for å

forstå på hvilken måte forskningsdata skiller seg fra andre typer offentlige data, og som illustrerer hvorfor temaet kan være mer komplekst enn tilgjengeliggjøring av data fra for eksempel offentlig forvaltning. Hovedlinjene tar for seg forskningsdata sett i lys av livssyklus og egenskaper ved dataene, eierskap og rettigheter/plikter i gjeldende lovverk samt akademisk frihet og forskningsetikk.

Bruk av begrepet forskningsdata i dette notatet

Det finnes ingen entydig definisjon av begrepet forskningsdata, men en fellesnevner i de fleste definisjoner er at det avgrenses til digitale data i maskinlesbart format som er knyttet til forskningsprosesser i en eller annen form. [OECDs](#) definisjon fra 2007 avgrenser begrepet til data som er brukt som primærkilde for forskningen: *"factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings"*. I Forskningsrådets [policy for tilgjengeliggjøring av forskningsdata](#) skiller man i tillegg mellom data som finnes allerede, uavhengig av forskningen som skal gjennomføres (kildedata), og data som genereres underveis i forskningsprosessen. Kildedata kan være informasjon som er innhentet til andre formål enn forskning, for eksempel data fra offentlig forvaltning, helsevesenet, privat virksomhet eller data fra samlinger, litterære tekster eller annen kunstnerisk produksjon.

Skillet mellom kildedata og data som er generert i forskningen er vesentlig ved spørsmål om rettigheter til data og bruk av lisenser. Forskere bruker ofte kildedata til videre forskning, og må da forholde seg til lisenser, avtaler og lovverk som gjelder for kildedataene. Dette kan begrense videre deling.

Forskningsdata som er generert uten bruk av kildedata vil kunne friere deles videre av forskeren selv eller forskningsinstitusjonen i henhold til gjeldende lover og retningslinjer. I mange tilfeller vil forskning innebære en kombinasjon av bruk av kildedata og data generert underveis i prosjektene. Dette kan være med på å komplisere vurderingen av videre deling ytterligere. Dette diskuteres videre i [kapittel 2](#) om det juridiske landskapet.

Åpne data versus lukkede data

Åpen forskning har vært et viktig policy-område de senere årene og er en sentral strategi for å styrke etterprøvnbarheten av forskningen, og samtidig gjøre gjenbruk av data mulig for andre forskere, næringsliv og samfunnet. I takt med den raske teknologiske utviklingen har mengden data som innhentes i forskningsmiljøene og samfunnet for øvrig, økt vesentlig. Dette har ført til nye utfordringer knyttet til forvaltning av data, og vi har sett en stadig utvikling av politikken på området. I den tidlige fasen var det viktig å få med seg alle miljøer på idéen om at deling av data var nødvendig og riktig i moderne forskning, og man innførte konseptet "open by default". Dette har blitt møtt med en del kritikk, blant annet fordi det i liten grad tar hensyn til utfordringer knyttet til personvern, sikkerhet, kommersialisering og kostnader knyttet til å gjøre data åpne. Konseptet har derfor blitt noe moderert og man snakker nå heller om å gjøre data "så åpne som mulig og så lukket som nødvendig", noe som også gjenspeiles i Forskningsrådets [policy for åpen forskning](#).

Samtidig ser man et skifte av fokus fra selve tilgjengeliggjøringen og delingen, til den faktiske gjenbruken av datasettene. Det er ikke delingen i seg selv som er målet, delingen er snarere et middel for å nå målet om bedre utnyttelse av offentlig finansierte data til ny kunnskapsutvikling og utvikling av teknologi,

produkter og tjenester til nytte for samfunn og enkeltindivider. Betegnelsen "[FAIR data](#)" har fått fotfeste for data som er tilrettelagt for gjenbruk etter "FAIR-prinsippene" (*Findable, Accessible, Interoperable, Reusable*).

I EU-kommisjonens [Final Report of the Open Science Policy Platform](#) kommer dette skiftet tydelig fram. Rapporten peker på at det må være en balanse mellom krav om åpenhet og vern om immaterielle rettigheter i tråd med prinsippet "*så åpne som mulig, så lukkede som nødvendig*". Rapporten understreker også at det er FAIR-prinsippene som bør være målet, og ikke nødvendigvis at dataene skal være åpne. Rapporten etterlyser økt bevissthet rundt verdien av IPR (Intellectual Property Rights, norsk "immaterielle rettigheter") og de individuelle bidragene forskere gir til å løse de store samfunnsutfordringene.

At data er åpne betyr at de deles uten begrensinger på hvem som kan bruke dataene (tilgangsbegrensninger) eller hva dataene skal brukes til (formålsbegrensninger). Sett i lys av at offentlige data er innhentet i virksomheter eller prosjekter som er finansiert av skattebetalernes penger, er det ønskelig at så mye data som mulig gjøres åpne. Det er imidlertid visse restriksjoner som kan komplisere dette bildet, for offentlige data generelt, og for forskningsdata spesielt. Når det gjelder forskningsdata er det nyttig å skille mellom data som er underlagt lovpålagte delingsrestriksjoner og data som av andre grunner ikke deles. Lovbestemmelser som begrenser hvilken tilgang som kan gis til ulike typer data, og hvordan de kan (gjen)brukes, er spredt i lovverket, fordi de har til formål å verne ulike interesser. Det gjelder lover fra så ulike områder som helseforskningsloven, sikkerhetsloven og åndsverksloven. Det er krevende for forskeren eller institusjonen som skal sørge for at delingen av dataene skjer i henhold til loven, å finne frem til og sette seg inn i begrensningene. Det kan skape usikkerhet og gi grunnlag for ulik praksis for deling og lisensiering av datasett. Vi kommer nærmere inn på dette i [kapittel 4](#).

