



Infrastruktur og tjenester for FAIR forskningsdata – Status og forslag til videre arbeid

Sluttrapport fra utredningsgruppen 15.3.2022
(med feilrettinger pr 31.3.)

«Det må være enklere å dele enn ikke å gjøre det»

Innholdsfortegnelse

1. Sammendrag	3
2. Innledning	6
2.1. Mandat og tolkning av dette	7
2.2. Prosess og metode.....	10
2.2.1. Organisering av arbeidet	10
2.2.2. Beskrivelse av de enkelte leveransene	11
2.2.3. Datainnsamling og kilder	12
3. Et rammeverk for FAIR forskningsdata	15
3.1. Mer data, mer kompleksitet	15
3.2. Mer fokus på deling av forskningsdata	16
3.3. European Open Science Cloud (EOSC).....	17
3.4. Endringer i retningslinjer, lover og regler	18
3.5. Livssyklus for forskningsdata	19
3.6. FAIR data i et økosystem for forskning	20
3.7. Datahåndteringsplaner	23
3.8. Klassifisering av data	24
3.9. Lagring og arkivering av forskningsdata	25
3.9.1. Håndtering og lagring av data i aktive prosjekter	25
3.9.2. Arkivering og langtidsbevaring av data etter prosjektslutt	25
3.10. Arkivering og gjenbruk av data med behov for beskyttelse	26
4. Behov for videre arbeid	28
4.1. Kunnskapsbehov og kompetanse	29
4.2. Infrastrukturer og tilhørende tjenester for datahåndtering etter FAIR-prinsippene	35
4.3. Kultur for deling	41
4.4. Regulatoriske virkemidler	43
4.5. Andre temaer.....	45

1. SAMMENDRAG

Denne rapporten er en av flere leveranser fra utredningen «Felles infrastruktur og tjenester for FAIR forskningsdata», på oppdrag fra Digitaliseringsstyret for høyere utdanning og forskning.

Rapporten publiseres på nettstedet opscience.no som er under utvikling som et samlingssted for informasjon om åpen forskning. Hovedredaktør for nettstedet er Sikt – Kunnskapssektorens tjenesteleverandør, og redaksjonsrådet har også medlemmer fra UHR og fra Forskningsrådet. Hovedmålgruppe for nettstedet er de norske forskningsinstitusjonene. Rapporten retter seg mot både forskningsinstitusjoner i universitets- og høyskolesektoren, helsesektoren og instituttsektoren, myndighetsorganer, finansiører og de enkelte forskningsmiljøene.

De andre leveransene fra utredningen (også publisert på opscience.no) er:

1. [En norsk terminologi for begreper knyttet til området forskningsdata](#). Denne er utarbeidet av en arbeidsgruppe i regi av den norske noden av Research Data Alliance (RDA) og vil bli forvaltet av denne gruppen. Terminologien bør også brukes som oppslagsverk for begreper brukt i denne rapporten.
2. [Rapport fra spørreundersøkelse blant norske forskere](#) gjennomført sommeren 2021 om deres bruk av infrastrukturer og tjenester for håndtering av forskningsdata.
3. [Rapport fra spørreundersøkelse blant norske forskningsinstitusjoner](#) (universiteter og høyskoler, forskningsinstitutter og helseforetak) gjennomført høsten 2021 om deres arbeid med håndtering av forskningsdata. Svarene fra helseforetakene er skilt ut i en [egen rapport](#).
4. **Oversikt over (primært generiske) forskningsinfrastrukturer og tjenester** som er i vanlig bruk blant norske forskere i dag. Dette er løst ved å utvide flere av de eksisterende sidene på opscience.no med ekstra informasjon. Se [utredningens sider](#) på opscience.no for mer informasjon om leveransen.

I denne utredningen har formålet primært vært å kartlegge og synliggjøre de tjenester for håndtering av forskningsdata som per i dag er tilgjengelige for norske forskere. Vi har blant annet gjennom spørreundersøkelsene som er gjennomført, identifisert barrierer mot datadeling, og behov for tjenester, infrastruktur og kompetanse. Vi gir anbefalinger om veien videre når det gjelder tjenester som gir muligheter for tilgjengeliggjøring og deling av forskningsdata.

Økt deling og gjenbruk av forskningsdata som er finansiert med offentlige midler er et viktig og riktig fokus i norsk og internasjonal forskningspolitikk. Diskusjonen om kunnskapsdeling i samfunnet pågår både nasjonalt og internasjonalt, og handler i stor grad om deling av data. I den globale økonomien er dataflyt et viktig premiss, og økt tilgang til data endrer hvordan samfunnet fungerer. Datadelingen skaper også helt nye muligheter og gevinster av data, for eksempel gjennom stordata og kunstig intelligens. Som et eksempel har det vist seg at deling av data om Covid-pandemien på tvers av databaser og landegrenser har ført til en mer effektiv håndtering av nasjonale og internasjonale smittevernstiltak.

Samtidig er ønsket om økt deling og gjenbruk av data forbundet med en rekke utfordringer. Først og fremst er det krevende å sikre tilgjengeligheten til de data som samles inn, slik at det faktisk er mulig å gjenbruke dem. For at data enkelt skal kunne gjenbrukes, bør de ideelt sett oppfylle det vi omtaler som FAIR-prinsippene. De fire bokstavene står for Findable, Accessible, Interoperable og Re-usable. Data som skal gjenbrukes må man kunne finne frem til, dataene må konkret kunne hentes opp, de må være forståelige og logisk kunne settes sammen med andre data, og de må faktisk være mulig å benytte til nye formål. Grad av FAIR («FAIRness») er et kontinuum, der prinsippene bak hver av de fire bokstavene kan være oppfylt i større eller mindre grad. De datainfrastrukturer og tilhørende datahåndteringstjenester vi omgir oss med kan legge til rette for FAIR på ulike måter og utvikle seg gradvis mot full FAIRness. Mange datakilder i samfunnet er i utgangspunktet ikke FAIR, og det må en innsats til for å gjøre dem mer FAIR. Selv det første steget fra helt lukket til Findable er viktig og kan gi store positive effekter, blant annet for å unngå at noen unødig samler inn data som allerede finnes. I 2017 da FAIR-prinsippene ble publisert, estimerte man at så mye som 80 % av alle datasett var «re-useless». Selv om mye har endret seg på 5 år, har vi fortsatt en lang vei å gå.

Denne utredningen dokumenterer stor heterogenitet med hensyn til hvor langt norske forskningsinstitusjoner har kommet i arbeidet med å dele forskningsdata. Forskjellen mellom store og etablerte forskningsinstitusjoner og de mindre miljøene kan være et viktig tema for videre arbeid. Samtidig indikerer vår kartlegging at det er store interne variasjoner både ved institusjonene og mellom ulike fagområder og disipliner i arbeidet med datadeling. Det fremkommer at det er en rekke utfordringer som på ulike måter representerer barrierer mot datadeling. Det synes generelt å mangle kultur for, kompetanse om og erfaring med deling av data, og det ser ut å være behov for bedre og mer sammenhengende tekniske løsninger for å støtte datadeling. Det hersker usikkerhet om juridiske forhold ved datadelingen, og det mangler insentiver for forskningsmiljøene til å dele data.

Opprinnelig var en av leveransene fra dette utredningsarbeidet å etablere retningslinjer med kriterier som infrastrukturer og tjenester for håndtering av forskningsdata bør oppfylle for at forskerne enklere kan dele sine data FAIR. Utredningen har imidlertid vist at det på dette området er mange gode pågående initiativer og samarbeidsarenaer i Europa – med nettopp dette for øyet. Det ville virke mot sin hensikt å utvikle nasjonale retningslinjer og kriterier spesielt for Norge. Vår anbefaling er derfor heller at norske tjenesteleverandører for forskningsdatahåndtering engasjerer seg i eksisterende nasjonale og internasjonale samarbeidsinitiativer som EOSC og RDA der slike retningslinjer for FAIR-forskningsinfrastrukturer og relaterte tjenester og datahåndteringsverktøy allerede utvikles. Samtidig er det helt nødvendig at gjeldende lovgivning følges ved den datadeling som foregår. Særlig for data som krever beskyttelse kan det være behov for løsninger som spesifikt tar høyde for nasjonalt lovverk.

Utredningen har avdekket at manglende kompetanse om og kultur for deling og gjenbruk av forskningsdata sannsynligvis er det største hinderet for å komme videre. For å dekke behovet for kompetanseheving og veiledning på en effektiv måte for alle forskere i Norge, vil det være nødvendig med en betydelig ressursinnsats og god koordinering av kompetansehevingstilbud mellom lokale- institusjonelle og nasjonale miljøer. Generelt må kompetansen på alle nivåer i

forskningssystemet heves, og tilgjengelig kapasitet for veiledning og operativ datahåndtering må økes betydelig. I EU-systemet vises det i ulike sammenhenger at i snitt 5 % av budsjettet for et forskningsprosjekt bør avsettes til datahåndtering.

En utfordring ved deling og gjenbruk av data er hvordan data med beskyttelsesbehov bør håndteres i en FAIR-sammenheng. Dette kan være ulike typer personidentifiserende data (herunder helsedata), data som berører forhold som er regulert av sikkerhetslovgivningen, industrihemmeligheter og opplysninger som er beskyttet etter annen lovgivning. Våre funn bekrefter at det er stor usikkerhet i forskningsmiljøene om hvordan slike data bør håndteres, gitt at behovet for deling og gjenbruk av slike data skal ivaretas. Usikkerheten er knyttet til både juridiske og etiske aspekter. Vi har i løpet av arbeidet sett svært gode eksempler på nyttevirksomheter av FAIRifisering av data der selve dataene ikke kan deles åpent, men der enten metadata eller aidentifiserte versjoner av dataene fortsatt kan deles.

Mangelfull kompetanse om juridiske forhold og de muligheter og begrensninger som gjelder, er en barriere for økt gjenbruk av forskningsdata. I mange tilfeller har forskningsmiljøene lite empiri å støtte seg på. Regelverket er komplekst, og frykten for å tråkke feil er stor blant forskerne. Det er naturligvis bekymringsfullt om forskningsmiljøer tenderer til å ikke dele sine data som følge av opplevde juridiske barrierer mot arkivering og gjenbruk av forskningsdata. Kompetanseheving innen både juridiske og forskningsetiske forhold blir derfor viktig fremover, på alle nivåer. Temaet er imidlertid såpass komplekst at vi anbefaler at det utredes videre hvordan et slikt kompetanseløft kan håndteres.

En funksjonell, tilgjengelig infrastruktur for arkivering og gjenbruk er essensielt, blant annet er det behov for infrastrukturer med tilgangsstyring for arkivering og deling av data som krever beskyttelse. Her er det flere gode nasjonale og europeiske initiativ. Vi anbefaler å arbeide videre på dette området.

Vi håper denne utredningen og oppdragets ulike leveranser vil være et viktig bidrag til økt deling slik at vi kommer nærmere at deling av forskningsdata blir en standard blant våre forskere. Vi håper også at utredningen er et bidrag i det mer langsiktige arbeidet med å sikre hensiktsmessige og samfunnsnyttige strukturer som understøtter den store transisjonen som både norsk og internasjonal forskning nå går gjennom.

2. INNLEDNING

Økt deling og gjenbruk av forskningsdata har potensial til å bidra til vitenskapelige fremskritt, innovasjon og smartere tjenesteutvikling i offentlig sektor¹. Gjenbruk av data fra forskningen er slett ikke noe nytt. Norsk senter for forskningsdata (NSD) feiret nettopp sitt 50-årsjubileum. En av deres sentrale oppgaver har vært å gjøre data fra samfunnsvitenskapelig forskning tilgjengelige for andre forskere og for samfunnet generelt. Flere andre fagmiljøer har også lang tradisjon for å arkivere og dele data. Likevel er det først de siste par tiårene at arbeidet med deling av forskningsdata virkelig har fått fart, ikke minst som resultat av de store utfordringene vi står overfor når det gjelder å opprettholde et bærekraftig samfunn. Innen flere fagdisipliner har tilgang til stadig større datamengder bidratt til at man ser et potensial for å oppnå bedre og raskere resultater av forskningsinnsatsen når forskere kan gjenbruke tidligere innsamlede data.

I løpet av de siste årene har stadig flere instanser også innført krav om at data som samles inn eller produseres som del av et forskningsprosjekt, gjøres tilgjengelig for gjenbruk når prosjektet er avsluttet. I 2016 ble [FAIR-prinsippene](#) publisert i tidsskriftet Scientific Data². Disse innebærer at data skal være FAIR, dvs. gjenfinnbare (Findable), tilgjengelige (Accessible), kompatible (Interoperable) og gjenbrukbare (Reusable). Samme år vedtok EU-kommisjonen å støtte disse prinsippene for å sikre en fremtidsrettet og mer effektiv bruk av offentlig finansierte forskningsdata på tvers av land, forskningsdomener og sektorer. I forsknings- og innovasjonsprogrammet Horisont 2020 ble det etablert en pilot for frivillig deling av data – først innen utvalgte deler og deretter i hele programmet. I videreføringen Horisont Europa er den frivillige piloten videreutviklet til et generelt krav om at forskningsdata som er finansiert gjennom rammeprogrammet skal gjøres FAIR. Data som er FAIR er ikke nødvendigvis åpent tilgjengelige. De vil derimot være maskinlesbare og interoperable, noe åpne forskningsdata ikke nødvendigvis er. Tilgangen til dataene kan begrenses dersom det er gode grunner for dette, for eksempel dersom det dreier seg om persondata, data som er knyttet til kommersielle interesser eller andre former for sensitive opplysninger. EUs krav er blant annet formulert som at data skal være «as open as possible, as closed as necessary».

Da EU-kommisjonen vedtok å støtte FAIR-prinsippene, lanserte de også "European Cloud Initiative" som et viktig steg på veien mot målet. Initiativet inneholdt planer om et felleseuropeisk økosystem for forskning og innovasjon på tvers av offentlig og privat sektor. Dette inkluderte datainfrastrukturer for forskningsdata, nettverkstjenester for mer effektiv dataflyt og superdatamaskiner for å behandle de stadig større og mer komplekse mengder av forskningsdata som genereres. For å kunne implementere et slikt økosystem foreslo kommisjonen blant annet å utvikle og implementere «European Open Science Cloud» (EOSC). Ambisjonen med EOSC er et nettverk for FAIR data og tjenester, hvor man bygger på og samordner allerede eksisterende europeiske datainfrastrukturer og relaterte tjenester og ressurser.

Høsten 2017 vedtok den norske regjeringen først [Digitaliseringsstrategi for universitets- og høyskolesektoren](#) og deretter [Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata](#). I denne strategien stadfestes blant annet at Norge vil følge EUs oppfordring om

¹ Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata

² DOI til artikkelen: <https://doi.org/10.1038/ng.3544>

å støtte FAIR-prinsippene og utviklingen av EOSC. Strategien understreker også betydningen av langsiktig arkivering og kuratering av forskningsdata, og at dette arbeidet må være basert på internasjonale standarder.

I 2021 vedtok EU også [Open Data Directive](#) som vedrører deling og viderebruk av åpne data finansiert av offentlig sektor, herunder også åpne forskningsdata. Direktivet vil ventelig bli implementert i norsk rett i løpet av kort tid. Et eget lovutvalg er nedsatt av regjeringen for å vurdere hvorvidt gjeldende bestemmelser om viderebruk hjemlet i offentlighetsloven er tilstrekkelig, eller om det bør utarbeides en egen, helhetlig lovgivning knyttet til viderebruk av offentlige data. Dette baserer seg blant annet på det faktum at deling og viderebruk av offentlige data i dag er regulert ikke bare gjennom offentlighetsloven, men i flere andre lover – herunder sektorspesifikke lover (for eksempel geodataloven). Lovutvalget skal etter planen avgi sin innstilling i februar 2023.

2.1. MANDAT OG TOLKNING AV DETTE

Denne rapporten er en del av utredningen som er gjennomført på oppdrag fra Digitaliseringsstyret for høyere utdanning og forskning. Hovedhensikten med oppdraget har vært å fremskaffe informasjon som kan gi veiledning til både institusjonene og den enkelte forsker om relevante infrastrukturer og tjenester som finnes i dag, og til tjenestetilbyderne gjennom felles retningslinjer for hva slags funksjonalitet tjenestene bør tilby. Utredningen skal også identifisere eventuelle behov for nye eller forbedrede infrastrukturer og tjenester, med spesiell vekt på langtidsbevaring av forskningsdata. Mandatet har også omfattet å etablere en norsk terminologi for begreper knyttet til forskningsdata, kartlegge eksisterende infrastrukturer og tjenester og å synliggjøre disse på nettsiden [opscience.no](#). [Mandatet for arbeidet](#) er i sin helhet publisert på [opscience.no](#).

