

## Arbeidstittel: Rettighets- og lisensspørsmål ved tilgjengeliggjøring av forskningsdata

Utvalget gjør oppmerksom på at rapporten du nå skal lese, er et utkast. Det er flere av temaene som tas opp som ikke er fullt ut diskutert og gjennomarbeidet. Dette gjelder spesielt anbefalingene som presenteres underveis i dokumentet. Fullstendige kildehenvisninger, språkvask og design er heller ikke helt på plass. Utvalget har likevel et ønske, i åpenhetens ånd, om å dele rapporten på nåværende tidspunkt slik at vi rekker å få innspill fra dere som har førstehåndskunnskap om temaene som diskuteres. Vi ser fram til å få høre hva dere har å si, slik at vi kan innarbeide dette i en endelig rapport som vi tror vil bli et godt og viktig verktøy for å oppnå mer deling og gjenbruk av forskningsdata i framtiden.

## Forord

\*Kommer i endelig rapport

## Innhold

Forord.....	1
Ordliste.....	5
Sammendrag.....	5
INNLEDNING.....	5
Bakgrunn .....	5
Utvalgets mandat.....	7
Sammensetning av utvalget.....	8
Andre bidragsyttere .....	8
Prosess .....	9
Om rapportens avgrensning og oppbygging.....	9
DEL 1 – Egenskaper ved forskningsdata .....	11
1. Hva ligger i begrepet forskningsdata? .....	11
1.1 Definisjon av forskningsdata .....	11
1.2 Rådata vs. bearbejdet data og dataprodukter.....	12
1.3 Kildedata vs. data generert underveis i forskningsprosessen .....	12
1.4 Kvalitative vs. kvantitative data .....	12
1.5 Persondata vs. data om verden for øvrig.....	13
1.6 Stordata vs. "den lange halen av forskningsdata" .....	13
2. Åpne data og FAIR data.....	14
2.1 Åpne data.....	14
2.2 FAIR data .....	15
3. Forskningsdataenes livssyklus.....	15
4. Tilgjengeliggjøring og publisering av data.....	16
5. Forskningsdata som er kvalifisert for deling og gjenbruk.....	16
DEL 2 – Juridisk rammeverk og bruk av lisenser og frigivelseserklæringer på forskningsdata.....	17
6. Juridisk rammeverk for deling av forskningsdata .....	17
6.1 Internasjonalt lovverk .....	17
6.2 Nasjonalt lovverk .....	17
6.3 Begrensninger på deling av data.....	23
6.4 Begrensninger som skal beskytte andre interesser .....	25
6.5 Om eierskap til forskningsdata – kan data eies? .....	26
6.6 Utvalgets vurdering av det juridiske rammeverket .....	26

6.7	Utvalgets anbefalinger om juridisk rammeverk for deling av forskningsdata .....	28
7.	Om lisenser og frigivelseserklæringer for deling og gjenbruk av forskningsdata .....	28
7.1	Om lisenser og deres funksjon .....	28
7.2	Om forutsetninger og vurderinger for deling og gjenbruk av forskningsdata .....	29
7.3	Om vilkår og tillatelser i lisenser og Public Domain frigivelseserklæringer .....	31
8.	Konsekvenser av lisensvalg .....	35
8.1	Mer lukket enn nødvendig .....	35
8.2	Navngivelsesstabling .....	37
8.3	Åpen deling = ansvarsfraskrivelse? .....	37
8.4	Hinder for kommersialisering .....	38
9.	Utvalgets oppsummering av bruk av lisenser på forskningsdata .....	38
10.	Premisser for tilgjengeliggjøring av forskningsdata .....	39
10.1	Tydelig ansvarsfordeling mellom den enkelte forsker og institusjon .....	39
10.2	Juridisk kompetent forskningsstøttesystem .....	39
10.3	Avklaring av rettigheter til, og delingsbegrensninger for, forskningsdata .....	39
11.	Utvalgets anbefaling .....	40
DEL 3 – Ikke-juridiske barrierer for deling og gjenbruk av forskningsdata .....		41
12.	Koordinert innsats for mer deling og gjenbruk av forskningsdata .....	41
12.1	Behov for harmonisering av datapolitikken ("policy alignment") .....	41
12.2	Internasjonal orientering .....	42
13.	Insentivmekanismer for mer deling og gjenbruk av data .....	42
14.	Investering i FAIR forskningsdata .....	44
14.1	Investering i infrastrukturer for FAIR data .....	44
14.2	Investering i kompetanse om FAIR data .....	45
14.3	Prioriterte data for FAIR deling .....	45
15.	Verdiskaping og innovasjon basert på forskningsdata .....	46
16.	Utvalgets anbefalinger .....	47

## Ordliste

\*Kommer i endelig rapport

## Sammendrag

\*Kommer i endelig rapport

## INNLEDNING

### Bakgrunn

Åpenhet og kunnskapsdeling står i sentrum for forskning og innovasjon i store deler av verden. Samtidig står samfunnet overfor store utfordringer som må løses blant annet for å sikre global bærekraftig utvikling og kunnskapsbasert næringsliv. 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. Effektive strukturer for åpenhet og deling av forskningsdata er nødvendig for å sikre både datakvalitet, effektiv utnyttelse av ressursene i forskningsmiljøene og økt bruk av forskningsfunn i samfunnet.

Internasjonalt er ønsket om mer åpenhet i forskning – og ikke minst åpen deling av data – ikke nytt, men vi opplever at det vektlegges i større grad enn tidligere. Den pågående koronapandemien har forsterket dette ytterligere. OECD arbeider med medlemsland og ikke-medlemmer for omforent politikk for å fremme åpen forskning og åpen innovasjon og har i 2021 oppdatert sine anbefalinger om åpen deling av data og digitale objekter som framkommer fra offentlig finansiert forskning. Fra [UNESCO](#) varsles det om oppdaterte anbefalinger om åpen forskning. Målet er å definere delte verdier og prinsipper for åpen forskning og identifisere konkrete tiltak som det forventes at de 193 medlemslandene stiller seg bak. Anbefalingene vil bygge på UNESCOs strategi for åpen tilgang til vitenskapelig informasjon og forskning og deres nye anbefalinger om åpne undervisningsressurser.

Europakommisjonen tar også sterkere grep om åpenhet i forskning. I Horisont Europa vil åpen forskning utgjøre regelen framfor unntaket. Søknader vil bli vurdert på åpen vitenskapelig praksis både under kriteriene “excellence” og “implementation” (unntatt ERC og EIT). Det stilles dessuten tydeligere krav til at data tilgjengeliggjøres i tråd med FAIR-prinsippene. Et sentralt grep fra EU for å bedre tilrettelegge for åpen deling av data er etableringen av European Open Science Cloud (EOSC). Denne felles infrastrukturen skal samle institusjonelle, nasjonale og europeiske interessenter, initiativer og datainfrastrukturer for å utvikle et inkluderende økosystem for åpen vitenskap i Europa. I norsk sammenheng er åpen forskning signalisert som en tydelig ambisjon gjennom blant annet Forskningsrådets hovedstrategi og policy for åpen forskning. Åpen forskning antas også å få en sentral plass i Kunnskapsdepartementets kommende langtidsplan for forskning og høyere utdanning.

Norge har store ambisjoner om å ta en sterk og ledende rolle i dataøkonomien for å realisere verdiene som ligger i data og gjennom dette skape flere arbeidsplasser for framtiden. Dette er tydelig signalisert gjennom blant annet stortingsmeldingen [Data som ressurs – Datadrevet økonomi og innovasjon \(2021\)](#) og [Nasjonal strategi for kunstig intelligens \(2020\)](#). For universitets- og høyskolesektoren (UH-sektoren) er det i tillegg lagt inn en rekke høye forventninger til institusjonene og tjenesteapparatet gjennom blant annet [ny digitaliseringsstrategi](#) og gjennom [styringsmeldingen for statlige universiteter og høyskoler](#). For å innfri ambisjonene er det essensielt med et godt rammeverk for effektiv deling av og gjenbruk av data,

i tillegg til god praksis og kultur for slik aktivitet. For at mer effektiv og sømløs tilgang til data skal bli en realitet, må det avklares hvem som kan rå over dataene, om dataene er underlagt eksklusive rettigheter, hvem som kan bruke dem og eventuelt på hvilke vilkår. Riktig og hensiktsmessig bruk av datalicenser og andre avtaler for gjenbruk vil være et viktig virkemiddel i denne sammenhengen.

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 kunne 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 innenfor noen forskningsfelt mange forskningsfunn som ikke kan reproduseres (og eventuelt tilbakevises) på grunn av data med lav kvalitet. Samtidig er det en utfordring at det samles inn og utarbeides mye forskningsdata som enten ikke blir brukt, eller som kunne blitt brukt videre i forskning eller på andre måter i 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.

Deling av data bidrar også til økt verdiskaping ved at både andre forskere og samfunnet for øvrig kan utnytte dataene på nye måter og i kombinasjon med andre datasett. Gode infrastrukturer for datadeling vil blant annet kunne legge til rette for langt mer effektive oppsummeringer og metastudier av det samlede kunnskapsgrunnlaget om ulike felt. Det er også store muligheter for å øke samfunnets utnyttelse av forskningsdata til innovasjon og kommersialisering. For å oppnå dette vil det også være viktig med en harmonisering av politikk på tvers av sektorer og en koordinering av ulike finansieringsordninger og andre virkemidler. Mer deling av forskningsdata vil fremme utvidet tverrfaglig forskning og nyskaping i samfunns- og næringsutvikling. Samtidig vil det 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 datainfrastrukturer, bærekraftige finansieringsmodeller, samt kompetanseheving og insentiver på både individ- og institusjonsnivå, er avgjørende for å oppnå mer deling og gjenbruk av data. Det kan imidlertid være noen ekstra utfordringer knyttet til deling av forskningsdata, både på grunn av usikkerhet rundt de rettslige rammene for deling av data, og manglende delingsinsentiver i forskningssystemet. 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.

Lisenser brukes i økende grad for å sikre kreditering og regulere gjenbruk. Hensiktsmessig bruk av lisenser kan være et viktig virkemiddel for å oppnå mer deling og gjenbruk av forskningsdata. Per i dag er ikke bruken av lisenser konsistent, og det mangler gode retningslinjer for norske forskere. Til en viss grad brukes også lisenser med begrensninger for gjenbruk av dataene, og det er lite kunnskap om

konsekvensene av slike 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 i større grad skal bli et verdifullt bidrag i verdiskapingen i samfunnet.

Lisenser har blitt mye diskutert i forbindelse med krav om åpen publisering av artikler både i Norge og internasjonalt. UNITs rapport fra desember 2019 om [Rettighets- og lisensspørsmål ved åpen publisering](#), skrevet av førsteamanuensis ved UiB Torger Kielland, behandler særlig Creative Commons-lisenser og håndtering av opphavsrettslig vernede publikasjoner ved forskningsinstitusjonene. Kiellands rapport behandler ikke forskningsdata, og innspill fra forskningssektoren pekte på et behov for videre utredningsarbeid om rettighetshåndtering og lisensiering også når det gjelder forskningsdata.

Som en oppfølging av dette satte Forskningsrådet, på vegne av Kunnskapsdepartementet, ned dette utvalget som skulle utrede rettighets- og lisensspørsmål i forbindelse med deling av datasett i forskningen. Denne rapporten er den endelige leveransen **[versjonen du leser nå er et utkast til den endelige leveransen]** fra utvalget. Her gis en systematisk gjennomgang av relevant lovverk som regulerer deling og gjenbruk av data, hvilke typer data som genereres gjennom forskning, hvilke delingsrestriksjoner de ulike datatypene eventuelt er underlagt og hvilke lisenser som kan brukes for ulike datatyper uten at delingsrestriksjonene kompromitteres. Det er viktig for utvalget å understreke at rapporten ikke gir noen endelige svar på de mange utfordringene vi står overfor på dette feltet, men den gir et oversiktsbilde og noen klare anbefalinger til alle deler av sektoren om hva som kan gjøres for å oppnå mer deling og gjenbruk av forskningsdata.

### Utvalgets mandat

[Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata \(2017\)](#) slår fast at forskningsdata skal være så åpne som mulig, men så lukkede som nødvendig. Det er en målsetting at resultatene fra offentlig finansiert forskning skal bidra til verdiskaping og komme allmennheten til gode. Økt etterprøvnbarhet, bedre kvalitetssikring og mindre duplisert datafangst er også viktige motiver for mer tilgjengeliggjøring av forskningsdata. Hensiktsmessig og konsistent bruk av lisenser, som understøtter enklere deling på tvers av fagfelt og sektorer og som er i samsvar med internasjonale anbefalinger og praksis, vil være av stor betydning.

Utvalgets mandat har vært å utrede rettighets- og lisensspørsmål i forbindelse med deling av datasett i forskningen ved å gjennomgå relevant lovverk, beskrive ulike typer lisenser og vurdere konsekvenser av forskjellige lisensvalg – både for forskeren, institusjonene og samfunnet. Utvalget har også blitt bedt om å diskutere hvilke forutsetninger som må være på plass for å nå målene om mer deling og gjenbruk av forskningsdata.

Oppdraget har vært todelt, med to leveranser som resultat.

1. I fase 1 ga utvalget sine anbefalinger om hvilke prinsipielle diskusjoner som burde tas, hvilke strukturelle endringer som burde vurderes og eventuelt hvilke lovverk som burde gjennomgås på nytt for at Norge skal oppnå mer deling og gjenbruk av forskningsdata. Utvalget ble bedt om å se sitt arbeid i sammenheng med andre pågående initiativer som er iverksatt for å bygge kunnskap om data og datahåndtering i ulike samfunnssektorer. Første leveranse ble materialisert i et notat fra utvalget til Kunnskapsdepartementet (høst 2020) som dannet grunnlag for neste fase i utvalgsarbeidet.

- I fase 2 har utvalget invitert til en åpen prosess med ulike møteplasser, diskusjonsarenaer og innspillmuligheter, slik at alle relevante interessenter har fått muligheten til å presentere sine synspunkter. Den endelige leveransen fra utvalget skal bli en oppsummerende rapport fra arbeidet og et sett nasjonale anbefalinger om bruk av lisenser på forskningsdata (sommer 2021).

### Sammensetning av utvalget

Utvalget representerer bredden av fagområder i forskningssektoren, deriblant to jurister innen immaterialrett og personvernrett. Medlemmene representerer både UH-sektor og instituttsektor, og skal også representere alle landsdeler. I tillegg er næringslivs- og samfunnsperspektivet representert i utvalget. Utvalget består av følgende medlemmer:

Rolle	Navn	Stilling	Arbeidsgiver, geografisk område
Leder	Magnus Aronsen, utvalgsleder	Førsteamanuensis	UiO, Institutt for medisinske basalfag/Oslo Universitetssykehus
Medlem	Koenraad De Smedt	Professor	UiB, Institutt for lingvistiske, litterære og estetiske studier
Medlem	Ingrid Heggland	Førstebibliotekar	NTNU, Bibliotekseksjon for samlinger og digitale tjenester
Medlem	Liv Dingsør	Daglig leder	DigitalNorway
Medlem	Stein Tronstad	Seksjonsleder	Norsk polarinstitutt
Medlem	Ole Petter Pedersen	Redaktør	Teknisk Ukeblad
Medlem	Inger Berg Ørstavik	Professor	UiO, Institutt for privatrett
Medlem	Tobias Mahler	Professor	UiO, Institutt for privatrett
Observatør	Nina Karlstrøm	Seksjonsleder	UNIT
Observatør	Katrine Weisteen Bjerde	Fagdirektør	UNIT
Sekretariatsleder	Siri Lader Bruhn	Spesialrådgiver	Norges forskningsråd
Sekretariat	Nenitha C. Dagslott	Rådgiver	Norges forskningsråd
Sekretariat	Jon Ø. Flæten	Spesialrådgiver	Norges forskningsråd
Sekretariat	Christine Daae Olseng	Spesialrådgiver	Norges forskningsråd

### Andre bidragsytere

For gjennomgangen av det juridiske rammeverket for deling og gjenbruk av forskningsdata har utvalget fått god hjelp fra eksperter på norsk og europeisk regelverk på området. Stipendiat Kristina Stenvik ved Institutt for privatrett (UiO) har foretatt en gjennomgang av norsk regelverk og systematisert dette. Denne gjennomgangen har gitt grunnlag for utvalgets diskusjoner om lovverket som regulerer deling og gjenbruk av forskningsdata i Norge. Gjennomgangen kan leses i sin helhet i vedlegg 1 til denne rapporten.

Heather Broomfield, seniorrådgiver i Digitaliseringsdirektoratet og stipendiat ved institutt for offentlig rett (UiO), og utredningsleder i Kommunal- og moderniseringsdepartementet Espen Dennis Kristoffersen har bidratt med en gjennomgang av relevant europeisk regelverk som påvirker og regulerer deling av forskningsdata i Norge. Denne gjennomgangen ligger ved rapporten i vedlegg 2 (**\*kommer i endelig rapport**).



Utvalget takker for disse bidragene, som har gitt et godt utgangspunkt for diskusjoner og grunnlag for flere av anbefalingene som utvalget gir i denne rapporten.

## Prosess

Utvalget startet sitt arbeid i juni 2020. Det har i alt vært 10 utvalgsmøter, som har blitt gjennomført jevnlig i løpet av perioden. I oktober 2020 overleverte utvalget notatet [Forskningsdata i et datadrevet samfunn](#) som et resultat av utvalgsarbeidets første fase. Utvalget inviterte til en åpen innspillrunde etter notatets publisering. Utvalget fikk mange gode innspill, som ble tatt med i det videre arbeidet i andre fase.

I løpet av første halvår i 2021 har utvalget i tillegg hatt en rekke møter med relevante aktører, blant andre departementenes forskningsutvalg, UHR Forskning, den norske RDA-noden, Abelias dataforum, Forskerforbundets opphavsrettsutvalg og ledelsen ved enkelte universiteter. Utvalgsleder og flere utvalgsmedlemmer har også holdt innlegg på ulike seminarer og konferanser underveis i perioden. I mars 2021 gjennomførte utvalget en åpen innspillkonferanse, med muntlige innspill og paneldiskusjoner.

*I juni 2021 ble utkast til rapport publisert. Utvalget inviterte på nytt til en åpen innspillrunde, med frist i august. Innspillene er behandlet og innarbeidet i denne endelige rapporten fra utvalget.*

Det har vært viktig for utvalget at prosessen fram til endelig rapport skulle være så åpen og involverende som mulig. Oppdraget til utvalget omfatter komplekse spørsmål og problemstillinger som berører alle deler av forskningssektoren, men også andre sektorer og aktører. Den åpne prosessen har ført til mange interessante og gode diskusjoner, og kontakt og dialog på tvers miljøer har blitt etablert. Utvalget takker for alle bidrag og ser fram til videre samarbeid for å nå målet om en mer åpen forskningssektor.

## Om rapportens avgrensning og oppbygging

Utvalgets arbeid har fått mye oppmerksomhet og skapt engasjement. Dette tyder på at tematikken er aktuell for mange. Det viser også at det er stort behov for å finne gode løsninger for at forskningssektoren skal kunne etterfølge politiske krav og for at forskere og andre skal få tilgang på gode og verdifulle datasett. Utvalgets mandat er i utgangspunktet tydelig avgrenset til å gjelde forskningsdata og medfølgende rettighets- og lisensspørsmål, men det er viktig å poengtere at det finnes en rekke prosesser og initiativer som omhandler deling av data generelt og som det er viktig å se dette arbeidet i sammenheng med. Denne rapporten må derfor sees på som ett av flere bidrag for å oppnå mer deling og gjenbruk av offentlig finansierte data i et større perspektiv, men med et spesielt fokus på data som produseres gjennom forskningsaktivitet.