Når det gjelder data som ikke er underlagt lovpålagte restriksjoner, men som likevel ikke deles, er det nødvendig å undersøke hvilke underliggende faktorer som påvirker denne praksisen. I disse tilfellene er det forskeren/institusjonen selv som må gjøre vurderinger om i hvilken grad dataene skal gjøres åpne eller ikke. Er det manglende infrastrukturer for deling, er det forskningsetiske begrunnelser eller er det mangel på incentiver som gjør at dataene ikke deles? Dette blir også nærmere diskutert i [kapittel 4](#) i dette notatet.

Åpne data versus FAIR data

Begrepene "åpne data" og "FAIR data" er begge begreper som brukes for å beskrive tilgang og vilkår for gjenbruk av data. Det er imidlertid store forskjeller i hva begrepene innebærer rent praktisk. Åpne data sier hovedsakelig at data kan brukes av enhver uten restriksjoner (beskrevet i forrige avsnitt), mens FAIR data legger vekt på teknisk tilgjengelighet (maskinlesbarhet). FAIR data skal være gjenfinnbare, tilgjengelige, kompatible og gjenbrukbare, uten at dette nødvendigvis betyr fri tilgang for alle. Kostnadene knyttet til å gjøre forskningsdata FAIR overgår ofte kostnadene for å gjøre forskningsdata åpne. For å bli fullt ut FAIR må data gjennomgå en høy grad av standardisering, som ofte krever vesentlig investering av tid og ressurser, blant annet som følge av at forskningsdata ofte er heterogene og lite standardiserte. Det finnes en rekke datasett som genereres innenfor fagområder med internasjonalt etablerte datastandarder, avanserte datainfrastrukturer og administrative støttesystemer, men det

finnes også små forskningsmiljøer som mangler denne infrastrukturen. Det kan være enkeltforskere som studerer en ny trend blant ungdommer, språkvansker hos en utvalgt gruppe informanter eller hittil ukjente miljøgifter som analyseres med nye metoder. For at disse forskerne skal gjøre dataene sine FAIR, vil det kreve mye ekstra ressurser, som i ytterste konsekvens kan gå på bekostning av verdifull forskningstid. Forskningsdata av denne typen betegnes ofte som "den lange halen av forskningsdata", og gir en viktig påminnelse om at forskningsdata er en svært heterogen og vag kategori av data. Strukturer som kan sikre effektive omforminger til FAIR data, og som samtidig kan skaleres til miljøer som ikke har slike strukturer tilgjengelig, vil derfor bli verdifulle framover.

Forskningsdataenes livssyklus og egenskaper

Data vil gjerne gå gjennom et dynamisk livsløp i et forskningsprosjekt, fra innsamling av rådata til ulike former for strukturering og analysering. (Figur 1, mørkeblå). I motsetning til offentlige data, som gjerne innhentes automatisk eller systematisk som en del av offentlig forvaltning, innhentes som regel forskningsdata i forbindelse med at man ønsker å teste hypoteser eller frembringe ny kunnskap om spesifikke tema. Forskningsdata er ofte knyttet sammen med rike metadata som er nødvendig for at dataene skal gi mening og kunne gjenbrukes av andre. Dette gjelder særlig for forskningsdata som ikke er en del av store datainnsamlinger hvor infrastruktur og forente standarder og formater er på plass. Forskningsrelevante digitale objekter, som har persistente ID-er, assosierte metadata og inkluderte relasjoner til for eksempel script og metodeprotokoller (Figur 1, blå), vil i mange tilfeller også eksistere utover prosjektperioden, sammen med dataene (Figur 1, oransje). Dette stiller ytterligere ressurskrav til vedlikehold og oppdatering av format og standarder i tråd med teknologisk utvikling, slik at det er mulig å gjenbruke og videreutvikle dataene også i fremtiden.



Figur 1 Livsløpet i et forskningsprosjekt. Skjematisk fremstilling av et typisk dynamisk livsløp i et forskerprosjekt hvor egenskaper ved dataene endrer seg over tid. Illustrasjon metadata: CC-BY Zehl et al. (2016).

<https://doi.org/10.3389/fninf.2016.00026>. Illustrasjon database: Joe Techapanupreeda (Shutterstock). Illustrasjon dataprodukter og tjenester: NicoElNino (Shutterstock).

Vitenskapelige publikasjoner, slik som artikler, bøker, monografier etc., inngår ofte som en del av livssyklusen i et forskningsprosjekt (**Figur 1**, lyseblå). Men det er ikke alltid forskningsdata resulterer i publikasjoner, enten fordi de blir sett på som negative eller sekundære resultater sett i lys av hypotesen eller forskningsspørsmålet. Slike data kan også være nyttige for forskere og tjeneste- og produktutviklere som ser et uutnyttet potensial for kunnskapsutvikling, gjenbruk eller videreutvikling.

Spørsmål om *når* man skal dele data i et forskningsprosjekt blir juridisk viktig fordi forskningsdata endrer status som rettsobjekt¹ etter hvor i livsløpet til prosjektet dataene befinner seg. Utvalget kommer nærmere inn på dette i neste kapittel.

2. Forskningsdata i et uoversiktlig juridisk landskap

Det er tydelige samfunnspolitiske forventninger om at forskningsdata skal deles i størst mulig grad, men det finnes ikke noe helhetlig lovverk som regulerer rettigheter til, og deling og gjenbruk av, slike data. For enkeltforskere, forskningsmiljøer og institusjoner er det vanskelig å forholde seg til et fragmentert og uoversiktlig juridisk landskap. Det er derfor viktig å definere hvem som er rettssubjekt², slik som forsker eller institusjon, og hva som er rettsobjekt i juridisk sammenheng, og å avklare hvilke rettigheter som knytter seg til de ulike rettsobjektene.