I samarbeid med forskningsinstitusjonene fikk Unit i 2018 oppdrag fra Kunnskapsdepartementet å utarbeide en handlingsplan for digitalisering i høyere utdanning og forskning. Digitaliseringsstyret for høyere utdanning og forskning ble etablert for å operasjonalisere de forannevnte strategiene, og det ble oppnevnt seks fagutvalg under digitaliseringsstyret for å foreslå tiltak på de ulike områdene. Strategiene nevnt i innledningen til dette kapitlet la dermed et viktig grunnlag for handlingsplanen. [En samlet handlingsplan](#) ble vedtatt av Digitaliseringsstyret i 2019, og UH-institusjonene besluttet å sette av et årlig beløp til finansiering av fellesprosjekter for å realisere den. Den foreliggende utredningen er ett av tiltakene som er finansiert via disse fellesmidlene. Hovedansvaret for tiltaket har ligget i Fagutvalg for forskning.

Utredningen har fokusert på generiske forskningsinfrastrukturer og tjenester for forskningsdata, fordi det også å kartlegge alle fagspesifikke infrastrukturer og tjenester ville vært for omfattende innenfor tidsrammene for utredningen. Dessuten ville det å opprettholde en slik oversikt på [opscience.no](#) i ettertid vært nesten umulig. Resultatet ville blitt en oversikt som innen kort tid ville hatt dårlig kvalitet. I arbeidet har vi likevel trukket frem noen fagspesifikke infrastrukturer eller tjenester som gode eksempler, og fordi vi har sett at noen infrastrukturer eller tjenester som i dag er brukt fagspesifikt, har potensial til å løftes opp til bruk innenfor flere fagområder.

Utredningen skal bidra til å få på plass et nasjonalt tilbud av forskningsinfrastrukturer og relaterte datahåndteringstjenester som legger til rette for at forskningsdata ikke bare blir tilgjengelige, men også mest mulig automatisk gjenbrukbare (se mer detaljert beskrivelse av FAIR data i kapittel 3.6). Det er imidlertid en erkjennelse at det er en lang vei frem til et slikt målbilde.

Diskusjonene i Fagutvalg for forskning har gjort det tydelig at det er ønskelig ikke bare å legge vekt på tekniske løsninger, men også på den kompetansen og de menneskelige ressursene som er nødvendig å mobilisere for å nå målet.

Utredningen fokuserer på data generert som del av et forskningsprosjekt³ som er finansiert med offentlige midler. Det har ikke ligget innenfor vårt mandat å se på data fra forvaltningen som er interessante for forskere å få tilgang til (for eksempel ulike typer klimadata, data i helseregistrene, og offisiell statistikk fra SSB). Dette er et tema som behandles av Datainfrastrukturutvalget (omtales lenger ned i dette kapitlet). Det er ikke alltid lett eller hensiktsmessig å sette et klart skille mellom data skapt som del av forskningsprosjekter og data innsamlet som del av forvaltningsoppdrag. De samme forskningsinfrastrukturene benyttes også flere steder til å lagre data av begge typer, og datasett fra forskningsmiljøer spilles tilbake til forvaltningen og blir en del av forvaltningens datakilder. I oversikten over infrastrukturer har vi derfor tatt med et par eksempler der forvaltningens data har stor relevans også for vårt arbeid.

Utredningen har fokusert på tilgjengeliggjøring av forskningsdata for gjenbruk. For å oppnå faktisk gjenbruk, må det også finnes gode tjenester for å finne og søke i dataene. Hvilke tjenester og verktøy som benyttes eller mangler for dette har vi ikke kartlagt, men dette bør være et tema for videre arbeid.

Opprinnelig var en av leveransene å etablere retningslinjer med kriterier som infrastrukturer og tjenester for håndtering av forskningsdata bør oppfylle for at de skal kunne betegnes som "FAIR". Utredningen har imidlertid vist at det på dette området finnes mange gode initiativer og samarbeidsarenaer i Europa. Vi mener det ville virke mot sin hensikt å lage særskilte nasjonale retningslinjer og kriterier for Norge. Vår anbefaling er derfor heller at norske tjenesteleverandører for forskningsdatahåndtering engasjerer seg i eksisterende nasjonale og internasjonale samarbeidsinitiativer som EOSC og RDA der slike retningslinjer for FAIR-forskningsinfrastrukturer og relaterte datahåndteringsverktøy allerede utvikles. Samtidig er det helt nødvendig at gjeldende norsk lovgivning følges ved all datadeling, herunder for håndtering av data som krever en eller annen form for beskyttelse.

Innledningsvis diskuterte utredningsgruppen hva som kan gjøres for å motivere flere forskere til å dele sine data. Dette er et viktig tema som i noen grad omtales, men vi mener at det ligger

³ I denne rapporten forstår vi forskningsdata som *alle de registreringer, nedtegnelser, og/eller rapporteringer i form av tall, tekster, bilder og lyder som genereres eller oppstår underveis i forskningsprosjekter*. Dette kan for eksempel være data som er generert gjennom ny analyse og sammenstilling av eksisterende data, eller helt nye data generert gjennom ny datainnsamling. Forskningsdata er i denne forståelsen av begrepet alltid et direkte resultat av forskningen, uavhengig av om dataene baserer seg på eksisterende data (kildedata) eller om de er samlet inn på nytt (resultatdata).

utenfor vårt mandat. Vi har i hovedsak valgt å fokusere på hva som kan støtte de som allerede er motivert for å dele, men som opplever at det er krevende, enten fordi de mangler kunnskap om hvordan de skal få det til, eller fordi de opplever ikke å ha tilgang til de verktøy som trengs for å gjøre det. Det å finne gode løsninger for denne målgruppen er en formidabel oppgave i seg selv, og som på lengre sikt også vil hjelpe de som ennå ikke har startet å dele sine data.

Utredningen henger nært sammen med to andre arbeider, som vist i Figur 1. Datalisensutvalget leverte 1. oktober 2021 sin sluttrapport. Vårt arbeid baserer seg delvis på diskusjoner, konklusjoner og anbefalinger derfra. Samtidig er også Datainfrastrukturutvalget i regi av Forskningsrådet, i gang med sitt arbeid. Sammenhengen mellom de tre utvalgene er illustrert nedenfor. Vi har tett kontakt med Datainfrastrukturutvalget og utveksler informasjon løpende. Sluttdato for denne utredningen er 15. mars 2022, mens Datainfrastrukturutvalget leverer sin rapport før 1. mai 2022. De kan dermed ta resultatene fra vår utredning med i det avsluttende arbeidet med sin leveranse.



FIGUR 1 TRE DELVIS PARALLELLE FORSKNINGSDATAUTVALG

Vi berører i liten grad problemstillinger knyttet til opphavsrett, lisensiering av forskningsdata, etikk og finansiering av videre arbeid. Opphavsrett, lisensiering og etikk er behandlet i Datalisensutvalgets rapport, som vi bygger videre på. Finansieringsspørsmålene behandles av Datainfrastrukturutvalget. Vi har imidlertid gitt løpende innspill til Datainfrastrukturutvalget spesielt på områder hvor våre anbefalinger vil kreve omfattende finansiering.

2.2. PROSESS OG METODE

Arbeidet med utredningen startet opp i april 2021. Datainnsamlingen pågikk i hovedsak i perioden juni til desember 2021.

2.2.1. Organisering av arbeidet

Utredningsarbeidet er utført av en utredningsgruppebestående av forskere, forskningsledere og ansatte fra forskningsstøtte/bibliotek. Medlemmene er listet opp i tabellen under.

Magnus Aronsen, Universitetet i Oslo (leder)	Førsteamanuensis, Institutt for medisinske basalfag, UiO
Erik Ø. Sørensen, Norges Handelshøyskole	Professor, Institutt for samfunnsøkonomi, NHH
Karoline Daugstad, NTNU	Professor, Institutt for geografi, NTNU
Susan Nacey, Høgskolen i Innlandet	Prodekan forskning, Fakultet for lærerutdanning og pedagogikk, Høgskolen i Innlandet
Valeria Tafintseva, NMBU	Postdoc, Institutt for fysikk, NMBU
Audun Beyer, Høgskolen Kristiania	Professor, Institutt for kommunikasjon, Høgskolen Kristiania
Wenche Reed, Oslo universitetssykehus	Avdelingsleder, Stab forskning, innovasjon og utdanning, OUS
Haakon Skogseth, St. Olavs hospital	Forskningsjef, St. Olavs hospital Førsteamanuensis, Institutt for klinisk og molekylær medisin, NTNU
Frode Eika Sandnes, OsloMet	Professor, Institutt for informasjonsteknologi, OsloMet
Kjell Morten Stormark, NORCE	Forsker 1, Gruppeleder Psykisk Folkehelse, Regionalt kunnskapssenter for barn og unge, NORCE
Kirsti Klette, Universitetet i Oslo	Professor, Institutt for lærerutdanning og skoleforskning, UiO
Arne Solli, Universitetet i Bergen	Førsteamanuensis, Institutt for arkeologi, historie, kultur- og religionsvitenskap, UiB
Ane Gabrielsen, NTNU	Førstebibliotekar, Bibliotekseksjon for samlinger og digitale tjenester, NTNU
Elin Stangeland, Universitetet i Oslo	Seniorrådgiver, Samlinger og digitale tjenester, Universitetsbiblioteket, UiO

TABELL 1: OVERSIKT OVER MEDLEMMENE I UTREDNINGSGRUPPEN

Utredningsgruppen har gjennom arbeidet bidratt med sin betydelige erfaring og kjennskap til tematikken, innsikt i personvernproblematikk og andre relevante – både generelle og fagdomenespesifikke – problemstillinger knyttet til deling av data, kjennskap til interessante case og annet som har relevans for arbeidet. Sist, men ikke minst, har diskusjonene i

utredningsgruppen bidratt til fortolkningen av funnene fra kartleggingen, og til å rette fokus på de mest sentrale utfordringene på området. Det ble gjennomført 6 ordinære møter i utredningsgruppen, samt ett ekstramøte før innlevering av førsteutkast til rapport.

Utredningen har hatt et sekretariat bestående av representanter fra HK-dir, Forskningsrådet og Sikt (tidligere NSD) forsterket med to eksterne konsulenter fra Agenda Kaupang.

Utredningsgruppen har også kunnet spille på to referansegrupper, den ene bestående av forskere og den andre av representanter fra tjenesteleverandører og forskerstøtte. Referansegruppen av forskere har vært invitert til å gi innspill i 2-3 runder i løpet av prosessen både på anbefalingene og på innholdet i sluttrapporten. Det har også vært avholdt 2 egne møter med gruppen i løpet av utredningen. Referansegruppen av tjenesteleverandører og forskerstøtte har vært invitert til å gi sine innspill til sluttrapporten. Det har vært avholdt 3 møter med gruppen underveis: ett møte om med innspill til spørreundersøkelsen til forskerne, et etterfølgende møte med diskusjon av funn fra spørreundersøkelsen samt ett møte med innspill til arbeid med retningslinjer for fellestjenester. Enkelte av medlemmene i referansegruppene har også vært konsultert underveis på områder de har hatt spisskompetanse på.

I regi av den norske noden av RDA (Research Data Alliance) er det etablert en egen arbeidsgruppe for arbeidet med norsk terminologi, og som skal ha leveranser som inngår som del av denne utredningen. Arbeidsgruppen koordineres av Sikt (tidligere NSD).

En styringsgruppe har hatt overordnet ansvar for mandat, økonomi og framdrift i utredningsarbeidet.

Detaljert liste over deltagerne finnes i [mandatet for utredningen](#).

2.2.2. Beskrivelse av de enkelte leveransene

Alle leveransene fra utredningen er publisert på opscience.no. I tillegg til denne sluttrapporten består leveransene av følgende:

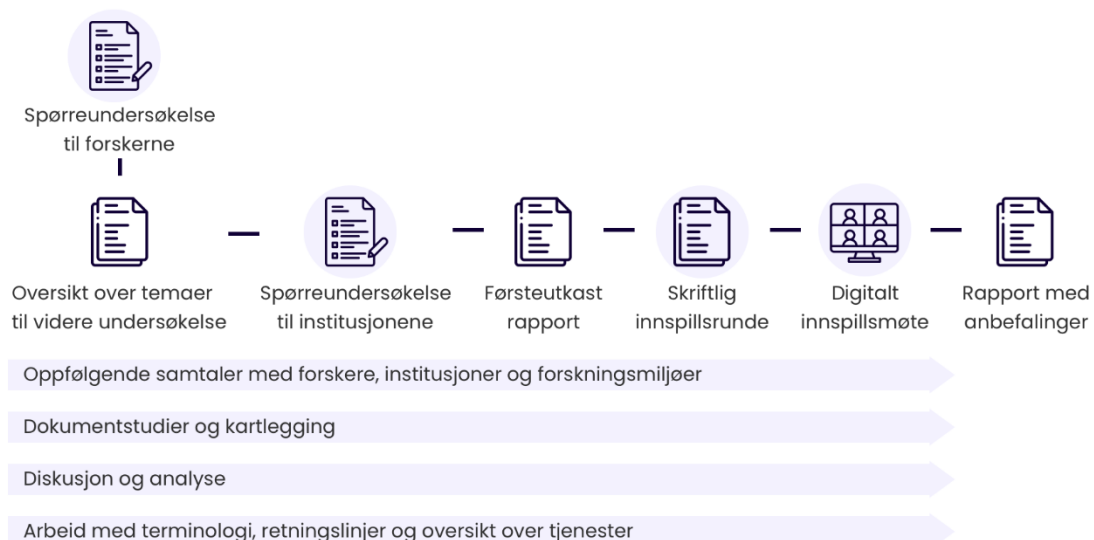
1. [En norsk terminologi for begreper knyttet til området forskningsdata](#). Denne er utarbeidet av en arbeidsgruppe i regi av den norske noden av Research Data Alliance (RDA) og vil bli forvaltet videre av denne gruppen. Leveransen er en første versjon av det som tar mål av seg til å bli en nasjonal ordliste som oppdateres jevnlig. Denne bør også brukes som oppslagsverk for begreper brukt i denne rapporten.
2. [Rapport fra spørreundersøkelse blant norske forskere](#) gjennomført sommeren 2021 om deres bruk av infrastrukturer og tjenester for håndtering av forskningsdata.
3. [Rapport fra spørreundersøkelse blant norske forskningsinstitusjoner](#) (universiteter og høyskoler, forskningsinstitutter og helseforetak) gjennomført høsten 2021 om deres arbeid med håndtering av forskningsdata. Svarene fra helseforetakene er skilt ut i en [egen rapport](#).
4. **Oversikt over forskningsinfrastrukturer og tjenester** som er i vanlig bruk blant norske forskere i dag. Dette er ment som veiledning for institusjoner som ønsker informasjon om tilgjengelige tjenester. Oversikten inneholder primært informasjon om generiske

tjenester/tjenester ment for bruk på tvers av fagdisipliner. Det er imidlertid også tatt med enkelte fagspesifikke infrastrukturer og tjenester, der disse enten har potensial til å tas i bruk bredere, eller der utredningsgruppen anser at disse tjener som gode eksempler til etterfølgelse. I hovedsak henvises til Forskningsrådets veikart for infrastrukturer innen de enkelte fagdisiplinene. Oversikten er publisert som utvidelser av flere eksisterende deler av openscience.no og vil forvaltes som del av det videre arbeidet med dette nettstedet. Se [utredningens sider](https://openscience.no) på openscience.no for mer informasjon om leveransen.

Mye godt arbeid ble gjort i prosjektet Open Science Toolbox, og materialet derfra ble inkorporert i openscience.no i 2021. Våre leveranser bygger videre på dette.

2.2.3. Datainnsamling og kilder

En sentral del av vår datainnsamling for å fremskaffe innsikt har vært to spørreundersøkelser, én rettet mot forskere og én rettet mot institusjonene. Spørreundersøkelsene er supplert med intervjuer og presentasjoner av funn i ulike fora der deltakerne har fått mulighet til å kommentere og gi ytterligere innspill. Vi gjennomførte også et åpent innspillsmøte og et dialogmøte om helseforskning. Begge møtene ble gjennomført digitalt, i lys av Covid-pandemien.



FIGUR 2: OVERSIKT OVER AKTIVITETENE I UTREDNINGEN

Spørreundersøkelse til forskere: I juni 2021 ble det sendt ut en spørreundersøkelse rettet mot forskere, og som lå åpen til 31. august. Undersøkelsen ble sendt ut via UH-institusjonene, medlemsinstitutter i Forskningsinstituttene Fellesarenas (FFA) og helseforetakene, og totalt 531 forskere deltok. I undersøkelsen stilte vi spørsmål om hvilke verktøy og tjenester for ulike deler av forskningsprosessen forskerne hadde erfaring med og om eventuelle behov som ikke dekkes av det eksisterende tilbudet. Relativt lav deltagelsen gjør at dataene skal tolkes med noe forsiktighet, men god spredning blant respondentene i fagfelt, type data de jobbet med og

kompetanse innen datahåndtering gjorde at undersøkelsen ga et godt grunnlag for å finne temaer for videre datainnsamling, intervjuer og undersøkelser.