Gjennom innspillsrunden har det kommet tydelig fram at temaer knyttet til deling og gjenbruk av forskningsdata opptar mange i forskningssektoren, og det har kommet en rekke gode innspill til hva som bør omtales i denne rapporten. Utvalget opplever at det er et stort behov for mange typer avklaringer og det er veldig mange og viktige problemstillinger som bør løses i tiden vi går inn i. Utvalget omtaler problemstillingene underveis, men har ikke tatt mål av seg å gi svar eller konklusjoner i denne rapporten. Dette er spørsmål som det vil kreve en felles og langsiktig innsats fra mange aktører for å finne gode svar på.

Rapporten bygger til dels videre på utvalgets notat fra oktober 2020, men har tatt opp i seg diskusjoner innad i utvalget og den omfattende og verdifulle dialogen utvalget har hatt med ulike aktører i løpet av utvalgsperioden. I tillegg inneholder rapporten en mer grundig gjennomgang av det juridiske rammeverket for deling og gjenbruk av forskningsdata.

Rapporten består av tre deler. Den første delen gjør rede for hva utvalget har lagt i begrepet forskningsdata, hva som i noen tilfeller kan skille denne typen data fra andre typer data, forskningsdatas livssyklus og betydningsforskjellen mellom åpne data og FAIR data. I del to gjennomgås det juridiske rammeverket som er relevant ved deling og gjenbruk av forskningsdata og utvalgets vurdering av behov for avklaringer og eventuelle endringer. I denne delen gis det også en gjennomgang av ulike typer lisenser og deres funksjon, samt hvilke konsekvenser ulike lisensvalg vil ha for gjenbruk av forskningsdata. Denne delen avsluttes med utvalgets overordnede anbefalinger om bruk av lisenser. I siste del presenteres de viktigste faktorene som utvalget mener påvirker forskeres og institusjoners mulighet og vilje til å dele forskningsdata. Til slutt gir utvalget sine anbefalinger til henholdsvis regjeringen og departementene, finansiører og virkemiddelapparatet, samt forskningsinstitusjonene. Utvalget mener tiltakene som anbefales vil kunne bidra til at norske forskningsdata vil bli mer delt og mer gjenbrukt i framtiden.

## DEL 1 – Egenskaper ved forskningsdata

Store deler av politikktutforming, og løsningene som utvikles, for å stimulere til mer deling av offentlige data generelt, vil også være relevant for deling av forskningsdata. For å utnytte forskningsdataenes potensial, er det imidlertid noen særskilte egenskaper og trekk som bør tas i betraktning når det skal avgjøres hvilke retningslinjer som skal gjelde. I denne delen av rapporten gir vi en beskrivelse av ulike varianter av forskningsdata som vil kunne kreve ulike vurdering og hensyn knyttet til deling og gjenbruk. Vi gir videre en gjennomgang av forskningsdataenes livssyklus, og beskriver hvordan ulike stadier i syklusen kan stille ulike krav til slike vurderinger. Til slutt angir vi hvilke typer data denne rapporten i hovedsak vil omhandle.

### 1. Hva ligger i begrepet forskningsdata?

Hva forskningsdata er forstås ulikt på tvers av fagdisipliner og forskningsmiljøer. Det finnes derfor ingen entydig definisjon, og begrepet brukes om mange forskjellige typer data og datarelaterte objekter. Dette kan gjøre det utfordrende å snakke om forskningsdata på tvers av områder, og ikke minst finne gode felles løsninger for håndtering av forskningsdata. I dette kapitlet presenteres noen definisjoner som vi mener er ryddige og gode. Vi vil også gi en oversikt og ulike egenskaper og kvaliteter ved forskningsdata som kan ha innvirkning på hvordan man skal forholde seg til regelverk og retningslinjer.

#### 1.1 Definisjon av forskningsdata

Det finnes mange til dels overlappende definisjoner av forskningsdata, men en fellesnevner er at det dreier seg om informasjon som på en eller annen måte brukes eller oppstår i forbindelse med forskningsaktivitet. De fleste definisjoner trekker fram at data er informasjonen som brukes som underlag eller grunnlag for vitenskapelige funn eller analyser, og at det også kan brukes for å etterprøve og validere resultater fra forskning. Forskningsdata kan være i enten digitale eller fysiske format, men i forbindelse med datadeling er det i all hovedsak digitale data som er mest aktuelt å diskutere.

Siden det allerede finnes mange definisjoner av forskningsdata, har ikke utvalget ønsket å lage en egen definisjon. Vi vil heller vise til eksisterende definisjoner, som har lagt seg på ulikt detaljeringsnivå, men som vi mener er nyttige på hver sin måte. Definisjonen fra EOSC er kortfattet, noe som kan være nyttig når man skal diskutere overordnede problemstillinger knyttet til forskningsdata. Definisjonen fra OpenAire er mer detaljert, noe som kan være klargjørende når man skal gå litt mer i dybden av problemstillingene og finne konkrete løsninger og tiltak.

Definisjon av Forskningsdata av [EOSC](#):

*Data collected or produced in the course of scientific research activities and used as evidence in the research process, or commonly accepted in the research community as necessary to validate research findings and results.*

Definisjon av Forskningsdata av [OpenAire](#):

*Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence. Data may be raw or primary (e.g. direct from measurement or collection) or derived from primary data for subsequent analysis or interpretation (e.g. cleaned up or as an extract from a larger data set), or derived from existing sources where the rights may be held by others. Data may be defined as 'relational' or 'functional' components of research, thus signalling that their identification and value lies in whether and how researchers use them as evidence for claims. They may include, for*

*example, statistics, collections of digital images, sound recordings, transcripts of interviews, survey data and fieldwork observations with appropriate annotations, an interpretation, an artwork, archives, found objects, published texts or a manuscript.*

Omforente definisjoner er nyttige når man skal diskutere temaer på tvers av sektorer og på tvers av forvaltningsnivåer. Nettopp fordi en diskusjon rundt forskningsdata favner så bredt i denne rapporten ønsker imidlertid utvalget å omtale forskningsdata på en måte som rommer flere ulike egenskaper. Derfor er det ikke hensiktsmessig å ha én generell og omforent definisjon. Utvalget mener det er viktig å være klar over ulike aspekter ved data som kan ha innvirkning på hvordan man skal ta fornuftige valg om deling, eventuelle restriksjoner og valg av lisenser. I det følgende har vi tatt for oss flere av disse egenskapene, og vi håper det vil gi et bilde av hvor heterogene forskningsdata faktisk er, og at det ikke vil være fornuftig å omtale forskningsdata som én ting med bestemte egenskaper når man skal legge til rette for tilgjengeliggjøring av forskningsdata.

### 1.2 Rådata vs. bearbejdet data og dataprodukter

Rådata brukes om data i sin opprinnelige form, slik de er når de samles inn. Når slike data systematiseres, sorteres eller analyseres, går de over til å bli bearbejdet data. Hvis dataene organiseres og struktureres i et system som man kan få tilgang til elektronisk, kalles det gjerne for en database. I juridisk sammenheng kan databaser få rettslig vern dersom visse vilkår er oppfylt i loven. Dette stiller seg annerledes for ustrukturerte rådata (dette kommer vi nærmere inn på under [DEL 2](#)). Under visse omstendigheter kan bearbejdelse av data gi eksklusive rettigheter, slik som opphavsrett, og denne kan i noen tilfeller også gjelde for databaser.

### 1.3 Kildedata vs. data generert underveis i forskningsprosessen

I forbindelse med deling og gjenbruk av forskningsdata vil det ofte være relevant å skille mellom allerede eksisterende data som gjenbrukes i forskning (gjærne kalt kildedata), og nye 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 i form av fysiske samlinger, litterære tekster eller annen kunstnerisk produksjon. Det kan også være snakk om data fra annen forskning. 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 gjøre det utfordrende eller umulig med videre deling.

Forskningsdata som er generert uten bruk av kildedata kan friere deles videre av den ansvarlige forskeren, forskningsgruppen 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. En slik blanding kan være med på å komplisere vurderingen av videre deling ytterligere. Dette diskuteres videre i [DEL 2](#).

### 1.4 Kvalitative vs. kvantitative data

En annen dimensjon ved forskningsdata som det er relevant å trekke fram, er hvorvidt dataene er innhentet ved bruk av kvantitativ eller kvalitativ metode. Hvilken metode som er brukt, kan ha innvirkning på hvor objektive dataene er, og hvorvidt det faktisk er mulig å skille "rene" data fra selve forskningsaktiviteten. Det kan også ha innvirkning hvor mye metadata, annen dokumentasjon og

informasjon om kontekst som er nødvendig for å kunne forstå og fortolke dataene. 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. Dataene fra slik forskning lar seg ofte kvantifisere gjennom tallstørrelser og kan brukes til statistisk analyse. Ved bruk av kvalitativ metode er ofte innhenting av dataene en mer integrert del av selve forskningsaktiviteten, noe som kan gjøre det vanskelig å skille klart mellom hva som utgjør dataene (det som analyseres eller fortolkes) og hva som er forskerens intellektuelle bidrag. Slike prosesser er ofte hermeneutiske, hvor nye forståelser og nye elementer trekkes inn underveis. Dette kan være med på å gjøre skillet enda mindre klart.

### 1.5 Persondata vs. data om verden for øvrig

Innenfor flere forskningsfelt, som for eksempel medisin, bioteknologi, samfunnsvitenskapene og humaniora, vil ofte informasjon om personer utgjøre en del datagrunnlaget for forskningen. Slike data må behandles på en annerledes måte enn data som gir informasjon om verden for øvrig. Informasjon og data som kan knyttes til enkeltpersoner, vil falle innunder virkeområdet til personopplysningsloven, og for å kunne behandle personopplysninger må man ha et lovlig grunnlag (behandlingsgrunnlag). Innenfor forskningen innhentes et slikt behandlingsgrunnlag gjerne i form av et samtykke. Dersom personopplysningene er sensitive, må det i tillegg innhentes et såkalt uttrykkelig samtykke, som innebærer at samtykket er gitt på en ekstra tydelig måte. Det er videre et krav at samtykket er spesifikt og informert, noe som innebærer at formålet med samtykket er klart og presist formulert. Det er som hovedregel ikke rom for å bruke personopplysninger til andre formål enn det som er beskrevet i samtykket. I utgangspunktet kan man derfor ikke dele persondata åpent. Men med et eksplisitt samtykke om deling, så vil det være mulig etter de vilkår som er beskrevet. Da vil det være aktuelt å bruke en lisens eller en frigivelseserklæring.

Dersom man anonymiserer dataene, det vil si at det ikke lenger er mulig å identifisere enkeltpersoner i datasettet, så vil ikke lenger personlovgivningen gjelde, og dataene kan deles med andre. Likevel er det et forskningsetiske spørsmål hvorvidt man bør dele anonymiserte data uten at de registrerte er informert når data er innsamlet basert på samtykke.

### 1.6 Stordata vs. "den lange halen av forskningsdata"

Betegnelsen stordata beskriver samlinger av data som er så store og komplekse at det ikke er mulig å bruke tradisjonell dataprosessering for å utnytte dem, som for eksempel værdata eller transportdata. Dataene kommer som regel fra mange ulike kilder, men er ofte standardiserte og kan derfor gi ny og viktig informasjon når de sammenstilles og analyseres. I forbindelse med dataøkonomi, datautvinning og kunstig intelligens, er det ofte stordata man snakker om. Data som befinner seg på den andre enden av skalaen, kan man for enkelhets skyld kalle smådata, selv om dette ikke alltid er en treffende betegnelse. Dette kan også være relativt store datasett, men de kjennetegnes ved at de ofte inneholder mer presis og detaljert informasjon enn stordata, og de trenger ikke veldig avanserte datasystemer eller stor regnekraft for å kunne analyseres. Stordata genereres ofte innenfor fagområder med internasjonalt etablerte standarder, avanserte infrastrukturer og administrative støttesystemer, mens smådataene ikke nødvendigvis har samme infrastruktur på plass. Smådata er også mer heterogene i sin natur og vil i mange tilfeller trenge individuell vurdering og tilpasning for at de skal kunne tilgjengeliggjøres for gjenbruk av andre. Slike data omtales også ofte som "den lange halen av forskningsdata"<sup>1</sup>

---

<sup>1</sup> [LongTailOfData2016 \(e-irg.eu\)](http://LongTailOfData2016(e-irg.eu))

## 2. Åpne data og FAIR data

Å dele resultater fra forskning er et viktig prinsipp for å sikre transparens og etterprøvbarehet. De senere årene har åpen forskning blitt et spesielt viktig policyområde, blant annet fordi digitaliseringen har muliggjort gjenbruk av resultater på helt nye måter for forskere, næringslivet og samfunnet. I takt med den raske teknologiske utviklingen har mengden data som innhentes i forskningsmiljøene og samfunnet for øvrig, økt vesentlig. Dette har også 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 begrepet "åpne data" og konseptet "open by default". Dette har i ettertid 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 og andre resultater fra forskning "så åpne som mulig og så lukket som nødvendig". Dette gjenspeiles også i regjeringens strategi og 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 [Open Science Policy Platform: final report](#) 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, på norsk "immaterielle rettigheter") og de individuelle bidragene forskere gir til å løse de store samfunnsutfordringene.

Siden "åpne data" og "FAIR data" begge er begreper som brukes for å beskrive tilgang til og vilkår for gjenbruk av data, og til dels brukes om hverandre, er det viktig å poengtere at de ikke betegner det samme. I det følgende går vi litt nærmere inn på hva de to begrepene innebærer og hvilken praktisk betydning de to måtene å dele data på kan ha.

### 2.1 Åpne data

At data er åpne, betyr at de deles uten begrensninger 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 offentlige midler, 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. For 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 åndsverkloven. Det kan være krevende for forskeren, forskergruppen eller institusjonen som skal sørge for at delingen av dataene skjer i henhold til loven, å finne fram 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 del 2 av denne rapporten og i vedlegg 1.

Når det gjelder data som ikke er underlagt restriksjoner i lov, så er det fortsatt slik at en del data ikke deles. For å finne en forklaring på hvorfor dette skjer, er det nødvendig å undersøke hvilke underliggende faktorer som påvirker 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, manglende juridisk bistand og kompetanse eller er det mangel på insentiver som gjør at dataene i noen tilfeller ikke deles? Dette blir også nærmere diskutert i kapittel 4 i dette notatet.

## 2.2 FAIR data

Mens åpne data hovedsakelig betyr at data kan brukes av enhver uten restriksjoner, legger betegnelsen FAIR data vekt på teknisk tilgjengelighet, noe som forutsetter gode maskinlesbare metadata, inkludert lisensinformasjon. 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 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 er imidlertid viktig å understreke at man kan vurdere dataenes FAIRness langs en akse og ikke ut fra en enten/eller-tilnærming.

FAIR-prinsippene innebærer ikke nødvendigvis at data skal være åpne, og åpne data innebærer ikke nødvendigvis at data skal være FAIR – selv om begge deler som regel er ønskelig.

## 3. Forskningsdataenes livssyklus

**\*figur 1 kommer i endelig rapport**

Data vil gjerne gå gjennom et dynamisk livsløp i et forskningsprosjekt, fra innsamling av rådata til ulike former for strukturering, analyse og prosessering. (ref figur 1). I motsetning til offentlige data, som gjerne innhentes automatisk eller systematisk som en del av offentlig forvaltning, innhentes som regel forskningsdata når man ønsker å teste hypoteser eller frambringe ny kunnskap om spesifikke tema. Forskningsdata må ofte knyttes 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 (ref figur 1), vil i mange tilfeller også eksistere utover prosjektperioden, sammen med dataene (ref figur 1). Det stiller ytterligere ressurskrav til vedlikehold og oppdatering av format og standarder og migrering til nye medier i tråd med teknologisk utvikling, slik at det er mulig å gjenbruke og videreutvikle dataene også i framtiden.

Vitenskapelige publikasjoner, som for eksempel artikler, bøker eller monografier, inngår ofte som en del av livssyklusen i et forskningsprosjekt (ref figur 1). Det er imidlertid ikke alltid forskningsdata resulterer i publikasjoner, for eksempel fordi de blir sett på som negative eller lite interessante resultater i lys av hypotesen eller forskningsspørsmålet (ofte kalt publiseringsbias). Slike data kan likevel være nyttige for forskere og tjeneste- og produktutviklere som ser et uutnyttet potensial for kunnskapsutvikling, gjenbruk eller videreutvikling. Samtidig vil slike data være nødvendige for å kunne gjøre metaanalyser av

det samlede kunnskapsgrunnlaget i et felt, og manglende åpenhet rundt dette vil kunne gi systematiske skjevheter i senere kunnskapsoppsummeringer. Preregistrering av studier og hvilke hypoteser man ønsker å teste, kan være et godt virkemiddel for å unngå publiseringsbias, samt bidra til robust og reproduserbar forskning

Spørsmål om *når* man skal dele data i et forskningsprosjekt blir juridisk viktig fordi forskningsdata endrer status som rettsobjekt<sup>2</sup> etter hvor i livsløpet til prosjektet dataene befinner seg. Utvalget kommer nærmere inn på dette i [DEL 2](#).

#### 4. Tilgjengeliggjøring og publisering av data

Forskningssektoren har på lik linje med andre sektorer og bransjer endret arbeidsmåter og prosesser som et resultat av omfattende digitalisering de siste årene. Måten data blir tilgjengeliggjort på er intet unntak. For ikke mange år siden ble data stort sett plassert i institusjonsinterne filarkiver og databaser, hvis de i det hele tatt ble tatt vare på etter at forskningen var ferdig og vitenskapelige artikler var publisert. I dag blir data i dag i større grad aktivt tilgjengeliggjort gjennom API-er og gjennom publisering av datasett som selvstendige, vitenskapelige produkter med siterbare titler, navngivelse og metadata. I tillegg har innføring av universelt unike identifikatorer, som DOI-er, gitt nye muligheter for sitering, mens bruk av lisenser gir gode muligheter for å sette vilkår for gjenbruk. Innføring av denne typen digitale løsninger har ført til et økt behov for standardlisenser, til erstatning for tidligere former for individuelle avtaler om manuell tilgjengeliggjøring.

#### 5. Forskningsdata som er kvalifisert for deling og gjenbruk

Som vi har vist i det foregående er forskningsdata en heterogen kategori som rommer mange ulike typer informasjon og kilder. Noen vitenskapsgrener arbeider i all hovedsak med fortolkninger og vil i liten grad bedrive empirisk forskning som gir resultater i form av delbare data. En ny fortolkning av et filosofisk verk eller av en rettsregel er eksempler på elementer som ikke vil kunne deles som andre typer data. Fortolkninger som dette vil som regel allerede inngå i en vitenskapelig publikasjon, for å underbygge svaret på forskningsspørsmålet. Når denne rapporten diskuterer deling av forskningsdata, gir det lite mening å inkludere disse opplysningstyper som er lite egnet for deling. 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. Når utvalget i det videre gir sine anbefalinger, er det viktig å understreke at disse gjelder forskningsdata som er kvalifiserte for deling og gjenbruk. Utvalget er spesielt opptatt av en slik kvalifisering for at miljøer som ikke har forutsetninger for å dele data på samme måte som andre, skal forfordes i en satsing på mer deling av forskningsdata, med medfølgende tiltak, virkemidler og insentivmekanismer.

Det vil ikke være mulig for et relativt lite utvalg, med medlemmer som ikke representerer alle deler av forskningsområdene, å definere fullt ut hva som bør regnes som kvalifiserte data, men vi mener at gjennomgangen over kan gi et godt utgangspunkt for å komme fram til en slik kvalifisering innenfor de ulike forskerfellesskapene.