Tilgjengeliggjøring av resultater fra forskning er mer regulert enn andre offentlige data fordi de samles inn i en vitenskapelig kontekst, hvor både intellektuelle og finansielle investeringer kan være betydelige. I tillegg til generelle lovreguleringer som personopplysningsloven og sikkerhetsloven, reguleres forskningsdata av blant annet forskningsetikkloven, universitets- og høyskoleloven og i noen tilfeller åndsverkloven.

Innenfor academia kan det være utfordrende å vurdere om datasett omfattes av rettigheter, og hvem som eventuelt innehar rettighetene. Dette må sees i sammenheng med at datainnhenting ofte er en del av et større akademisk arbeid, som gjerne resulterer i vitenskapelige publikasjoner og dataprodukter, som programvare, som kan være vernet av opphavsrett etter åndsverkloven. Data er informasjon eller kunnskap, og det finnes ikke generelle lovregler for eierskap til data. Data kan imidlertid, avhengig av omstendighetene, være beskyttet av eksklusive rettigheter etter åndsverkloven. Åndsverksloven gir vern for strukturerte databaser, men ikke for ustrukturerte rådata.

Fra et rettslig perspektiv kan ingen "eie" enkeltdata i betydningen opplysninger, men samlinger av data eller datasett kan være underlagt eksklusive rettigheter. Åndsverkloven gir en eksklusiv rett til utnytting av en database som er et resultat av en "vesentlig investering" ([åndsverkloven](#) § 24). Dette må vurderes konkret. Det er innsamlingen eller struktureringen av databasen som må ha krevet en vesentlig investering, og det kan være vanskelig å skille mellom innsamling av data og produksjon av data.

Mange forskningsdatabaser vil fylle vilkårene for vern. Når en database fremstilles av ansatte i utføringen av arbeid for en arbeidsgiver, er nok utgangspunktet at databaseretten oppstår hos arbeidsgiver. Spørsmålet om fordelingen av rettigheter til digitale databaser mellom forskere eller akademisk ansatte og institusjonen, har ikke vært avgjort i praksis. Men uten andre holdepunkter, er det

¹ Eksempler på rettsobjekt kan være informasjon, dokumenter, databaser og publikasjoner.

² En som kan få rettigheter og plikter i tråd med gjeldende lovverk. Kilde: [Wikipedia](#)

naturlig å legge til grunn at det samme også gjelder for forskere og akademisk ansatte, slik at rettighetene til databasen etter åvl. § 24, som regel vil tilkomme institusjonen. Institusjonen kan i så fall bestemme at dataene skal deles, og hvordan. I særskilte tilfeller kan også en database utgjøre et åndsverk fordi utvalget og innsamlingen av dataene utgjør en kreativ og original innsats gjort av for eksempel en forsker. Opphavsretten til databasen vil da oppstå og beholdes hos forskeren, men denne regulerer imidlertid ikke dataene som en database inneholder, eller gjenbruk av noen av dataene. Mangel på rettslig praksis og avklaringer av opphavsrett og databasevern knyttet til forskningsdata gjør det utfordrende når disse skal deles. Det er også uavklarte spørsmål når det gjelder rekkevidden av forskernes rett til å bestemme om, når og hvordan forskningsresultater skal publiseres (den akademiske friheten, [UHL § 1-5](#)), sett opp mot institusjonens databaserettigheter. Det er behov for videre utredning omkring disse spørsmålene.

Eksempel fra forskning med rettighetsbelagte kildedata

Innenfor forskningen finner vi mange eksempler på at forskningsdata ikke er et entydig begrep, men et samlebegrep som omfatter data som samles inn i ulike fagmiljøer med særegne fagtradisjoner. I noen fagdisipliner er for eksempel skillet mellom inndata i et forskningsprosjekt og åndsverk uklart. Kunstnerisk utviklingsarbeid kan eksempelvis være en tett sammenvevd kombinasjon av kunstproduksjon og faglig refleksjon. I noen tilfeller kan dataene som samles inn være vernet av opphavsrett. Innenfor fagområder som studerer kunstneriske uttrykk (språk, litteratur, musikk, dans, teater etc.) er forskningsdata ofte basert på rettighetsbelagt materiale. Språklige korpus er for eksempel vitenskapelige annotasjoner av tekst eller tale, der hvert inngående bidrag kan være selvstendige åndsverk, vernet av opphavsrett. En musikkvitenskapelig database inneholder gjerne noter i tillegg til både lyd- og videoopptak av ulike fremføringer. Hver enkelt fremføring kan ha flere ulike opphavere (komponist, tekstforfatter, utøver, produsent, koreograf, m.m.), og det kan være ulike organisasjoner som håndterer opphavernes ideelle og økonomiske rettigheter. I så fall kan videre deling, gjenbruk og videreutvikling av forskningsdataene bare skje så langt det ikke krenker de opprinnelige rettighetene. Overlappende rettigheter gjør rettighetsklareringen mer komplisert før eventuell publisering av materialet. Ivaretagelse av fagmiljøers egenart og identifisering av fagspesifikke utfordringer er viktig i arbeid med å avklare rettighetsspørsmål knyttet til forskningsdata.