Spørreundersøkelse til institusjonene: I perioden september til november 2021 gjennomførte vi en spørreundersøkelse rettet mot institusjonene. Undersøkelsen ble sendt til postmottak til UH-institusjonene, instituttene og helseforetakene. Totalt 63 av 100 inviterte institusjoner besvarte undersøkelsen og det var god spredning i deltakelsen mellom type institusjon (UH, institutt, helseforetak), størrelse og geografi. Undersøkelsen stilte spørsmål om institusjonenes arbeid med politikk og retningslinjer for FAIR data, hva de anbefaler sine forskere å benytte av tjenester og verktøy, hvilke tilbud de eventuelt har utviklet, og hva de som institusjoner opplever som hindringer og behov/tiltak for mer FAIR forskningsdata. Vi opplevde også her god bredde i utvalget av besvarelser. Se også [egen rapport](#) fra helseforetakene.

Intervjuer og samtaler med forskere, ansatte ved institusjonene og tjenesteleverandører

Spørreundersøkelsene bidro med et godt oversiktsbilde av situasjonen og behovene i sektoren. For å skaffe til veie en dypere forståelse for problemstillingene gjennomførte vi intervjuer og samtaler med relevante aktører. Informantene var et strategisk utvalg av enkeltpersoner eller miljøer som vi, ut fra deres svar på undersøkelsene, antok kunne utdype temaer vi hadde behov for mer innsikt i – og som vi tok kontakt med. Det var blant annet forskere som hadde deltatt i spørreundersøkelsen og som hadde takket ja til å bli kontaktet for videre oppfølging.

Oversikt over gjennomførte aktiviteter:

- To gruppeintervjuer med totalt seks forskere som har erfaring med å håndtere persondata
- Fire individuelle intervjuer med forskere (ulike disipliner)
- Fire individuelle intervjuer med totalt tre personer i administrative stillinger ved institusjonene
- Fire intervjuer med tjenesteleverandører
- To intervjuer med miljøer som har kommet langt på området

Innspillsrunde, åpent innspillmøte og eget dialogmøte om helseforskning

16. desember la vi en første versjon av denne rapporten ut åpent tilgjengelig for innspill fra institusjoner, enkeltpersoner og andre aktører med interesse for temaet. Vi fikk totalt 42 skriftlige innspill til rapporten, 34 fra institusjoner/organisasjoner og 8 fra enkeltpersoner. 19. januar 2022 gjennomførte vi et åpent innspillmøte med presentasjon av innspill og åpen dialog. Totalt deltok (på det meste) 110 personer i møtet.

10. februar arrangerte vi et eget dialogmøte om helseforskning for å få dypere forståelse for spesielle problemstillinger knyttet til dette. Totalt deltok (på det meste) 41 personer i møtet.

Dokumentstudier

Som et utgangspunkt har vi gjennomgått en rekke dokumenter som til sammen har gitt et bilde av området og som har rettet vår oppmerksomhet mot eksisterende tjenester, regelverk, føringer og strategier, og ulike relevante problemstillinger knyttet til FAIR forskningsdata. Dokumentstudiene er en del av grunnlaget for diskusjonene og anbefalingene i denne

utredningen. Vi har gjennomgått en rekke relevante nasjonale strategier og planer, tilsvarende fra EU, retningslinjer, handlingsplaner, utredninger og sentrale prosjektbeskrivelser fra UHR, Forskningsrådet Sikt, Direktoratet for e-helse, RHFene, m.fl., tilsvarende fra enkelte sammenlignbare land i Skandinavia og Europa, samt en rekke nettsider fra forskningsinstitusjonene.

Presentasjoner

Utredningsleder og/eller medlemmer av sekretariatet har gitt en rekke korte presentasjoner i løpet av desember/januar 2022 for å gjøre oppmerksom på arbeidet og oppfordre til å gi innspill:

- UHR-Bibliotek
- UHR-Forskning
- RHFenes strategigruppe for forskning
- Nasjonalt forum for åpen forskning
- Forskningsrådets datagruppe
- UHR-Kunst, design og arkitekturs arbeidsgruppe for kunstnerisk utviklingsarbeid

Samkjøring med Datainfrastrukturutvalget

Det har vært lagt vekt på god samkjøring med det (i stor grad parallelle) arbeidet i Datainfrastrukturutvalget. Samkjøringen har vært gjennomført via flere tiltak:

- Katrine Weisteen Bjerde (HK-dir) har deltatt i sekretariatet i begge oppdragene
- Det har vært avholdt to samkjøringsmøter mellom ledelse og sekretariat i de to gruppene
- Materiale har hele veien blitt utvekslet mellom de to gruppene, og de har gitt innspill til hverandres rapporter og konklusjoner

3. ET RAMMEVERK FOR FAIR FORSKNINGSDATA

I dette kapitlet beskriver vi noen viktige trender knyttet til forskning og forskningsdata de siste årene, vi beskriver hvordan et økosystem for forskningsdata som følger FAIR-prinsippene ideelt sett bør være, og vi utdyper noen problemstillinger som har vært mye diskutert i løpet av utredningen.

3.1. MER DATA, MER KOMPLEKSITET

De siste tiårs digitale omstilling har gitt en voldsom vekst i mengden data som årlig skapes, oppstår, kopieres og brukes på global basis. Enkelte estimat⁴ antyder en eksponentiell vekst som forventes å øke ytterligere de nærmeste årene. Dataene oppstår og konsumeres kontinuerlig innenfor ulike deler av samfunnet og til ulike formål. Data kan for eksempel oppstå i det private næringsliv, i forvaltningen, eller samles inn og benyttes til forskningsformål. Noen eksempler på data som genereres kontinuerlig er data fra sosiale medier eller andre kilder på internett i form av tekst, bilder, lyd og video, og registre over banktransaksjoner, kjøpstransaksjoner, vannstandsmålinger og GPS-sporinger. Innen offentlig forvaltning genereres det saksbehandlingsdata og ulike typer registerdata, herunder helseregistre som oppdateres fra pasientjournaler og andre kliniske og administrative systemer. Innenfor fagfelt som for eksempel meteorologi, geologi, og klima-, atmosfære- og havfysikk, er overvåking og kontinuerlig datafangst gjennom for eksempel sensorer eller satellittbilder viktige kilder både for forvaltningens beslutninger og for forskning. Offentlige forskningsinstitusjoner genererer store mengder data både gjennom sin egen forskningsaktivitet og gjennom samarbeid med nasjonale og internasjonale samarbeidspartnere og næringsliv. Samtidig som vi ser en enorm vekst i mengden data, har stadig kraftigere datamaskiner bidratt til langt større kapasitet til å prosessere og analysere de dataene vi har. Maskinlæring og kunstig intelligens har satt ytterligere fart i utnyttelsen av data både for forskning og for andre formål.

Alle de datakilder som enten oppstår som følge av forskningsaktivitet, eller har potensiale til å kunne brukes til forskning, representerer også nye muligheter til å kombinere datakilder og koble data på tvers av fagfelt – og det har oppstått nye forskningsmetoder og analyseteknikker som gjør at kvantitative og kvalitative data kan brukes og kombineres på nye måter. Digitaliseringen har gjort det generelt enklere å samarbeide på tvers av landegrensene, og det har blitt mer nødvendig med et internasjonalt perspektiv både på samarbeidsformer og når det gjelder utformingen av verktøy og tjenester. Den heterogeniteten som alle nye datatyper, nye forskningsmetoder og nye verktøy innebærer, stiller i sin tur nye krav til kompetanse både hos forskere, forskningsinstitusjoner og tjenesteleverandører.

⁴ Statista: *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025:* <https://www.statista.com/statistics/871513/worldwide-data-created/>

3.2. MER FOKUS PÅ DELING AV FORSKINGSDATA

Samtidig med veksten både i mengden data og heterogeniteten i datatyper har det blitt mer fokus på åpen forskning og deling av forskningsresultater og forskningsdata. Initiativer fra internasjonale forskermiljø, som for eksempel [Research Data Alliance \(RDA\)](#) og [CODATA](#) har, sammen med nye retningslinjer og føringer fra både nasjonale og internasjonale forskningsmyndigheter og forskningsfinansierer, ført til økt bevissthet om behovet for å tilgjengeliggjøre og dele data. Både EU⁵, OECD⁶ og nasjonale myndigheter vektlegger nå dette på en helt annen måte enn før. Særlig EU har et betydelig fokus på deling av offentlige data, herunder også forskningsdata finansiert over offentlige budsjetter. EU vedtok i 2021 [Open Data Directive](#), som vil regulere viderebruk (av offentlige data) som et sentralt premiss for verdiskapning i et digitalt samfunn. Dette direktivet blir sannsynligvis implementert i norsk rett, og vil sammen med andre reguleringer (Data Governance Act⁷ og Data Act⁸) og bestemmelser om at spesielt viktige data i samfunnet (såkalt High Value Data⁹) skal deles gratis og være tilgjengelig for verdiskapende aktiviteter, føre til et ytterligere økt fokus på deling fra offentlige forskningsinstitusjoner.

I Regjeringens [Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata](#) heter det blant annet at: "...bedre tilgjengeliggjøring og mer gjenbruk av forskningsdata¹⁰ kan bidra til vitenskapelige fremskritt ved at hver enkelt forsker får et større datagrunnlag å bygge på, til økt etterprøvnbarhet og bedre kvalitetssikring av tidligere forskningsfunn, og til å unngå at samme type datainnsamling må finansieres flere ganger".

Strategien inneholder også tre grunnprinsipper:

- Grunnprinsipp 1: Forskningsdata skal være så åpne som mulig, så lukkede som nødvendig.
- Grunnprinsipp 2: Forskningsdata bør håndteres og tilrettelegges slik at verdiene i dataene kan utnyttes best mulig.
- Grunnprinsipp 3: Beslutninger om arkivering og tilrettelegging av forskningsdata må tas i forskerfellesskapene.

⁵ The EU's open science policy: https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en

⁶ OECD: *Making Open Science a Reality* (2015): <http://dx.doi.org/10.1787/5jrs2f963zs1-en>

⁷ <https://www.regjeringen.no/no/sub/eos-notatbasen/notatene/2021/juli/datastyringsforordningen-dga-/id2889711/>

⁸ <https://digital-strategy.ec.europa.eu/en/policies/data-act>

⁹ <https://data.europa.eu/en/datastories/high-value-datasets>

¹⁰ I strategien fokuseres det primært på *offentlige data* og på *offentlig finansierte* forskningsdata. Offentlige data er da definert som all informasjon som offentlige organer produserer, samler inn eller betaler for, mens offentlig finansierte forskningsdata er mer spesifikt data som er samlet inn eller frembrakt til bruk for eller som et resultat av offentlig finansiert forskning. I denne rapporten er vi mindre opptatt av distinksjonen mellom offentlig (finansierte) og ikke-offentlige data; vi forstår forskningsdata slik begrepet er definert blant annet hos Forskningsrådet, nemlig som *alle de registreringer, nedtegnelser, og/eller rapporteringer i form av tall, tekster, bilder og lyder som genereres eller oppstår underveis i forskningsprosjekter*, uavhengig av om det er data som er generert gjennom ny analyse, eller om det er sammenstilling av eksisterende data.

Strategien vektlegger med andre ord at det må gjøres en vurdering både av hvor åpne data kan være og av hvilken verdi det vil ha å gjøre dem tilgjengelige for gjenbruk og viderebruk. Disse beslutningene bør tas så nært forskermiljøene som mulig.

I denne sammenheng er det også verdt å nevne Stortingsmelding nr. 22, [Data som ressurs – Datadrevet økonomi og innovasjon](#) der det heter det at: “... regjeringens ambisjon er å få til økt deling av data innad i næringslivet og mellom offentlig og privat sektor”. Her påpekes det at åpenhet og deling av data ikke bare er viktig for forskning, men at datadelingen også vil fungere som kilde til innovasjon, nye forretningsmodeller og forbedring av eksisterende digitale tjenester.

3.3. EUROPEAN OPEN SCIENCE CLOUD (EOSC)

European Open Science Cloud (EOSC) er EU-kommisjonens viktigste tiltak for å etablere et økosystem for FAIR forskningsdata. Ambisjonen er å skape et nettverk for FAIR-kompatible data og tjenester ved å bygge på og samordne allerede eksisterende europeiske datainfrastrukturer og relaterte tjenester og ressurser.

EOSC vil være en sammenslutning av systemer som legger til rette for at forskere, via sine egne institusjonelle infrastrukturer, skal kunne finne, få tilgang til, dele og gjenbruke forskningsdata arkivert på hvilken som helst datainfrastruktur som er integrert i EOSC. Dette skal fungere på tvers av fagdisipliner, sektorer og landegrenser. For å nå dette ambisiøse målet kreves det betydelig standardisering og godt integrerte tekniske løsninger, retningslinjer og styrende prinsipper, samtidig som at fagdisiplinenes egenart og behov blir ivarettatt.

I årene siden 2016 har EU-kommisjonen i samarbeid med medlemslandene finansiert pilotprosjekter, gjennomført ekspertutredninger og holdt dialogmøter med en rekke aktører for å kartlegge hva som skal til for å realisere EOSC. Det har kommet klart frem at forskningspolitiske føringer, delingskultur og tekniske løsninger må utvikles i takt med hverandre i årene fremover for at EOSC skal fungere godt for forskerne og skape den ønskede merverdi for forskning, forvaltning og næringsliv.

Videreutvikling av EOSC ble i 2021 formalisert gjennom et såkalt 'co-programmed' europeisk partnerskap. En del av partnerskapsstrukturen er EOSC Association, der europeiske forskningsmiljøer og institusjoner kan bli medlemmer. EOSC Association involverer i dag et bredt spektrum av interesserte aktører med sine over 200 medlemmer og observatører, fem rådgivende grupper og 13 arbeidsgrupper. I disse gruppene kan enkeltpersoner søke om å delta, og slik være med på å påvirke den videre utviklingen. Det er viktig at utviklingen av nasjonale datainfrastrukturer og tjenester for håndtering av forskningsdata i årene fremover skjer i takt med utviklingen av EOSC, dersom nasjonale investeringer skal kunne skape verdier i samspill med den europeiske utviklingen. Det er også viktig at norske forskningsmiljøer bidrar aktivt i dette arbeidet og sikrer at norsk kompetanse innen datadeling blir brukt til å videreutvikle EOSC. En slik aktiv deltakelse vil også styrke implementeringen av EOSC og økosystemene for FAIR-kompatible forskningsdata ved norske forskningsinstitusjoner.

3.4. ENDRINGER I RETNINGSLINJER, LOVER OG REGLER

Parallelt med - og til dels som en konsekvens av - større datavolum, mer komplekse data, samt økt søkelys på åpenhet på deling - har det juridiske og forskningsetiske rammeverket også endret og utviklet seg. I takt med at mulighetene til å samle inn store mengder persondata har økt, har også lovgivningen som regulerer dette endret seg. Den nye personopplysningsloven som ble innført i 2018, implementerer EUs personvernforordning (GDPR) i Norge. Endringene har innvirkning på forskningsaktivitet, blant annet på grunn av de strenge sanksjonsmulighetene som den nye loven innebærer. Samtidig har det, dukket opp nye problemstillinger knyttet til opphavsrett¹¹, personvern og forskningsetikk. Teknologiutviklingen, nye metoder og mulighetene for å kombinere datakilder setter press på personvernet og på datasikkerhet generelt.

Mange forskere opplever en krevende balansegang mellom ønsker og krav om åpenhet og datadeling på den ene siden, og etterlevelse av lover, regler og retningslinjer på den andre siden. Vektingen av disse ulike hensynene kan for noen oppleves som et hinder for effektiv utøvende forskning, spesielt dersom både forskeren selv og forskningsinstitusjonen de tilhører mangler tilstrekkelig juridisk og annen relevant kompetanse. For at bruken av data skal skje på en lovlig og ansvarlig måte er det avgjørende at kompetansen for å foreta de formelle juridiske og forskningsetiske avklaringer er tilgjengelig.

Noen av de mest sentrale spørsmålene som må besvares er:

- Hvilke data har verdi for fremtiden; hva skal lagres, hva skal tas vare på?
- Hvordan ivareta sikkerhet, personvern og etterlevelse av lovverk både når data genereres og når data skal arkiveres?
- Hvor og når skal data lagres og arkiveres?
- Når kan data deles åpent og når bør data deles med begrensninger? Og når bør - eller kan - data ikke deles?
- Hva må til for at data rent faktisk kan gjenfinnes, forstås og brukes av andre og til andre formål enn det dataene opprinnelig ble brukt til?
- Hvem skal betale? Det koster å finansiere forskning og å finansiere utstyr som brukes til å generere, samle inn og strukturere data.
- Hvordan sikre at forskere og forskningsinstitusjoner prioriterer deling av forskningsdata i henhold til FAIR-prinsippene?

Disse utfordringene imøtekommes gjennom et samspill mellom den enkelte forsker og øvrige aktører og tjenestetilbydere i økosystemet for forskning. Vi skal i det følgende kort drøfte dette samspillet ved å se på hvordan man i forskningsdataenes livssyklus kan jobbe for å gjøre data FAIR.