---

<sup>2</sup> Eksempler på rettsobjekt kan være informasjon, dokumenter, databaser og publikasjoner.



## DEL 2 – Juridisk rammeverk og bruk av lisenser og frigivelseserklæringer på forskningsdata

I denne delen av rapporten gir vi først en gjennomgang av det juridiske rammeverket som regulerer deling og gjenbruk av forskningsdata. Denne teksten er et sammendrag av en mer utfyllende tekst som kan leses i vedlegg 1 (**\*kommer i endelig rapport**). Deretter diskuterer vi ulike typer lisenser og frigivelseserklæringer, deres funksjon og hvilke konsekvenser ulike valg av lisens eller frigivelseserklæring vil ha for gjenbruk av forskningsdata. Avslutningsvis presenteres utvalgets overordnede anbefalinger om bruk av lisenser.

### 6. Juridisk rammeverk for deling av forskningsdata

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, eller deling og gjenbruk av, slike data. For enkeltforskere, forskningsmiljøer og institusjoner er det vanskelig å forholde seg til et fragmentert og uoversiktlig juridisk landskap.

Formålet med dette kapitlet er å gi en oversiktlig og kortfattet redegjørelse for relevante internasjonale og nasjonale rettsregler som er aktuelle for deling av forskningsdata, både når det gjelder rettsregler som gir grunnlag for deling, og rettsregler som setter eventuelle begrensninger for deling. Kapitlet er en oppsummering av en mer omfattende gjennomgang av regelverket som finnes i vedlegg 1 (**\*kommer i endelig rapport**). Siden utvalgsarbeidet er begrenset i tid og omfang tar ikke gjennomgangen sikte på å gi en uttømmende eller bred redegjørelse.

Inngåtte avtaler og lisenser kan også gi rettigheter til, legge begrensninger på eller stille krav om deling av forskningsdata. Det rettslige regelverket setter rammene for avtaler og lisenser på forskningsområdet. I den grad lovbestemmelsene er fravikelige, kan man avtale seg bort fra reglene. En del av regelverket kan man ikke avtale seg bort fra, herunder reglene om innsyn i blant annet offentlighetsloven og reglene om forskningsetikk og reglene om personopplysninger, nasjonal sikkerhet og taushetsplikt. Regelverkene kan riktignok selv tenkes å oppstille unntak, for eksempel ved samtykke. På immaterialrettsområdet er det en viss frihet til å inngå avtaler som avviker fra lovens reguleringer, og det er også i utgangspunktet anledning til å overdra rettighetene. Denne friheten oppstår imidlertid først etter at rettighetene har oppstått – bestemmelser som oppstiller krav for å oppnå rettigheter etter åndsverkloven og patentloven er absolutte, herunder nyhetskravet i patentretten. Og selv om rettighetene kan overdras, kan man i begrenset grad frasi seg de ideelle rettighetene etter åndsverkloven § 5, herunder retten til å navngis som opphavsmann. Arbeidstakeroppfinningsloven som regulerer arbeidsgivers rett til arbeidstakers oppfinnelser er som hovedregel fravikelig, med noen unntak. Lisensiering av, og bruk av frigivelseserklæring ("Public Domain") på forskningsdata utdypes i [kapittel 7](#).

#### 6.1 Internasjonalt lovverk

**\* kommer i endelig rapport**

#### 6.2 Nasjonalt lovverk

Det finnes ikke et enhetlig lovverk for forskningsdata i Norge, dette er regulert gjennom fragmentariske deler av en rekke ulike lover. Regelverket kan deles inn i tre hovedkategorier: (1) immaterialrettslig regelverk som gir rettigheter til data, samling av data eller oppfinnelser som bygger på data, (2)

regelverk som tilrettelegger for, regulerer eller krever deling av data, og (3) regelverk som setter begrensninger på deling av data.

### 6.2.1 Immaterialrettslig regelverk

#### Forskningsdata som utgjør kunstneriske eller litterære verk (opphavsrett)

Opphavsretten gir vern for kunstneriske og litterære verk som er «uttrykk for original og skapende åndsinnsetning», jf. åndsverkloven § 2 annet ledd. Dette omfatter blant annet faglitterære tekster og muntlige foredrag, jf. åvl. § 2 annet ledd bokstav a og b. Slike tekster eller foredrag kan inneholde forskningsdata. Men selv om man kan få rettigheter til teksten eller foredraget, er det bare den konkrete og individuelle utformingen man får opphavsrettigheter til, og ikke det faktiske innholdet. Man får ikke opphavsrett til forskningsdata som inneholder rene fakta eller informasjon og heller ikke ideer eller vitenskapelige teorier. Ulike publikasjoner som inneholder forskningsdata, vil altså kunne være underlagt opphavsrett, men omfatter ikke forskningsdata som ligger til grunn for innholdet som sådan. Opphavsretten kan dermed gi opphavspersonen enerett til en artikkel eller annen publikasjon, men selve forskningsdataene kan fritt benyttes av andre uhindret av opphavsretten.

I den grad forskningsdata inneholder kunstneriske eller litterære verk som er underlagt opphavsrett, vil dette selvsagt stille seg annerledes. I slike tilfeller kan videre deling, gjenbruk og videreutvikling av forskningsdataene bare skje så langt det ikke krenker de opprinnelige rettighetene.

Under visse omstendigheter kan man ha opphavsrett til en database. Det forutsetter at vilkårene for åndsverk er oppfylt ved utformingen av databasen – nemlig at utvalget og innsamlingen er uttrykk for individuell skapende åndsinnsetning. Det er sammenstillingen som eventuelt kan ha vern, ikke de enkelte data. Vitenskapelige databaser vil nok sjelden kvalifisere som åndsverk, ettersom utformingen og sammenstillingen typisk ikke vil bero på kreative valg, og man må derfor som regel basere seg på databasevernet.

Opphavsretten oppstår hos opphavspersonen, den som har utøvet den individuelle skapende innsatsen, i motsetning til databasevernet som oppstår hos den som har gjort investeringen. Opphavsrett vil derfor typisk oppstå på forskerens hånd, mens databaserettigheter kan oppstå hos institusjonen. Flere personer kan i fellesskap ha utøvet en slik innsats at opphavsrett oppstår. Hverken veiledning, laboratoriearbeid, analyse av data eller finansiering kan imidlertid gi opphavsrett. Det understrekes også at det kun er fysiske personer som kan være opphavere. Dersom institusjonen skal få opphavsrettigheter, må opphavsretten overdras gjennom avtale.

#### Vern for samling av forskningsdata (databasevern)

Etter åndsverkloven § 24 kan den som framstiller en «database» få enerett til å råde over denne databasen, dersom innsamling, kontroll eller presentasjon av innholdet innebærer en «vesentlig investering». En «database» er definert i databasedirektivet som en samling av selvstendige verk, data eller annet materiale, som er strukturert systematisk eller metodisk, og som er tilgjengelig individuelt ved bruk av elektroniske eller andre midler.<sup>3</sup> Databasevernet gir altså vern for en samling av data som er strukturert systematisk eller metodisk, ikke for rådata eller enkeltdata som sådan. Det er selve *samlingen* som er vernet.

---

<sup>3</sup> Databasedirektivet 96/9/EC, Artikkel 1 nr. 2.

Vilkåret for vern av databaser er at innsamlingen, kontrollen eller presentasjonen av innholdet utgjør en «vesentlig investering». 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.

Rettighetshaver til databaser er den som gjør den «vesentlige investeringen». Det vil typisk være en institusjon. Rettighetene vil her oppstå hos institusjonen selv om det for eksempel var en ansatt som kom på at en database var den beste måten å strukturere forskningsdata på. Selv om det er ansatte som legger ned tid og arbeid i innsamlingen eller presentasjonen av databasen, vil det være arbeidsgiver som gjør investeringen, forutsatt at arbeidet foregår i arbeidstiden. Ved eksternt finansierte prosjekter kan det tenkes at rettighetene oppstår i sameie mellom institusjonen og den eksterne finansiør.

#### [Enerett til oppfinnelser og offentliggjøring av patentsøknader \(patentrett\)](#)

Patentretten kan gi en oppfinner (eller andre som rettighetene er overdratt til) enerett til å utnytte en teknisk løsning. Der forskningsdata gir grunnlag for å utvikle en oppfinnelse, kan man altså få enerett til utnyttelse av oppfinnelsen, men man får ikke enerett til dataene som sådan. For å oppnå patentrettigheter, er det et krav at informasjonen om oppfinnelsen ikke blir kjent før det søkes patent. Det kan medføre at forskningsdata holdes hemmelig inntil oppfinnelsen eventuelt er patentsøkt. Det er imidlertid også et krav at informasjonen om oppfinnelsen offentliggjøres gjennom offentliggjøringen av patentsøknaden. Forskningsdata som inngår i en patentsøknad, vil dermed bli offentlig tilgjengelig og kan benyttes i videre forskning og utvikling, så lenge man ikke krenker patentretten gjennom å utnytte den patentbeskyttede oppfinnelsens økonomiske verdi. Patentretten er ikke til hinder for bruk av oppfinnelsen som kunnskapskilde, herunder til forskning og eksperimentelle formål, inkludert kommersiell forskning og utvikling. Patentretten setter i utgangspunktet altså ingen hindringer for deling av forskningsdata, men kan tvert imot innebære at forskningsdata må deles i den grad de inngår i en patentsøknad. I utgangspunktet innebærer patentretten derfor ingen begrensning for deling av forskningsdata, men kan snarere bidra til informasjonsspredning.

#### [Arbeidsgivers rett til oppfinnelser og forbud mot informasjonsdeling](#)

I situasjoner der arbeidsgiver har rett til å overta rettighetene til en oppfinnelse etter arbeidstakeroppfinnelsesloven, kan loven sette begrensninger på oppfinnerens rett til å dele informasjon om oppfinnelsen forut for inngivelse av en patentsøknad. Lærere og vitenskapelig personale ved universiteter og høyskoler har imidlertid en publiseringsrett etter lovens § 6 tredje ledd, som går foran arbeidsgivers rett til oppfinnelsen. I tilfeller der patentsøknad inngis, vil denne offentliggjøres på vanlig måte, inkludert eventuelle forskningsdata som søknaden inneholder. Etter at patentsøknad er inngitt, er det dessuten ingen begrensninger på oppfinnerens deling av informasjon.

#### *6.2.2 Regler som tilrettelegger for, regulerer eller krever deling av forskningsdata*

Formålet her er å redegjøre for de regelverk som i utgangspunktet tilrettelegger for og krever offentliggjøring av forskning. Vi har blant annet offentlighetsloven som gir rett til å kreve innsyn i offentlige organers dokumenter, herunder dokumenter som inneholder forskningsdata. Videre har vi forskningsetikkloven og universitets- og høyskoleloven som stiller visse krav til åpenhet og tilgjengeliggjøring av forskning, både av institusjonene og av forskerne selv. Helseforskningsloven stiller dessuten strenge krav til blant annet offentliggjøring av resultater på sitt område. Bestemmelser om tilgjengeliggjøring og offentlighet må imidlertid sees i lys av de begrensninger og unntak som følger av

den enkelte lov, men også de begrensninger som legges på deling av data gjennom annet regelverk. Regler som hovedsakelig legger begrensninger på deling av data, er tema [i delkapittel 6.3](#).

### Offentlighetsloven

Offentlighetsloven gir rett til innsyn i dokumenter som gjelder et offentlig organs virksomhet. Slike dokumenter kan tenkes å inneholde forskningsdata, og offentlighetsloven kan gi rett til innsyn også i disse. Offentlighetsloven krever således deling av forskningsdata under visse omstendigheter.

Retten til innsyn er betinget av at noen framsetter et krav om innsyn, jf. § 3 («alle kan krevje innsyn»). Innsyn kan kreves skriftlig eller muntlig, men det kravet må i utgangspunktet gjelde «ei bestemt sak», jf. § 28 annet ledd. Dette innebærer at kravet må være tilstrekkelig spesifisert – den som krever innsyn må vite hva vedkommende ønsker innsyn i. Forskningsdata som enda ikke er offentliggjort, kan derfor være vanskelig tilgjengelig for et innsynskrav. Der det foreligger en artikkel eller annen publikasjon, kan det være enklere å formulere et tilstrekkelig presist innsynskrav tilknyttet de underliggende data.

Innsynsretten gjelder i utgangspunktet for «saksdokument» (i tillegg til «journalar» og «liknande register»), jf. § 3. Både «dokument» og «saksdokument» er definert i lovens § 4. Et «dokument» er en logisk avgrenset informasjonsmengde som er lagret på et medium for senere lesing, lytting, framvisning, overføring eller lignende. Dette er en vid definisjon, og omfatter mer enn det man gjerne tenker på som et tradisjonelt tekstdokument. Kravet er at det er en logisk avgrenset informasjonsmengde som kan lagres og leses eller lignende. Dette kan være et alminnelig tekstdokument, men omfatter også lyd, bilde, tegninger, modeller og så videre. Begrepet er teknologinøytralt, og det har ikke noe å si hvordan informasjonen kommer til uttrykk.<sup>4</sup>

Hovedregelen om innsyn følger av offentlighetsloven § 3. Der heter det at saksdokumenter mv. er åpne for innsyn dersom «ikkje anna følgjer av lov eller forskrift». Retten til innsyn kan altså begrenses av andre lovbestemmelser som setter begrensninger på deling av informasjon. Dette kan være for eksempel regler om taushetsplikt, personvern eller regler om vern av forretningshemmeligheter.

### Forskningsetikkloven

Det følger av forskningsetikklovens formålsbestemmelse i § 1 at loven «skal bidra til at forskning i offentlig og privat regi skjer i henhold til anerkjente forskningsetiske normer». Loven regulerer forebygging og brudd på *anerkjente forskningsetiske normer for god vitenskapelig praksis*.<sup>5</sup> Lovfestede etiske krav som blant annet finnes i helseforskningsloven og personopplysningsloven vil ved motstrid gå foran forskningsetikkloven.<sup>6</sup> Loven bestemmer også at det oppnevnes nasjonale forskningsetiske komiteer, jf. § 9, og regionale komiteer for medisinsk og helsefaglig forskningsetikk (REK), jf. § 10.

I de Generelle forskningsetiske retningslinjer, som er utarbeidet av De nasjonale forskningsetiske komiteene, er det lagt til grunn i punkt 11 at forskningsresultater som hovedregel skal tilgjengeliggjøres.<sup>7</sup> Manglende tilgjengeliggjøring kan derfor tenkes å utgjøre brudd på forskningsetiske normer. Øvrige forskningsetiske normer i forbindelse med innsamling og publisering av forskningsdata,

---

<sup>4</sup> Ibid., note 68.

<sup>5</sup> Lovkommentar til forskningsetikkloven, Rettsdata, ved Bjørn L. Zwiilmeyer, note 1.

<sup>6</sup> Ibid.

<sup>7</sup> <https://www.forskningsetikk.no/retningslinjer/generelle/>

omfatter krav til redelighet (punkt 7) og plikt til å følge god henvisningsskikk ved bruk av andres materiale (punkt 8).

Forskningsetikkloven kan altså kreve tilgjengeliggjøring av forskningsdata i den grad det følger av anerkjente forskningsetiske normer, og dersom man benytter andres forskningsdata må man følge normer om redelighet og god henvisningsskikk. I den grad et brudd på en forskningsetisk norm kan sies å være alvorlig, og forskeren har opptrådt forsettlig eller grovt uaktsomt, kan saken tenkes å resultere i en uttalelse om at forskeren har opptrådt uredelig. Brudd på forskningsetiske normer i forbindelse med forskningen kan også resultere i en uttalelse om at et vitenskapelig arbeid bør korrigeres eller trekkes tilbake.

#### Universitets- og høyskoleloven

Universitets- og høyskoleloven skal blant annet tilrettelegge for at universiteter og høyskoler utfører forskning og faglig og kunstnerisk utviklingsarbeid på høyt internasjonalt nivå, jf. universitets- og høyskoleloven § 1-1 bokstav b. Loven skal også tilrettelegge for at disse «formidler kunnskap om virksomheten og utbrer forståelse for prinsippet om faglig frihet og anvendelse av vitenskapelige og kunstneriske metoder og resultater», jf. § 1-1 bokstav c.

I § 1-3 er det listet opp ulike måter hvorigjennom universiteter og høyskoler skal fremme lovens formål. Her er det eksplisitt uttalt at institusjonene skal «bidra til å spre og formidle resultater fra forskning ...» (bokstav d). Dette kan sies å innebære en forpliktelse til å oppmuntre og legge forholdene til rette for slik formidling.<sup>8</sup> Det følger imidlertid også av bestemmelsen at institusjonene skal «bidra til innovasjon og verdiskapning basert på resultater fra forskning» (bokstav e). I Norsk Lovkommentar er det framhevet at det er ønskelig at resultater av forskning og faglig utviklingsarbeid som har et kommersielt potensial skal formidles til virksomheter som kan bidra til innovasjon og verdiskapning – slik kunnskapsoverføring kan skape nye arbeidsplasser og styrke norsk kunnskapsindustri.<sup>9</sup> En forutsetning for slik kommersialisering er at forskningens frihet og integritet ivaretas, og at forskningsresultatene offentliggjøres innen rimelig tid.<sup>10</sup> Ønsket om kommersialisering og kravet til offentliggjøring kan imidlertid tenkes å komme i konflikt dersom kommersialisering forutsetter inngivelse av patentsøknader, hvilket legger begrensninger på offentliggjøring av forskningsresultatene forut for inngivelse av slik søknad.

Det følger videre av § 1-5 (6) at universiteter og høyskoler «skal sørge for åpenhet om resultater fra forskning ...». I denne bestemmelsen er også *forskernes rett og plikt til offentliggjøring* utførlig regulert. Den som er ansatt i stilling hvor forskning inngår i arbeidsoppgavene, har *rett til å offentliggjøre* sine resultater og *skal sørge for* at slik offentliggjøring skjer. Institusjonen kan imidlertid ikke pålegge offentliggjøring på et bestemt tidspunkt eller med en bestemt hyppighet, slik at forskeren selv bestemmer når og hvordan resultatene offentliggjøres.<sup>11</sup> Det følger av § 1-5 (5) at forskeren har rett til å velge emne og metode for sin forskning og sitt utviklingsarbeid innenfor de rammer som følger av ansettelsesforholdet eller særskilt avtale. Videre heter det i § 1-5 (6) at «det relevante forskningsgrunnlaget skal stilles til rådighet i overensstemmelse med god skikk på vedkommende fagområde». Her er det altså spesifisert at ikke bare skal resultatene offentliggjøres, men også

<sup>8</sup> Lovkommentar til universitets- og høyskoleloven, Rettsdata, ved Jan Fridthjof Bernt, note 25.

<sup>9</sup> Ibid., note 26.

<sup>10</sup> Ibid.

<sup>11</sup> Ibid., note 67.

«forskningsgrunnlaget» skal stilles til rådighet. Dette kan være statistiske grunndata eller fysisk materiale.<sup>12</sup> At det skal stilles til rådighet, innebærer at det skal gjøres tilgjengelig ved forespørsel.<sup>13</sup> Opplysninger som er underlagt taushetsplikt, må imidlertid som utgangspunkt skjermes eller anonymiseres.<sup>14</sup> Styret ved institusjonen kan samtykke til «utsatt offentliggjøring» av forskningsresultatene dersom det foreligger legitime hensyn. Slike legitime hensyn kan være beskyttelsen av patentrettslige eller konkurransemessige interesser eller hensynet til løpende forskningsarbeid.<sup>15</sup> Det er imidlertid uttrykkelig fastslått at det ikke kan avtales eller fastsettes «varige begrensninger» i retten til offentliggjøring utover det som følger av lov.

