

# Felles infrastruktur og tjenester for FAIR forskningsdata

Utkast til rapport fra utredningen 16.12.21

---

**Utarbeidet av:** Sekretariatet for utredningen  
**Kontakt:** Kristin Brænden - [kristin.braenden@agendakaupang.no](mailto:kristin.braenden@agendakaupang.no)  
**Tilhørighet:**



## Innholdsfortegnelse

<b>1. Sammen drag</b> .....	<b>4</b>
<b>2. Innledning</b> .....	<b>5</b>
2.1. Bakgrunn.....	5
2.2. Mandat .....	7
2.3. Prosess og metode .....	8
2.3.1. Organisering av arbeidet .....	8
2.3.2. Status for leveransene .....	9
2.3.3. Datainnsamling og kilder .....	10
2.4. Avgrensninger og tolkning av mandatet.....	12
2.5. Definisjoner .....	12
<b>3. Hvilke tjenester og verktøy brukes i dag?</b> .....	<b>17</b>
3.1. Introduksjon - Livssyklusen til forskningsdata .....	17
3.2. Innledende fase .....	18
3.2.1. Søke etter eksisterende data.....	18
3.2.2. Lage datahåndteringsplan .....	19
3.3. Håndtere forskningsdata i pågående prosjekter .....	20
3.4. Arkivere og publisere forskningsdata .....	23
3.5. Tilgang til kompetanse og bistand.....	26
3.6. Muligheter for finansiering av datadeling .....	28
<b>4. Hva mangler?</b> .....	<b>29</b>
4.1. Verktøy og tjenester .....	29
4.2. Kunnskapsbehov og kompetanse .....	31
4.3. Håndtering av data med behov for beskyttelse .....	34
4.4. Kultur for deling.....	35
4.5. Regulerende virkemidler .....	36
4.6. Oppsummering av dagens utfordringsbilde .....	37
<b>5. Diskusjon</b> .....	<b>40</b>
<b>6. Anbefalinger</b> .....	<b>44</b>
6.1. Forskningens Bruk av verktøy og tjenester .....	44
6.2. forbedring og utvikling av verktøy og tjenester .....	46
6.3. Kompetanse.....	48



6.4. Kultur for deling.....	49
<b>7. Vedlegg .....</b>	<b>51</b>



# 1. SAMMENDRAG

Kommer i neste versjon



## 2. INNLEDNING

### 2.1. BAKGRUNN

Gjenbruk av data fra forskningen er slett ikke et behov som er nytt. Norsk senter for forskningsdata (NSD) feiret nettopp sitt 50-årsjubileum. En av NSDs sentrale oppgaver har vært å gjøre data fra samfunnsvitenskapelig forskning tilgjengelig. Men flere andre fagmiljøer har også lang tradisjon for å arkivere og dele data. Likevel er det først de siste par tiårene at arbeidet virkelig har fått fart, ikke minst som resultat av de store utfordringene vi står overfor når det gjelder å opprettholde et bærekraftig samfunn. Tilgang til stadig større datamengder og teknologi for å håndtere data har i stor grad bidratt til at man ser et potensial for å oppnå bedre og raskere resultater av forskningsinnsatsen når forskere kan gjenbruke innsamlede data ikke bare fra egne, men også fra andres tidligere arbeid.

I løpet av de siste årene har stadig flere instanser innført krav om at data som samles inn eller produseres som del av et forskningsprosjekt, også gjøres tilgjengelig for gjenbruk når prosjektet er avsluttet. I 2016 ble [FAIR Guiding Principles for scientific data management and stewardship](#) publisert i tidsskriftet Scientific Data<sup>1</sup>. Prinsippene stadfester at forskningsdata skal være FAIR, dvs. gjenfinnbare (Findable), tilgjengelige (Accessible), samhandlende (Interoperable) og gjenbrukbare (Reusable). Samme år vedtok EU-kommisjonen å støtte disse prinsippene, og 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. Under videreføringen Horisont Europa er den frivillige piloten omgjort til et krav om at data samlet inn gjennom forsknings finansiert av programmet skal gjøres FAIR<sup>2</sup>.