¹¹ Se lisensutvalgets rapport: *Hvordan skal vi dele forskningsdata? Utredning og anbefalinger om lisensiering og tilgjengeliggjøring* (2021): https://www.forskningsradet.no/siteassets/publikasjoner/2021/rapport_rettighets-og-lisenssporsmal-ved-tilgjengeliggjoring-av-forskningsdata.pdf

Det er også et spørsmål hvor man trekker grensen mellom forskningsdata og forvaltningens dataressurser. I økende grad blir datasett en del både av forskningsdomenet og av statlige etaters forvaltningsansvar (for eksempel innen klima- og miljøforskningen), og en del forskningsdata smelter sammen med forvaltningens dataressurser (for eksempel Statistikkbanken til SSB, registerforskningen i helsesektoren og planene om at forskningsdata og data fra næringslivet skal berike forvaltningens data slik det er planlagt i Helseanalyseplattformen). Selv om det fortsatt er temaer som er spesielle for forskningssektoren (blant annet forholdet til publisering og merittering), vil de fleste prinsipper og retningslinjer for datadeling i stor grad være de samme i forskning og forvaltning. Derfor er det viktig med et samarbeid på tvers av sektorer.

3.5. LIVSSYKLUS FOR FORSKNINGSDATA

Forskning organiseres på ulike måter i ulike fagdisipliner og ved ulike institusjoner. Likevel kan det være nyttig å tenke på forskningsarbeidet og forskernes håndtering av data som en syklus med ulike stadier der det må tas ulike hensyn i de ulike fasene, for å sikre god håndtering av forskningsdata. Modellen under illustrerer de ulike fasene av forskningsprosessen fra innledende idé til avsluttet prosjekt.



FIGUR 3: LIVSSYKLUS FOR FORSKNINGSDATA. CC-BY NTNU

En slik skjematisk fremstilling er naturligvis en sterkt forenklet og teoretisk modell. Underveis i et prosjekt vil det ikke nødvendigvis være klare skiller mellom de ulike stegene, og det vil også være forskjeller på hva som inngår i de ulike stegene – avhengig av prosjekters spesifikke utforming. Det vil også være forskjellene som kan tilskrives variasjon i praksis innenfor ulike fagområder og forskningstradisjoner. Det vil for eksempel kunne være forskjeller mellom

forskning som primært baserer seg på eksisterende datakilder, og forskning der datainnsamlingen utgjør en sentral del av arbeidsprosessen. Likevel kan slike syklusmodeller fungere som nyttige rammeverk i arbeidet med å gjøre data FAIR.

3.6. FAIR DATA I ET ØKOSYSTEM FOR FORSKNING

Underveis i forskningsprosessen må det tilrettelegges for at alle som ønsker å bruke data møter så få hindringer som mulig når de skal finne, forstå og få tilgang til data. FAIR-prinsippene er utviklet med dette for øyet: å bygge ned tekniske og praktiske barrierer slik at data blir gjort tilgjengelige og kan lokaliseres, forstås og brukes av andre. Prinsippene ble presentert første gang i artikkelen [The FAIR Guiding Principles for scientific data management and stewardship](#). FAIR data er data som kan gjenfinnes (Findable), er tilgjengelige (Accessible), er kompatible (Interoperable) - altså at de kan samhandle eller samvirke med andre data - og som kan gjenbrukes (Reusable).

Som EU-kommisjonens ekspertgruppe for FAIR data påpeker i rapporten [Turning FAIR into reality](#), handler FAIR om at data består av flere lag med informasjon som gir kontekst og til sammen gir data merverdi og muliggjør gjenbruk. For å gjøre data FAIR, må de utstyres med unike og varige (persistente) identifikatorer (PID), de må benytte relevante og fagspesifikke standarder, formater, ontologier og kontrollerte vokabular og de må være utstyrt med rike metadata og lisenser som tydelig definerer vilkår for bruk. FAIR-prinsippene innebærer også at både data og metadata skal være maskinlesbare og maskinhåndterbare.

Som Datalisensutvalget påpekte i sin rapport, gjelder FAIR-prinsippene for alle data, uavhengig av om de kan deles åpent eller ikke. FAIR-prinsippene innebærer ikke nødvendigvis at data skal være åpne, og åpne data er ikke nødvendigvis FAIR – selv om begge deler som regel er ønskelig¹². FAIR-prinsippene gjelder med andre ord også for data der det må legges begrensninger på tilgang og for data som kun brukes internt i en organisasjon eller i et prosjekt. Data vil uansett ha høyere bruksverdi om de er FAIR. Blant annet gjør FAIR metadata at det er mulig å finne frem (Findable) til at dataene faktisk finnes, selv om det ikke uten videre er mulig å få tilgang til dem.

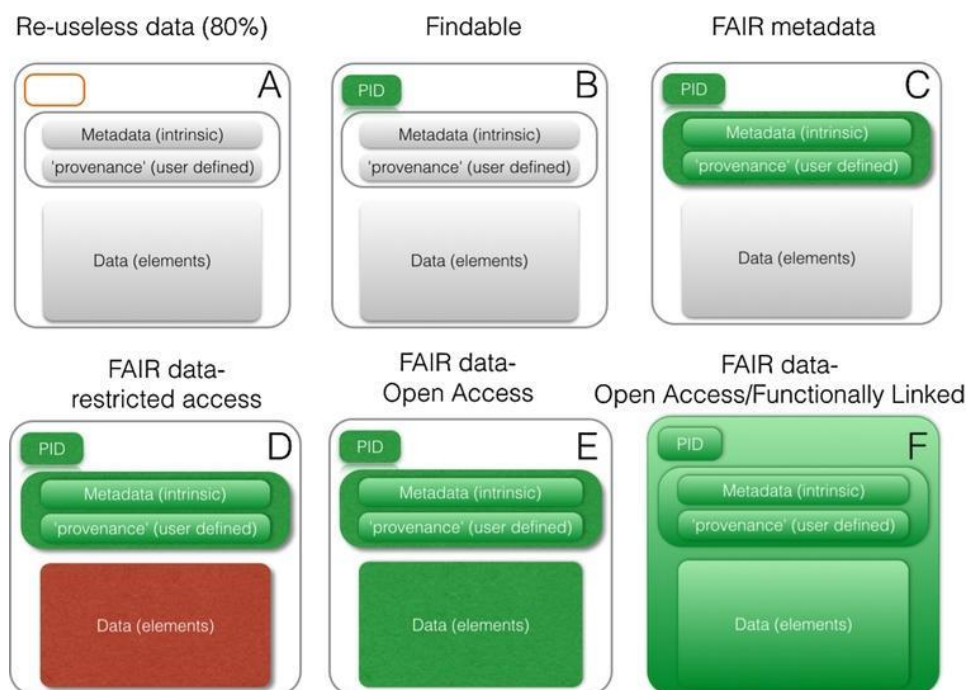
Figuren under er hentet fra [Mons et.al \(2017\)](#) og illustrerer hvordan FAIR og åpenhet består av flere ulike nivåer eller trinn, med ulike grader av FAIR og ulike grader av åpenhet. I kategori (A) mangler både data og metadata en PID og er dermed utilgjengelige og 'ugjenbrukelige' (re-useless). Ifølge samme artikkel er det anslått at rundt 80 % av data som genereres i dag faller i denne kategorien. En overveldende majoritet av de data som skapes, kan med andre ord aldri gjenbrukes.

Det absolutte minimum av tilrettelegging for at data kan lokaliseres, er å knytte en varig eller persistent identifikator (PID) til dataene, slik at det er mulig å finne dem (kategori B). En PID er nødvendig, men i seg selv ikke tilstrekkelig. Det må også være maskinlesbare metadata knyttet

¹² Lisensutvalget, se lenke i fotnote 9.

til dataene (kategori C) for at de skal kunne regnes som FAIR. I kategori D er dataene maskinlesbare og utstyrt med både PID og maskinlesbare metadata, mens tilgangen til selve dataene er begrenset. Dersom også dataene gjøres åpent tilgjengelige og utstyres med en lisens som definerer vilkår for gjenbruk av andre (kategori E), får vi det som per i dag kan regnes som det høyeste nivået av FAIR for de fleste data. Når alle disse dataene i tillegg kan knyttes til og funksjonelt behandles av et stort antall av relevante tjenester og applikasjoner, kan vi begynne å snakke om et fungerende økosystem av FAIR data og tjenester (kategori F).

Data as increasingly FAIR Digital Objects



FIGUR 4: GRADER AV FAIR DATA. KILDE: BAREND MONS ET AL. (2017). CC-BY 4.0

Både prosjektet [FAIRsFAIR](#) og den nevnte rapporten [Turning FAIR into reality](#) identifiserer essensielle komponenter i et FAIR økosystem. Sistnevnte definerer dette økosystemet som en modell¹³ som består av alle de komponenter som er nødvendig for å danne et system der FAIR data kan skapes, forvaltes og gjenbrukes på en effektiv og bærekraftig måte. I et slikt system er forskningsdataene understøttet og ledsaget av følgende essensielle komponenter: Politikk og retningslinjer, datahåndteringsplaner, identifikatorer, standarder og dataarkiver og andre relevante tjenester.

I økosystemet vil retningslinjer defineres av forskningsfelleskapene, forskningsinstitusjoner, finansører, forlag, tidsskrift, og andre aktører som setter rammene for forskning. Datahåndteringsplaner utarbeides av forskere eller dataforvaltere og beskriver og skaper koblingspunkter for de ulike FAIR-komponentene i prosjektet (data, metadata og annen

¹³ Se vedlegg for en skematisk fremstilling (under definisjon av FAIR økosystem).

dokumentasjon). Alle elementene i systemet har unike identifikatorer, inkludert data, applikasjoner, metadata, datahåndteringsplaner, forskere, prosjekter, forskningsinstitusjoner og finansører. Standarder og maler vil være relevante både for metadata, vokabularer, ontologier, lisenser, datahåndteringsplaner og for tjenesteleverandører (gjennom for eksempel sertifiseringer eller standarder for informasjonssikkerhet). Tanken er også at i et slikt økosystem skal det være mulig å kunne måle og vurdere pålitelighet både til tjenester og dataobjekter, gjennom evalueringsverktøy og andre kvalitetsindikatorer. Alle elementer må være minst på nivå C for å kunne inngå i et slikt fullintegrert økosystem.

Selv om er slikt økosystem foreløpig er en framtidsvisjon, finnes det allerede mange initiativ og tjenester som drar i retning av integrering og interoperabilitet mellom data og mellom tjenester. [EOSC](#), [FAIRsFAIR](#) og [OpenAIRE](#) er sentrale internasjonale initiativer og det finnes internasjonale fellesregistre for informasjon om policyer, standarder og dataarkiv¹⁴. I tillegg er det etablert rammeverk for sertifisering av dataarkiv (som for eksempel [CoreTrustSeal](#)) og det finnes verktøy som kan brukes til å evaluere hvor FAIR data er¹⁵. EOSC Executive Board FAIR working group har også publisert rapporten [Recommendations on FAIR Metrics for EOSC](#) med anbefalinger for hvordan man skal måle FAIRness. Det er også verdt å nevne TRUST-prinsippene (TRUST Principles for digital repositories) som er tenkt som et rammeverk for å implementere beste praksis innen bevaring av digitale ressurser (inkludert forskningsdata).

Utviklingen av et FAIR økosystem for det offentlige forskningssystemet er i en tidlig fase og i stadig utvikling. Mange av de tjenestene og verktøyene som er i bruk eller under utvikling, er nært tilknyttet forskermiljøene. Få av løsningene er sentraliserte (gjennom for eksempel nasjonale fellesløsninger). Snarere er det slik at løsninger utvikles som et distribuert nettverk av tjenester der hver løsning utvikles med fokus på lokale og fagspesifikke utfordringer. Dette har man god erfaring med, og den videre utviklingen planlegges på samme måte.

Samtidig innbefatter utviklingen av økosystemet et samspill mellom mennesker, prosesser og teknologi, og det vil stilles høye krav til kompetanse og kunnskap for å få det til å fungere godt. Dette er områder der det vil være nyttig og kostnadssparende å samarbeide både nasjonalt og internasjonalt.

FAIR-prinsippene fokuserer utelukkende på egenskaper ved data som vil gjøre det lettere å dele, finne og gjenbruke data. De bør derfor komplementeres med andre retningslinjer og rammeverk som tar hensyn til forskningsetikk og mulige bakenforliggende maktstrukturer og historiske kontekster. Et eksempel er CARE-prinsippene (CARE Principles for Indigenous Data Governance) som er utarbeidet for å ta hensyn til ulike aspekter ved forvaltning og bruk av data som er tilknyttet urbefolkning og urfolk.

¹⁴ Se for eksempel <https://fairsharing.org/> og <https://www.re3data.org/>

¹⁵ Se for eksempel [FAIR Evaluation Service](#) og [F-UJI Automated FAIR Data Assessment Tool](#)

3.7. DATAHÅNTERINGSPLANER

Med datahåndtering mener vi aktiv planlegging, organisering og vedlikehold av data gjennom hele forskningsprosessen og livssyklusen for forskningsdata, inkludert planlegging, datainnsamling, prosessering, analyse, lagring, deling, arkivering og gjen- og viderebruk. En datahåndteringsplan (ofte kalt DMP - Data Management Plan) er et dokument som beskriver hvordan håndteringen av data i et bestemt forskningsprosjekt skal foregå, både underveis i prosjektet og etter at prosjektet er avsluttet.¹⁶

Datahåndteringsplaner er et helt sentralt element i økosystemet for FAIR forskningsdata. I tillegg til at datahåndteringsplanene kan brukes til å skape koblingspunkter for de ulike FAIR-komponentene i et prosjekt, er de også et nyttig verktøy og virkemiddel som bidrar til både til planlegging og bevisstgjøring rundt forvaltning og håndtering av forskningsdata i prosjektfasen og etter prosjektets slutt.

En datahåndteringsplan kan bidra til økt kvalitet og kontroll over forskningsdataene, og god planlegging bidrar også til at det er lettere å etterleve alle relevante lovverk og forskningsetiske retningslinjer og ivareta informasjonssikkerhet. Mange forskere i Norge opererer i internasjonale miljøer og samarbeidsprosjekter. I slike sammenhenger er det ekstra stort behov for å planlegge og gjøre nødvendige avklaringer knyttet til deling av data både underveis i prosjektet og etter at det er avsluttet. Konsepter som personvern, anonymisering og etikk kan ha ulik betydning og meningsinnhold for forskere i ulike land. I tillegg kan forskningen være underlagt ulike lovverk og ulike nasjonale retningslinjer og forskningsetiske rammeverk. En datahåndteringsplan kan bidra til avklaringer rundt dette. I tillegg kan den brukes til å planlegge innhenting av gyldige behandlingsgrunnlag og nødvendige tillatelser fra relevante nasjonale tilsyn.

En datahåndteringsplan kan også bidra til å planlegge for eventuelle kostnader, for eksempel knyttet til innsamling, lagring eller arkivering, samt avklare roller og ansvarfordeling. Planen bør være et 'levende' dokument som oppdateres og brukes aktivt underveis i prosjektet.

En datahåndteringsplan kan i seg selv betraktes som et digitalt objekt i et økosystem av mer eller mindre FAIR digitale objekter. Merverdien av en datahåndteringsplan avhenger også av hvor FAIR planen i seg selv er. Datahåndteringsplaner som er utarbeidet i for eksempel Word-dokumenter vil normalt ha liten grad av FAIRness. For å oppnå høyere grad av FAIRness må datahåndteringsplanene gjøres maskinhåndterbare (machine-actionable). Dette innebærer å benytte standardiserte maler for datahåndteringsplaner samt benytte standardisert semantikk og formater der det er mulig og relevant. Dette vil gjøre at planens innhold er søkbart, tilgjengelig og mulig å inkorporere i eller utveksle med andre systemer på en god måte. Dette vil gi merverdi for alle aktører som behøver å utveksle informasjon, slik som forskere, finansører og dataarkiv. For forskere vil merverdien være at datahåndteringsplanen også blir et reelt

¹⁶ En datahåndteringsplan må ikke forveksles med en personvernkonsekvensvurdering (DPIA – Data Protection Impact Assessment), som er en prosess som skal beskrive behandlingen av personopplysninger, og vurdere om den planlagte behandlingen er nødvendig og proporsjonal.

arbeidsdokument og informasjonskilde for datahåndteringen i forskningsprosjektet. Det vil fasilitere senere kuratering og deling av dataene, og gjøre datahåndteringsplanen til en integrert del av forskningsprosessen og eksisterende arbeidsflyter.

For å lage maskinhåndterbare datahåndteringsplaner kreves blant annet en felles datamodell¹⁷. En arbeidsgruppe i regi av [Research Data Alliance](#) (RDA) har for eksempel utarbeidet en standardmodell som brukes av flere verktøy og tjenester for datahåndteringsplaner, både nasjonalt og internasjonalt.

Siden 2017 er åpen tilgang til forskningsdata og krav om datahåndteringsplaner standard i EUs forsknings- og innovasjonsprogram Horisont 2020. Dette tiltaket videreføres i Horisont Europa, der det er krav om å etablere og jevnlig oppdatere prosjekttilknyttede datahåndteringsplaner.

I Kunnskapsdepartementets [Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata](#) ligger det en forventning om at forskningsinstitusjonene bidrar til å heve kompetansen hos ansatte og studenter gjennom opplæring i datahåndtering og gjenbruk av data, samt en forventning om at forskningsinstitusjonene utvikler prosedyrer for godkjenning av datahåndteringsplaner.