### Helseforskningsloven

Helseforskningsloven har til formål å fremme god og etisk forsvarlig medisinsk og helsefaglig forskning, jf. helseforskningsloven § 1. Loven gjelder for medisinsk og helsefaglig forskning på mennesker, humant biologisk materiale eller helseopplysninger, jf. § 2.

Lovens § 8 bestemmer at «kommersiell utnyttelse av forskningsdeltakere, humant biologisk materiale eller helseopplysninger som sådant er forbudt». Bestemmelsen er vag, og den nærmere grensedragningen er overlatt til praksis.<sup>16</sup> Bestemmelsen er ikke ment å forhindre forskning i næringsformål, for eksempel legemiddelutprøving.<sup>17</sup> Omsetning og kommersialisering av forskningsresultater og bearbeidet materiale faller også utenfor bestemmelsen.<sup>18</sup>

Det stilles krav til at den regionale komiteen for medisinsk og helsefaglig forskningsetikk (REK) godkjenner forskningsprosjektet etter søknad, jf. §§ 9 og 10, jf. forskningsetikkloven § 10 første ledd. I § 12 er det bestemt at når forskningsprosjektet avsluttes, skal det sendes en sluttmelding til REK. I denne skal resultatene presenteres på en objektiv og etterrettelig måte, som sikrer at positive og negative funn framgår. Formålet med bestemmelsen er å bidra til kontroll og åpenhet rundt forskningen.<sup>19</sup> Samfunnet skal ha mulighet til å følge med på hva slags medisinsk og helsefaglig forskning som foregår, og det må sikres at resultater ikke underslås dersom de for eksempel er til ugunst for den som finansierer forskningen.<sup>20</sup>

Helseforskningsloven har også regler i kapittel 6 om forskningsbiobanker som opprettes i forbindelse med innsamling, oppbevaring og bruk av humant biologisk materiale. I § 31 er det bestemt at den forskningsansvarlige skal gi andre forskere tilgang til humant biologisk materiale i virksomhetens forskningsbiobanker, med mindre den ansvarlige selv har behov for materialet eller det foreligger andre særlige grunner.

I kapittel 7 er det regler om behandling av helseopplysninger, som blant annet skal være i samsvar med prinsippene i personvernforordningen artikkel 5 og ha uttrykkelig angitte formål, jf. § 32. Det følger av § 34 at helseopplysninger kan sammenstilles, tilgjengeliggjøres og behandles i tråd med

---

<sup>12</sup> Ibid., note 68.

<sup>13</sup> Ibid., note 69.

<sup>14</sup> Ibid.

<sup>15</sup> Ibid., note 72.

<sup>16</sup> Lovkommentar til helseforskningsloven, Rettsdata, ved Sigmund Simonsen, note 40.

<sup>17</sup> Ibid.

<sup>18</sup> Ibid.

<sup>19</sup> Ot.prp. nr. 74 (2006-2007) pkt. 11.3.4.2.

<sup>20</sup> Ibid.

forskningsprosjektets formål, eventuelle samtykker, forhåndsgodkjenningen etter § 33 og i samsvar med forskningsprotokollen. REK kan imidlertid nekte slik sammenstilling, tilgjengeliggjøring mv. dersom denne anses som medisinsk eller etisk uforsvarlig, § 34 annet ledd. Det understrekes at sammenstilling og tilgjengeliggjøring av helseopplysninger kan skje til dataansvarlige eller forskningsansvarlige som har adgang til å behandle opplysningene etter personvernforordningen artikkel 6 og 9.

Kapittel 8 har regler om *åpenhet og innsyn i forskningen*. Det følger av § 39 at den forskningsansvarlige og prosjektlederen skal sørge for åpenhet rundt forskningen. Allmennhetens rett til innsyn er regulert i § 41. Enhver skal ved henvendelse til REK få vite hvilke forskningsprosjekter etter helseforskningsloven som en forskningsansvarlig eller prosjektleder er eller har vært involvert i, samt formålet med prosjektet. Unntakene i personopplysningsloven §§ 16 og 17 gjelder tilsvarende her, jf. § 42. Endelig følger det av § 44 at REK skal føre en systematisk fortegnelse over innmeldte og avsluttede forskningsprosjekter basert på opplysninger i søknad og sluttmelding, og at registrene skal være offentlig tilgjengelig. Man har imidlertid mulighet til å søke om utsatt offentliggjøring etter § 45, der dette er nødvendig for å beskytte legitime patentrettslige eller konkurransemessige interesser eller av hensyn til et løpende forskningsarbeid. REK kan dessuten bestemme at sensitive opplysninger om et prosjekt ikke skal inngå i registeret eller ikke kunne kreves innsyn i, for et klart avgrenset tidsrom, dersom offentlighet kan skade vesentlige private eller offentlige interesser.

### 6.3 Begrensninger på deling av data

Her skal vi se nærmere på lovverk som legger begrensninger på deling av data på grunn av *innholdet* i dataene. Dette omfatter begrensninger som skal beskytte datasubjektet, herunder personvernregelverket og GDPR; begrensninger som skal beskytte andre interesser, for eksempel forretningshemmeligheter; og begrensninger av hensyn til nasjonal sikkerhet.

Begrensninger på deling av data som skyldes *rettigheter* til data, samling av data eller oppfinnelser som bygger på data, er redegjort for. Forbudet mot informasjonsdeling etter arbeidstakeroppfinnelserloven er også redegjort for ovenfor – dette knytter seg som nevnt til informasjon om en oppfinnelse, ikke enkeltdata som sådan.

#### 6.3.1 Begrensninger som skal beskytte datasubjektet

##### Personopplysningsloven

Personopplysningsloven gjennomfører EUs personvernforordning og regulerer behandlingen av personopplysninger. Loven gjelder behandling av personopplysninger også i forbindelse med vitenskapelig forskning. Der personopplysninger inngår i forskningsdata, vil personopplysningsloven altså stille krav til behandlingen av slike opplysninger. Ved offentliggjøring av forskningsresultater og forskningsdata hvor personopplysninger inngår, vil personopplysningsloven kunne sette begrensninger og krav.

Personvernforordningen har til formål å verne fysiske personer i forbindelse med behandling av personopplysninger, og sikrer deres rett til vern av personopplysninger, samtidig som fri utveksling av personopplysninger i EØS ikke skal begrenses eller forbys (se artikkel 1). Personopplysninger er enhver opplysning om en identifisert eller identifiserbar fysisk person («den registrerte»), jf. artikkel 4. En identifiserbar fysisk person er definert som en person som direkte eller indirekte kan identifiseres, særlig ved hjelp av en identifikator. En slik identifikator kan være navn, identifikasjonsnummer,

lokaliseringsopplysninger, nettidentifikator eller lignende I fortalepunkt 34 og 35 er det uttalelser om personopplysninger i form av genetiske opplysninger og opplysninger om helsetilstand.

Behandling av personopplysninger må oppfylle et av vilkårene i artikkel 6 nr. 1, herunder (a) at det foreligger (frivillig og informert) samtykke fra den registrerte, (b) at behandlingen er nødvendig for å oppfylle en avtale som den registrerte er part i, (e) at behandlingen er nødvendig for å utføre en oppgave i allmennhetens interesse, eller (f) behandlingen er nødvendig for formål knyttet til de berettigede interessene som forfølges av den behandlingsansvarlige eller en tredjepart.

Om forskning heter det i forordningen fortalepunkt 33 at det for formål knyttet til vitenskapelig forskning ofte ikke er «mulig fullt ut å identifisere formålet med behandlingen av personopplysninger på tidspunktet for innsamlingen av opplysningene. De registrerte bør derfor kunne gi sitt samtykke til visse områder innen vitenskapelig forskning når dette er i samsvar med anerkjente etiske standarder for vitenskapelig forskning». Personopplysningsloven § 8 bestemmer at personopplysninger kan behandles på grunnlag av personvernforordningen artikkel 6 nr. 1 bokstav e), dersom det er nødvendig for «formål knyttet til vitenskapelig eller historisk forskning eller statistiske formål».

Personopplysningsloven § 8 bestemmer som nevnt at personopplysninger kan behandles på grunnlag av personvernforordningen artikkel 6 nr. 1 bokstav e, dersom det er nødvendig for «formål knyttet til vitenskapelig eller historisk forskning eller statistiske formål». Det er uttalt i personvernforordningen fortalepunkt 159 at «I denne forordning bør behandling av personopplysninger i forbindelse med formål knyttet til vitenskapelig forskning tolkes vidt og f.eks. omfatte teknologisk utvikling og demonstrasjon, grunnleggende forskning, anvendt forskning og privatfinansiert forskning».<sup>21</sup> Behandlingen skal i slike tilfeller være omfattet av nødvendige garantier i samsvar med personvernforordningen artikkel 89 nr. 1, det vil si nødvendige garantier for å ivareta den registrertes rettigheter og friheter. Det følger også av fortalepunkt 159 at «for å ta hensyn til de særlige forholdene som gjelder ved behandling av personopplysninger i forbindelse med vitenskapelig forskning, bør særlige vilkår få anvendelse, særlig med hensyn til publisering eller annen utlevering av personopplysninger i forbindelse med nevnte formål. Dersom resultatet av den vitenskapelige forskningen, særlig på området helse, berettiger ytterligere tiltak i den registrertes interesse, bør de allmenne bestemmelsene i denne forordning få anvendelse med henblikk på nevnte tiltak».

Det kan nevnes at Datatilsynet gir nærmere informasjon på sine sider om behandling av personopplysninger i forbindelse med forskning: <https://www.datatilsynet.no/regelverk-og-verktoy/sporsmal-svar/forskning/jeg-skal-bruke-personopplysninger-i-forskningen-min.-hva-skal-jeg-gjore/>. Datatilsynet understreker at det er den behandlingsansvarlige (for eksempel en forskningsinstitusjon) som må vurdere og dokumentere at den planlagte behandlingen av personopplysninger oppfyller kravene i loven. Kravet omfatter all forskning hvor det brukes personopplysninger. Det gjøres ikke unntak fra kravet selv om forskningsprosjektet må ha etisk godkjenning fra REK, og deres vedtak gir ikke rettslig grunnlag for behandling av personopplysninger. Datatilsynet framhever også at det i mange tilfeller kreves dispensasjon fra taushetsplikt for behandling av personopplysninger, og at REK vurderer om det kan gis dispensasjon fra taushetsplikt i pasientjournaler. Personvernrådet (EDPB) har dessuten vedtatt retningslinjer om behandling av

---

<sup>21</sup> Norsk Lovkommentar til personopplysningsloven, note 8-7.



helsesdata til forskningsformål og bruk av lokasjonsdata og kontaktsporingsverktøy i forbindelse med COVID-19 pandemien: <https://www.datatilsynet.no/regelverk-og-verktoy/internasjonalt/retningslinjer-og-uttalelser-fra-personvernradet/nye-retningslinjer-fra-personvernradet-om-helsesdata-og-lokasjonsdata/>.

#### Regler om taushetsplikt

Regler om taushetsplikt finnes blant annet i forvaltningsloven §§ 13 flg. Forvaltningsloven § 13 pålegger offentlig ansatte taushetsplikt om opplysninger som gjelder (1) noens personlige forhold, og (2) tekniske innretninger og fremgangsmåter samt drifts- eller forretningsforhold som det vil være av konkurransemessig betydning å hemmeligholde av hensyn til den som opplysningen angår. § 13 e pålegger forskere taushetsplikt om visse opplysninger, blant annet opplysninger undergitt taushetsplikt og opplysninger som er mottatt fra private under taushetsløfte i forbindelse med forskningsarbeidet. Taushetsplikten etter fvl. § 13 er ikke til hinder for at opplysninger gjøres kjent dersom den som har krav på taushet samtykker, jf. fvl. § 13 a nr. 1. Taushetsplikten er heller ikke til hinder for at opplysningene brukes når behovet for beskyttelse ivaretas ved at de gis i statistisk form eller ved at individualiserende kjennetegn utelates på annen måte, jf. § 13 a nr. 2.

#### 6.4 Begrensninger som skal beskytte andre interesser

##### Lov om vern av forretningshemmeligheter

Lov om vern av forretningshemmeligheter skal sikre innehavere av forretningshemmeligheter vern mot urettmessig tilegnelse, bruk og formidling av hemmeligheten, jf. forretningshemmelighetsloven § 1. «Forretningshemmeligheter» er definert i lovens § 2, og omfatter opplysninger som (a) er «hemmelige», det vil si at de ikke er allment kjent eller lett tilgjengelig, (b) har «kommersiell verdi fordi de er hemmelige», og (c) innehaveren har «truffet rimelige tiltak for å holde hemmelige». Etter forarbeidene har opplysninger kommersiell verdi der inngrep vil kunne skade interessene til innehaveren av forretningshemmeligheten «på en måte som undergraver personens vitenskapelige og tekniske potensiale, forretningsmessige eller finansielle interesser, strategiske plasseringer eller evne til å konkurrere».<sup>22</sup> Dersom man i innhenting av forskningsdata eller i publisering av forskningsresultater gjør inngrep i en forretningshemmelighet, kan dette være i strid med forretningshemmelighetsloven og resultere i sanksjoner etter lovens §§ 5 til 10.

##### Sikkerhetsloven

Sikkerhetsloven skal bidra til å trygge nasjonale sikkerhetsinteresser og forebygge sikkerhetstruende virksomhet, jf. sikkerhetsloven § 1-1. Loven gjelder for statlige, fylkeskommunale og kommunale organer, jf. § 1-2. Nasjonale sikkerhetsinteresser omfatter overordnede sikkerhetspolitiske interesser knyttet til blant annet de øverste statsorganers virksomhet, forsvar, sikkerhet og beredskap, forholdet til andre stater, økonomisk stabilitet, samfunnets grunnleggende funksjonalitet og befolkningens grunnleggende sikkerhet, jf. § 1-5. ). I forbindelse med forskning som omfatter informasjon som kan være vernet etter sikkerhetsloven, kan bestemmelsene i sikkerhetsloven sette begrensninger i tilknytning til innhenting og publisering av forskningsdata og forskningsresultater.

---

<sup>22</sup> Prop. 5 LS (2019-2020), punkt 5.1.3.

## 6.5 Om eierskap til forskningsdata – kan data eies?

Diskusjoner om deling av forskningsdata sirkler ofte rundt spørsmålet om hvem som “eier” dataene, se også stortingsmeldingen [Data som ressurs – Datadrevet økonomi og innovasjon \(2021\)](#) . Dette spørsmålet er det vanskelig å gi noe klart svar på. Som juridisk begrep, er “eier” den som har eiendomsrett til en fysisk ting, for eksempel en bil. Eiendomsretten er eksklusiv, slik at andre ikke kan bruke eller råde over bilen uten eierens samtykke. Data, derimot, er informasjon. Informasjon kan brukes av mange samtidig, og det gjør ofte informasjonen mer verdifull. Rettsreglene legger derfor opp til at data eller informasjon i størst mulig grad skal være fritt tilgjengelig og legger kun opp til eksklusive rettigheter til data der det er særlig behov for det. Eksempler er reglene om opphavsrett, databasevern og bedriftshemmeligheter, som på strenge vilkår gir en tidsbegrenset eksklusiv rett til å utnytte informasjonen. Formålet med reglene er å oppmuntre til innovasjon.

Diskusjonen om “eierskap” til data kommuniserer imidlertid en forventning om en rett til å råde over og kontrollere bruken av dataene. For forskningsdata, er dette knyttet til forskerens oppfatning om “eierskap” til forskningen sin, institusjonens oppfatning om arbeidsgivers rettigheter til ansattes arbeidsresultater, og i tillegg eventuelle datasubjekters oppfatning av en rett til å kontrollere videre bruk av personopplysninger. Disse forventningene motsvares bare delvis av juridiske rettigheter. Forholdet mellom forskeren og institusjonen som arbeidsgiver må løses etter arbeidsrettslige og immaterialrettslige regler. Dersom dataene eller databasen er vernet av opphavsrett, oppstår opphavsretten hos forskeren personlig, mens databaserettigheter og rett til bedriftshemmeligheter oppstår hos institusjonen. For data som ikke er vernet av noen regler, er rådende oppfatning at det er institusjonen som har råderett over dataene. Når det gjelder datasubjektenes forventninger, ivaretas disse av personopplysningsloven, eventuelt av andre lover.

Det kan altså være et visst gap mellom forventninger om “eierskap” og de rettigheter som følger av rettsreglene. Det medfører usikkerhet om hvem som har rett til – og ansvar for – å ta stilling til om og hvordan forskningsdata kan deles. Denne usikkerheten avhjelpes ved at interne retningslinjer og styringsdokumenter på institusjonene fastslår at data som ikke er underlagt opphavsrett “eies” eller “tilhører” institusjonen. I den grad man bruker begrepet “eie”, videreføres det en viss usikkerhet til hva denne posisjonen juridisk sett innebærer.

## 6.6 Utvalgets vurdering av det juridiske rammeverket

Deling av forskningsdata skjer innenfor det komplekse regelverket som er skissert ovenfor og videre beskrevet i vedlegg 1. Utvalget oppfatter at intensjonen og retningen i lovverket er tydelig: data skal i minst mulig grad være underlagt eksklusive rettigheter, og det skal legges til rette for størst mulig åpenhet. Den største utfordringen for forskere og institusjoner som skal dele forskningsdata, er, slik utvalget ser det, at regelverket er fragmentarisk. Dette fører til at det kan være mange ulike spørsmål som må besvares før en kan enten kreve eller sørge for at data kan deles og deles på en lovlig måte. Kort oppsummert så må den som skal ta stilling til om og hvordan et datasett skal deles, vurdere om det er eksklusive rettigheter knyttet til dataene (som opphavsrett eller databaserett), hvem som eventuelt innehar slike rettigheter, om dataene er av en slik art at det gjelder begrensninger for delingen, og om tilstrekkelige samtykker i så fall er innhentet. Alle disse vurderingene krever kjennskap til lovreglene på området.

Den juridiske kompetansen hos den som gjør vurderingen må være bred nok til å dekke mange rettsområder og dyp nok til å kunne gjøre de vurderingene som er nødvendige for å anvende reglene riktig. Slik kompetanse kan vanskelig opparbeides hos den enkelte forsker, men må foreligge hos institusjonen. Etter utvalgets oppfatning medfører det fragmentariske regelverket et særlig behov for at institusjonene sørger for å ha tilstrekkelig kompetanse til å dele data, og at de sørger for et system for deling som gjør det enkelt for forskerne å dele data på en trygg og lovlig måte. Systemene må tilpasses at ulike fagdisipliner produserer data av svært ulik karakter, som derfor kan være underlagt nokså ulike regler for om og hvordan dataene kan deles.

Selv om det hadde vært praktisk med et enkelt og enhetlig regelverk som regulerer forskningsdata og deling av forskningsdata, er det utfordringer med å etablere det. Siden data er så heterogene, vil et regelverk for forskningsdata også bli fragmentarisk sett opp mot regelverk for andre data. Det stiller også krav til samordning av regelverk på tvers av sektorer. Utvalget oppfatter derfor at det antagelig ikke er hensiktsmessig å lovregulere forskningsdata særskilt i dag.

Utvalget vil framheve at løsningene som følger av dagens regulering, legger til rette for nokså utstrakt deling av forskningsdata. Lovreglene setter noen grenser for deling for å beskytte andre interesser, som personvern, forretningshemmeligheter og sikkerhet, og særlig for persondata settes det også grenser av etiske hensyn. For regler som retter seg mot spesielle typer data, som personvernreglene, er en vurdering av dataenes karakter (er det tale om personopplysninger?) avgjørende for om reglene får anvendelse. Det er et begrenset og i utgangspunktet oversiktlig rettsanvendelsesspørsmål. For regler som beskytter interesser knyttet til dataene som er uavhengig av dataenes karakter, som for eksempel reglene om forretningshemmeligheter, kan vurderingen av om reglene får anvendelse, være noe mer kompleks.