Det er viktig å presisere at data som er FAIR ikke nødvendigvis er åpent tilgjengelige. Tilgangen kan begrenses dersom det er gode grunner for dette, f.eks. dersom det dreier seg om persondata eller data som er knyttet til kommersielle interesser. EUs krav er at data skal være «as open as possible, as closed as necessary».

Høsten 2017 vedtok den norske regjeringen først [Digitaliseringsstrategi for universitets- og høyskolesektoren](#) og deretter [Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata](#). I strategien for tilgjengeliggjøring og deling av forskningsdata stadfestes blant annet at Norge vil følge EUs oppfordring om å følge FAIR-prinsippene. Strategien understreker også betydningen av langsiktig arkivering og kuratering (eller «røkt») av forskningsdata, og at dette arbeidet må være basert på internasjonale standarder.

På oppdrag fra KD utarbeidet Unit i samarbeid med forskningsinstitusjonene en handlingsplan for å gjennomføre disse strategiene. Digitaliseringsstyret for høyere utdanning og forskning ble etablert for å operasjonalisere strategiene, og det ble oppnevnt seks fagutvalg under digitaliseringsstyret for å foreslå tiltak på de ulike områdene. Handlingsplanen ble vedtatt i 2019, og UH-institusjonene besluttet å sette av et årlig beløp til finansiering av fellesprosjekter

---

<sup>1</sup> DOI til artikkelen: <https://doi.org/10.1038/ng.3544>

<sup>2</sup> Horisont Europas mal for datahåndteringsplan

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 henger nært sammen med to andre arbeider, som vist i Figur 1. Datalisensutvalget leverte 1. oktober sin [sluttrapport](#), og vi bygger deler av vårt arbeid på diskusjoner, konklusjoner og anbefalinger derfra. Samtidig er også arbeid i gang i Datainfrastruktur-utvalget i regi av Forskningsrådet. Sammenhengen mellom disse tre er illustrert nedenfor. Datainfrastruktur-utvalget og denne utredningen har tett kontakt og utveksler informasjon løpende. Sluttdato for denne utredningen er 1. mars 2022, mens Datainfrastruktur-utvalget har sluttdato 1. mai. De kan dermed ta resultatene fra denne utredningen med i arbeidet med sin leveranse.



FIGUR 1 TRE DELVIS PARALLELE FORSKNINGSDATAUTVALG

Mengden data som samles inn til forsknings- og analyseformål har økt dramatisk de siste tiårene. Mange disipliner har tatt i bruk nye metoder både for datainnsamling (bl.a. sensorer, droner, IoT) og behandling av data. Stadig kraftigere datamaskiner har bidratt til langt større kapasitet til å prosessere og analysere data. Maskinlæring og kunstig intelligens har satt ytterligere fart i utnyttelsen av data i forskningssammenheng.

En del forskningsdisipliner har allerede en godt innarbeidet tradisjon for å dele data åpent. Dette gjelder for eksempel innen samfunnsvitenskap, lingvistikk, havforskning, meteorologi, nevrovitenskap, m.fl. Innen disse disiplinene har man tidlig sett nytteverdien av å dele data og har startet arbeidet med å gjøre data tilgjengelige. Innen andre disipliner er det fortsatt mer vanlig at forskningsmiljøene anser data som konkurransefortrinn og helst vil holde dem tett til brystet.

FAIR-prinsippene innebærer en policyendring hvor data ikke bare skal gjøres tilgjengelige, men at dette skal foregå på en måte som legger til rette for at datamaskiner kan søke, sammenstille og gjenbruke dataene. Gjenbruk skal altså kunne skje med minst mulig behov for tidkrevende manuelle prosesser.