I sin [policy for tilgjengeliggjøring av forskningsdata](#) krever Norges forskningsråd at godkjente datahåndteringsplaner skal foreligge ved kontraktsinngåelse for prosjekter som genererer data.

3.8. KLASSIFISERING AV DATA

Som ledd i planleggingen og for å ivareta informasjonssikkerhet og personvern er det viktig å klassifisere dataene i henhold til hvordan de må behandles. Den vanligste måten å klassifisere data på er gjennom ulike konfidensialitetskategorier, nemlig *åpen* (grønne data), *intern* (gule), *fortrolig* (røde) og *strengt fortrolig* (sorte)¹⁸. En klassifisering basert på konfidensialitet er et viktig grep både for å kunne velge riktige løsninger for lagring og prosessering av data underveis i et prosjekt, og for å velge riktige løsninger for arkivering og deling av data etter prosjektslutt.

Normalt vil prosjektleder stå for de vurderinger som ligger til grunn for å plassere forskningsdata i en gitt kategori, men siden det i de aller fleste tilfeller vil være institusjonen som er behandlingsansvarlig og har den formelle styringsretten over dataene, er det viktig at klassifiseringen skjer i samråd mellom institusjon og forsker. Klart definerte institusjonelle retningslinjer med veiledere for informasjonsklassifisering og rutiner for behandling og oppbevaring av data som trenger beskyttelse, er i så måte essensielt. Slike institusjonelle policyer og retningslinjer bør, som alt annet i et velfungerende økosystem for forskningsdata, være tilrettelagt for å kunne være maskinhåndterbare, slik at informasjonen i dem kan gjenbrukes og innpasses med andre systemer.

¹⁷ Forslag til prinsipper for dette finnes for eksempel i følgende artikkel:

<https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1006750&type=printable>

¹⁸ Se blant annet Uninetts veileder for [klassifisering av informasjon](#).

3.9. LAGRING OG ARKIVERING AV FORSKNINGSDATA

Et annet viktig element i et økosystem for forskningsdata er tjenester som tilbyr lagring og arkivering av data. Vi skiller mellom lagring av data underveis i aktive forskningsprosjekter og arkivering/langtidsbevaring av data etter endt prosjekt.

3.9.1. Håndtering og lagring av data i aktive prosjekter

Lagring handler i denne sammenheng om oppbevaring av data underveis i et prosjekt. Data kan prosesseres, analyseres, endres, deles med samarbeidspartnere eller på andre måter brukes operativt underveis i prosjektet og kan bestå av både rådata og prosesserte data. En slik form for lagring handler også om å oppbevare sikkerhetskopier som kan brukes til gjenoppretting av tidligere versjoner dersom data går tapt eller på andre måter blir ødelagt underveis i prosessen.

Lagringstjenester kan tilrettelegge for FAIR forskningsdata ved å tilby funksjonalitet for å berike dataene med metadata underveis i pågående prosjekter. Det vil både gi forskerne bedre oversikt over egne data underveis i prosjektet og forenkle arbeidet dersom dataene skal arkiveres og/eller publiseres på et senere tidspunkt. Tjenestene bør da også sikre interoperabilitet ved at dataene på en enkel måte kan overføres til andre tjenester med sine tilhørende metadata. I samarbeidsprosjekter er det også viktig at tjenestene har funksjonalitet for å dele data underveis i prosessen.

3.9.2. Arkivering og langtidsbevaring av data etter prosjektslutt

Arkivering innebærer også lagring og oppbevaring, men omfatter i tillegg et eksplisitt langtidsperspektiv. Dette handler om å ivareta, opprettholde og tilgjengeliggjøre data for relevante grupper eller (fag)felleskap over en lang periode, gjerne på ubestemt tid. Arkivering skjer normalt i avslutningsfasen av et forskningsprosjekt. Man gjør da et utvalg av de dataene som er samlet inn og håndtert gjennom prosjektet, basert på avveininger og vurderinger av deres verdi for fremtiden. De dataene som arkiveres anses altså å ha verdi både for fremtidig forskning og/eller som grunnlag for validering av forskningsresultater. Ofte utarbeides det også statiske («kalde») versjoner av materialet som skal arkiveres¹⁹. Arkivering innebærer også ofte at forvaltningsansvaret for dataene overføres til arkiveier, men det finnes også arkiver der forskermiljøet fortsetter å ha ansvar for (deler av) forvaltningen.

Etter hvert som både volum, kompleksitet og heterogenitet i forskningsdata har økt, har det dukket opp flere dedikerte plattformer for publisering og deling av data og metadata. Der det tidligere fantes få eller ingen egnede plattformer for publisering eller deling av data, finnes det i dag en rekke løsninger; fra de generiske tjenestene som er åpne for alle typer data fra et bredt spekter av fagområder, til de mer fag- og domenespesifikke som er spesialisert for visse typer data.

¹⁹ Det er imidlertid ikke alltid et absolutt skille mellom 'varme' og 'kalde' data: i noen tilfeller kan for eksempel også arkiverte data oppdateres med nye versjoner av datafiler eller med oppdateringer i metadata. I tillegg kan man også i den aktive fasen av prosjektet opprette ulike versjoner av data som har ulik grad av permanens.

Forskningssektoren produserer stadig økende mengder forskningsdata. Noe av dette er rådata som ikke kan eller bør beholdes blant annet som følge av stort volum, mens det i andre sammenhenger er viktig å beholde originaldata som dokumentasjon på de mer bearbejdede dataene. Forskningsdata kan ha langsiktig verdi både for forskning og til andre formål, også lenge etter prosjektslutt. Klart definerte prioriteringer og seleksjonskriterier er derfor viktig for å kunne avgjøre hvorvidt data har verdi for langtidsbevaring og dermed bør gjøres FAIR. Grunnlaget for utvelgelse av data og fastsetting av kriterier vil i stor grad variere mellom ulike fagfelt, så dette er beslutninger som i all hovedsak bør gjøres av fagmiljøene. I en del tilfeller vil det være fornuftig at beslutningene tas i samråd med institusjonene eller et relevant dataarkiv, og deling kan være forpliktet gjennom betingelser i tildeling for eksterntfinansierte prosjekter. I et økosystem for forskningsdata er et dataarkiv forstått som en tjeneste som sørger for langtidsbevaring og tilgjengeliggjøring av digitale objekter med forskningsverdi²⁰.

Noen arkivløsninger kan være basert på selv-deponering (altså at forskere selv laster opp data og legger på metadata og annen dokumentasjon), mens andre løsninger tilbyr veiledning og aktiv kuratering av data. Kuratering vil si å innordne materiale i vitenskapelige eller andre faglige system, og gjøres ofte i for eksempel et dataarkiv. Kuratering handler i så måte om kvalitetssikring, dokumentering, standardisering og formatering av data. Det innebærer ofte også å legge på metadata, lage og ha kontroll over ulike versjoner av data, aggregering eller omkodning av data, og danne nye datasamlinger med utgangspunkt i ulike datakilder. Langtidsbevaring innebærer dessuten å ta hensyn til endringer og skiftende teknologier, nye (oppbevarings)medier og nye formater, og å holde data og tilhørende informasjon vedvarende tilgjengelig og forståelig for relevante brukergrupper – uavhengig av disse endringene.

3.10. ARKIVERING OG GJENBRUK AV DATA MED BEHOV FOR BESKYTTELSE

Ikke alle data kan deles fritt og åpent, men må beskyttes i større eller mindre grad eller ikke kan deles i det hele tatt. Årsaker kan for eksempel være personvern hensyn, kontraktsforpliktelser overfor oppdragsgiver, ønske om å sikre personvern, å beskytte kommersiell, forretningsmessig eller annen virksomhetssensitiv informasjon, og sikkerhetshensyn.

Det er viktig at forskere og forskningsinstitusjoner har et bevisst forhold til slike beskyttelsesbehov og at det planlegges for at data håndteres riktig i henhold til gjeldende lover og regelverk. Data med beskyttelsesbehov må plasseres i sikre arkiver med tilgangsstyring. Selv om ikke data med beskyttelsesbehov kan deles åpent, vil det likevel i mange tilfeller være mulig å gjøre dataene FAIR og tilgjengelige for gjenbruk ved at metadataene gjøres åpent tilgjengelige. Det vil gjøre at dataene kan gjenfinnes, og de som har legitime grunner, kan så søke om å få tilgang. På denne måten sikrer man at samfunnet får merverdi også av data som har beskyttelsesbehov.

²⁰ Det innebærer at arkiver som oppbevarer andre ting enn forskningsdata og/eller som håndterer og tar vare på gjenstander og fysiske objekter, er med andre ord ikke er omfattet av denne definisjonen og er derfor heller ikke omtalt i denne rapporten.

For data med personopplysninger må det tas spesielle hensyn. Personvernforordningen setter rammer for deling av personopplysninger, og det er egne bestemmelser for overføring av personopplysninger utenfor EU/EØS-området. Fallers prosjektet inn under helseforskningsloven, er normalen at data med personopplysninger skal anonymiseres eller slettes når prosjektet er avsluttet. Oppbevaring etter prosjektslutt er i utgangspunktet kun lov for etterprøving av resultatene, men det er mulig å søke om unntak, for eksempel ved å be om samtykke til et bredere formål enn dette ene prosjektet.

For å få samtykke til gjenbruk må man planlegge for dette fra starten og ta det med i samtykkeskrivet som informantene signerer på. Her kan det være en balanse mellom ønsket om å kunne tilgjengeliggjøre data for gjenbruk og bekymring for at informantene da vegrer seg for å gi samtykke²¹. Innen medisinsk forskning er det imidlertid indikasjoner på at pasientene ønsker at data blir utnyttet så godt som mulig. Gjenbruk har også en etisk side ved at data fra informantene kommer mer til nytte og reduserer belastningen på fremtidige informanter.

Kvalitative data er ofte både komplekse og personidentifiserende, for eksempel i form av detaljerte feltnotater, intervjuer, video, lyd, biologiske parametere og medisinske data. Det å tilgjengeliggjøre kvalitative data for gjenbruk reiser dermed ofte en del problemstillinger, samtidig som det nettopp er ønskelig med gjenbruk – siden kvalitative data ofte er kostnadskrevenende å innhente og lagre. Igjen vil deler av løsningen kunne finnes i god planlegging, innhenting av samtykker, anonymisering og i å lage gode metadata som kan deles.

Deling av data fra kunstnerisk utviklingsarbeid og andre data som inneholder bilder, video og lyd er ofte regulert av opphavsrett, som i mange tilfeller kan gjøre gjenbruk utfordrende.

²¹ Dette kan løses ved for eksempel å gi informantene valget mellom to samtykkeskjemaer, der det ene gir åpning for gjenbruk, mens det andre er avgrenset til angjeldende studie.

4. BEHOV FOR VIDERE ARBEID

I dette kapitlet beskrives behov for og forslag til videre arbeid knyttet til ulike aspekter ved håndtering av forskningsdata, og som har blitt identifisert gjennom arbeidet med denne utredningen.

I slutten av hvert underkapittel oppsummeres behovene og det gis anbefalinger som i så stor grad som mulig er adressert til de vi mener er rette mottakere. Disse er:

- **Kunnskapsdepartementet**, spesielt med tanke på den kommende Langtidsplan for forskning og høyere utdanning (heretter kalt LTP), men også med tanke på annen bruk av virkemiddelapparatet og styring av universiteter og høyskoler.
- **Helse- og omsorgsdepartementet** som ansvarlige for helseforskningslovgivingen, men også med tanke på annen bruk av virkemiddelapparatet og for styring av helseforetakene.
- **Direktoratet for høyere utdanning og kompetanse (heretter kalt HK-dir)** som ansvarlige for å lage handlingsplan for implementering av Strategi for digital omstilling i universitets- og høyskolesektoren og i rollen som samordner og pådriver for åpen forskning på tvers av sektorer.
- **Sikt** som tjenesteleverandør til høyere utdanning og forskning og som koordinatører for den norske noden av RDA.
- **Forskningsrådet** som den viktigste finansiøren av forskningsinfrastruktur og forskningsprosjekter, som norsk «mandated organisation» i EOSC-stiftelsen og som tildeler av grunnbevilgning til forskningsinstituttene.
- **Forskningsinstitusjonene** (UH-institusjoner, forskningsinstitutter og helseforetak), eventuelt via deres interesseorganer UHR, FFA og RHFenes strategigruppe for forskning
- **Tjenesteleverandørene**. I noen tilfeller er det mer naturlig å henvise til den samlede gruppen av tjenesteleverandører heller enn å peke på en enkelt av dem. Dette kan være Sikt, Sigma2, institusjoner som tilbyr tjenester, mm.

Vi omtaler også noen interessenter og potensielt viktige samarbeidspartnere utenfor sektoren – slik som Digitaliseringsdirektoratet (Nasjonalt ressurscenter for deling av data), Direktoratet for e-helse, Datatilsynet og de regionale og nasjonale etiske komiteene.

I tillegg til juridiske organisasjoner som kan ta formelt ansvar for tiltakene, nevner vi her noen relevante diskusjons- og samarbeidsarenaer der tiltak også kan diskuteres og videreutvikles:

- **Nasjonalt forum for åpen forskning**. Formålet er å samle sentrale aktører jevnlig (ca 4 ganger i året) for gjensidig oppdatering på hva som skjer på området åpen forskning. Forumet er etablert som et samarbeid mellom UHR, HK-dir (tidligere Unit) og Forskningsrådet, og mandatet for forumet er forankret i ledelsen i de tre organisasjonene. I møtene deltar representanter for disse tre organisasjonene og i

tillegg representanter for institusjonene (UH, institutt og helseforetak), KD, HOD, Sikt, Forskerforbundet og NSO.

- Persondataforum. Formålet er å samle ledere tilknyttet infrastrukturer som håndterer persondata for å diskutere relevante felles problemstillinger og temaer for videreutvikling. Forumet ble etablert i et samarbeid mellom Forskningsrådet, HK-dir (tidligere Unit) og Direktoratet for e-helse. I starten var mange av diskusjonene knyttet til utviklingen av Helseanalyseplattformen, men etter hvert har andre temaer også blitt mer aktuelle. Forumet har vekselvis møter på ledernivå og operativt nivå. Lederne samles normalt 2 ganger pr år, og det operative nivået 2 ganger ekstra pr år.
- RDA (Research Data Alliance) og RDA Norge. RDA er en internasjonal organisasjon, primært for forskere, der alle mulige temaer knyttet til forskningsdata tas opp. Det etableres jevnlig arbeidsgrupper som over en kort periode arbeider med et tema og skal frembringe en eller flere leveranser. Det er også mulig å delta i interessegrupper rundt spesifikke temaer.
- Nasjonalt EOSC-forum. Dette er et forum under etablering. Forumet drives av Forskningsrådet i rollen som «mandated organisation» i EOSC Association. Planen er å samle

4.1. KUNNSKAPSBEHOV OG KOMPETANSE

På spørsmål om hva de mente var de viktigste tiltakene for å oppnå mer deling av forskningsdata, plasserte langt de fleste institusjonene kompetanseheving og økt tilgang på veiledningstjenester på toppen av listen.

Grunnleggende kunnskap om FAIR og verdien av å gjøre forskningsdata gjenbrukbare

Innen flere fagfelt er det å tilgjengeliggjøre data for gjenbruk allerede en innarbeidet tradisjon, og det er godt etablert kunnskap blant aktørene at dette gir gevinster. I offentlig forvaltning er deling av data i ferd med å bli et grunnleggende premiss og en integrert del av virksomhetenes samfunnsoppdrag – selv om deler av forvaltningen fortsatt sliter med begrensende lovverk, manglende kompetanse og ressurser, og ikke minst mangelfulle tekniske løsninger for å imøtekomme forventningene som blant annet er formulert i EUs nye data-relaterte direktiver. Innen forskningen ser vi at arkivering og gjenbruk av forskningsdata er en helt ny problemstilling i mange miljøer. Det er derfor behov for bevisstgjøring om samfunnsnyten av å dele og tilgjengeliggjøre forskningsdata etter FAIR-prinsippene og gode eksempler på hvordan dataene de håndterer kan være til nytte for andre og – ikke minst – hvordan deling kan gi gevinst for hele forskningsfeltet og for det aktuelle forskningsmiljøet selv. Det etterlyses informasjonsmateriell og beskrivelser av gode eksempler fra ulike fagfelt.

I undersøkelsene kommer det frem at noen forskere og institusjoner ser på FAIR data og åpne data som det samme. Det er et klart behov for å øke kunnskapen om at FAIR data kan lagres og arkiveres med tilgangsbegrensninger, og om mulighetene for å dele metadata uten nødvendigvis å gjøre selve datasettene åpent tilgjengelige.

Viktigheten av å planlegge datahåndtering

For å sikre at forskningsdata i stadig større grad kan gjøres FAIR er det behov for økt bevissthet om viktigheten av å tenke på dette allerede under planleggingen av forskningsprosjekter. Forskerne må aktivt planlegge hvordan dataene kan beskrives, organiseres og oppbevares på en måte som gjør at de får verdi for andre, og slik at det ikke oppstår unødvendige begrensninger i mulighetene for deling. Datahåndteringsplaner (DMPer) er et sentralt verktøy i denne sammenheng.