Den største usikkerheten knyttet til lovverket synes å være knyttet til hvorvidt ulike involverte har noe "eierskap" til dataene. Som forklart over er det en diskrepans mellom forventningen om "eierskap" og lovverkets forutsetning om at data, i betydningen informasjon, som utgangspunkt ikke skal være underlagt eksklusive rettigheter. I forholdet mellom forskeren og institusjonen innebærer dette en viss usikkerhet om hvem som har rett til, og eventuelt ansvar for, å gjøre forskningsdata tilgjengelige. I forholdet mellom forskeren og datasubjektene, for eksempel informanter i et forskningsprosjekt, innebærer dette at det må avklares at det er gitt tilstrekkelig samtykke til deling av eventuelle personopplysninger.

Når det gjelder andre involverte, det være seg andre institusjoner, finansører, private eller offentlige partnere i et forskningsprosjekt, kan det være ulike forventninger om en form for kontroll med videre bruk av dataene. En forventning under stikkordet "eierskap" er ikke nødvendigvis til en eksklusiv råderett, men kan være en forventning om rett til å motsette seg videre deling av "sine" data, eller en rett til en form for avkastning dersom dataene gir grunnlag for økonomisk gevinst hos et videre ledd. Slike forventninger kan skape usikkerhet for den enkelte forsker om data kan deles, om det er nødvendig med begrensninger og hvem som eventuelt må samtykke til deling. Usikkerhet knyttet til slike forventninger kan dermed indirekte motvirke at forskningsdata deles. Denne usikkerheten kan ikke løses gjennom lisenser, fordi det at data deles på lisens nettopp forutsetter at den som deler, har tilstrekkelige rettigheter til dataene til å dele dem på de vilkårene lisensen setter. 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.

## 6.7 Utvalgets anbefalinger om juridisk rammeverk for deling av forskningsdata

### Utvalget anbefaler

- at institusjonene sørger for at forskningsstøttefunksjonen også inkluderer ressurser for å bistå med deling av forskningsdata, og har juridisk kompetanse til å legge til rette for lovlig og hensiktsmessig deling, og at det opparbeides relevant kompetanse for aktuelle fagdisipliner
- at myndigheter økonomisk legger til rette for at institusjonene kan tilby gode støttetjenester til forskere og for at denne kompetansen kan utvikles
- At myndigheter utreder rettighetsforhold til og i data, for å avklare forventninger om "eierskap"
- at institusjonene tydeliggjør fordelingen av rettigheter til forskningsdata mellom forskeren og institusjonen, særlig med sikte på deling av forskningsdata
- At institusjonene utarbeider klare rutiner for innhenting av tilstrekkelige samtykker fra datasubjekter, for eksempel informanter i et forskningsprosjekt, hvor dette er påkrevet
- at institusjonene utarbeider klare og tydelige retningslinjer for om, hvordan og når forskningsdata skal deles
- at det utredes og utarbeides klare rammeverk for datahåndtering i samarbeidsprosjekter som involverer private partnere i tillegg til offentlig finansierte institusjoner

## 7. Om lisenser og frigivelseserklæringer for deling og gjenbruk av forskningsdata

Formålet med dette kapittelet er å gi en oversikt over lisenser og frigivelseserklæringer for deling og bruk av forskningsdata. Forskningsdata, slik det er brukt av utvalget i dette kapittelet, inkluderer ikke forskningsrelaterte objekter slik som kildekode, programvarer, modeller og metodeprotokoller m.m. Denne avgrensningen er gjort fordi utvalgsarbeidet er begrenset i tid og omfang, men rapporten vil kort adressere lisensiering av forskningsrelaterte objekter under delkapittel [3.7.5](#).

### 7.1 Om lisenser og deres funksjon

Når forskningsdata gjøres tilgjengelige for andre, er det vanlig at bruken av dataene knyttes til en lisens. En lisens betyr tillatelse. Lisensen som knyttes til dataene er en avtale mellom den som har gjort dataene tilgjengelige og den som gjør bruk av dataene. Lisensen regulerer eventuelle vilkår for bruken av dataene, som at de kun skal brukes til forskning eller at den som har produsert datasettet, skal navngis. En lisens er nødvendig dersom de dataene som skal deles er undergitt eksklusive rettigheter i lov, for eksempel opphavsrett eller at bruken av dataene er begrenset av tidligere kontrakter, som samtykkebegrensninger. Lisenser brukes også i dag som et virkemiddel for å forsikre seg om at brukeren av dataene tilfredsstillers forskningsetisk "beste praksis". Dersom dataene som skal deles verken er beskyttet i lov eller underlagt lovmessige eller andre delingsbegrensninger (som kontrakter eller samtykke), er det ikke nødvendig med en lisens for at bruken av dataene skal være lovlig. Dersom den som gjør dataene tilgjengelig ønsker å sette begrensninger på bruken av dataene, for eksempel at de bare skal brukes i forskning eller at det skal oppgis hvor dataene kommer fra, er en lisens nødvendig. Selv om man ikke ønsker å begrense bruken av dataene, er det likevel ryddig, praktisk og anbefalt å knytte en lisens til bruken av dem for å klargjøre at forskeren og institusjonen gir avkall på eventuelle rettigheter til dataene.

En lisens er en kontrakt. Det vil si at den gir grunnlag for og angir innholdet i rettigheter og plikter mellom den som deler dataene og den som vil bruke dem. Begrensninger på bruk av forskningsdata som

følger av lovregler som har som formål å verne andre interesser, slik som personopplysningsloven, sikkerhetsloven, patentloven og helseregisterloven, kan ikke endres eller reguleres i en lisens. Det samme gjelder stort sett tredjeparters rettigheter til dataene: Disse må respekteres og ivaretas i en lisens. Det innebærer at dersom dataene er samlet inn basert på et samtykke som er begrenset, må denne begrensningen respekteres og videreføres i en lisens dersom dataene skal deles videre.

En lisens helt uten bruksbegrensninger er strengt tatt ikke en lisens, men heller som en frigivelseserklæring (til allemannseie), kalt *Public Domain* på engelsk. Det fins i hovedsak to typer frigivelseserklæringer: Public Domain Mark (PDM) og CC0. Der hvor PDM brukes for å dokumentere at det ikke er noen kjente rettigheter tilknyttet dataene, brukes CC0 for å vise at ingen opphavsrettslige rettigheter er forbeholdt, og forsker gir dermed avkall på alle eventuelle rettigheter. En lisens tjener altså flere formål: Den kan være nødvendig for å gjøre bruk av dataene lovlig, den kan sette vilkår for bruken og den kan dokumentere avkall på rettigheter.

Basert på om man bruker en frigivelseserklæring eller en lisensvariant kan graden av åpenhet variere fra data i det fri (f.eks. PDM eller CC0) som ikke har bruksbegrensninger til lisenser med vilkår knyttet til navngivelse, bearbeidelse, deling på samme vilkår og kommersialisering (f.eks. CC 4.0-BY). Dette går nærmere inn på i delkapittel [7.3](#).

## 7.2 Om forutsetninger og vurderinger for deling og gjenbruk av forskningsdata

Det er ønskelig at offentlige finansierte forskningsdata skal deles åpent med så få tilgangs- og formålsbegrensninger som mulig. Deling og gjenbruk må i imidlertid skje innenfor et komplekst juridisk regelverk og forskningsetiske normer som kan komplisere deling og påvirke hvilke data som skal deles og i så fall når. I det følgende beskrives ulike forutsetninger og vurderinger som bør ligge til grunn for valg av lisens eller frigivelseserklæring på forskningsdata.

### 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. Lisensgiver har ansvar for å avgjøre hvilken lisens som skal benyttes og tildele den. Før man tildeler en lisens, er det viktig for lisensgiver å ha avklart om og eventuelt i hvilket omfang forskningsdataene er gjenstand for juridiske rettigheter som opphavsrett og databasevern, og om det foreligger rettigheter hos en tredjepart. En lisensgiver trenger ikke nødvendigvis være rettighetshaver til forskningsdataene, der rettigheter foreligger av lover eller avtaler, men da må lisensgiver ha tillatelse fra rettighetshaver til å lisensiere.

I tillegg må den som gjør dataene tilgjengelig forsikre seg om at dataene ikke er underlagt rettslige delingsbegrensninger som for eksempel etter personopplysningsloven eller av sikkerhetshensyn. Det bør også vurderes om institusjonen eller forskeren selv vil komme til å bruke dataene på en måte som kan forstyrres av at dataene deles, for eksempel dersom dataene inngår i en oppfinnelse som skal patenteres eller de av andre grunner er å anse som forretningshemmeligheter.

Skillet mellom kildedata og data som er generert i et forskningsprosjekt, er vesentlig ved spørsmål om rettigheter til data og bruk av lisenser. [OpenAire](#) har en utfyllende guide for bruk av andre sine forskningsdata. Forskere bruker ofte kildedata til videre forskning og må da forholde seg til lisenser og avtaler som regulerer bruken av disse. I mange tilfeller vil forskning innebære en kombinasjon av bruk av kildedata og data generert i prosjektene. Det kan være med på å komplisere vurderingen av videre deling ytterligere, der lisenser brukt på kildedata ikke samhandler eller er kompatible med lisenser man

ønsker å bruke på avledete data, såkalte dataderivater. Dette diskuteres nærmere under [delkapittel 7.3.3](#).

I tillegg til å vurdere hvilke data som skal deles, må en lisensgiver ta stilling til når disse dataene skal deles. Dette er et komplekst spørsmål, uten en klar fasit. Forskningsdata som ligger til grunn for en vitenskapelig publikasjon, slik som en fagfelle-vurdert vitenskapelig artikkel, bør som hovedregel deles senest ved publiseringstidspunktet. Med forskningsdata menes her et minimum av data som trengs for å validere og bygge videre på konklusjonene i publikasjonen. Bruk av embargo på forskningsdata som ligger til grunn for vitenskapelige publikasjoner, bør derfor unngås, fordi dette svekker forskningens integritet. Flere vitenskapelige tidsskrifter har deling av forskningsdata på publiseringstidspunkt, samt forskningsrelevante objekter, som ligger til grunn for en vitenskapelig publikasjon som prinsipp.

Forskningsdata som ligger til grunn for en patentsøknad om rettigheter til en oppfinnelse, kan ikke være allment tilgjengelig forut for offentliggjøring av søknaden. Hovedformålet med patentretten er å skape insentiver for videre forskning og utvikling. Informasjon som ligger til grunn for en oppfinnelse det søkes patentrettigheter på, blir offentliggjort med patentsøknaden. I bytte mot å dele informasjon får rettighetshaver eneretten til å utnytte oppfinnelsen og skal derfor ikke hindre at forskningsdata blir delt, i den grad de ligger til grunn for oppfinnelsen. UH-loven §1-5 snakker om faglig frihet og ansvar, og presiserer forholdet mellom institusjon og ansatte i stillinger hvor forskning eller faglig eller kunstnerisk utviklingsarbeid inngår i arbeidsoppgavene, når det gjelder dette. Den som er ansatt i stilling hvor forskning eller utviklingsarbeid inngår i arbeidsoppgavene, har *rett til å offentliggjøre* sine resultater og *skal sørge for* at slik offentliggjøring skjer. Institusjoner kan imidlertid ikke pålegge offentliggjøring av resultater på et bestemt tidspunkt eller med en bestemt hyppighet<sup>23</sup>, men loven sier at det ikke kan "avtales eller fastsettes varige begrensninger i retten til å offentliggjøre resultater utover det som følger av lov eller i medhold av lov."<sup>24</sup>

Hva som utgjør legitime, berettigede og ikke-legitime hensyn for bruk av embargo på forskningsdata, er ikke alltid tydelig og er et komplekst tema som blant annet knytter seg til en forventning om "eierskap" til forskningsdata. Usikkerhet knyttet til forventninger om "eierskap" og en forventning om rett til å motsette seg videre deling av "sine" data, eller en rett til en form for avkastning dersom dataene gir grunnlag for økonomisk gevinst hos et videre ledd, kan ikke løses ved å bruke en lisens. Utvalget har derfor i sine [anbefalinger om juridisk rammeverk for deling av forskningsdata](#) skrevet at spørsmål knyttet til "eierskap" bør avklares i størst mulig grad.

Endelig må det vurderes om det er etiske problemstillinger knyttet til deling av dataene. Etiske vurderinger er en iboende del av et forskningsprosjekt, fra planlegging til prosjektslutt. Innenfor EU har RESPECT-prosjektet utarbeidet [profesjonelle og etiske retningslinjer](#) for sosioøkonomisk forskning basert på tre underliggende prinsipper: opprettholdelse av vitenskapelige standarder, overholdelse av juridiske lovverk og unngåelse av sosial og personlig skade. I tillegg har EU utviklet dokumentet [The European Code of Conduct for Research Integrity](#) (2017) som baserer seg på prinsipper om reliabilitet, ærlighet, respekt og ansvarlighet. Dette er overordnede prinsipper som skal ligge til grunn for gjennomføringen av forskningsprosjekter, sammen med [nasjonale forskningsetiske retningslinjer](#), og som dermed også skal ligge til grunn for valg av lisens.

---

<sup>23</sup> Lovkommentar til universitets- og høyskoleloven, Rettsdata, ved Jan Fridthjof Bernt, note 67.

<sup>24</sup> <https://lovdata.no/lov/2005-04-01-15/§1-5>

### 7.2.2 Lisenstakers ansvar – bruk og utnyttelse av forskningsdata i henhold til lisenskontrakt og forskningsetiske normer

En lisenstaker er i dette tilfellet en hvilken som helst person som i egenskap av lisensen får tillatelse til å utnytte forskningsdata, slik lisenskontrakten stipulerer. Dette innebærer at lisenstaker må forstå vilkårene og tillatelsene som ligger til grunn for å utnytte dataene og om dette er forenelig med tiltenkt bruk. I noen tilfeller kan det lisensierte datamaterialet være gjenstand for ulike lisenser og dermed ulike vilkår, og det er viktig at lisenstaker er oppmerksom på dette. En database vil for eksempel ha en type lisens som vil gjelde for selve sammenstillingen og struktureringen av datasamlingen, mens innholdet i databasen kan være lisensiert med andre vilkår og tillatelser.

Noen lisenser har tre nivåer: en menneskelesbar kunngjøring, en maskinlesbar kode og en juridisk kode, hvor den juridiske koden er det primære nivået av lisensen som beskriver det juridiske grunnlaget for å utnytte det lisensierte materialet. Dette nivået av lisensen er viktig å sette seg inn i for en lisensbruker og vil kreve juridisk kompetanse. Det er også nødvendig å merke seg om lisensen er ment for internasjonal bruk, uavhengig av jurisdiksjon, slik som siste versjon av CC-lisensene (4.0), eller om lisensen er en såkalt "ported" lisens, hvor de juridiske vilkårene i lisensen er tilpasset, for eksempel opphavsrett, i ulike jurisdiksjoner. Dette kan få konsekvenser for den tiltenkte bruken og utnyttelsen av de lisensierte forskningsdataene, ved at lisensen blir tildelt i en jurisdiksjon under visse juridiske vilkår, mens brukeren forholder seg til en annen jurisdiksjon hvor de samme vilkårene kanskje ikke gjelder.

I tillegg til juridisk forståelse av vilkår og tillatelser i lisensen forplikter lisensbruker seg til å bruke og utnytte forskningsdata i henhold til forskningsetiske normer. Forskningsetikkloven, som er utdypet over, skiller ikke mellom forskning i offentlig og privat regi og gjelder for forskere og forskning i Norge. Krav om å følge anerkjente forskningsetiske normer gjelder både forsker og institusjon. Redelighet og god henvisningsskikk er en viktig del av forskningsintegriteten. Selv om ikke lisensen eksplisitt setter vilkår om navngivelse, skal lisensbruker, i kraft av å bruke det lisensierte materialet i forskningsøyemed, sørge for å etterleve forskningsetiske normer og fagspesifikke standarder for god henvisningsskikk. [Se kapittel 7.3.4](#)

Det er dessuten viktig å være klar over at selv om forskningsdata er utstyrt med en lisens eller en frigivelseserklæring, så forteller dette ikke noe om kvalitet og er ikke ensbetydende med at dataene er pålitelige, gyldige eller gjenbrukbare. God datakvalitet er et flerdimensjonalt begrep og kan inkludere hvor gjenbrukbare dataene er, hvor pålitelig og gyldige de er, om det er brukt åpne standardformater for metadata og data som er maskinlesbare med mer. Datakvalitet kan ivaretas ved for eksempel bruk av FAIR sertifiserte arkiv (repository), slik [European Open Science Cloud planlegger for](#), for datakuratering og dokumentasjon av tilblivelsesprosessen for dataene (data proveniens).

### 7.3 Om vilkår og tillatelser i lisenser og Public Domain frigivelseserklæringer

Begrepet "åpen kunnskap" er definert av [Open Knowledge Foundation](#): "Knowledge is open if anyone is free to access, use, modify, and share it – subject, at most, to measures that preserve provenance and openness." Videre omtales "open work", tolket her som åpent forskningsarbeid, hvor et av prinsippene for at forskningsarbeid kan kalles åpent er bruk av åpne lisenser eller frigivelseserklæringer.

#### 7.3.1 Åpne standardlisenser

Lisenser som stipulerer [vilkår og tillatelser](#) som er i overensstemmelse med åpen kunnskap, slik det er definert over, kalles åpne standardlisenser. Open Knowledge Foundation har utarbeidet en [oversikt over](#)

[slike lisenser](#). I ytterste konsekvens betyr åpne forskningsdata at de deles uten begrensinger på hvem som kan bruke dataene (tilgangsbegrensninger) eller hva dataene skal brukes til (formålsbegrensninger).

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. Når man produserer forskningsdata man ønsker å dele, bør lisensgiver eksplisitt oppgi om lisensbruker kan remikse, endre, tilpasse og legge til data og under hvilke vilkår. Det finnes lisenser som er mye brukt på internasjonal basis og som er betraktet som god praksis for lisensiering av forskningsdata.

Standardlisenser, slik som Creative Commons 4.0 (CC)-lisenser<sup>25</sup>, er utarbeidet med sikte på å regulere bruk av opphavsrettslig vernet materiale, slik som musikk, bilder, litterære tekster og vitenskapelige artikler med mer, men brukes også mye for å lisensiere forskningsdata. To av CC-lisensene, CC-BY-4.0 og CC-BY-SA-4.0, står også oppført i oversikten over lisenser som er i overensstemmelse med åpen kunnskap, slik det er definert av Open Knowledge Foundation over. I den grad forskningsdata inneholder kunstneriske eller litterære verk som oppfyller vilkår om opphavsrett, vil CC-lisensene være et naturlig valg. Forskningsdata som inneholder rene fakta eller objektiv informasjon om verden er imidlertid ikke underlagt opphavsrett eller tilgrensende vern. Da blir bruken av CC-lisenser problematisk.

Disse lisensene er mye brukt blant forskere på grunn av vilkåret om attribusjon<sup>26</sup> 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 tilgrensende 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.