Ikke alle typer data kan gjøres åpent tilgjengelige for gjenbruk, uten spesielle vurderinger og ulike manuelle rutiner. Det åpenbare eksempelet er data som inneholder personidentifiserbar informasjon, hvor det hersker spesielle betingelser til bruken av dataene. Men også personidentifiserbar informasjon kan gjøres mer tilgjengelig ved å koble beskrivende data (ofte kalt metadata) til de aktuelle datasettene. Slike beskrivende data kan i sin tur gjøres søkbare, slik at forskere kan oppdage at datasettene finnes, og deretter kan søke om å få tilgang. Ved både UiO og NTNU har man for eksempel utviklet metodikk for å legge til beskrivende informasjon til videoopptak av personer, slik at det skal bli mulig å søke i disse, selv om ikke selve videoen uten videre kan er åpent tilgjengelig. Bruk av relevante metadata for søk og gjenfinning er en viktig del av FAIR-prinsippene.

## 2.2. MANDAT

Oppdraget om denne utredningen er gitt av Digitaliseringsstyret for høyere utdanning og forskning. Mottakere av utredningens funn og anbefalinger er altså formelt sett UH-institusjonene, men de vil også være relevante for alle institusjoner som driver med forskning, f.eks. forskningsinstituttene og helseforetakene og for leverandører av relevant infrastruktur. Resultatet av utredningen vil også bli presentert for Kunnskapsdepartementet (KD).

I styringsdokumentets kapittel om hensikten med utredningen står følgende:

*Hovedhensikten med utredningen er å fremskaffe informasjon som gir veiledning til både institusjonene og den enkelte forsker om relevante infrastrukturer og tjenester som finnes i dag, samt gir veiledning til tjenestetilbyderne gjennom felles retningslinjer for hvilken funksjonalitet tjenestene bør tilby. Utredningen skal også identifisere eventuelle behov for nye eller forbedrede infrastrukturer og tjenester. Dette beskrives i mer detalj videre i dette kapitlet.*

Videre i kapitlet sies det:

*For å nå målet om at forskningsdata skal behandles forsvarlig i alle faser av et forskningsprosjekt, men samtidig være så åpent tilgjengelig som mulig, er det nødvendig gjennom hele forskningsprosessen å gi forskerne tilgang til tjenester som bidrar til å nå begge disse målene på en hensiktsmessig og minst mulig arbeidskrevende måte.*

*Utredningen er et første steg på veien for å identifisere hva som mangler for å dekke disse behovene.*

*Følgende aktiviteter skal gjennomføres:*

- 1. Etablere et sett med norske begreper for bruk knyttet til forskningsdata. Oppgaven med å foreslå et slikt begrepsapparat foreslås satt ut til en arbeidsgruppe under den norske noden av RDA (Research Data Alliance).*
- 2. Gjennomføre en oppdatert kartlegging av eksisterende infrastrukturer og tjenester for håndtering av forskningsdata i alle faser av arbeid med forskningsdata. Kartleggingen skal dekke internasjonale tjenester, nasjonale tjenester og lokale tjenester som tilbys til brukere utenfor egen institusjon. Kartleggingen skal dekke både persondata og andre typer data. Det skal legges mest vekt på generiske tjenester, men fagspesifikke tjenester*

*kan være aktuelt å ta med dersom de har potensial til å benyttes utenfor det opprinnelige fagområdet. Kartleggingen skal være et supplement til tidligere kartlegginger som er gjort og til nasjonale og internasjonale veikart for forskningsinfrastrukturer og skal bygge på disse.*

3. *Synliggjøre resultatene av kartleggingen på en nettside i tilknytning til openscience.no. Nettsiden skal inneholde en oversikt over anbefalte tjenester og eksemplifisere og visualisere disse på en måte som gjør det enklest mulig å finne frem i informasjonen.*
4. *Definere retningslinjer med et sett kriterier som tjenester knyttet til forskningsdata bør oppfylle. Informasjon om hvilke kriterier den enkelte tjeneste oppfyller, kan publiseres på nettsiden i pkt 3.*
5. *Gi anbefalinger for eventuelle videre prosjekter for å komplettere kartlagte tjenester, med spesiell vekt på langtidsbevaring av forskningsdata.*

Av tidligere relevante arbeider nevnes blant annet Forskningsrådets veikart, Datalisensutvalgets arbeid og Open Science Toolbox. Utredningsgruppen er også forutsatt å skulle samarbeide nært med Datainfrastruktur-utvalget som nå er i gang i regi av Forskningsrådet.