I kartleggingen svarer i overkant av en tredjedel av forskere at de ikke har brukt tjenester for datahåndteringsplaner. Blant årsakene som nevnes er lite kjennskap til slike tjenesters eksistens eller at det ikke har vært behov for en slik tjeneste i prosjektet. Av dette kan vi se et tydelig behov for bedre veiledning til forskere, både om viktigheten av å lage en datahåndteringsplan, hvordan denne kan være til nytte for prosjektet, og om de verktøyene som fins for å hjelpe dem i prosessen. Slik veiledning er primært institusjonenes ansvar. Både EU og Forskningsrådet krever at alle prosjekter med finansiering fra dem skal ha en datahåndteringsplan. [Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata](#) tilkjenner en klar forventning om at «*forskningsinstitusjonene utvikler prosedyrer for (i) godkjenning av datahåndteringsplaner og (ii) beslutninger om at et forskningsprosjekt eventuelt er av en slik art at egen datahåndteringsplan ikke er nødvendig eller hensiktsmessig*». Samarbeid institusjonene imellom kan gjøre det å etablere slike prosedyrer til en enklere oppgave. Datahåndteringsplaner er ment som et hjelpemiddel i håndteringen av alle typer forskningsdata. I spørreundersøkelsene finner vi imidlertid eksempler på forskere og institusjoner som knytter behovet for datahåndteringsplaner spesifikt til håndteringen av personopplysninger²². Dette indikerer et behov for kompetanseheving og veiledning på området, slik at flere forskere og forskningsinstitusjoner skal erkjenne nytten og viktigheten av å lage datahåndteringsplaner for sine prosjekter uavhengig av hva slags data prosjektet håndterer.

Håndtering av persondata og andre data med behov for beskyttelse

Flere beskriver utfordringer og uavklarte problemstillinger knyttet til data som krever beskyttelse. Eksempler på slike data er persondata, data som er vernet gjennom åndsverksloven eller på annen måte er beskyttet av immaterielle rettigheter (forretningshemmeligheter) og data med sikkerhetsrelevans.

Både spørreundersøkelsen og den senere innspillsrunden avdekker et stort behov for mer informasjon om hvordan data som inneholder personopplysninger kan lagres og arkiveres for deling og gjenbruk.

Noen beskriver en frykt for å gjøre feil som kan føre til at man, for å være på den sikre siden, velger å ikke dele. Dette forsterkes av et lovverk med kraftige sanksjonsmuligheter. Manglende tilgang på juridisk kompetanse og støttetjenester får da som konsekvens at data som kunne vært delt gjennom FAIR-prinsippene ikke kommer til nytte gjennom gjenbruk. En av institusjonene i undersøkelsen beskriver blant annet et ønske om en fellestjeneste som kan gjennomgå datasett før publisering for å vurdere om åpen arkivering av data er lovlig i forhold

²² Vi antar at dette skyldes sammenblanding av datahåndteringsplaner med personvernkonsekvensvurdering (DPIA).

til personvern og sensitivitet, for eksempel slik NSDs meldeskjema fungerer i dag ved prosjektstart.

Det etterspørres eksempler på samtykkeskjemaer og veiledning rundt hvordan man kan utforme samtykkeskjemaer på en måte som legger til rette for senere deling.

Ulik modenhet i institusjonenes retningslinjer og politikk for FAIR datahåndtering

Forskerne forholder seg til ulike retningslinjer når de planlegger datahåndteringen i sine prosjekter. I spørreundersøkelsen oppga mange at de forholdt seg til nasjonale retningslinjer, retningslinjer fra de forskningsinfrastrukturene de brukte og fra deres finansører. Det vanligste svaret var likevel at de forholdt seg til retningslinjer fra egen institusjon, noe som viser viktigheten av institusjonenes arbeid på området.

I spørreundersøkelsen ble institusjonene bedt om å vurdere egen modenhet i arbeidet med politikk og retningslinjer for å følge FAIR-prinsippene. Svarene viste stor spredning, der omkring halvparten av institusjonene enten ikke hadde startet eller var i gang med å utarbeide politikk og retningslinjer. Gitt at forskerne primært forholder seg til egen institusjons retningslinjer, er det viktig i det videre at forskningsinstitusjonene prioriterer utvikling av retningslinjer. Det er en rekke institusjoner som har gjort godt arbeid. Vi mener man kan komme langt ganske raskt ved å dele på den kompetansen som allerede er bygget opp. Samtidig kan det være store interne variasjoner også ved de institusjonene som oppgir å ha kommet langt med arbeid knyttet til datadeling.

Det vil være en fordel for forskerne om institusjonenes retningslinjer for datahåndtering ligner hverandre der det ikke er gode grunner til annet (for eksempel på grunn av institusjonenes egenart). Mange institusjoner både i Norge og i Europa har viktige erfaringer og har utviklet gode eksempler på retningslinjer. Det er behov for samarbeid og koordinering mellom institusjonene her i Norge for å sikre at disse erfaringene kan komme flere til gode.

Behov for mer kunnskap om metadata og samarbeid om utvikling av standarder

Som tidligere beskrevet er gode metadata og bruk av anerkjente standarder for metadata helt avgjørende for å gi mulighet for gjenfinning og gjenbruk, ikke minst for at datamaskiner skal kunne utveksle informasjon uten manuell bistand. Utredningen viser at både forskerne og institusjonene har behov for mer kunnskap om dette. Flere uttrykker også selv at de ønsker bistand og opplæring på området.

Det er nødvendig at alle institusjoner har grunnleggende kompetanse om viktigheten av metadata, om de standarder for metadata som gjelder, og om hvor de kan søke assistanse for å lære mer.

Innen mange fagdisipliner er det utviklet internasjonale standarder for metadata som forenkler samhandling mellom de ulike forskningsinfrastrukturene og datahåndteringstjenestene, også på tvers av landegrensene. Det er viktig at man bruker slike internasjonale standarder der de finnes. Norske forskere fra ulike fagmiljøer bør også bidra til at standardene videreutvikles og forbedres. RDA har for eksempel en oversikt over arbeid for en del spesifikke disipliner (<https://www.rd-alliance.org/rda-disciplines>). En lignende oversikt finnes også på [FAIRsharing.org](https://www.fairsharing.org).

I andre fagmiljøer har man ikke den samme tradisjonen for å gjøre forskningsdata tilgjengelige for gjenbruk, og det er kanskje ikke ennå utviklet hverken nasjonale, eller internasjonale standarder. Det er derfor fortsatt et behov for at flere fagmiljøer får kjennskap til og kompetanse om det arbeidet som gjøres for standardisering for eksempel gjennom RDA, og lar seg inspirere til å delta i et koordinert samarbeid om videreutvikling av standarder for metadata innen sine respektive fagfelt.

Behov for mer kapasitet og bistand til veiledning

Flere forskere etterlyser mer veiledning og bistand i praktiske spørsmål som omhandler datahåndtering. I spørreundersøkelsen til institusjonene vurderte institusjonene sin kapasitet til å bistå forskerne innen juridiske, etiske og ulike fagområdespesifikke problemstillinger, og til å besvare tekniske spørsmål og spørsmål om informasjonssikkerhet.

Juridiske spørsmål skiller seg ut som et område som mange institusjoner opplever begrenset kapasitet på, med hensyn til å bistå forskere. Nesten 50 %²³ vurderer egen kapasitet som begrenset eller svært begrenset på dette området. Dette oppfattes som en så spesialisert kunnskap at flere har anbefalt at tilgang til slik kompetanse bør organiseres på nasjonalt nivå og/eller gjennom et nasjonalt samarbeid. Det er viktig at nettopp veiledning på det juridiske området kommer fra miljøer med god kompetanse og høy tillit blant forskerne.

Internasjonalt beregnes at arbeid med datahåndtering og kuratering vil kreve ca. 3-5 årsverk (fordelt på alle involverte, inkludert arbeid med datahåndtering utført av forskere) per 100 forskerårsverk²⁴. Vi ser også fra forskningsmiljøer som i dag har kommet langt med sitt arbeid med tilgjengeliggjøring av data at dette kostnadsestimatet stemmer ganske godt. Dersom dette skal gjennomføres i alle fagområder, står vi foran et stort finansieringsløft.

Det er verdt å peke på at mange av de kompetanseutfordringene og veiledningsbehovene som berøres her, ikke er unike for forskningen. Både i forvaltningen og i næringslivet er tekniske og juridiske forhold ved datadeling og viderebruk krevende. I forvaltningen er det etablert et Nasjonalt ressurscenter for datadeling (Digitaliseringsdirektoratet) som har som formål å bidra med kompetanse og veiledning på området.

Behov for oversikt over hva som finnes av FAIR- forskningsinfrastrukturer og datahåndteringstjenester

Både forskerne og institusjonene etterlyser bedre oversikter over eksisterende forskningsinfrastrukturer og tilgjengelige datahåndteringsverktøy for FAIR forskningsdata. Blant institusjonene er det primært de mindre uten spesialiserte forskerstøttestillinger som etterlyser dette.

Oppsummering og anbefalinger

Oppsummert ser vi at det er behov for både kompetanseheving om FAIR forskningsdatahåndtering og økt kapasitet til veiledning på alle nivåer, dvs. enkeltforskere, dataforvaltere tilknyttet forskningsmiljøene, sentral forskningsstøtte i institusjonene,

²³ 30 av 63

²⁴ [Barend Mons 2020, Nature](#)

forskningsledere og opp til nasjonale aktører og myndigheter. Dette inkluderer generell kunnskap om FAIR-prinsippene, retningslinjer, datahåndteringsplaner, metadata, kuratering og andre praktiske forhold rundt deling av data etter FAIR-prinsippene. Samtidig bør bevisstheten bli større om samfunnsnyttene av å dele og tilgjengeliggjøre forskningsdata som er finansiert med offentlige midler etter FAIR-prinsippene.

Det er gjort og gjøres fortsatt mye godt arbeid både i Norge og på europeisk nivå når det gjelder kompetanseheving i forskningsmiljøer. Flere av institusjonene har laget gode retningslinjer og veiledninger for egne ansatte, det er etablert kurs som i noen tilfelle er åpne for alle som ønsker det, og det er startet utdanning av dataforvaltere. Gjennom etableringen av den norske RDA-noden har vi fått en god arena for dialog mellom forskere og institusjoner. De anbefalte tiltakene legger vekt på å bygge videre på det som allerede finnes og på å koble det norske arbeidet til det som skjer i Europa.

Utredningsgruppen anbefaler:

1. Det må etableres en koordinert innsats på finansiering, organisering og samarbeid for nasjonal og institusjonell kompetanseheving for FAIR- forskningsdata. Dette bør skje i et samarbeid mellom institusjonene, HK-dir, Sikt, Forskningsrådet og andre relevante nasjonale finansiører og produsenter av forskningsdata som er finansiert med offentlige midler.
Ansvarlig: RDA Norge v/Sikt
2. Det anbefales at nettstedet opscience.no videreutvikles til en velkjent, prioritert nasjonal ressurs med informasjon om temaer relevante for åpen forskning. Hovedmålgruppe for informasjonen er institusjonene. Sekundær målgruppe er forskere, som forventes primært å forholde seg til informasjon fra egen institusjon. Nettstedet forvaltes av Sikt i samarbeid med HK-dir, Forskningsrådet og UHR. Det anbefales at det etableres en brukergruppe med bred representasjon fra institusjonene og relevante aktører, og som bidrar til videreutviklingen av nettstedet.
Ansvarlig: Sikt
3. Det må etableres nasjonalt samarbeid om kompetanseheving og veiledningstjenester på juridiske, etiske og tekniske sider av håndtering av personopplysninger og andre typer data som av ulike årsaker trenger beskyttelse. Det anbefales også å videreutvikle eksisterende veiledninger for lagring, arkivering og deling av sensitive data (både persondata og data som av andre grunner trenger beskyttelse) og gjøre dette tilgjengelig fra opscience.no. Det bør konkretiseres hva som er mulig og ikke mulig for deling av data som krever beskyttelse under gjeldende regelverk, og det bør gis eksempler på hvordan man i praksis kan legge til rette i starten av et prosjekt for å kunne arkivere og tilgjengeliggjøre slike data for gjenbruk etter prosjektslutt. Det er også behov for bedre veiledning om aidentifisering, fødererte analyser og syntetiske data. Arbeidet på disse områdene må samordnes med tilsvarende arbeid på internasjonalt, særlig europeisk nivå. Det anbefales at dette gjøres i samarbeid mellom Sikt, Nasjonalt ressurscenter for deling av data, Direktoratet for e-helse og de institusjonene som ligger langt fremme på området. Datatilsynet og de regionale og

nasjonale etiske komiteene må involveres.

Ansvarlig: RDA Norge v/Sikt

Temaet spilles også inn til diskusjon i Persondataforum.

4. Det bør etableres kontakt med Nasjonalt ressurscenter for deling av data (Digitaliseringsdirektoratet) for å sikre koordinering av innsatsen på relevante områder, på tvers av forskning og forvaltning. Det bør tilsvarende etableres et samarbeid om kompetanseheving også med næringslivet. Det anbefales at HK-dir initierer kontakten med Nasjonalt ressurscenter for deling av data og samarbeider med Sikt og Direktoratet for e-helse.
Ansvarlig: HK-dir
5. Den enkelte institusjon bør prioritere å skaffe seg enten egen kompetanse eller tilgang til kompetanse for å jobbe med datahåndtering. For større institusjoner anbefales det å bygge opp kompetanse både på generisk nivå og innen de ulike fakulteter/institutter/fagområder. De institusjonene som ikke har mulighet til å ha intern kompetanse anbefales å følge arbeidet i anbefaling 1.
Ansvarlig: Forskningsinstitusjonene
6. Det må utdannes et betydelig antall dataforvaltere/datakuratorer i Norge. Kostnadene til nødvendig kapasitetsutvidelse og kompetanseheving forventes å være på et slikt nivå at temaet bør inn i neste versjon av regjeringens langtidsplan for forskning og høyere utdanning. Anbefalingen videreføres også til Datainfrastruktur-utvalget.
Ansvarlig: Forskningsrådet via Datainfrastruktur-utvalget og videre til Kunnskapsdepartementet
7. Alle forskningsinstitusjoner bør utvikle retningslinjer som legger til rette for at forskningsmiljøene ved institusjonen som hovedregel kan håndtere sine forskningsdata i tråd med FAIR-prinsippene.
Ansvarlig: Forskningsinstitusjonene
8. Det anbefales at institusjonene (UH, institutt og helseforetakene) etablerer dialog og erfaringsutveksling om retningslinjer for datahåndtering for å sikre en mest mulig effektiv vei til målet om retningslinjer for FAIR-forskningsdata. Det er naturlig at dette skjer i dialog mellom UHR v/ledende institusjoner på området, FFA, RHFenes strategigruppe for forskere, Sikt, Forskningsrådet og HK-dir og man kan vurdere om det etableres et eget forum for FAIR-forskningsdata som del av det nasjonale forum for åpen forskning.
Ansvarlig: HK-dir tar ansvar for å finne en ansvarlig
9. Det anbefales at institusjonene innarbeider opplæring i FAIR datahåndtering i PhD- og master-utdanningene. God opplæring bør også gis til de som veileder kandidatene. Flere av universitetene har allerede startet dette arbeidet, og det finnes godt materiale å bygge videre på.
Ansvarlig: Den enkelte institusjon

4.2. INFRASTRUKTURER OG TILHØRENDE TJENESTER FOR DATAHÅNTERING ETTER FAIR-PRINSIPPENE

Både enkeltforskere og institusjoner har beskrevet behov for videreutvikling i eksisterende forskningsinfrastrukturer og andre verktøy for håndtering av forskningsdata etter FAIR-prinsippene, samt behov for utvikling av nye.

Behov for FAIRifisering

Vår kartlegging avdekker et behov for en generell «FAIRifisering» av de forskningsinfrastrukturene og tekniske løsninger for håndtering av forskningsdata i tråd med FAIR-prinsippene som er tilgjengelige i dag. Mange eksisterende tjenester muliggjør publisering av forskningsdata og gjør data gjenfinnbare og tilgjengelige (“F” og “A” i FAIR-prinsippene). Tjenester som sikrer samhandling og gjenbrukbarhet (“I” og “R” i FAIR-prinsippene) oppleves imidlertid å mangle i det nåværende økosystemet for forskningsdata. Dette omfatter metadatastandarder som kan sikre semantisk interoperabilitet²⁵, støtte til og kompetanse om å legge på gode metadata på forskningsdataene, og digitale strukturer for informasjonsutveksling, såkalte APIer²⁶, for å legge data inn i og hente dem ut av tjenestene. Mangel på slike funksjoner gjør det i praksis krevende å gjøre dataene gjenbrukbare.