Noen lisenser har også "del på samme vilkår", såkalte copyleft- eller "Share Alike"-lisenser. Det gir lisensbruker mulighet til å remikse, endre og legge til data samt distribuere dem, men bare hvis man deler med samme tillatelse som de opprinnelige, lisensierte dataene. Lisenser med dette vilkåret blir også kalt virale lisenser<sup>27</sup>, da copyleft-lisenser legger restriksjoner på lisensbrukere om å lisensiere bearbeidet data, såkalte dataderivater, med en kompatibel copyleft-lisens med samme vilkår. Dette bidrar dermed til kontinuerlig spredning – videreføring – av lisensen. Dette er problematisk med hensyn til juridisk interoperabilitet og gjenbruk av lisensierte datasett. Hvis man ønsker å kombinere eller remikse to datasett hvor det ene er lisensiert med vilkår om attribusjon og "del på samme" vilkår, for eksempel CC BY-SA 4.0, og det andre med vilkår om navngivelse, "del på samme" vilkår samt vilkår om ikke-kommersiell bruk, for eksempel CC BY-SA-NC 4.0, er ikke dette mulig fordi lisensbruker ikke vil ha mulighet til å dele sitt dataderivat på samme vilkår som de opprinnelige datasettene, da de ikke er like. Dette vil rapporten komme tilbake til i [kapittel 8](#) som omhandler konsekvenser av lisensvalg.

---

<sup>25</sup> Med CC-lisenser menes lisensene fra Creative Commons 4.0, totalt seks i antall, som alle inneholder vilkår om attribusjon. CC0 som omtales som et dedikasjonsverktøy i rapporten, og ikke en lisens, er ikke inkludert. Distinksjonen mellom lisens og dedikasjonsverktøy uten bruksbegrensninger brukes gjennomgående i rapporten.

<sup>26</sup> **Å tilskrive opphav.**

<sup>27</sup> [https://en.wikipedia.org/wiki/Viral\\_license](https://en.wikipedia.org/wiki/Viral_license)



Lisenser som inneholder vilkår om ikke-kommersiell bruk gir kun tillatelse til å bruke det lisensierte materialet for ikke-kommersielle formål. I tillegg inneholder noen lisenser vilkår om "ikke-derivater", der hvor lisensbruker har lov til å kopiere eller distribuere det lisensierte materialet, men ikke avlede data med utgangspunkt i opprinnelige datasett. Disse problemstillingene vil også utvalget komme tilbake til i [kapittel 8](#).

### 7.3.2 CC0 og Public Domain Mark – data i det fri

Opphavsrett og databasevern oppstår automatisk dersom vilkårene i åndsverkloven er oppfylt. Reglene som gjelder rettigheter til samlinger av data, er i mindre grad harmonisert internasjonalt enn reglene om opphavsrett til tradisjonelle åndsverk, i forskningsmessig sammenheng typisk bøker og artikler. Reglene om databasevern er harmonisert i EØS, men det er til dels store forskjeller til andre jurisdiksjoner, derunder USA. En oversikt over mangfoldet knyttet til opphavsrett innenfor ulike europeiske jurisdiksjoner er utarbeidet av [CESSDA](#).

Når rettighetshaver benytter seg av CC0 eller lignende erklæring om avkall på rettigheter, innebærer det at rettighetshaver frafaller retten til å håndheve opphavsrett eller databaserett, og dataene gjøres tilgjengelige i det offentlige rom, uten vilkår. Dataene kan dermed gjenbrukes, bearbeides og kommersialiseres fritt og slik CC0 er formulert, uten vilkår om at lisenstaker må kreditere opphavet til dataene. For datasamlinger som er vernet som åndsverk, følger det av norsk rett at enkelte rettigheter kan opphaveren ikke fraskrive seg. Det gjelder de rettighetene som kalles opphaverens "ideelle rettigheter", altså retten til navngivelse og respektretten. Åndsverkloven § 5 første ledd sier at "*Ved fremstilling av eksemplar av åndsverk og når verket gjøres tilgjengelig for allmennheten, har opphaveren krav på å bli navngitt slik som god skikk tilsier såfremt navngivelse er praktisk mulig*". Videre sier tredje ledd: "*Opphaveren kan ikke fraskrive seg retten etter første og andre ledd, med mindre det gjelder en bruk av verket som er avgrenset etter art og omfang*." 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. Utvalget vil komme tilbake til god henvisningsskikk for forskningsdata i delkapittel [delkapittel 7.3.4](#).

Der hvor data ikke er vernet av opphavsrett eller databaserett eller beskyttet i samsvar med andre lovbestemmelser, for eksempel regler om taushetsplikt, personvern eller regler om vern av forretningshemmeligheter, kan man bruke et "Public Domain Mark" for å vise at det ikke er knyttet noen kjente opphavsrettslige rettigheter til materialet og at dataene derfor er i det fri.

### 7.3.3 Juridisk interoperabilitet og forskningsdata

I rapporten [EOSC Interoperability framework](#) handler juridisk interoperabilitet om å sørge for at organisasjoner som opererer under ulike juridiske rammeverk, politikk og strategier, kan samarbeide. Juridisk interoperabilitet for forskningsdata handler således om det internasjonale forskningsnettets evne til å dele, få tilgang til og gjenbruke forskningsdata fra ulike kilder som faller inn under ulike jurisdiksjoner, politikk og strategier. RDA-CODATAs interessegruppe for juridisk interoperabilitet har listet opp seks overordnede prinsipper for juridisk interoperabilitet for forskningsdata i sine [retningslinjer](#) for implementering av dette: 1) fremme lovlig tilgang til og gjenbruk av forskningsdata, 2) fastslå rettigheter til og ansvar for forskningsdata, 3) balansere juridiske interesser, 4) oppgi rettigheter tydelig og transparent, 5) promotere harmonisering av rettigheter til forskningsdata og 6) oppgi riktig navngivelse og kreditering for forskningsdata.

Kompatible lisenser er lisenser man kan bruke for å dele kombinerte eller sammenstilte datasett lovlig. Disse bør i størst mulig grad benyttes for å lisensiere forskningsdata for i størst mulig grad fremme prinsipper om juridisk interoperabilitet for forskningsdata. Det innebærer å bruke lisenser som har så få tilgangs- og formålsbegrensninger som mulig, slik at man ikke får utilsiktet begrensning på derivater av datasett i framtiden. Spørsmål man må stille seg når man skal bruke andres lisensierte datasett er om den tiltenkte bruken er i tråd med lisensvilkår, hvilken lisens man selv må bruke for å dele avledet data og hvordan lisensen vil påvirke framtidige dataderivater. Et datasett-derivat, som er et avledet resultat av kombinasjonen av to eller flere datasett med ulike grader av lisensrestriksjoner, vil i utgangspunktet måtte lisensiere med den lisensen som har de mest restriktive vilkår og betingelser, etter prinsippet om minste felles multiplum. I ytterste konsekvens kan ikke datasett kombineres i det hele tatt, fordi lisensvilkårene for bruken av dataene er i konflikt. Dette kan illustreres med et tenkt eksempel, brukt tidligere i rapporten. Hvis man ønsker å kombinere eller remikse to datasett hvor det ene er lisensiert med vilkår om attribusjon og "del på samme" vilkår, for eksempel CC BY-SA 4.0, og det andre med vilkår om navngivelse, "del på samme" vilkår samt vilkår om ikke-kommersiell bruk, for eksempel CC BY-SA-NC 4.0), er ikke dette mulig fordi lisensbruker ikke vil ha mulighet til å dele sitt dataderivat på samme vilkår som de opprinnelige datasettene, da vilkårene ikke er like.

Det er vanskelig å avgjøre om lisenser er compatible eller ikke, uten tilstrekkelig juridisk forståelse og kompetanse. Creative Commons har laget en [liste](#) over lisenser som er compatible med CC BY-SA og CC BY NC-SA, men bruk av lisenser med "del på samme" vilkår er problematisert i det tenkte eksempelet over og nedenfor under [konsekvenser av lisensvalg](#). [OER Africa](#) og [OpenMinted](#), en e-infrastruktur for tekst- og datautvinning av vitenskapelig innhold, har hver for seg også utviklet matriser over compatible Creative Commons lisenser.

#### *7.3.4 God henvisningsskikk for forskningsdata*

God henvisningsskikk ved bruk av andres forskningsdata er lovbestemt i forskningsetikkloven gjennom forskerens aktsomhetsplikt (§ 4) og krav om at institusjoner skal sikre at forskning som utføres ved institusjonen skjer i henhold til anerkjente forskningsetiske normer. Dette kravet innebærer at institusjonen har ansvar for nødvendig opplæring i, og at alle som deltar i eller utfører forskning er kjent med, anerkjente forskningsetiske normer. I de [generelle forskningsetiske retningslinjer](#) står det presisert at "forskere skal følge god henvisningsskikk som sikrer krav til etterprøvbarehet og gir grunnlag for videre forskning". Navngivelseslisenser blir i praksis ofte brukt av forskere for å sikre vilkår om å bli kreditert. CC-lisenser, som alle har vilkår om navngivelse, er problematisk og forvirrende der det blir brukt på data som ikke oppfyller vilkår om opphavsrett eller databasevern. Da kan i prinsippet brukere av forskningsdataene se bort ifra dette vilkåret, uten å gjøre noe ulovlig, fordi navngivelse bare gjelder for data som oppfyller vilkår i åndsverkloven, ifølge lisensen. I den grad forskningsdata oppfyller vilkår om opphavsrett og databasevern kan man ikke etter norsk lov fraskrive seg ideelle rettigheter, deriblant navneretten. Derfor vil brukere av forskningsdata med opphavsrett eller databasevern bryte loven ved å ikke navngi kilder i sin forskning. Utvalget mener at navngivelseslisenser ikke bør brukes som et verktøy for å oppnå sitering, da dette er nedfelt i lovbestemmelser, både i åndsverkloven, for data som oppfyller vilkår om opphavsrett og tilgrensende vern spesielt og i forskningsetikkloven, for forskningsresultater generelt.

[The Dryad Digital Repository](#) tar til orde for å bruke frigivelseserklæringen CC0 på forskningsdata, selv om denne ikke inneholder vilkår om navngivelse. CC0 er laget for å redusere både tilsiktede og utilsiktede juridiske og tekniske hindringer for gjenbruk av forskningsdata. CC0 kan ikke endre den

juridiske statusen til dataene ved å oppheve vern eller beskyttelse som er gitt i lovbestemmelser, siden fakta og informasjon i seg selv ofte ikke har opphavsrett innenfor de fleste jurisdiksjoner. CCO tar derimot høyde for at man *kan* ha opphavsrett eller tilgrensende vern i jurisdiksjoner man ikke kjenner til og at man derfor frigrir eventuelle rettigheter knyttet til lovbestemmelser, i den grad loven tillater det, ved å bruke CCO. Bruk av CCO på forskningsdata fratrukker heller ikke lisensbrukere et forskningsetisk ansvar om god henvisningsskikk. Dette prinsippet er best ivaretatt gjennom normative og fagspesifikke retningslinjer og god praksis for kildebruk i forskning, heller enn gjennom bruk av lisensvilkår om navngivelse som må håndheves rettslig. En god praksis når datasett skal tilgjengeliggjøres er å henvise tydelig til normer og retningslinjer, såkalte "community norms", for kildehenvisning. Dataverse-prosjektet, en webapplikasjon basert på åpen kildekode for deling og gjenbruk av forskningsdata, har utviklet slike [normative retningslinjer](#) for datasitering samt en automatisk standardiseringsprosess for [sitering av datasett](#), som er tilgjengeliggjort i et Dataverse arkiv (repository). Fagområder og forskningsmiljøer bør etterstrebe å utvikle normative retningslinjer for god henvisningsskikk. Dette vil tydeliggjøre forventninger om god henvisningsskikk, både for de som tilgjengeliggjør datasett og de som er databrukere.

### 7.3.5 Lisensiering av forskningsrelaterte objekter

Utvalget har ikke fokusert på lisensiering av forskningsrelaterte objekter, som kildekode, algoritmer, modeller, metodeprotokoller og elektroniske laboratoriebøker, i denne rapporten. Allikevel utgjør dette i mange forskningsprosjekter en stor del av forskningsresultatene. God praksis for datahåndtering kan i mange tilfeller overføres til for eksempel kildekode, men skiller seg fra tilgjengeliggjøring av forskningsdata på noen områder. [Open Science Toolbox](#) viser til nyttige verktøy utviklet for publisering og versjonering av kildekode, slik som [GitHub](#), og gir en oversikt over de mest vanlige lisensene for publisering av kode utviklet i forskningsprosjekt. Å tilgjengeliggjøre forskningsrelaterte objekter under åpne lisenser ivaretar integriteten til forskningsresultatene blant annet ved å sikre etterprøvbarehet og gyldighet til analyser, modeller og metoder. [Open Source Initiative](#) har definert [ti prinsipper](#) som kildekode må oppfylle for å bli betraktet som åpen kildekode og har listet opp [lisenser](#) som er i tråd med prinsippene.

## 8. Konsekvenser av lisensvalg

Valg av lisens, når man deler og tilgjengeliggjør forskningsdata, kan få både tilsiktede og utilsiktede konsekvenser, både for lisensgiver og lisensbruker. Derfor er det viktig å ha en juridisk forståelse av lisenskontrakten og dens vilkår for bruk av forskningsdataene. Konsekvensene ved å ikke bruke lisens når man deler forskningsdata er usikkerhet rundt dataenes juridiske status og hvilke rettigheter som gjelder for enten hele eller deler av materialet som deles, og hvilke formålsbegrensninger dataene måtte ha. Dette vil bidra til å redusere gjenbruk, blant annet fordi databruker ikke ønsker å løpe noen juridisk risiko, eller ulovlig bruk av dataene fordi det ikke er tydelig hvilken jurisdiksjon dataene er delt under. Nedenfor vil utvalget diskutere noen potensielle konsekvenser ved valg av lisens.

### 8.1 Mer lukket enn nødvendig

Mangel på juridisk kompetanse og forståelse for det juridiske rammeverket dataene opererer i kan bidra til at forskningsdata begrenses mer enn nødvendig, ofte utilsiktet, og dermed redusere gjenbruk ved at andre ikke har lov til å bearbeide dataene og distribuere videre.

Bearbeidede data som er et resultat av kombinasjonen av to eller flere datasett med ulike grader av lisensrestriksjoner vil i utgangspunktet måtte lisensiere med den lisensen som har de mest restriktive

vilkår og betingelser, etter prinsippet om minste felles multiplum. I ytterste konsekvens kan ikke datasett kombineres i det hele tatt. Det er viktig at både den som tildeler lisens og lisenstaker er oppmerksom på bruk av kompatible lisenser, der det er mulig, slik at man ikke utilsiktet hindrer andre i å gjenbruke forskningsdata. Kombinasjon av data som omfatter en sammenstilling av forskjellige datasett som allikevel holdes separat og for seg selv, mellom et datasett som oppfyller vilkår for rettslige rettigheter og vern, og data som ikke gjør det, kan komplisere deling. En løsning på dette kan være å bruke differensierte lisenser for å unngå at hele datasettet må deles med den mest begrensende lisensen. Lisensen man tildeler dataene sine får altså konsekvenser for de som ønsker å gjenbruke dataene, og det er derfor anbefalt å bruke en lisens med så få vilkår og begrensninger for bruk, som mulig.

Deling av forskningsdata som inneholder sensitive data<sup>28</sup>, slik som personopplysninger, involverer både juridiske, tekniske og etiske hensyn og problemstillinger. Før man tildeler lisens på slike data må man avklare om det i hele tatt er lovlig å dele dataene slik de er, eller om man må for eksempel ha tilgangsbegrensninger eller dele data på et annet detaljnivå. Innenfor EØS er data som inneholder personopplysninger strengt regulert i [General Data Protection Regulation](#) (GDPR). Relevante norske lovbestemmelser som også kan begrense deling av forskningsdata er redegjort for i [DEL 2](#). Rent teknisk så er kravet til datasikkerhet og arkivering av data som inneholder, for eksempel personopplysninger, strengere enn ved arkivering av forskningsdata som regnes som ikke-sensitive. Etiske hensyn, slik som å hindre bruk og deling av informasjon som kan skade mennesker, dyr, miljø og samfunn, er også tungtveiende.

Krav om å dele data åpent og vern av data som inneholder personopplysninger kan virke motstridende. [CESSDA](#) og [ELIXIR](#) har begge utfyllende beskrivelser for håndtering av forskningsdata som inneholder sensitive data, som personopplysninger. [OpenAire](#) har også en guide for hvordan man kan håndtere data som inneholder personopplysninger. Hovedbudskapet er at det finnes måter man kan balansere kravet om åpenhet og krav om å verne data som inneholder sensitiv informasjon. Anonymisering av dataene er et slikt tiltak, men innebærer at forskningsdataene kan miste noe av sin verdi. Anonymisering innebærer at dataene er prosessert på en slik måte at det ikke er mulig å identifisere dem tilbake til enkeltpersoner, og er irreversibel. Kryptering, bruk av koblingsnøkkel eller tilleggsinformasjon for å reidentifisere data, som holdes separat og sikkert fra selve dataene, kalles pseudoanonymisering. GDPR skiller mellom anonymisering og pseudoanonymisering av data som inneholder personopplysninger, hvor anonymisering innebærer at GDPR ikke betrakter data som personopplysninger lengre, mens pseudoanonymiserte data fortsatt er det. Der hvor anonymisering av data som inneholder personopplysninger forringer dataenes verdi og gjør dem ubrukelige, er bruk av frivillig, informert samtykke der datasubjektene, informantene, samtykker til at dataene kan deles, en mulighet. Formålet med forskningsprosjektet må presenteres på en slik måte at det er helt tydelig for datasubjektet hvordan dataene skal prosesseres, hva dataene skal brukes til og hvordan de skal tilgjengeliggjøres. 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. CESSDA sine ressursider [for informert samtykke](#) beskriver eksempler på "granular consent", detaljert samtykke.

---

<sup>28</sup> Sensitive data kan være personopplysninger, helseopplysninger, forretningshemmeligheter, informasjon som begrenses av sikkerhetsloven med mer.

Dataplattformer med tilgangsbegrensninger og dedikert analyseinfrastruktur for å ivareta personvern og god informasjonssikkerhet er også en måte å dele sensitive forskningsdata på, om enn ikke helt åpent. Dette gir muligheter for å gjenbruke forskningsdata som inneholder sensitive data, på en kontrollert og sikker måte.

## 8.2 Navngivelsesstabling

Navngivelsesstabling, kalt "attribution stacking" på engelsk, er et teknisk, juridisk og etisk problem, uten en tydelig løsning. Utvalget ønsker her å løfte frem mulige problemstillinger knyttet til navngivelsesstabling.

Det kan være nyttig i denne sammenheng å skille mellom navngivelse, ofte brukt for å anerkjenne arbeidet som er gjort, og sitering, ofte brukt for å referere til dataene i seg selv, via en lenke eller en identifikator, slik som en "digital object identifier" (DOI). Sitering bidrar til å dokumentere dataenes tilblivelsesprosess, fremmer gjenbruk og kan også brukes som en nøkkeltallsindikator for å måle effekten og innflytelsen, "impact", dataene har.

Navngivelsesstabling kan oppstå når man ønsker å kombinere mange datasett fra ulike kilder, eller gjenbraker forskningsdata som baserer seg på mange kilder. Navngivelseslisenser stipulerer navngivelse som et vilkår for gjenbruk av forskningsdata. 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.

Uklare og ulike vilkår om navngivelse og sitering kan gå på bekostning av en strømlinjeformet datadelingsprosess og er spesielt krevende når man ønsker å kombinere mange datasett automatisk. Her kan det oppstå problemer med maskinlesbarhet og overholdelse av vilkår for bruk av dataene.

Navngivelsesstabling er ikke bare den enkelte forskers problem og utfordring. Fag- og forskningsmiljøer bør diskutere bruk av felles standarder og konvensjoner, som harmoniserer med internasjonale standarder, for sitt felt. Standarder bør balansere utfordringer ved navngivelsesstabling og forskningsetiske prinsipper om redelighet, etterrettelighet og god henvisningsskikk. Dette er spesielt viktig for fagfelt som ofte benytter seg av store mengder ulike datasett i forskningen og er avhengig av en automatisert prosess for å kombinere og gjøre nytte av ulike datakilder.