## 2.3. PROSESS OG METODE

Arbeidet med utredningen startet opp i april 2021. Datainnsamlingen har pågått siden juni og pågår fortsatt (desember 2021).

### 2.3.1. Organisering av arbeidet

Selve utredningen er utført av en utredningsgruppe med et sekretariat:

Utredningsgruppen	Sekretariat
Magnus Aronsen, Universitetet i Oslo (leder)	Kristin Brænden, Agenda Kaupang
Erik Ø. Sørensen, Norges Handelshøyskole	Gjermund Lanestedt, Agenda Kaupang
Karoline Daugstad, NTNU	Trond Kvamme, NSD
Susan Nacey, Høgskolen i Innlandet	Ulrike Jaekel, Forskningsrådet
Valeria Tafintseva, NMBU	Nenitha Charlotte Dagslott, Forskningsrådet
Audun Beyer, Høgskolen Kristiania	Katrine Weisteen Bjerde, HK-dir
Wenche Reed, Oslo universitetssykehus	Siri Lader Bruhn, Forskningsrådet (til juni)
Haakon Skogseth, St. Olavs hospital	
Frode Eika Sandnes, OsloMet	
Kjell Morten Stormark, NORCE	
Kirsti Klette, Universitetet i Oslo	
Arne Solli, Universitetet i Bergen	
Ane Gabrielsen, NTNU	
Elin Stangeland, Universitetet i Oslo	

Utredningsgruppen har også kunnet spille på to referansegrupper, der den ene bestående av forskere og den andre av representanter fra tjenesteleverandører og forskerstøtte.



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 utredningen. Denne arbeidsgruppen koordineres av NSD.

En styringsgruppe har hatt overordnet ansvar for mandat, økonomi og framdrift i utredningsarbeidet. Styringsgruppen har hatt følgende medlemmer:

Tomas Willebrand, prorektor for forskning, Høgskolen i Innlandet (leder)
Jørgen Berge, prodekan for forskning, UiT Norges arktiske universitet
Maria-Carme Torras, bibliotekdirektør, Universitetet i Bergen
Gard Thomassen, underdirektør, Universitetet i Oslo
Solveig Flock, avdelingsdirektør, Forskningsrådet
Nina Karlstrøm, seksjonsleder, Unit
Katrine Weisteen Bjerde, fagdirektør, HK-dir

### 2.3.2. Status for leveransene

Utredningen skal være slutført innen 1. mars 2022. Ved utsending av dette utkastet er status for de ulike leveransene som følger:

- **Norske begreper/terminologi**  
Det er etablert en liste med ca. 100 begrepene som skal defineres. Noen av de mest sentrale forslagene er tatt inn i dette utkastet til rapport med foreløpige forslag til definisjoner.
- **Kartlegging**  
Denne er i all hovedsak gjennomført, men det gjenstår fortsatt en del analyse av det datamaterialet som er kommet inn gjennom svarene fra spørreundersøkelsene blant forskere og forskningsinstitusjoner, og de intervjuer som er gjennomført. Denne rapporten gjengir de viktigste funnene. Kunnskapsgrunnlaget vil oppdateres helt frem mot sluttleveransen. En viktig milepæl blir innarbeiding av innspillene til den foreløpige rapporten i januar.
- **Nettsider under openscience.no**  
Her foreligger det skisser til innhold og utforming, men mye gjenstår å utvikle. Open Science Toolbox, som allerede ligger på disse sidene, danner et godt grunnlag for deler av arbeidet. Primær målgruppe for nettsidene er institusjonene og underliggende enheter. Vi forventer at enkeltforskere i første omgang forholder seg til informasjon fra egen institusjon.
- **Retningslinjer/kriterier som tjenester knyttet til forskningsdata bør oppfylle**  
Et første utkast er diskutert med referansegruppen for tjenesteleverandører og forskerstøtte. Dette blir nå tatt videre til diskusjon med institusjoner som ikke selv er tjenesteleverandører og som dermed må antas å være de mest aktuelle brukerne av tjenestene og den informasjonen som skal ligge på nettsidene. Noen av forslagene er også med som anbefalinger i dette rapportutkastet.
- **Anbefalinger for eventuelle videre tiltak og prosjekter**  
Dette rapportutkastet viser status for denne delen av arbeidet. De anbefalingene som



er gitt er foreløpig ikke formulert som forslag til konkrete tiltak eller prosjekter. Dette vurderes løpende frem mot endelig leveranse.

Ambisjonen har vært å ha et utkast til alle leveransene senest 23.12.2021. Når dette rapportutkastet er publisert, flyttes fokus de siste dagene frem mot jul til arbeidet med terminologi, nettside og retningslinjer.

### **2.3.3. Datainnsamling og kilder**

Den sentrale delen av vår datainnsamling er to spørreundersøkelser, én rettet mot forskere og én rettet mot institusjonene. Spørreundersøkelsene er supplert med intervjuer med forskere og presentasjoner av funn i ulike fora der deltakerne har fått mulighet til å kommentere og gi innspill. Datainnsamling og analyse pågår fortsatt på tidspunktet dette rapportutkastet distribueres. Informasjonen som fremgår må derfor ses på som foreløpig.

#### **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 og til Forskningsinstituttene Fellesarenas (FFA)<sup>3</sup> medlemsinstitutter. Totalt deltok 531 personer.

Undersøkelsen som er gjengitt i vedlegg 1 (separat vedlegg) handlet om tjenester og verktøy for FAIR forskningsdata, med spørsmål om hvordan forskerne håndterte sine forskningsdata i dag, hvilke tjenester og verktøy de hadde erfaring med å bruke, og om det var behov som dagens tilgjengelige tjenester og verktøy ikke dekker.

#### **Spørreundersøkelse til institusjonene**

I perioden september til november gjennomførte vi en spørreundersøkelse rettet mot institusjonene. Undersøkelsen ble sendt til postmottak til UH-institusjonene, instituttene og helseforetakene. Totalt har vi per 29.11 fått svar fra 60 institusjoner, og vi mottar fortsatt svar.

Undersøkelsen som er gjengitt i vedlegg 2 (separat vedlegg) handlet om institusjonenes arbeid med policies 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 opplever god bredde i utvalget av besvarelser.

#### **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 har vi gjennomført intervjuer og samtaler med relevante aktører. Dette er et pågående arbeid.

Per 13. desember har vi gjennomført to gruppeintervjuer med til sammen seks forskere som har erfaring med å håndtere persondata, og fire enkeltintervjuer med forskere. Vi har også snakket

---

<sup>3</sup> Forskningsinstituttene fellesarena: <https://www.abelia.no/bransjeforeninger/ffa-forskningsinstituttene-fellesarena/>



med en leder ved en institusjon, en infrastrukturleverandør og to miljøer som har kommet langt på området.

### 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 policies, og ulike relevante problemstillinger knyttet til FAIR forskningsdata.