Behov for samhandlende tjenester og datahåndteringsplaner for forskningsdata

I underkant av halvparten av forskerne i vår undersøkelse svarer at de bruker en mal for datahåndteringsplanen, ofte i Word-format, som deres institusjon har utarbeidet. For forskernes datahåndtering i enkeltprosjekter, spesielt for de som ikke er vant til dette, er en slik Word-mal fra institusjonen bedre enn ingenting, men vil ofte ikke gi mye mer enn å oppfylle institusjonens overordnede behov. For at en datahåndteringsplan skal være til noen stor nytte også for de spesifikke forskningsmiljøene, må malene tilpasses fagområdene. For å unngå en uoversiktlig flora av tilpasninger, bør slike fagspesifikke utvidelser utarbeides i en internasjonal kontekst. Gode eksempler finnes i dag i ESFRI-infrastrukturene, som for eksempel ELIXIR og CEESDA. Også [Science Europe](#), [European Open Science Cloud](#) (EOSC), og EOSC-prosjekter slik som FAIRsFAIR har gode veiledninger om datahåndteringsplaner.

For at verktøy for datahåndteringsplaner totalt sett skal bli effektive verktøy for FAIR datahåndtering er det behov for større konsensus om utformingen enn i dag. Forskningsmiljøer, institusjoner og finansiører må samordne seg og utarbeide maler og innhold som både støtter det overordnede samfunnsålet om at offentlige forskningsmidler blir brukt effektivt og ansvarlig, samtidig som malene oppleves som relevante for det spesifikke fagområdet.

Nasjonale initiativ bør ses i sammenheng med tilsvarende initiativ internasjonalt som [Science Europe](#), ESFRI-infrastrukturer hvor norske fagmiljøer er tilknyttet og ledende innen FAIR datahåndtering, [European Open Science Cloud](#) (EOSC), og EOSC-prosjekter slik som [FAIRsFAIR](#), samt føringer i rammeprogrammet Horisont Europa.

Som vist i kapittel 0 bør datahåndteringsplaner ideelt sett være maskinhåndterbare. Det er behov for videreutvikling av brukervennlige løsninger for dette. For å realisere

²⁵ Dvs datamaskiner kan utveksle informasjon basert på felles forståelse av innholdet

²⁶ API = Application programming interface

maskinhåndterbare datahåndteringsplaner er det behov for at alle relevante aktører som skal utveksle informasjon, slik som fagmiljøer, institusjoner, tjenesteleverandører og finansierer, koordinerer seg.

Teknologisk støtte for å berike med gode og standardiserte metadata tidlig i forskningsprosessen

Tradisjonelt tar mange forskere først mot slutten av prosjektet stilling til hvordan dataene skal arkiveres, med beriking med passende metadata. På dette tidspunktet har forskeren ofte begynt å fokusere på nye oppgaver, det er lenge siden dataene aktivt ble håndtert, og stipendiater og andre midlertidig ansatte som kanskje håndterte dataene har sluttet. I en slik situasjon kan oppgaven med å arkivere data forenlig med FAIR-prinsippene ofte bli unødvendig krevende. Dersom verktøyene utformes slik at det ble naturlig å påføre metadata allerede ved innsamlingstidspunktet, ville arkivering av data forenlig med FAIR-prinsippene forenkles betydelig.

For forskere som gjennomfører laboratorieforsøk er elektroniske lab-journaler et nyttig hjelpemiddel. Lab-journaler er generelt viktige for å dokumentere hvordan eksperimenter er gjennomført og under hvilke forutsetninger de resulterende forskningsdata kan tolkes. Ideelt sett bør digitale journaler foretrekkes selv om mange fortsatt bruker analoge systemer. I digitale lab-journaler kan metadata legges på automatisk. Denne type automatiserte metadata-generatorer er nyttige verktøy, som antagelig kunne forenkle senere datadeling. En utfordring er at systemene ofte er knyttet til den enkelte leverandøren og at man dermed ender med å bli låst til en leverandørs systemer for datahåndtering. Det ville vært til stor hjelp om brukerne av lab-utstyr samordnet seg om felles krav til bruk av felles metadatastandarder, og som kunne benyttes i anskaffelser av nye lab-systemer.

Det er viktig at de metadata som brukes for å berike og beskrive datasettene, og som påføres allerede i datahåndteringsplanen eller i prosjektets tidlige faser, følger de metadatastandarder som er utbredt i det aktuelle fagområdet. Samtidig blir det stadig viktigere å legge til rette for tverrfaglig samarbeid. Det er behov for gode fellesstandarder for generiske metadata (dette finnes i stor grad) og for fagspesifikke tillegg (dette finnes på en del områder), men også for at ulike fagområder som benytter samme type data, samordner sine standarder eller lager oversettelser mellom dem. For eksempel fant vi i løpet av utredningen at både medisinere og arkeologer bruker standarder for gendata. Det er viktig at ulike fagdisipliner som bruker samme type data, men ofte til ulike formål, samordner sine standardiseringsbestrebelsler.

Med tanke på at forskningens og forvaltningens data ofte smelter sammen, er det også viktig å se forskningens og forvaltningens metadatastandarder under ett, altså at det legges til rette for hele livsløpet til dataene. Innen visse fagområder har forvaltningen lagt sterke føringer på hvilke metadata som bør benyttes, for eksempel når det gjelder data i helseregistre²⁷.

Standarder må også følges når det gjelder formatet på selve datasettet. Standardisering av variabeldefinisjoner, kodeverk og formater er viktig for at data enkelt skal være gjenbrukbare og

²⁷ Se for eksempel <https://www.ehelse.no/standardisering/standarder/nasjonal-spesifikasjon-for-metadata-om-helsedata>

interoperable. Derfor har flere av tjenestene krav om at spesifikke formater skal brukes ved arkivering og publisering av data. Spesielt gjelder dette dersom tjenestene skal garantere langtidsbevaring av dataene. Fra forskernes side er det imidlertid enkelte som etterlyser større fleksibilitet i slike krav.

Sammenhengende infrastrukturer og tjenester

Flere forskere opplever at dagens tilbud av verktøy for håndtering av forskningsdata er fragmentert og savner at forskningsinfrastrukturer og datahåndteringstjenester for de ulike delene av forskningsprosessen henger bedre sammen. Ideelt sett ønsker de for eksempel at man, når man er ved slutten av et forskningsprosjekt, har lagret data underveis, har lagt på gode metadata og har vurdert hvilke data som har verdi for ettertiden, for eksempel gjennom en funksjon «arkiver mine data».

Informantene i denne utredningen har ulike tilnærminger til hva som er mulige løsninger på utfordringene. Noen etterlyser tjenester som dekker flere faser, for eksempel ved at en forsker kan lage en datahåndteringsplan, lagre data underveis og til slutt arkivere dataene i samme tjeneste. Andre etterlyser API-er og standardisering av formater for data og metadata som sikrer et mangfold av tjenester for ulike fagområder, men på en måte som sikrer interoperabilitet mellom tjenestene.

Enkle, fleksible tjenester for mindre datamengder

Forskere håndterer svært ulike mengder data. En del forskere etterlyser tjenester tilsvarende Sigma2 (lagrings- og beregningskapasitet) for prosjekter som ikke er store nok til å få tilgang til Sigma2 sine tjenester, som er ment for datamengder over 10 terrabyte. Flere innspill peker også på behovet for brukervennlige tjenester for studentprosjekter, som normalt har svært begrenset med tid å bruke på avansert datahåndtering. Enkelte institusjoner oppgir at de har slike tilbud lokalt, men dette ser ikke ut til å gjelde alle.

Noen av innspillene peker på et spesifikt behov for lavterskeltilbud for FAIRifisering av data fra forskere som har samlet verdifulle datasett og som slutter ved institusjonen, for eksempel ved pensjon eller overgang til annen jobb i eller utenfor institusjonen.

Informasjonssikkerhet

Enkelte forskere stiller spørsmål ved informasjonssikkerheten i de lokale infrastrukturene for lagring og arkivering av data som de har tilgang til i dag. Det handler både om frykt for at uvedkommende får tilgang til dataene og for at data skal forsvinne. Forskerne som løftet utfordringene håndterte både sensitive data og data som kan ligge åpent tilgjengelig.

HK-dir gjennomfører en årlig kartlegging av informasjonssikkerheten ved de offentlige forskningsinstitusjonene. Ett av funnene fra årets undersøkelse var at bruk av servere som ikke er underlagt det normale driftsregimet til institusjonen, for eksempel ved at den heller driftes av et institutt eller en forskergruppe ved et universitet, kan medføre økt risiko for brudd på informasjonssikkerheten.

Etablering av nye infrastrukturer – udekkede behov?

Gjennom utredningen er vi gjort kjent med institusjoner og forskningsmiljøer som arbeider med å etablere egne infrastrukturer for åpen publisering av forskningsdata. Dette kan tyde på at

eksisterende tjenester ikke dekker institusjonenes behov. Samtidig kan det imidlertid se ut til at beslutningen om å etablere nye infrastrukturer i en del tilfelle springer ut av manglende kjennskap til eksisterende infrastrukturer og tjenester og/eller en – noen ganger feilaktig – antagelse om at egne data er helt annerledes enn alle andres. En tredje begrunnelse for å etablere nye infrastrukturer er ønsket om å ha lokal styring på både infrastrukturen og kompetansen til å yte brukerstøtte.

En utfordring med å etablere nye infrastrukturer i hvert enkelt nytt forskningsprosjekt er den manglende finansieringen etter prosjektslutt. For noen infrastrukturer innebærer det manglende videreutvikling og drift. Utredningsgruppen er også bekymret for de totale kostnadene ved at en rekke institusjoner skal utvikle den – etter hvert ganske krevende – funksjonaliteten som skal til for å tilby løsninger for deling og tilgjengeliggjøring av FAIR forskningsdata. Vi har gjennom utredningen sett flere eksempler på at fagmiljøer har fått til en god kombinasjon ved at de har etablert lokale infrastrukturer på toppen av en større, generisk infrastruktur. Slik har de sikret lokal styring samtidig som de deler på kostnadene for utvikling av den grunnleggende funksjonaliteten med et større miljø. Dette er en mulighet som bør gjøres kjent for flere.

Infrastrukturer og verktøy for lagring og arkivering av data med behov for beskyttelse

For data med behov for beskyttelse beskrives behov for tjenester som muliggjør deling med samarbeidende forskere utenfor institusjonen i pågående prosjekter. Noen etterlyser også flere alternativer for lagring og arkivering av data med behov for beskyttelse. Forskerne savner også gode tjenester og verktøy for å anonymisere eller pseudonymisere personopplysninger. Også syntetisering av datasett (data generert av en maskinlæringsmodell basert på reelle data) er et tema som løftes i denne sammenheng. Flere av informantene beskriver også et behov for oversikt over hvilke typer data som kan lagres og arkiveres hvor. Slike veiledninger finnes ved noen av institusjonene, men langt fra alle.

Flere institusjoner oppgir UiOs Tjenester for Sensitive Data (TSD) som anbefalt tjeneste for arkivering/langtidslagring av sensitive data. Det er også flere forskere som oppgir at de har arkivert data i TSD, men TSD er per i dag ikke ment for arkivering av data etter prosjektslutt. Den nåværende bruken indikerer at det er behov for å videreutvikle tilbudet av tjenester for sensitive forskningsdata i Norge. Det etterlyses arkivløsninger med fleksibel tilgangskontroll, der man kan gi ulik tilgang til data basert på for eksempel gruppetilhørighet eller rolle. Sikt (tidligere NSD) har tilbud om å motta persondata dersom den som avleverer dem har behandlingsgrunnlag for dataene (oftest gjennom samtykke fra informantene). Tilgang til slike data kan gis ved bestilling og i samråd med den som har avlevert dataene. Det er ikke mulighet for en forskerstyrt tilgangskontroll, slik det etterlyses i spørreundersøkelsen.

Flere av helseforetakene har lokale eller regionale løsninger under utvikling, og Direktoratet for e-helse har sin store satsing gjennom Helseanalyseplattformen. Selv om det i første omgang tas sikte på å tilgjengeliggjøre data fra helseregistrene, har Helseanalyseplattformen ambisjoner om på sikt også å kunne lagre data fra forskningsprosjekter. Ferdigstillingen av Helseanalyseplattformen ble før jul satt på vent på grunn av juridiske utfordringer. Tidsplanene for eventuell arkivering av helserelaterte forskningsdata i denne infrastrukturen er dermed svært usikre.

Behov for forutsigbare kostnader

Noen av tjenestene har betalingsmodeller der forskerne etterskuddsvis betaler for faktisk bruk av tjenester og kapasitet. For noen oppleves dette som uforutsigbart og det er ikke alle som helt forstår hva de faktisk betaler for. Det synes å være behov for mer tydelighet i de forretningsmodeller som benyttes, herunder prismodeller og -eksempler, og bestemmelser om avregning og betalingsvilkår. Det er også institusjoner som fremhever at det må finnes infrastrukturer for arkivering/publisering av data som for det enkelte forskningsprosjekt er rimelige og helst gratis å benytte. Det fremheves at dette er spesielt viktig for å sikre arkivering av data som det angjeldende prosjektet kanskje ikke selv er interessert i for ettertiden, men som kan være til nytte for andre.

Det er her også verdt å peke på at EUs nye reguleringer gjennom Open Data Directive vil forventes å gjøre det vanskeligere for offentlig finansierte datainfrastrukturer å kreve betaling for offentlig finansierte forskningsdata utover det å dekke inn marginalkostnadene for tilgangen (det kan eventuelt også tas betalt for tilleggstjenester). For de som skal finansiere forskningsinfrastrukturer vil dette gjøre det mer krevende, samtidig som det vil gjøre kostnadene for tilgang/viderebruk mer forutsigbare.

Oppsummering og anbefalinger

I videreutvikling og eventuell nyutvikling av infrastrukturer og tilhørende verktøy er det sterkt behov for nasjonal og internasjonal samordning. Leverandørene må sikre god dialog med relevante brukermiljøer, det må legges til rette for koordinering og samarbeid mellom leverandører av institusjonelle og nasjonale løsninger og for aktiv deltagelse i europeiske og internasjonale initiativer som EOSC og RDA. Forskningsrådet ble i 2021 oppnevnt som norsk hovedmedlem (“mandated organisation”) i EOSC Association og koordinerer innsatsen blant norske organisasjoner gjennom det nyetablerte Nasjonalt EOSC-forum.

Institusjonene opplever at de i dag ikke har brukervennlige nok verktøy for forskernes arbeid med datahåndteringsplaner, og det er behov for at eksisterende løsninger og verktøy samordnes bedre. Det er ønskelig å gradvis bevege seg bort fra Word-maler til mer strukturerte, maskinlesbare informasjonsbærere, og det er nødvendig med både generiske og disiplin-spesifikke datahåndteringsplaner som kan benyttes på tvers av forskningsinstitusjoner, finansører og ulike systemer, og være kompatible med internasjonale standarder.

10. Det anbefales en gjennomgang av tilgjengelige verktøy for å lage datahåndteringsplaner med sikte på å videreutvikle brukervennlig funksjonalitet til både forskermiljøer som er godt i gang med datadeling og miljøer som ikke har tradisjon for dette. I så stor grad som mulig er det ønskelig å tilby verktøy for å lage planer som inneholder strukturert, maskinhåndterbar informasjon. Arbeidet bør gjennomføres i samarbeid mellom Sikt, institusjonene og Forskningsrådet.

Ansvar: Sikt

11. Det anbefales en videre utredning av behov for infrastrukturer for arkivering/langtidsbevaring og tilgjengeliggjøring av forskningsdata med behov for beskyttelse (persondata og andre data i klassene gule og røde data). Det er spesielt behov for god funksjonalitet for publisering av åpne metadata for slike data, og fleksibel

tilgangsstyring til selve dataene. Tiltaket vil kreve betydelig finansiering og bør inn i neste versjon av langtidsplanen for forskning og høyere utdanning. Anbefalingen henvises derfor videre til Datainfrastrukturutvalget. Forslaget spilles også inn som tiltak under Handlingsplan for digital omstilling.