## 8.3 Åpen deling = ansvarsfraskrivelse?

En bekymring for mange vedrørende åpen deling av forskningsdata, og bruk av så få tilgangs- og formålsbegrensninger som mulig, er opplevelsen av å miste kontroll over brukertilgang og hva dataene brukes til. Å gå fra den mer tradisjonelle delingen med tilgangsbegrensning, ved at man for eksempel kontakter den eller de som har produsert forskningsdata personlig, til å dele forskningsdata så åpent som mulig, i dedikerte dataarkiv (repository), kan utfordre det etiske ansvaret en forsker har for "sine" data og en forventning om "eierskap". Det er derfor viktig at forskningsstøtteapparatet har tilstrekkelig, juridisk kompetanse til å bistå forskere med avklaring av juridiske problemstillinger slik at deling av data kan skje på en lovlig og ansvarlig måte, i en internasjonal kontekst. Rammeverket for etisk forsvarlig og åpen deling av forskningsdata bør utvikles i fagspesifikke miljøer, men henblikk på at åpen deling skjer i en internasjonal kontekst.

En annen bekymring knyttet til åpen deling er forvitring eller utvanning av dataene. Som utvalget har pekt på ovenfor bør åpen datadeling skje i tråd med god henvisningsskikk og kildehenvisning slik at det er mulig for brukere å finne tilbake til opprinnelige kildedata.

#### 8.4 Hinder for kommersialisering

Bruk av lisenser som inneholder vilkår om ikke-kommersiell bruk begrenser gjenbruk av data for kommersialiseringsformål. Hva som utgjør kommersielle og ikke-kommersielle formål kan i mange tilfeller være tydelig, men i noen tilfeller kan det være store gråsoner og vanskelig å skille. Et eksempel på gråsoner er vitenskapelige tidsskrifter som baserer sin finansieringsmodell på abonnement, det vil si at brukeren/leseren betaler for tilgang. I de tilfeller lisensgiver har tildelt forskningsdata en lisens som inneholder begrensning av databruk til kommersielle formål kan ikke forskningsdataene, strengt tatt, ligge til grunn for figurer og analyser som utgjør en del av publiseringen, i dette tilfellet.

Utvalget vil diskutere verdiskaping og innovasjon basert på forskningsdata nærmere i [kapittel 15](#).

### 9. Utvalgets oppsummering av bruk av lisenser på forskningsdata

\*kommer i endelig rapport

#### Nyttige ressursider

##### Åpen forskning

<http://opencscience.prototyp.io/>

##### Interoperabilitet

<https://op.europa.eu/en/publication-detail/-/publication/d787ea54-6a87-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-190308283>

[https://www.codata.org/uploads/Legal%20Interoperability%20Principles%20and%20Implementation%20Guidelines\\_Final2.pdf](https://www.codata.org/uploads/Legal%20Interoperability%20Principles%20and%20Implementation%20Guidelines_Final2.pdf)

##### Research Data Management verktøy

<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide>

<https://rdmkit.elixir-europe.org/index.html>

##### Navngivelse og sitering

<https://uc3.cdlib.org/2016/09/19/cc-by-and-data-not-always-a-good-fit/>

<https://www.library.ucdavis.edu/news/cc0-and-data-citation/>

<https://blog.datadryad.org/2011/10/05/why-does-dryad-use-cc0/>

<https://dataverse.org/best-practices/data-citation>

##### Lisensvelger

<https://ufal.github.io/public-license-selector/>

<https://chooser-beta.creativecommons.org/>

<https://choosealicense.com/>

<https://opensource.org/licenses/>

<https://docs.data.world/en/59261-59714-2--Common-license-types-for-datasets.html>

<https://help.data.world/hc/en-us/articles/115006114287-Common-license-types-for-datasets>

##### Oversikt over compatible lisenser

<https://openminded.github.io/releases/license-matrix/>

<https://www.oerafrica.org/FTPFolder/Website%20Materials/UnderstandingOER/licwiz/english.html>

## 10. Premisser for tilgjengeliggjøring av forskningsdata

Utvalget mener at det er en del premisser som må ligge til grunn for bruk av lisenser og frigivelseserklæringer av forskningsdata.

### 10.1 Tydelig ansvarsfordeling mellom den enkelte forsker og institusjon

Ansvarsfordelingen mellom institusjon og forsker når det gjelder åpenhet om forskningsresultater er tydelig og klar i UH-loven (§ 1-5(6)). Den enkelte forskeren har rett til å offentliggjøre sine resultater og en plikt til å sørge for at dette skjer. Universiteter og høyskoler har en plikt og et ansvar for å sørge for åpenhet om resultater fra forskning eller faglig eller kunstnerisk utviklingsarbeid. Utvalget mener at denne ansvarsfordeling bør gjelde for alle institusjoner og forskere, uavhengig av sektor, som har forskning som del av sin virksomhet eller sitt virke. Ansvarsfordelingen bør tydelig komme frem i institusjonelle retningslinjer for datahåndtering.

### 10.2 Juridisk kompetent forskningsstøttesystem

Åpenhet om forskningsresultater og lisensiering av forskningsdata skjer til dels i et fragmentarisk og uoversiktlig juridisk landskap. Det kan ikke forventes at den enkelte forsker skal ha tilstrekkelig juridisk kompetanse for å legge til rette for lovlig tilgang og gjenbruk av forskningsdata og valg av lisens eller frigivelseserklæring. Denne kompetansen bør institusjonen sørge for er tilgjengelig for forskere som skal dele dataene sine ved å tilrettelegge for forskningsstøttesystemer som forenkler datahåndtering. Dette innebærer blant annet å bistå med juridisk kompetanse, for eksempel ved avklaring av rettigheter til, og lovbestemt vern av, forskningsdata. Institusjonene bør også legge til rette for at forskningsrettede utdannelser har tilstrekkelig kunnskap om rettigheter og ansvar, knyttet til forskningsdata, som en del av utdannelsen. Forskningsstøttesystem som forenkler datahåndtering for den enkelte forsker krever ressurser og behovet vil være økende framover. Utvalget har, som en av anbefalingene om juridisk rammeverk for deling av forskningsdata, anbefalt at myndigheter økonomisk legger til rette for at institusjonene kan tilby gode støttetjenester til forskere og for at denne kompetansen kan utvikles.

### 10.3 Avklaring av rettigheter til, og delingsbegrensninger for, forskningsdata

Før man deler forskningsdata må eventuelle rettigheter til, og lovmessige eller andre delingsbegrensninger, avklares og tydeliggjøres. Flere norske institusjoner har retningslinjer for datahåndtering eller IPR politikk, men det er forskjell i detaljeringsgrad og tydelighet. Institusjoner bør ha tydelige retningslinjer for avklaring av eventuelt hvem som har rettigheter til data produsert ved institusjonen og i hvilken grad rettighetene kan kontrolleres av andre, for eksempel ansatte ved institusjonen. Det må også være tydelig hvem som skal anerkjennes og krediteres ved gjenbruk av dataene i tilfeller der rettighetshaver er for eksempel en institusjon. Der forskningsdata som skal deles utgjør en kombinasjon av flere datasett fra ulike kilder bør rettighetsstatus til dataene være nøyaktig og transparent. Statusen bør si noe om 1) hvem som har eventuelle rettigheter til deler av datasettet, og 2) hvem som er rettighetsholder til hele datasettet. Tydelig og transparent avklaring av rettigheter til forskningsdata er i tråd med et av de overordnede prinsippene for juridisk interoperabilitet for forskningsdata, skrevet av [RDA-CODATA](#). Utredning av rettighetsforhold til og i forskningsdata, for å avklare forventninger til "eierskap", er en av utvalgets anbefalinger om juridisk rammeverk for deling av forskningsdata.

## 11. Utvalgets anbefaling

### UTVALGETS ANBEFALINGER OM BRUK AV LISENER OG FRIGIVELSESERKLÆRING PÅ FORSKNINGSDATA SOM SKAL PUBLISERES

1. Forskningsdata bør utstyres med vilkår og tillatelser for videre bruk, typisk en lisens eller frigivelseserklæring.
2. Forskningsdata bør utstyres med internasjonale standardlisenser som:
  - a. har så få tilgangs- og formålsbegrensninger som mulig
  - b. i størst mulig grad er kompatible med andre lisenser og fremmer prinsippene om juridisk interoperabilitet for forskningsdata
  - c. er menneske- og maskinlesbare
3. Metadata bør alltid utstyres med en frigivelseserklæring, typisk CCO eller PDM.
4. Bruk av navngivelseslisenser bør begrenses til de tilfeller hvor dataene oppfyller vilkår i åndsverkloven.
5. I de tilfeller der ulike rettslige vilkår og betingelser gjør seg gjeldende på ulike deler av datasett, bør det brukes differensierte lisenser etter anbefaling 2, for å unngå at alle dataene får begrensende lisens fordi noen av dataene krever det.
6. Øvrige begrensninger på tilgang og viderebruk av data bør bare settes i tilfeller hvor gjeldende lovverk eller etiske grunner tilsier det (f.eks. sikkerhetshensyn, vern av personopplysninger, patentering, forretningshemmeligheter og uakseptable etiske konsekvenser for mennesker, dyr, samfunn og miljø).
7. Institusjonene bør ha et overordnet ansvar for at forskningsdata som er produsert av ansatte gjennom sitt virke, blir lisensiert.

Bruk av andres data bør for øvrig alltid skje i tråd med forskningsetiske normer og fagspesifikke retningslinjer.



## DEL 3 – Ikke-juridiske barrierer for deling og gjenbruk av forskningsdata

I de foregående delene av denne rapporten har utvalget gjennomgått det juridiske rammeverket for deling og gjenbruk av forskningsdata, gitt en oversikt over ulike aspekter ved lisensiering av forskningsdata og kommet med sine overordnede anbefalinger om bruk av lisens. Vi tror at dette kan gjøre det lettere for forskere og institusjoner å ta veloverveide og hensiktsmessige valg når de skal lisensiere sine forskningsdata. Samtidig er det utvalgets klare oppfatning at det er mange andre faktorer som spiller inn, som i minst like stor grad vil påvirke hvorvidt vi oppnår mer deling og gjenbruk i framtiden. I denne delen av rapporten vil vi derfor gjenoppta de politiske, strukturelle og kulturelle faktorene som ble beskrevet i utvalgets notat som ble overlevert til Kunnskapsdepartementet i oktober 2020. Disse faktorene er ikke direkte knyttet til juridiske forhold, men spiller en viktig rolle når forskere skal dele dataene sine for gjenbruk. Ved å finne gode løsninger som reduserer de ikke-juridiske barrierene for deling, i kombinasjon med at norske forskere får en bedre oversikt over det juridiske rammeverket, tror utvalget at dette vil bidra til en god delingskultur i det norske forskningssystemet.

I løpet av innspillperioden har vi mottatt mange nyttige og relevante innspill og kommentarer til utfordringene som ble diskutert i notatet, og disse innspillene er nå innarbeidet i denne delen av rapporten. Vi takker for det store engasjementet og tror at vi gjennom en åpen prosess har klart å fange opp ikke bare de mange utfordringene vi står overfor, men også gode og reflekterte tanker om hvordan utfordringene kan løses.

Avslutningsvis har utvalget formulert anbefalinger som kan bidra til å imøtekomme og løse utfordringene. Anbefalingene er basert på de mange gode diskusjoner utvalget har hatt, i kombinasjon med innspillene som har kommet underveis, og er rettet mot henholdsvis regjeringen og departementene, finansierer og virkemiddelapparatet og forskningsinstitusjonene. **[anbefalingene i denne versjonen er foreløpige, og enkelte av dem er ikke ferdig diskutert i utvalget – vi velger likevel å dele dem på nåværende tidspunkt for å kunne få så gode innspill som mulig]** Det er viktig for utvalget å påpeke at anbefalingene må sees som et skritt på veien innenfor et område som er i kontinuerlig utvikling, og er på ingen måte ment som en fasit. Dette er ikke slutten, men starten, på en prosess for å oppnå en mer åpen forskningssektor rustet for framtiden.

### 12. Koordinert innsats for mer deling og gjenbruk av forskningsdata

De siste to–tre årene har det blitt satt i gang en rekke initiativer i Norge som alle har samme formål – å få til mer deling og gjenbruk av data for å oppnå mer kunnskap, bedre tjenester og verdiskaping på tvers av fagområder og sektorer. Det er en gjengs oppfatning at dette vil bidra til å løse samfunnsutfordringene mer effektivt og gi Norge et godt utgangspunkt for å delta i den internasjonale dataøkonomien. Det er vel og bra at det satses på dette feltet, og utvalget mener det er en riktig prioritering i tiden vi er inne i. Det er likevel en utfordring at de mange initiativene ikke er godt nok koordinert fra øverste hold. Dette fører til delvis overlappende og delvis motstridende konklusjoner, som kan gjøre det vanskelig for de som er ansvarlige for datadelingen å navigere mellom alle gode råd og anbefalinger.

#### 12.1 Behov for harmonisering av datapolitikken ("policy alignment")

Det er utvalgets oppfatning at det er et stort behov for en harmonisering av datapolitikken, kalt "policy alignment" på engelsk. Det er viktig for at de som faktisk skal dele data for gjenbruk, har klare retningslinjer å forholde seg til. Hvis det gis delvis motstridende krav og forventninger fra ulike hold, kan

man i ytterste konsekvens risikere at den som skal dele data føler avmakt og velger å ikke forholde seg til noen av kravene. Et annet utfall kan være at man velger hvilke krav som skal innfris basert på egeninteresse, og ikke basert på hva som er intensjonen bak anbefalinger og krav.

Harmonisering av datapolitikk er nødvendig på tvers av sektorer, slik at deling mellom ulike forvaltningsområder og kunnskapssystemer kan gjennomføres. Potensielt kan ulike forvaltningsområder legge vekt på ulike deler av datadeling fordi de ikke er klar over behov som de andre forvaltningsområdene har. Utvalget mener at norske myndigheter bør koordinere politikken bedre på tvers av sektorer gjennom å legge til rette for gode samarbeidsarenaer for de som utarbeider policyer og retningslinjer innenfor de ulike forvaltningsområdene.

Det er også behov for harmonisering av datapolitikk på tvers av ulike nivåer innenfor et forvaltningsområde. For kunnskapssektoren innebærer dette at policyer og retningslinjer fra departementene, virkemiddelapparatet, tjenesteleverandørene, infrastrukturene og institusjonene tilpasses hverandre og trekker i samme retning. Hos store institusjoner vil det i tillegg være nødvendig å harmonisere retningslinjer internt mellom institutter, administrative avdelinger og fagområder. Det kan for eksempel forekomme at forskere opplever motstridende råd og krav fra for eksempel sin lokale instituttledelse og støttetjenester sentralt på institusjonen.

## 12.2 Internasjonal orientering

Forskning er internasjonal i sin natur. Mange norske forskere forholder seg i større grad til internasjonale samarbeidspartnere og finansierer enn til norske. Mange internasjonale forskerfelleskap har i en årrekke delt data på tvers av landegrenser, og norske forskere som søker midler fra EU, vil i tiden som kommer oppleve strengere krav om åpen forskning i sine prosjekter. Mange globale organisasjoner og sammenslutninger (deriblant OECD, UNESCO, ISC, WMO) har åpen forskning og deling av forskningsresultater på agendaen, og det vil være svært viktig framover at norske aktører er internasjonalt orientert i videre politikktutforming på datafeltet. Harmonisering mellom norske og internasjonale organer er med andre ord like viktig som mellom de nasjonale sektorene. Utvalget oppfordrer departementet til å prioritere norsk deltakelse i internasjonale fora hvor åpen forskning og deling og gjenbruk av forskningsdata diskuteres. Gjennom dette kan vi lære av andre for å bli bedre her hjemme, samtidig som vi kan bidra til at den internasjonale utviklingen går i en retning som passer med norske forhold og behov.

## 13. Incentivmekanismer for mer deling og gjenbruk av data

Publisering av forskningsresultater er en viktig del av den akademiske virksomheten, og en faktor som 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 academia, blant annet for å kunne disputere til doktorgrad eller få fast ansettelse. I tillegg gir tellekantsystemet, som er en del av grunnlaget for tildeling av midler til forskningsinstitusjoner, hovedsakelig uttelling for publiserte artikler, og ikke dataene som ligger til grunn for dem. Dette forsterker fokuset på vitenskapelige publikasjoner ytterligere. Kombinasjonen av at datasettene blir et verdifullt grunnlag for annen innbringende aktivitet og samtidig ikke gir uttelling i seg selv, kan med andre ord ha en kontraproduktiv effekt med tanke på deling av data, ved at de holdes tilbake lenger enn nødvendig. På denne måten kan krav og forventninger om deling av forskningsdata

bidra til verdimesseige konflikter mellom forsker, institusjon og samfunnets investering i og krav til offentlig finansierte data som et fellesgode.

Sett i lys av det ovenstående, kan det se ut til at dagens tellekantsystem ikke er tilpasset den nye forskningsstandarden hvor åpen deling av data og andre typer materiale og ressurser, er en klar forventning både fra finansiører og andre potensielle brukere av resultatene fra forskning. Utvalget mener at det er på tide å gjennomgå tellekantsystemet og vurdere om det bør justeres for å gjøre det åpne valget til det naturlige valget for forskningsmiljøene. I tillegg vil det være hensiktsmessig å gjennomgå resten av insentivsystemet i det norske forskningssystemet på en systematisk måte for å se hvor man kan gjøre justeringer og endringer.

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 eksterne midler til forskning og ved ansettelse og opprykk. DORA-erklæringen sier blant annet eksplisitt at også datasett og programvare skal være gjenstand for vurdering i slike tilfeller. Et kriterium for vurdering av datasettenes kvalitet, i tillegg til virkninger og effekter, kan i slike tilfeller være datasettenes vitenskapelige produktivitet (grad av gjenbruk og bidrag til nye forskningsresultater). Det er viktig at kriteriene er tydelige nok til å insentivere tilgjengeliggjøring, men samtidig fleksible nok til at de legger til rette for tverrfaglig forskning og annen forskning hvor data ikke er en naturlig "output". Den nylig publiserte NORCAM-rapporten fra UHR presenterer et nytt rammeverk for vurdering i akademiske karriereløp, som baserer seg på en mer fleksibel og helhetlig vurdering. Rammeverket legger opp til at det skal være mer åpenhet, større bredde i vurderingene og en mer helhetlig tilnærming enn tradisjonell vurdering hvor ensidig bruk av indikatorer står sentralt. Utvalget mener at dette rammeverket også kan ha en nytte i ulike typer for evaluering av forskningsmiljøer, fagområder og forskningsprosjekter.

Et sterkt insentiv for forskningsmiljøene er å motta ekstern finansiering for å gjennomføre forskningsprosjektene de har søkt om. Derfor kan kriterier som forskningsfinansiører setter for vurdering av prosjekter bli et kraftfullt virkemiddel som kan bidra til endring i praksis. I Horisont Europa har de nå innført åpen forskning-praksis som en del av vurderingsgrunnlaget ved tildeling av midler. I guiden til rammeprogrammet vises det til både obligatorisk og anbefalt åpen forskning-praksis, og hvordan søkerens beskrivelse av egen praksis vil påvirke hvor god karakter som prosjektene får. Det vil være naturlig at norske finansiører følger etter EU i denne sammenheng, men utvalget vil understreke at det i tillegg til insentiver av denne typen også er nødvendig med et fokusskifte. Det er behov for å dreie oppmerksomheten bort fra ulempene og hindringene for deling av data, og heller fokusere på det forskningsetiske ansvaret forskere har for å dele sine forskningsfunn med omverdenen og de store fordelene de selv vil få gjennom tilgang til mange flere datakilder i et mer åpent forskningssamfunn. En annen positiv effekt av bedre praksis for datapublisering er at dette vil stimulere til publisering av komplette, kvalitetssikrede og veldokumenterte dataprodukt.