Data generert i offentlig sektor er som hovedregel underlagt lovkrav om offentliggjøring basert på et innsynsregime ([Offentleglova](#)). Norske forskningsinstitusjoner er i all hovedsak offentlige. Offentleglova er ofte forbundet med dokumenter knyttet til saksbehandling i offentlig forvaltning. Det kan argumenteres at lovens hovedregel om innsyn i offentlig informasjon også gjelder for forskningsdata, men her er det ulike tolkninger av lovverket. Forholdet mellom forskningsdata og offentliglova bør tydeliggjøres i lys av visjonen om økt deling av forskningsdata i samfunnet.

Universitets- og høyskoleloven berører håndtering av forskningsdata som del av den større tematikken åpen forskning og kunnskapsdeling. UH-loven ble nylig gjennomgått av Aune-utvalget ([NOU 2020:3](#)).

Lovutvalget foreslår blant annet en sterkere betoning av institusjonenes ansvar for kunnskapsdeling og åpen forskning. Utvalget påpeker at det er en viktig forskningsetisk norm å gjøre forskningsresultater tilgjengelig, og viser til Lov om organisering av forskningsetisk arbeid ([forskningsetikkloven](#)) og institusjonenes og den enkeltes ansvar for å sikre at all forskning skjer i henhold til anerkjente forskningsetiske normer. Selv om Aune-utvalget i hovedsak diskuterer åpen publisering, er det nærliggende å forstå dette som et ansvar som også omfatter deling av forskningsdata. Som følge av endringer i publiseringsformater, ved at forlagene krever at data gjøres tilgjengelig som grunnlag for kvalitetskontroll, og at tekst i økende utstrekning er interaktiv med underliggende data, er det ikke mulig å skille klart mellom åpen publisering og åpne forskningsdata.

Flere av de aktuelle lovene er nylig endret eller under bearbeidelse, men det finnes i liten grad lovverk eller forslag til nye lover som tydelig gjenspeiler dagens situasjon, hvor håndtering og deling av til dels store mengder forskningsgenererte data er en stadig viktigere del av forskningsaktiviteten innenfor alle fagfelt.

3. Om lisenser for deling og gjenbruk av forskningsdata

Immaterialrettigheter ivaretar og beskytter for eksempel oppfinnelser, forretningshemmeligheter samt kunstneriske og litterære åndsverk. I dette kapitlet vil utvalget fokusere på lisenser som hovedsakelig befatter seg med opphavsretten og databasevernet. Dette innebærer både lisenser som ivaretar disse rettighetene, men også erklæringer som går ut på at rettighetshaver frasier seg alle rettigheter, slik at verket fullt ut "faller i det fri" – det offentlige rom. utfordringer knyttet til kommersialisering av forskningsdata og immaterialrettigheter vil bli skissert i [kapittel 4](#).

I tilfeller hvor forskningsdata er vernet av i lov, må rettighetshaver samtykke til at andre skal få tilgang til å for eksempel tilgjengeliggjøre, gjenbruke, bearbeide eller kommersialisere dataene. Det kan gjøres ved en lisens. Et udefinert lovverk når det gjelder rettigheter til forskningsdata medfører usikkerhet i forskningsmiljøene og ulik praksis rundt lisensiering ved de forskjellige forskningsinstitusjonene. Mens opphavsrett til forskningsartikler oppstår hos forskeren og ikke overføres til arbeidsgiver, er det institusjonene som i utgangspunktet har rettighetene til forskningsdata med de forbehold som er redegjort for over. Likevel er det i praksis ofte enkeltforskere, og ikke institusjonene, som tar avgjørelser om deling av dataene, og vilkårene for tilgang og bruk av disse.

Lisenser vil kunne være en form for regulering av vilkårene for tilgang og gjenbruk av forskningsdata, som gir et avtalemessig grunnlag mellom den som deler og den som gjenbraker dataene. Det finnes lisenser som er godt gjennomarbeidet og mye brukt på internasjonal basis. De fleste av disse standardlisensene, slik som Creative Commons (CC)-lisenser, er utarbeidet med sikte på å gi tilgang til, og regulere bruk av, opphavsrettslig vernet materiale, og er ikke nødvendigvis innrettet mot forskningsdata. I den grad forskningsdata er samlet i en database som er vernet etter åndsverkloven, vil for eksempel CC-lisensene være et nærliggende alternativ, men det bør vurderes nærmere hvordan lisensbetingelsene er tilpasset delingsrestriksjoner for data som følger av annen lovgivning, og sanksjonsmulighetene ved brudd på lisensvilkårene.

Det er behov for en undersøkelse av og bevisstgjøring om den rettslige reguleringen av data og datasett. Det bør også vurderes om noen av lisensene er utformet slik at de ivaretar nødvendige behov hos

forsker, akademisk institusjon, næringsliv og samfunnet for øvrig, uavhengig av om dataene er vernet eller ikke.

CC BY 4.0. er en lisensvariant som ivaretar forskerens behov for anerkjennelse for sin innsats gjennom kravet om at opphavspersonen skal siteres. Denne varianten ivaretar også samfunnets behov for å kunne gjenbruke forskningsresultater, ettersom lisensen tillater både bearbeidelse og kommersiell utnyttelse. CC BY 4.0 er imidlertid ikke uproblematisk, ettersom kravet om attribusjon (å tilskrive opphav ved gjenbruk) kan skape praktiske problemer med gjenbruk av store mengder data fra mange forskjellige opphavspersoner, ofte kalt "*attribution stacking*". Derfor vil en CCO-dedikasjon (public domain-dedikasjon) kunne være hensiktsmessig for noen typer data. Ved å bruke en fristatus-erklæring, slik som CCO, frasier lisensgiver seg alle rettigheter til dataene, slik at data blir gjort tilgjengelig i det offentlige rom, uten restriksjoner. CCO er i utgangspunktet ikke utviklet for materiale der det ikke foreligger opphavsrett eller databasevern til data, hvor et Public Domain merke vil være formelt riktig, og det bør etterstrebtes at det ikke legges restriksjoner på data som ikke er vernet i gjeldende lovverk. Sitering og navngivning kan oppfordres til uten å bruke lisens, og forskere har plikt til å opptre med aktsomhet og sikre at anerkjente forskningsetiske normer overholdes ved å blant annet henvise til kilder, i henhold til forskningsetikkloven og [generelle forskningsetiske retningslinjer](#). Forholdet mellom forskningsetiske plikter og andre regler som regulerer forskningsdata bør dermed avklares og utredes videre.