Ansvarlig: Forskningsrådet gjennom Datainfrastruktur-utvalget og videre til Kunnskapsdepartementet

12. Vi anbefaler at norske leverandører av løsninger for håndtering av forskningsdata deltar i relevante europeiske og internasjonale nettverk og arenaer for dialog og samarbeid, blant annet for å sikre utvikling mot samhandlende tjenester. Utredningsgruppen anser blant annet EOSC, ESFRI, GO FAIR og RDA som viktige arenaer i den sammenheng.
Ansvarlig: Tjenesteleverandørene

13. Tjenesteleverandørene må gjøre seg kjent med Referansearkitekturen for datadeling og Kunnskapssektorens datafelleskap (KUDAF).
Ansvarlig: Tjenesteleverandørene

14. Forskningsinstitusjonene og nasjonale aktører bør engasjere seg i arbeidet med European Open Science Cloud (EOSC) for å bidra til videre utvikling og implementering. Dette kan være gjennom å melde seg inn i EOSC-stiftelsen, (Forskningsrådet dekker inntil videre kontingenten), å engasjere seg i relevante EOSC-arbeidsgrupper eller å delta i EU-finansiert samarbeid som er relevant for EOSC. Det anbefales også at det etableres et nasjonalt EOSC-forum for relevante og interesserte aktører.
Ansvarlig: Forskningsinstitusjonene og nasjonale aktører

15. Forskningsinstitusjonene anbefales å oppfordre egne forskere til å melde seg inn i RDA, legge til rette for at forskere kan delta i aktiviteter i RDA for eksempel når det gjelder utvikling av metadatastandarder, standardiserte ontologier og kontrollerte vokabularer innen disipliner der dette ikke allerede eksisterer, eller videreutvikling av eksisterende standarder, ontologier og vokabularer.
Ansvarlig: Forskningsinstitusjonene

16. Gode metadata og veiledning om hvordan og når disse skal benyttes er et fundament for FAIR-prinsippene. Det er behov for nasjonal koordinering av arbeid med å utvikle standarder for metadata både innen og på tvers av fagområder. Arbeidet bør også skje i samarbeid med forvaltningen. Det trengs også standarder for koder og formater i selve datasettene, for å sikre gjenbruksverdi av de data som lagres.
Ansvarlig: RDA Norge v/Sikt

17. Norske myndigheter og finansører bør fortsette å støtte deltakelse i internasjonalt samarbeid rundt arbeidet med FAIR forskningsdata, med særlig vekt på EOSC.
Ansvarlig: Departementene, Forskningsrådet, de regionale helseforetakene

18. Forskningsinstitusjonene bør sikre

- 18.1. at deres forskere har tilgang til verktøy for datahåndteringsplaner som er i tråd med FAIR-prinsippene
- 18.2. at deres forskere har tilgang til infrastrukturer for lagring av forskningsdata i den aktive perioden, med god støtte for å berike datasett med FAIR metadata.
- 18.3. at deres forskere har tilgang til infrastrukturer for arkivering og langtidsbevaring av forskningsdata i henhold til FAIR-prinsippene etter prosjektslutt
- 18.4. at forskere og studenter har tilgang til relevante datahåndteringstjenester også for prosjekter med små datamengder.

Ansvarlig: Forskningsinstitusjonene

4.3. KULTUR FOR DELING

Vår kartlegging viser at det ikke bare er forskernes tilgang på infrastrukturer, verktøy og tjenester som er viktig for at de skal bidra til tilgjengeliggjøring og gjenbruk av forskningsdata. Det å tilgjengeliggjøre forskningsresultater er en grunnleggende del av forskningsetikken, og det å sørge for at data kan gjenbrukes må oppfattes som en naturlig del av det å være forsker. For mange er det nytt å dele forskningsdata på den måten det nå legges opp til, og det krever en kulturendring. Forskningsinstitusjonene må legge til rette for en slik delingskultur gjennom fokus på kompetanse, tekniske forhold rundt datahåndtering og det forskningsetiske i å gjøre ny kunnskap tilgjengelig for samfunnet. Kulturbygging handler om kunnskap, men også om hvorvidt institusjonen forskeren tilhører og forskningsmiljøet rundt dem forventer at de deler, og at forskeren selv på en eller annen måte opplever å ha nytte av å gjøre det.

Manglende tradisjoner for deling

I spørreundersøkelsen til institusjonene stilte vi spørsmål om hva de opplevde som de største hindringene for at deres vitenskapelig ansatte i større grad skal gjøre sine forskningsdata tilgjengelige for gjenbruk. Det var flere som trakk frem fagtradisjoner og manglende kultur for deling. For noen anses dataene de samler inn i forskningen som forskerens «egne» data som det tradisjonelt ikke har vært kultur for å dele med andre. Det er viktig at forskerne ser eksempler fra forskere i eget fagfelt som deler data. Prosjektet [QUALI-FAIR](#) ved UiO er et spennende eksempel²⁸.

Behov for insentiver for datadeling

Flere av forskerne og institusjonene peker på et behov for insentiver for datadeling. Det å gjøre data gjenbrukbare krever tid og krefter som mange i dag tenker kan anvendes bedre på annen aktivitet. Det må arbeides aktivt for å få stadig flere forskere til å anse dette som vel anvendt tid.

Gitt at infrastrukturen der dataene har blitt arkivert tilbyr mulighet for fast lenking (DOI eller tilsvarende) til datasett, vil det være mulig å sitere datasett. Statistikk fra internasjonale

NB! Feil rettet 31.03: Et nesten tomt tidligere kapittel 4.3 om data med behov for beskyttelse fjernet

²⁸ Miljøer som jobber mye med kvalitative og kontekst-sensitive data har slått seg sammen om å utvikle metoder for å gjøre slike data FAIR

databaser viser at dette skjer i stadig større grad, noe som gir synlighet til den innsatsen som er lagt ned.

Det pågår arbeid både innen EU og i Norge²⁹ med å videreutvikle metoder for evaluering av forskere og forskning og å inkludere både det å dele egne data og det å gjenbruke andres data som del av kriteriene man vurderes på. Også i [Strategi for digital omstilling i universitets- og høyskolesektoren](#) fremheves dette som ett av tiltakene.

Det å sikre at tilgjengeliggjøring av forskningsdata blir ansett som meritterende er åpenbart viktig, men vi har valgt å ikke ta dette videre i anbefalingene da det ikke er en eksplisitt del av mandatet til dette arbeidet og det allerede er igangsatt videre arbeid med dette nasjonalt blant annet via [UHR](#) og NFR og internasjonalt på EU-nivå.

Oppsummering og anbefalinger

19. Utredningsgruppen støtter Datalisensutvalgets anbefaling om at det bør settes av midler til å støtte opp om og initiere tiltak for å fremme delingskultur, både gjennom deling av egne og gjenbruk av andres forskningsdata. En mulighet er en ordning med såkornsmidler til å løfte lokale initiativer, slik Nederland har gjort.
Ansvarlig: KD, Forskningsrådet, institusjonene

20. Forskningsinstitusjonene oppfordres til å legge til rette for at forskerne opplever det som en integrert del av forskningsarbeidet å gjøre sine data tilgjengelig for gjenbruk og at slik aktivitet anses som meritterende, for eksempel ved at datadeling etter FAIR-prinsippene kan telle positivt ved vitenskapelige ansettelse, stillingsopprykk og tildeling av forskningstid. Det oppfordres til å følge arbeidet i regi av UHR med implementering av NOR-CAM³⁰. Det samme gjelder karrierevurderinger ved tildelinger av forskningsmidler fra Norges forskningsråd og andre finansiører.
Ansvarlig: Forskningsinstitusjonene

²⁹ [NOR-CAM – Veileder for vurdering i akademiske karriereløp](#)

³⁰ <https://www.uhr.no/en/front-page-carousel/nor-cam-a-toolbox-for-recognition-and-rewards-in-academic-careers.5780.aspx>

4.4. REGULATORISKE VIRKEMIDLER

Både i spørreundersøkelsene og i de påfølgende innspillene har flere institusjoner og forskere etterlyst klare krav og retningslinjer.

Krav fra finansørene

Tydligere krav og oppfølging fra finansørene trekkes frem som et mulig regulerende virkemiddel fra informanter i utredningen. I dag stiller både Forskningsrådet og EUs Horisont Europa-program krav om at data fra prosjektene de finansierer skal gjøres så åpent tilgjengelige som mulig. I Horisont Europa er det å kunne vise til tidligere deling av data og gjenbruk av andres data tatt inn som del av vurderingskriteriet for forskningskvalitet (excellence) i prosjektsøknadene.

I undersøkelsen til institusjonene nevnes det at forskerne «gjør det de er kontraktsforpliktet til å gjøre». Strengere kontraktsfestede krav fra finansørene og i de kontrakter som inngås mellom offentlige kontraktører og oppdragsforskningsmiljøer, og en konkret oppfølging av disse i etterkant, foreslås derfor som relevante tiltak for å sikre at data fra tildelte prosjekter faktisk blir håndtert i tråd med kontraktsforholdene. Noen av innspillene i kartleggingen har påpekt at dette også betyr at finansørene må sette av de nødvendige midler (standardtakster, eller etter forskningsmiljøets spesifikasjoner) til å dekke de faktiske ekstra kostnadene FAIR måtte innebære. Fra Forskningsrådets side skjer dette allerede i stor grad, men dette kan muligens styrkes ytterligere.

Krav fra institusjonene

Forventningen om å tilgjengeliggjøre forskningsdata for gjenbruk stilles fra myndighetenes side til institusjonene, og det må primært være institusjonenes ansvar å sikre at det enkelte forskningsprosjekt følger dette. Igjen er det viktig med institusjonelle retningslinjer med krav om å lage datahåndteringsplaner og anbefalinger for hvor og hvordan data skal lagres og arkiveres.

Et forslag som er fremmet i løpet av utredningen er at institusjonene stiller klare krav til både studenter³¹ og PhD-kandidater om datahåndteringsplaner og redegjørelser for arkivering av forskningsdata ved innlevering av mastergrader og doktorgradsarbeider. Samtidig vil forskningsinstitusjonen ha handlingsrom til å sikre langt større deling av forskningsdata gjennom sin styringsrett over data som utarbeides ved institusjonen, og utvikling av kultur for datadeling, kompetanse og støttetjenester bør være fokus i tiden framover.

Krav fra tidsskrifter om deling av data

Krav fra forlag, tidsskrifter og fagkonferanser om publisering av data kan være et sterkt regulerende virkemiddel, ved at det er direkte knyttet til forskernes muligheter til å publisere deres artikler og konferansebidrag. Et eksempel er innen samfunnsøkonomi hvor flere tidsskrifter i dag krever at data gjøres tilgjengelige enten før publisering eller ved senere

³¹ Studenter har styringsrett over egne data, men vi mener det må være mulig å kreve en redegjørelse.

anmodning fra tidsskriftet. American Economic Association krever for eksempel nå pre-aksept «computational reproducibility» og FAIR data (i alle fall til en viss grad) av alle artikler³².

Dersom forskerne pålegges å publisere sine data på denne måten, vil disse dataene også omfattes av åpne data-direktivet og offentlighetsloven.

Helsesdata

Helseforetak og andre som forsker på helsedata beskriver gjeldende regelverk som en barriere for å bevare data etter prosjektperioden. Det oppleves også at veiledninger legger ensidig vekt på beskyttelse i forhold til deling og gjenbruk. Både helseforetakene, Kreftforeningen og Kreftregisteret trekker frem at mange pasienter ønsker at deres data skal deles, men at gjeldende taushetsbestemmelser og lovverk gjør det svært krevende. I dag oppleves det at helseforskningsloven med sitt krav om normalt å anonymisere eller slette data etter prosjektslutt, setter strammere rammer for arkivering og gjenbruk av forskningsdata enn hva GDPR gjør.

Samarbeidsprosjekter som innebærer deling av persondata med forskere utenfor EU/EØS er særs krevende pga. lovverk som hindrer. For tiden er det også så godt som umulig å dele data med forskere og forskningsinfrastrukturer under USAs jurisdiksjon. For slikt samarbeid som omfatter helsedata har tre europeiske forskningsorganisasjoner gått sammen om et [innspill til EU](#). Her beskriver de konsekvensene dette i øyeblikket har for både forskningen og for pasientene, og det foreslås mulige løsninger.

Oppsummering og anbefalinger

21. Det anbefales at Forskningsrådet og andre viktige finansiører (som de regionale helseforetakene) oppfordrer til gjenbruk av både data og eksisterende forskningsinfrastrukturer der det er hensiktsmessig.
Ansvarlig: Forskningsrådet og andre finansiører
22. Utredningsgruppen ønsker å støtte anbefalingen fra Datalisensutvalget om at institusjonene og deres ansatte ikke signerer avtaler som gir kommersielle aktører noen form for rettigheter til data. Institusjonene anbefales også å råde sine ansatte til ikke å lagre data i infrastrukturer direkte tilknyttet forlag eller tidsskrift, men isteden lagre data på en infrastruktur der forsker eller institusjon har kontroll over dem, og så henvise hit fra tidsskriftet.
Ansvarlig: Institusjonene
23. Institusjonene bør legge til rette for at forskere - dersom dette er definert for deres disiplin - bruke standardiserte ontologier og kontrollerte vokabularer for å berike sine datasett med metadata. RDA har laget en [oversikt](#) over relevante standarder innen ulike disipliner.
Ansvarlig: Institusjonene

³² <https://www.aeaweb.org/articles?id=10.1257/pandp.111.808>

24. Utdanningsinstitusjonene bør stille krav til både studenter og PhD-kandidater om redegjørelser for arkivering av forskningsdata ved innlevering av mastergrader og doktorgradsarbeider.
Ansvarlig: Institusjonene
25. Det bør utredes hvordan standardkontrakter som sektormyndighetene benytter, for eksempel i forbindelse med tildeling av kartleggingsoppdrag og ulike former for oppdragsforskning, kan ta inn formuleringer om FAIR (krav til metadata, lisenser osv) for de data som samles inn. Likeledes kan det i slike utlysninger etterspørres tidligere bidrag til FAIR data. Ansvar kan plasseres for eksempel hos Digitaliseringsdirektoratet som forvalter av standardkontraktene, hos DFØ som eier av Doffin, og andre steder.
Ansvarlig: HK-dir bringer saken videre til relevante aktører
26. Det anbefales at man gjør en gjennomgang av regelverket for helseforskning i relasjon til personvernlovgivningen/GDPR og sett opp mot behovet for gjenbruk av data. Det kan også være behov for å se nærmere på helseforskningsloven i lys av implementeringen av EUs nye [viderebruksdirektiv](#), i norsk rett. Flere instanser må involveres, som HOD, REK/NEM, Direktoratet for e-helse og helseforetakene.
Ansvar: HK-dir bringer saken videre til relevante aktører

4.5. ANDRE TEMAER

Oppfølging og monitorering

Det må etableres verktøy og rutiner for å følge opp utviklingen på området FAIR forskningsdata. Tidligere Unit ble i *Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata* gitt i oppgave å etablere indikatorer og statistikk over åpne data. Monitorering bør planlegges i samarbeid mellom HK-dir, Forskningsrådet, Sikt og andre leverandører av tjenester der man kan hente relevante data.

Monitoreringen må følge standarder og indikatorer som utvikles i regi av EOSC. Parametere knyttet til monitorering bør brukes primært til oppfølging av aktivitet og utvikling i forskningssystemet, og ikke til å fordele økonomiske midler. I tråd med anbefalingene for FAIR indikatorer i EOSC bør monitorering og kvantifisering av FAIRness først og fremst brukes som inkluderende virkemidler for å få flere institusjoner, miljøer og forskere til å prioritere datadeling i tråd med FAIR-prinsippene.

Det kan også være synergier i å koordinere indikatorutviklingen med tilsvarende behov i forvaltningen og med de monitoreringsregimer som vil følge av EUs nye reguleringer på dataområdet (Open Data Directive, Data Governance Act og High Value Datasets) når disse blir innarbeidet i norsk lovgivning.

Utfordringer knyttet til håndtering av registerdata

Registre, som nasjonale kvalitetsregistre og andre store forskriftsbaserte registre, er i praksis databaser tilrettelagt for bruk og gjenbruk av data. Likevel opplever mange at det tar tid å søke om og å få tilgang til å bruke data fra både helseregistre, SSBs registre og andre relevante

personregistre som benyttes som kilder i forskningen. Registerne og bruken av dataene er ofte regulert gjennom egne forskrifter, som fastslår behandlingsansvaret og hvordan dataene kan brukes. Registerforvalterne kan noen ganger bruke svært lang tid på å avklare personvern hensyn, premisser for utlevering og annet, noe som gjør det krevende for forskerne å planlegge forskningsarbeidet. Selv om denne utredningen i utgangspunktet skal fokusere på tilgjengeliggjøring av nye forskningsdata, er tilgang til data fra slike registre så viktig at det løftes også i denne sammenheng.

En kunne tenke seg at institusjonene som gjør registerne tilgjengelig for forskere (som SSB) også har ansvar for å dokumentere hva som har vært utlevert av datasett. I dag er det ofte vanskelig for interesserte tredjeparter å vite hva det skal søkes om for å kunne reprodusere publisert forskning. Dersom slik dokumentasjon var tilgjengelig i etterkant, ville det bidra til bedre muligheter for reproduserbarhet.

Forskningsdata med potensiell kommersiell verdi

Det er lite informasjon og kompetanse tilgjengelig knyttet til forskningsdata med potensiell kommersiell verdi, samtidig som det er et offentlig mål å øke samspillet mellom offentlige forskningsinstitusjoner og næringsliv. Det anbefales derfor at det gjøres egne utredninger rundt kommersielle forhold for forskningsdata generert med offentlig finansiering og data generert i samarbeid mellom offentlige forskningsinstitusjoner og næringslivet.

Anbefalinger

27. Det anbefales å etablere monitorering og oppfølging av utviklingen mot større grad av FAIR forskningsdata og gjenbruk av disse. Arbeidet må følge de indikatorene og de standardene som nå utvikles og etableres i regi av EOSC.
Ansvarlig: HK-dir og Forskningsrådet i samarbeid med Sikt
28. Det anbefales at det gjøres egne utredninger rundt kommersielle forhold for forskningsdata generert med offentlig finansiering og data generert i samarbeid mellom offentlige forskningsinstitusjoner og næringslivet.
Ansvarlig: Kunnskapsdepartementet.