Utvalget vil også understreke viktigheten av at framtidens tellekantsystem, vurderingskriterier og andre insentivmekanismer må utformes slik at alle fagdisipliner får like muligheter. Mange fag vil sjelden produsere eller samle inn data som er egnet for deling og gjenbruk, og slike fag må ikke forfordes i system som favoriserer datatunge fagfelt.

## 14. Investering i FAIR forskningsdata

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 etterprøvbarhet, integritet og effektivitet, men vil ha begrenset verdi for samfunnet hvis dataene ikke kan brukes til å utforske nye ideer, skape ny innsikt eller utvikle nye produkter. For at data faktisk skal kunne gjenbrukes, må de gjøres FAIR, og dette krever ekstra ressurser. Dette må sees på som en investering som vil gi avkastning på sikt, og spesielt vil kvalitetsdata fra forskning kunne være en god investering i framtiden. Investering i FAIR forskningsdata må skje på flere nivåer for at Norge skal lykkes.

### 14.1 Investering i infrastrukturer for FAIR data

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 gi 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 og oppgraderinger. 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 infrastrukturene blir mer robuste. Per 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, nasjonalt og 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 dagsordenen. 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. Dette vil være spesielt viktig for kostbare, generiske forskningsinfrastrukturer som skal betjene hele forskningssektoren.

#### 14.2 Investering i kompetanse om FAIR data

Å gjøre data FAIR krever ikke bare ressurser i form av investeringer i gode infrastrukturer og systemer for deling og gjenbruk, men også i form av menneskelige ressurser. Forskningsdata er ofte komplekse, og det å gjøre dataene brukbare for andres bruk vil kreve både tid og kompetanse. Slik kompetanse er det ikke gitt at forskeren som har produsert dataene har. Det er derfor nødvendig med kompetent forskerstøtte, som kan veilede og gi råd. Innenfor alle fagområder vil det også bli et økende behov for personer med spesialistkompetanse i skjæringspunktet mellom jus, digitalisering og teknologi, åpen forskning og vitenskapelig metode. Samtidig ser man et økende behov for spesialistkompetanse på kunstig intelligens, maskinlæring og dyp læring ved bruk av tungregningsprosesser, når generiske e-infrastrukturer skal yte bistand til forskerprosjekt med slike behov.

Åpen forskning og datapubliserings 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. Det er viktig at myndighetene, som setter krav om en mer åpen forskningssektor, også legger til rette for at institusjonene kan tilby gode støttetjenester til forskerne. Det kan være vanskelig å få til en god dimensjonering mellom forskning og støttefunksjoner når viktig kompetanse blir omfattet av krav til effektivisering gjennom blant annet ABE-reformen.

#### 14.3 Prioriterte data for FAIR deling

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 trengs det imidlertid en prioritering av hvilke forskningsdata som skal gjøres fullt ut FAIR, og det kan være nødvendig å sette kriterier for hvilke data som betraktes som prioriterte i en første fase. En prioritering vil også føre til en mer effektiv og målrettet innsats for å gjøre så mye data som mulig tilgjengelig så raskt som mulig. I [Rapporten fra ekspertgruppen for datadeling i næringslivet \(2020\)](#) påpekes det at det er urealistisk og uhensiktsmessig å kreve at alle mulige datasett skal gjøres tilgjengelig på et nivå for gjenbruk for alle mulige 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.

På den andre siden argumenteres det for at vi ikke har råd til å *ikke* gjøre data FAIR. Ifølge en rapport som PWC skrev på oppdrag fra EU-kommisjonen i 2018, vil dette utgjøre et samlet tap på 10,2 milliarder euro hvert år for landene i EU. Regnestykket har kun tatt med det økonomiske tapet, og ikke medregnet

alt annet som man vil gå glipp av, som blant annet muligheter for mer kvalitetssikret datatilgang til store mengder maskinlesbare data.

Til tross for at det er mye verdi å hente i å investere i FAIR data, så er det naturligvis økonomiske begrensninger på dette området som på alle andre områder. Utvalget foreslår derfor at det i første omgang prioriteres forskningsdata med spesielt stor forskningsmessig- og samfunnsverdi etter modell fra Horisont Europa, slik at riktig mengde ressurser kan allokere til riktig sted for at målsettingen om mer gjenbruk skal bli realisert. En slik prioritering betyr imidlertid ikke at andre offentlig finansierte forskningsdata ikke skal gjøres offentlig tilgjengelig. Det betyr bare at de ikke nødvendigvis må tilfredsstille FAIR-prinsippene (men likevel deles med tilstrekkelig informasjon for at dataene kan etterprøves og kvalitetssikres).

En annen innfallsvinkel kan være å komme i gang med de enkle områdene først, og dele det som kan deles uten altfor store kostnader eller andre barrierer. På denne måten kan man synliggjøre mulighetene som deling av data gir, og vi kan ta lærdom på veien. 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.

## 15. Verdiskaping og innovasjon basert på forskningsdata

Deling av data er et viktig premiss for å legge grunnlaget for datadrevet verdiskaping og innovasjon. Dette er en begrunnelse for de tydelige føringene som gjennom det siste halvannet året er gitt fra regjeringen til direktorater, kommunal sektor og andre offentlige virksomheter. Dette er i tråd med de anbefalingene som er gitt i Digital21; det er sentralt i ekspertgruppen for deling av data sitt arbeid, og det gjentas i stortingsmeldingen [Data som ressurs – Datadrevet økonomi og innovasjon \(2020–2021\)](#): Deling og tilgang til data blir viktig framover.

Transparens betyr likevel ikke full åpenhet om alt i alle sammenhenger. Det er for eksempel opplagt at det må være transparens om hvordan evaluering av en søknad om forskningsmidler foregår, men det er ingen som forventer at innholdet i søknaden skal deles. Når en bedrift har fått støtte til et prosjekt om forskning og utvikling, må det være transparens om dette, hvor mye støtte de mottar og hva midlene skal brukes til. Men det er ikke nødvendigvis slik at alle resultatene fra det fullførte prosjektet må deles åpent. Åpenhet rundt offentlig finansierte (både helt og delvis) samarbeidsprosjekter mellom næringsliv og forskning er et viktig prinsipp, men åpenhet for å realisere den offentlig finansierte forskningens samfunnsoppdrag må avveies mot private aktørers konkurransevne.

Det offentlige finansierer forskning og utvikling for flere formål og gjennom ulike virkemidler, som også reflekteres i støttegrader og evalueringskriterier. Forskningsmidler går til universiteter og forskningsinstitutter for å bygge kunnskap til nytte for samfunnet og legge til rette for forskningsbasert utdanning, sosial innovasjon og innovasjon i offentlig forvaltning. Slike prosjekter evalueres etter kriterier som vitenskapelig kvalitet, prosjektledelse og forskningsgruppen, gjennomførbarhet, formidling og kommunikasjon og internasjonalt samarbeid. Dette er prosjekter med universiteter og institutter i førerretet, som typisk finansieres med 100 eller 80 prosent. Kravet om åpenhet og deling av resultater må være tydelig.

I tillegg bruker det offentlige midler for et annet formål: stimulere verdiskaping, innovasjon og styrket konkurransekraft i norsk næringsliv. Dette er avgjørende for et bærekraftig velferdssamfunn. For å løse denne delen av samfunnsoppdraget bidrar det offentlige gjennom Forskningsrådet og andre finansører med finansiell risikoavlastning. Dette dreier seg typisk om 25–50 prosent av prosjektkostnadene, som gis til bedrifter for å stimulere til at de tar i bruk forskning og utvikling i sitt innovasjonsarbeid. Her er det bedriftene som sitter i førersetet, tar den største risikoen og selv må sørge for at resultatene videreutvikles til kommersiell aktivitet og ny verdiskaping. Utvalget påpeker at det trengs klare retningslinjer for deling av data også for denne type prosjekter, slik at aktørene kan forholde seg til disse. Her vil det være en mulig konflikt mellom bedriftens interesser av å holde noe hemmelig, og samfunnets generelle interesse av å dele data fra offentlig delfinansierte forskningen.

Det er en villet politikk at norske (og utenlandske) bedrifter skal samvirke tett med de beste (norske) forskningsmiljøene. Det bidrar til å gjøre forskningsmiljøene relevante og til at kompetansen i sterke forskningsmiljøer kommer til nytte i norske bedrifter. Et spørsmål er om et krav om deling av resultater fra prosjekter der norske bedrifter benytter norske forskningsmiljøenes kompetanse vil kunne gjøre det uaktuelt for de norske bedriftene å bruke disse forskningsmiljøene. Det vil i så fall svekke innovasjonsarbeidet i bedriften og de norske forskningsmiljøene. Et krav om helt eller delvis deling av forskningsresultater fra prosjekter der bedriften selv finansierer den største delen, bør ideelt sett ikke være en enestående norsk politikk. Om Norge skiller seg vesentlig fra andre land, er det mulig å tenke seg at norske bedrifter i stedet søker prosjekter med utenlandske forskningsmiljøer og finansieringsmekanismer der slike krav ikke er til stede.

Det er lovbestemt at en del av virksomheten til UH-institusjonene skal bidra til innovasjon og verdiskaping basert på resultatene fra forskning og faglig og kunstnerisk utviklingsarbeid (UHL § 1-3 bokstav e). Samtidig er det sterke føringer om at UH-sektor skal sørge for åpenhet om resultater fra forskning (UH-lovens § 1-5 (6)). Ofte vil disse formålene gå hånd i hånd fordi åpenhet om resultater fra forskningen vil bidra til kunnskapsspredning og gi et grunnlag for videre forskningsaktivitet, samtidig som det vil bidra til innovasjon og verdiskaping ved at aktører i næringsliv og offentlig sektor kan utvikle produkter og tjenester basert på dem. Når innovasjonen har et kommersialiseringsaspekt knyttet til seg, kan det imidlertid oppstå noen utfordringer i skjæringsfeltet mellom forventning om åpenhet om resultater og forventninger om å bidra til innovasjon og verdiskaping i UH-sektor. Der institusjoner samarbeider med næringslivet, for eksempel, kan det oppstå spørsmål om i hvilken grad dataene skal holdes tilbake gjennom vern av forretningshemmeligheter. Andre samfunnsinteresser kan hevde at publisering av data er en rimelig forutsetning for offentlig støtte, fordi det gir noe konkret tilbake til samfunnet. Denne problemstillingen overlapper delvis med problemstillingen som er diskutert over, om samarbeidsprosjekter finansiert delvis med offentlige midler gjennom Forskningsrådet. Dette er kompliserte temaer, som bør avklares nærmere. Utvalget har hverken hatt tid, eller tilstrekkelig kompetanse, til å gå i dybden av denne tematikken, så en løsning kan være å gi et slikt oppdrag til noen andre.

## 16. Utvalgets anbefalinger

**[Utvalget vil presisere at disse anbefalingene ikke er ferdig diskutert og gjennomarbeidet. De har ulik detaljeringsgrad og muligens også ulike relevans for denne rapporten. Vi ønsker likevel å dele anbefalingene som foreløpig er diskutert, og tar gjerne imot konkrete innspill]**

## REGJERINGEN OG DEPARTEMENTENE

- Det bør være et gjennomgående forskningspolitisk mål å ta i bruk virkemidler som stimulerer til deling og gjenbruk av data fra offentlig finansiert forskning, blant annet gjennom langtidsplanen for forskning og utdanning, utviklingsavtaler mellom KD og sektor, og tildelinger gjennom Norges Forskningsråd.
- Regjeringen bør sørge for at strategier, regelverk og retningslinjer som omhandler og/eller regulerer deling av og gjenbruk av data, harmoniseres på tvers av sektorer, og på tvers av alle nivåer i kunnskapssektoren, slik at de alle trekker i samme retning.
- Tellekantsystemet bør justeres eller erstattes med et nytt system som, i samsvar med DORA-erklæringen, ivaretar bredden i resultater fra forskning, blant annet gjennom at produksjon og tilgjengeliggjøring av vitenskapelige data gir uttelling.
- Regjeringen bør initiere en prosess for å definere hva som er prioriterte datasett for å kunne løse nasjonale og internasjonale samfunnsoppdrag. Hva som skal kvalifisere til å være prioriterte datasett bør avgjøres i dialog med forskningsmiljøer, næringsliv, offentlig forvaltning og andre samfunnsaktører.
- Regjeringen bør legge til rette for norsk deltakelse i internasjonalt arbeid som søker å fremme åpen forskning og deling av forskningsdata i tråd med FAIR-prinsippene.
- Norge bør fortsatt øke investeringer i datainfrastrukturer i årene som kommer slik at norske forskningsmiljøer har de tjenestene og de verktøyene de trenger for å gjøre data FAIR.
- Det bør arbeides videre med å lage bærekraftige finansieringsmodeller for etablering og drift av datainfrastrukturer.
- Norge bør ta en aktiv rolle i EOSC slik at norske interesser ivaretas i utviklingen av den europeiske infrastrukturen for deling og gjenbruk av forskningsdata.
- Kompetansebehovet innenfor åpen forskning må synliggjøres og bli en del av oppfølgingen av [Nasjonal kompetansepolitisk strategi](#) og stortingsmeldingen [Lærekraftig utvikling — Livslang læring for omstilling og konkurranseevne](#)
- Institusjonene må få de ressursmessige forutsetningene som skal til for å heve kompetansen om åpen forskning og til å etablere nødvendige støttetjenester til forskerne.
- Det bør vurderes å sette ned en gruppe/utvalg med medlemmer som representerer næringslivet, instituttsektor, UH-sektor, virkemiddelapparatet og allmenheten, som får i oppdrag å diskutere utfordringer med deling av forskningsdata som kan ha kommersiell verdi og/eller som produseres i samarbeidsprosjekter som er delvis finansiert av næringslivet.
- Sørge for at krav og forventninger om deling av data hos institusjonene harmoniserer med andre krav og forventninger til de samme institusjonene.



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

## FINANSIØRER OG VIRKEMIDDELAPPARATET

- Åpen forskning bør være en del av kvalitetskriteriet ved tildeling av midler.
- Tilgjengelige datasett, og deres tilknyttede meritter som antall siteringer og omfang av gjenbruk, bør i tillegg til vitenskapelige artikler gis uttelling ved tildelinger av offentlige forskningsmidler (i tråd med DORA-erklæringen).
- Alle eksterne evalueringer av offentlig finansiert forskning (miljøer, institusjoner eller fagområder) bør i større grad innlemme åpen forskning og god datahåndtering som evalueringsparametere. Utvikling, vedlikehold og deling av verdifulle datasett bør inngå som parametere i slike evalueringer.
- FAIR datasett og andre dataprodukter som er tilgjengelig for gjenbruk av andre bør gi uttelling i offentlig finansierte forskningsprosjekter.
- Forskningsrådet bør følge opp prosjekters datahåndtering tettere enn det som gjøres i dag, og vurdere veiledning og tiltak dersom prosjektene ikke deler og forvalter data i tråd med intensjonene i prosjektene.
- Forskningsrådet og andre finansiører bør vurdere nye virkemidler som stimulerer til bruk av eksisterende data i forskning og innovasjon i større grad.
- Vektlegging av datadeling som vurderingskriterium må tilpasses virkemiddelet som benyttes og de miljøene som det er naturlig at søker om midler i ulike tilfeller. Det bør foretas en differensiering mellom datatunge forskningsfelt og forskningsfelt som produserer mindre data når en slik vektlegging skal gjennomføres.
- Forskningsrådet bør lyse ut egne midler til datainfrastrukturer som bidrar til mer deling og gjenbruk av forskningsdata.
- Insentiver og oppfølging av større tildelinger, slik som SFF, SFI og forskerskoler, bør også knyttes mot god dataforvaltning.
- Egne tildelinger bør gis til oppbygning, vedlikehold og deling av viktige datasett, for å stimulere til prioritering av slik aktivitet.
- Forskningsrådet bør vurdere å lyse ut midler til piloter/forprosjekter for å kunne investere i nyskapende og framtidrettede datainfrastrukturer som kan innebære risiko.

- Finansiører bør være tydelige på når resultatene av offentlig finansierte prosjekter skal deles, dette er ikke minst viktig for å ikke undergrave bedrifters interesse for FoU og samarbeidsprosjekter.
- Resultater fra prosjekter med universiteter og institutter som er offentlig finansierte skal som hovedregel deles helt åpent.
- I samarbeidsprosjekter mellom forskning og næringsliv, som er delvis finansierte av det offentlige og som har innovasjon, konkurransekraft og verdiskaping som formål, er det viktig å skille mellom prosjekterresultater og forskningsresultater som man ønsker å dele. Hvis forskningsdelen i et innovasjonsprosjekt resulterer i en publikasjon, bør man etterstrebe og publisere i et åpent tidsskrift, dele kode og dele data. Dette er viktig også for at forskningen skal være etterprøvbar. Hva som ellers kan deles åpent fra et slikt prosjekt, og vilkår for delingen, kan være avhengig av offentlig finansieringsgrad og konkurranseklausuler i avtaleverket. Vilklårene bør være offentlig kjent før prosjektet starter.

#### FORSKNINGSINSTITUSJONENE

- Institusjonene bør utarbeide langtidstrategier for deling og forvaltning av data, som beskriver hvordan institusjonen skal gå fram for å sikre et framtidig behov for deling av data, inkludert infrastruktur og kompetanse ved institusjonen.
- Institusjonene bør innføre helhetlige karrierevurderinger i tråd med DORA-erklæringen, som også gir uttelling for publisering av gjenbrukbare datasett og andre elementer knyttet til datadeling (som programvare og kildekode).
- Legge til rette for at kompetanse om datahåndtering og FAIR data økes gradvis og i alle deler av forskningssystemet, for eksempel gjennom opplæring i datadeling som obligatorisk del av doktorgradsutdanning, postdoktorprogrammer og kurs for forskningsledere.
- Institusjonene må sørge for at interne veiledere og retningslinjer for datadeling er i tråd med nasjonale krav og føringer, og sørge for at policyer i ulike deler av virksomheten er i samsvar med hverandre.
- Institusjoner og forskningsmiljøer som søker om eksterne midler, må synliggjøre kostnader knyttet til å gjøre data FAIR – og legge dette inn i sine prosjektsøknader.
- Åpen forskning generelt og datahåndtering spesielt, bør være en del av alle bachelorprogrammer og masterprogrammer. Dette bør kobles opp mot forskningsetikk og integritet i forskningen.
- Forskningsmiljøer bør delta aktivt i nasjonale og internasjonale debatter knyttet til datadeling i sitt fagfelt, med tanke på å sikre datadelingssystemer som er gunstige for eget fagfelt.

- Institusjonene bør sørge for at retningslinjer for IPR og kommersialisering harmoniserer med retningslinjer for åpen forskning og tilgjengeliggjøring av data.
- Institusjonene bør utarbeide statistikker over egen datadeling til intern monitorering av egen aktivitet. Denne bør inkludere andel av prosjekter hvor det er delt data eller eventuelt gjort rede for hvorfor data ikke kan deles, samt statistikker for publisering av data forenlig med FAIR-prinsippene.
- Institusjonene bør vurdere å ha egne utvalg/datadelingsombud som gir råd og anbefalinger om datadeling, for eksempel ved uklarhet om data kan deles ut fra gjeldende lovverk, utfordringer knyttet til lisensvalg og andre praktiske forhold.
- Institusjonene bør allokere tilstrekkelige midler til teknisk støtte/forskerstøtte som er nødvendig for å oppnå god forvaltning av forskningsdata.