Akademisk frihet, institusjonenes styringsrett og forskningsetikk

Akademisk frihet og institusjonell autonomi er akademiske kjerneverdier som står sterkt i Norge. Universitets – og høyskoleloven legger vekt på at institusjonene skal fremme og verne akademisk frihet og at de har rett til å utforme sitt faglige og verdimessige grunnlag innenfor gjeldende lovverk. Samtidig skal de ..."*sikre at undervisning, forskning og faglig og kunstnerisk utviklingsarbeid holder et høyt faglig nivå, og utøves i overensstemmelse med anerkjente vitenskapelige, kunstfaglige, pedagogiske og etiske prinsipper*". Forskningsetikkloven stiller krav til både enkeltforsker og institusjonene om å sikre at all forskning skjer i henhold til anerkjente forskningsetiske normer. De generelle forskningsetiske retningslinjer, som ble utarbeidet av De nasjonale forskningsetiske komiteer i 2014, vektlegger blant annet som hovedregel at forskningsresultater skal tilgjengeliggjøres, som et grunnleggende prinsipp. Dette skal sikre etterprøvbarehet og transparens rundt forskningsdata, og samtidig legge til rette for at samfunnet skal ha beste kunnskap tilgjengelig. Tilgjengeliggjøring av forskningsresultater har også en demokratisk funksjon ved å sikre dialog med offentligheten. Samtidig har forskningsinstitusjonene en rolle i å sikre forskerens frihet ved valg av tema, metode, gjennomføring av forskning og publisering av resultater. Frihet til å velge kanal for publisering har for eksempel vært et sentralt tema rundt debatten om [åpen publisering](#).

Politikk for åpen forskning og forskningsdata bygger på et premiss at forskningsresultater er fellesgoder, siden de bygger på felleskapets investeringer i forskning. Selv om det er stor tilslutning til prinsippet om åpen tilgang til forskningsresultater på et generelt nivå, er det flere eksempler på at forskere ikke fullt ut deler idealet om resultater som ikke-eksklusive fellesgoder. Diskusjonene som fulgte lanseringen av Plan S for åpen tilgang til publikasjoner, viste at mange er enige i målet om åpen tilgang, samtidig som kravet om åpne lisenser til publikasjoner ble møtt med motstand. Flere gav uttrykk for skepsis til at åpne

publikasjoner skal kunne utnyttes videre i kommersielle sammenhenger, og ønsket muligheten til å kunne bruke lisenser som begrenser dette. Når det gjelder data har vi ennå ikke sett en tilsvarende bred diskusjon om fellesgodeproblematikk knyttet til datadeling. Likevel er det eksempler på at forskere benytter restriktive lisenser som blokkerer gjenbruk i kommersielle sammenhenger når de deler datasett, simpelthen fordi en ikke ønsker at resultatene skal utnyttes kommersielt. En slik praksis er lite forenlig med idealet om at resultater skal kunne utnyttes som fellesgoder, også der hvor det er muligheter for kommersiell verdiskapning. En slik praksis med bruk av ikke-kommersielle lisenser kan dessuten legge langt større begrensninger på mulighetene til gjenbruk enn det som lisensgiver har intendert, fordi det er uklart hvordan en kommersiell sammenheng skal defineres. Det er viktig å skille mellom praksis som legger unødvendige begrensninger for gjenbruk på generell basis, og den mer legitime beskyttelsen av forskningsresultater innenfor institusjonenes rammeverk for kommersiell utnyttelse av forskningsresultater.

4. utfordringer knyttet til deling av forskningsdata spesielt

I dette avsnittet pekes det på hvilke politiske, strukturelle og kulturelle faktorer som kan være mulige barrierer for deling av forskningsdata. Under hvert punkt vil utvalget presentere forslag til hvordan man kan møte disse utfordringene.

Åpne forskningsdata i et resultatstyrt og merittbasert system

Publisering av forskningsresultater er en viktig del av den akademiske virksomheten, og en faktor forskere blir målt etter både i sitt ansettelsesforhold og i konkurranse med andre forskere om forskningsmidler. I mange tilfeller blir derfor den vitenskapelige publiseringen en helt sentral faktor i oppbyggingen av forskerkarrierer innenfor akademien, blant annet for å kunne disputere til doktorgrad eller få fast ansettelse. I tillegg er tellekantsystemet, som er en del av grunnlaget for tildeling av midler til forskningsinstitusjoner, basert på vitenskapelig produksjon, noe som forsterker fokuset på publikasjoner ytterligere. Verdifulle datasett ligger ofte til grunn for publikasjoner av høy kvalitet som får innpass i anerkjente tidsskrift. Når publikasjoner er en såpass viktig del av institusjonenes tildelinger og forskernes karriere, vil slike datasett bli enda mer verdifulle. Derfor kan deling av forskningsdata bidra til verdimeslige konflikter mellom forsker, institusjon og samfunnets investering i og krav til offentlig finansierte data som et fellesgode.

Det blir ofte påpekt at det mangler insentiver for å dele datasett, siden slik deling ikke gir uttelling i dagens tellekantsystem eller sjelden blir tillagt vekt i vurdering av søknader om finansiering av forskning. Dette kan ha en kontraproduktiv effekt i forbindelse med deling av data, fordi det kan føre til at forskere ønsker å beholde en eksklusiv tilgang til datasettene, for å kunne maksimere den personlige og institusjonelle uttellingen som vitenskapelige artikler gir. Av og til stilles det krav til at dataene som forskningen bygger på, publiseres sammen med de vitenskapelige artiklene, men det betyr nødvendigvis ikke at hele datasett blir åpent tilgjengelig for videre bruk av andre.

Utvalget mener at dagens tellekantsystem delvis undergraver målet om mer deling og gjenbruk av forskningsdata. Systemet baserer seg utelukkende på vitenskapelige artikler, og anerkjenner ikke verdien av (eller bredden av) vitenskapelig produksjon og materiale. Systemet bør derfor vurderes på nytt, slik at eksklusiv tilgang til datasett ikke gir noen fordel i form av grunnlag for publikasjoner som gir poeng og

økonomisk uttelling. Én løsning kan være at publiserte datasett gir uttelling i systemet på linje med vitenskapelige artikler. En annen løsning kan være å nedjustere, eller eventuelt fjerne, det økonomiske insentivet for høyt volum av publikasjoner, slik at motivasjonen for å ikke dele dataene blir mindre. En eventuell endring av systemet bør sikre at slike justeringer ikke forfordeler enkelte fagområder.

Utvalget mener videre at prinsippene i [DORA-erklæringen](#), som er signert av Norges forskningsråd og mange norske forskningsinstitusjoner, bør gjenspeiles bedre i vurderinger knyttet til tildeling av midler til forskning og ved ansettelse og opprykk. DORA-erklæringen sier eksplisitt at også datasett og programvare skal være gjenstand for vurdering i slike tilfeller. Et kriterium for vurdering av datasettenes kvalitet, samt virkninger og effekter, kan i slike tilfeller være datasettenes vitenskapelige produktivitet.

I tillegg til en endring i insentivstrukturen tror utvalget at det vil være nødvendig med et fokusskifte. Det er behov for å dreie oppmerksomheten bort fra ulempene forskningsmiljøene kan oppleve ved å ikke få mest mulig merverdi ut av egne data, og heller fokusere på de store fordelene de vil få gjennom tilgang til mange flere datakilder i et åpent forskningsnettverk.

Hva er formålet – deling eller gjenbruk?

Mer deling av offentlig finansierte forskningsdata er en ambisjon, men deling i seg selv vil ikke nødvendigvis føre til mer gjenbruk. Data man bare kan få innsyn i vil være nyttig med tanke på dokumentasjon og forskningens integritet, men vil ha begrenset verdi for samfunnet hvis dataene ikke kan brukes å utforske nye ideer, skape ny innsikt eller utvikle nye produkter. For at data faktisk skal kunne gjenbrukes, selv om eventuelle rettslige hindre for gjenbruk er håndtert, kreves det ekstra ressurser. Dette må sees på en investering som vil gi avkastning på sikt, og spesielt vil kvalitetsdata fra forskning kunne være en god investering i fremtiden. Data som er interoperable med andre data og som kan gjenbrukes av andre vil være en viktig forutsetning i en vellykket datadrevet økonomi. På grunn av store kostnader vil det imidlertid være et behov for en prioritering av hvilke data som skal gjøres FAIR, og det kan være nødvendig å sette kriterier for hvilke data som betraktes som "høyverdige" i en første fase. I [rapporten](#) fra ekspertgruppen for datadeling i næringslivet påpekes det at det er urealistisk og uhensiktsmessig å kreve at alle mulige datasett skal gjøres tilgjengelig på et nivå for gjenbruk for alle potensielle bruksområder. I rapporten foreslås det derfor at det etableres mekanismer som bringer aktører sammen rundt bruksområder der tilgjengeliggjøring og utnyttning av data vil være av stor samfunnsinteresse.

Utvalget mener at det må settes av ressurser for å sette forskningssektoren i stand til å gjøre data tilgjengelige for gjenbruk etter FAIR-prinsippene. Det foreslås at det i første omgang foretas en prioritering av forskningsdata med spesielt stor samfunnsverdi etter modell fra Horisont Europa, slik at riktig mengde ressurser kan allokeres til riktig sted for at målsetningen om mer gjenbruk skal bli realisert. På sikt vil det være nødvendig med en plan for hvordan "den lange halen av forskningsdata" skal bli et verdifullt bidrag til et datadrevet samfunn og en datadreven økonomi.

Infrastrukturer for deling og gjenbruk av forskningsdata og bærekraftige finansieringsmodeller

De siste årene har investeringene i generisk e-infrastruktur, som datanettverk med høy kapasitet, regneressurser for store beregninger og lagringsressurser, økt betraktelig. Slik e-infrastruktur er avgjørende for en rekke avanserte og ressurskrevende beregninger, simuleringer og analyser. I tillegg er

det gjort store investeringer i områdespesifikke datainfrastrukturer. Disse har som formål å samle inn data, gi dem merverdi og gjøre dem FAIR ved å bearbeide, lagre, kuratere og tilrettelegge for metaanalyser og gjenbruk. Sentrale miljøer ved forskningsinfrastrukturene bidrar også med kunnskap om lisensiering og med systemer for autorisert tilgang til data og aksept av lisenser. Slike infrastrukturer er imidlertid ikke på plass innenfor alle områder, og etablerte infrastrukturer blir ikke alltid utnyttet så godt som ønskelig. Det er derfor behov for et kontinuerlig investeringstrykk for å etablere, videreutvikle og promotere forskningsinfrastrukturer som sikrer datadeling, slik at forskere skal få gode verktøy og brukerstøtte som er nødvendig for deling og gjenbruk. Nasjonal og internasjonal infrastruktur som er godt koordinert og forankret i fagmiljøer og institusjoner vil bedre tilgang til, og muliggjøre passende lisensiering for sikker deling og gjenbruk av, forskningsdata.

Kunnskapsbaserte datainfrastrukturer må videreutvikles kontinuerlig for å møte samfunnets og forskningsmiljøenes behov, og det er dermed behov for regelmessige investeringer. I tillegg er det store kostnader knyttet til driften av slik infrastruktur. Langsiktige planer for finansiering av datainfrastrukturer basert på ulike inntektskilder kan bidra til at slike infrastrukturer blir mer robuste. Pr. i dag er det ulike oppfatninger av hvordan slik drift skal finansieres, og innenfor enkelte fagområder er det lang tradisjon med å få tilgang til data uten kostnad. Enkelte infrastrukturer har utviklet egne finansieringsmodeller med blant annet brukerbetaling, mens andre har store utfordringer med drift fordi viljen til brukerbetaling er liten.

Utvalget mener at behovet for videre investering i infrastruktur for deling, tilrettelegging og forvaltning av FAIR data (både generisk e-infrastruktur og områdespesifikke datainfrastrukturer) er viktig i tiden framover for å nå målet om mer deling og gjenbruk av forskningsdata. Forskningsinstitusjoner og relevante forvaltningsorganer og næringslivsaktører bør i større grad gå sammen og definere hvilke infrastrukturbehov som må dekkes innenfor de ulike fag- og forvaltningsområdene. Behovene bør beskrives både på institusjonelt nivå, på nasjonalt nivå og på internasjonalt nivå. Dette bør blant annet resultere i koordinerte søknader til for eksempel Forskningsrådets INFRASTRUKTUR-utlysning og utlysninger i forskningsinfrastrukturprogrammet i Horisont Europa. Slike søknader bør tydelig vise hvordan nasjonale og internasjonale infrastrukturer kan bidra til mer deling og gjenbruk innenfor ulike fagområder.

I tillegg må spørsmålet om bærekraftige finansieringsmodeller for drift av datainfrastrukturer settes enda tydeligere på dagsorden. For at infrastrukturene skal bidra til mer deling og gjenbruk av forskningsdata på en god måte, må det etableres gode finansieringsmekanismer som kan redusere infrastrukturenes utfordringer med å dekke kostnader knyttet til utvikling og drift.

Komplekse forskningsdata, ressursbehov og krav til kompetanse

Å gjøre data FAIR krever ressurser, ikke bare i form av investeringer i gode infrastrukturer og systemer for deling og gjenbruk, men også i form av menneskelige ressurser. Som det kommer fram av dette notatet, kan forskningsdata være mer komplekse enn andre typer offentlige data, og det vil i slike tilfeller kreve mer ressurser å tilrettelegge dem for deling og gjenbruk. Innenfor alle fagområder vil det også bli et økende behov for personer med spesialistkompetanse i skjæringspunktet mellom digital teknologi, åpen forskning og vitenskapelig metode (såkalte datarøkttere). Samtidig ser man et økende

behov for spesialistkompetanse på kunstig intelligens, oversettelse av kildekode (kilde-til-kilde kompilering) og maskinlæring og dyp læring ved bruk av tungregningsprosesser, når generiske e-infrastrukturer skal yte bistand til forskerprosjekt med slike behov.

Utvalget mener at det må legges stor vekt på kompetanseheving innenfor datahåndtering i tiden som kommer. Dette gjelder både kompetanse hos den enkelte forsker og hos ansatte ved infrastrukturer og i støttefunksjoner ved institusjonene. Åpen forskning bør få større plass i forskerutdanningene, slik at dette blir en integrert del av forskerkompetansen hos norske forskere. Samtidig må UH-sektoren utdanne flere spesialister på området, da behovet for kompetanse kommer til å bli stort i framtiden og vil være en kritisk faktor for å oppnå visjonene med åpen forskning.

Kommersialisering basert på forskningsdata

Det er et offentlig mål å skape nytt næringsliv blant annet gjennom økt innovasjon og kommersialisering av resultater fra offentlig finansiert forskning. Det kan oppstå motsetninger mellom dette målet og de insentivene som er gitt forskningsinstitusjonene for å selv gjennomføre kommersialisering av egne forskningsdata. Noen forskningsinstitusjoner overtar databaserettigheter fra ansatte etter avtale, i den grad institusjonen ikke allerede eier alle rettigheter. Mange institusjoner overfører alle kommersialiseringsrettigheter fra institusjonens forskning til teknologioverføringskontorer (TTOer), som kan være kommersielle tredjeparter. TTOene har mandat til å drive kommersialisering av rettighetene på vegne av institusjonen og forskeren, og vil derfor i mange tilfeller holde forskningsresultater tilbake og motvirke deling av disse. Hvorvidt dette er en effektiv struktur for å sikre utnyttelse av forskningsdata i samfunnet har vært gjenstand for offentlig debatt.

Kommersialiseringsaktivitet basert på databaserettigheter kan forutsette at tilgang til databasen er begrenset eller at dataene holdes hemmelige. Når det fastsettes politikk for deling av forskningsdata, bør det etter utvalgets skjønn ses hen til institusjonenes kommersialiseringspolitikk og målene for den, slik at forskningsdata gjøres tilgjengelige på en måte som best mulig bidrar til innovasjon og verdiskaping for samfunnet som helhet.

Mens innovasjon fra UH-institusjonene pr. nå i stor grad har vært basert på kommersiell utnyttelse av rettigheter til upubliserte forskningsdata og lisensiering av IPR, så er samspillet mellom forskningsinstitusjoner og næringsliv ventet å bli langt mer komplekst i framtiden enn det er i dag. Global datadrevet økonomi er i stor grad preget av nye forretningsmodeller der data fra mange ulike kilder settes sammen og i mange tilfeller skaper økonomiske verdier gjennom bruk som ikke var tiltenkt ved datainnhenting. Bruk av stordata i nye forretningsmodeller blir til i et komplekst samspill mellom datatilbydere, næringsliv og andre parter. Potensialet for samfunnsutnyttelse av slike data er stort i næringslivet, og det bør derfor diskuteres om forskningsinstitusjonenes kommersialiseringsstrategier kan virke begrensende for at data med kommersielt potensial deles med aktører utenfor institusjonen.

Utvalget mener at offentlige strategier for kommersialisering av forskningsresultater og strukturer rundt innovasjon i offentlige forskningsinstitusjoner bør revurderes for å stimulere til økt deling og bruk av forskningsresultater i samfunnet. Spørsmålet om hva som skal være den primære strategien for kommersialisering bør særlig diskuteres. Skal insentivet være direkte økonomisk inntjening for institusjonene/TTOene slik det er nå, eller skal målet for offentlige forskningsinstitusjoner heller være å stimulere til økt innovasjon direkte i næringslivet og samfunnet som helhet ved å i større grad dele data

til fri utnyttelse? Utvalget mener videre at det ikke bør legges generelle begrensninger mot å bruke data fra offentlige forskningsinstitusjoner til kommersielle formål.

Om lovverk og regulering for deling og gjenbruk av forskningsdata

Som det er påpekt i dette notatet er det juridiske landskapet knyttet til rettigheter til og deling og gjenbruk av forskningsdata uoversiktlig og komplekst. Lovverket er fragmentarisk, og mange ulike lover medfører begrensninger på deling og gjenbruk av data ut fra svært ulike interesser og hensyn. Lite av lovgivningen er tilpasset stordatasamfunnet, og tar derfor ikke høyde for at både forskning og samfunnsøkonomi i stadig større grad er datadrevet. Det er en fare for at usikkerhet og uklarhet med hensyn til rettslig regulering fører til mindre deling av data enn nødvendig.

Utvalget mener at det er behov for en systematisk gjennomgang av aktuelle lovverk og praksis for institusjonelle retningslinjer for å få klarhet i hvilke rettigheter som er knyttet til forskningsdata, og hvordan blant annet lisenser bør brukes for å regulere deling og gjenbruk av forskningsdata på en hensiktsmessig måte. En slik gjennomgang vil også kunne avdekke konkrete områder der det eventuelt eksisterer en reell usikkerhet om hvilke lover og retningslinjer som gjelder, både nasjonalt og internasjonalt. Resultatet av en slik gjennomgang vil kunne gi et utgangspunkt for å etablere nasjonale retningslinjer for bruk av lisenser på forskningsdata, eventuelt forslag til lovverk som bør gjennomgås på nytt.

Oppsummerende sluttkommentar fra utvalget

Det er utvalgets oppfatning at offentlig finansierte forskningsdata som hovedregel skal deles så åpent som mulig. Dette er informasjon som er innhentet med offentlige midler og bør følgelig komme allmenheten til gode. Det innledende arbeidet i utvalget har imidlertid gjort det tydelig at lovverket som regulerer deling av forskningsresultater er uoversiktlig og at det er behov for en gjennomgang. Dette er avgjørende for at lisensiering av data skal bli hensiktsmessig for å oppnå både høyere kvalitet på forskning, bedre utnyttelse av forskningsdata i andre sektorer og tilgjengeliggjøring av forskningsresultater for allmennheten.

Lisenser kan være et nyttig verktøy for å gi riktig tilgangs- og formålsbegrensninger til data der det er nødvendig. Bruken av lisenser slik det er i dag, er imidlertid for tilfeldig og i enkelte tilfeller ikke godt nok gjennomtenkt. På sikt bør det være en målsetning å lage nasjonale retningslinjer for bruk av lisenser, men dette kan først skje når det juridiske grunnlaget er gjennomgått. I en utarbeidelse av nasjonale retningslinjer er det viktig med bred involvering fra ulike fagområder, næringsliv og andre offentlige aktører. Før eventuelle retningslinjer for bruk av lisenser er på plass er det utvalgets generelle anbefaling at forskere bruker lisenser med så få tilgangsbegrensninger og formålsbegrensninger som mulig.

Utvalget mener videre at en opprydning i det juridiske landskapet og en tydeliggjøring av bruk av lisenser ikke vil svare på mange av de underliggende utfordringene knyttet til deling av forskningsdata. Det er i hovedsak politiske, økonomiske, strukturelle og kulturelle hindringer som står i veien for mer deling av offentlig finansierte forskningsdata. Utvalget ser fram til en åpen prosess hvor alle interessenter får mulighet til å få presentert sine synspunkter i denne sammenheng. Forhåpentligvis vil utvalgets arbeid sette i gang prosesser som vil resultere i en forskningssektor som bidrar til både bedre forskning, næringsutvikling og samfunnsutvikling gjennom å gjøre sine forskningsdata mer tilgjengelig for gjenbruk.