Dokumentstudiene er en del av grunnlaget for anbefalingene i denne utredningen. Blant de dokumentene vi har gjennomgått er:

- Damvad (2014) *Sharing and archiving of publicly funded research data*. Rapport til Norges forskningsråd.
- Norges forskningsråd (2014) *Open Access to Research Data*. Policy
- Norges forskningsråd (2021) *Hvordan skal vi dele forskningsdata?* Utredning og anbefaling om lisensiering og tilgjengeliggjøring.  
<https://www.forskningsradet.no/siteassets/publikasjoner/2021/rettighets--og-lisenssporsmal-ved-tilgjengeliggjoring-av-forskningsdata.pdf>
- Samarbeidsforum for samordning mellom universiteter med medisinstudium og helseforetak (2020) *Håndtering av helseforskningsdata – pseudonym data i klinisk forskning*. Rapport fra en arbeidsgruppe.
- Helsedataplattformen Midt-Norge
- SOU 2021:65 - Stärkt fokus på framtidens forskningsinfrastruktur
- Flere relevante dokumenter og pågående initiativer i utredningens [mandat](#)

I tillegg kommer enkeltartikler om tematikken; ulike fremstillinger av FAIR-prinsippene (slik som <https://www.go-fair.org/fair-principles/>); policies og retningslinjer knyttet til FAIR forskningsdata, presentert på norske institusjoners nettsider; tilgjengelig åpen informasjon og presentasjoner av relevante nasjonale infrastrukturer og tjenester; og tilsvarende om internasjonale forskningsinfrastrukturer og tilknyttede tjenester, slik som Elixir (<https://elixir-europe.org/>) og EBRAINS (<https://ebrains.eu/>).

### Presentasjoner og innspillmøter

Vi har gitt en rekke korte presentasjoner (allerede gitt eller avtalt i nær fremtid) for å gjøre oppmerksom på arbeidet og oppfordre til å gi innspill i løpet av desember/januar:

- UHR-Bibliotek (gjennomført)
- UHR-Forskning
- RHFenes strategigruppe for forskning
- Nasjonalt forum for åpen forskning
- UHR-Kunst, design og arkitekturs arbeidsgruppe for kunstnerisk utviklingsarbeid

### Diskusjoner i og innsiktsbidrag fra utredningsgruppen og sekretariatet

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.

## 2.4. AVGRENSNINGER OG TOLKNING AV MANDATET

I oppdraget har fokus vært på generiske infrastrukturer og tjenester. Unntaket er der fagspesifikke infrastrukturer eller tjenester er trukket frem som gode eksempler og der vi ser at infrastrukturer eller tjenester som i dag er brukt fagspesifikt, har potensial til å løftes til bruk på flere fagområder.

Hovedmålet med arbeidet er å bidra til å få på plass infrastruktur som gir forskningsdata som ikke bare er tilgjengelige, men også mest mulig automatisk gjenbrukbare (se mer detaljert beskrivelse av FAIR data i definisjonkapitlet). Det er imidlertid en erkjennelse at det er en lang vei frem dit. 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 for å nå målet.

Denne utredningen fokuserer på data generert som del av et forskningsprosjekt. Vi har ikke sett på data fra forvaltningen som er interessante for forskere å få tilgang til, og som er et tema som dekkes av Datainfrastruktur-utvalget. Vi har likevel tatt med et par eksempler der slike data har stor relevans også for vårt arbeid. Det er heller ikke alltid lett å sette et klart skille mellom henholdsvis forskningsdata og forvaltningens dataressurser.

Vi berører i liten grad problemstillinger knyttet til opphavsrett, lisensiering av forskningsdata og finansiering av videre arbeid. Det første er behandlet i Datalisens-utvalgets rapport, som vi bygger videre på. Det siste behandles nå, delvis i parallell, av Datainfrastruktur-utvalget. Vi gir imidlertid løpende innspill til dette utvalget på de områder hvor vi ser at våre anbefalinger vil kreve omfattende finansiering.

Innledningsvis diskuterte utredningsgruppen hva som kan gjøres for å motivere flere forskere til å dele sine data. Dette behandles i noen grad, men i hovedsak har vi besluttet å fokusere på de som allerede er motivert for å dele men som opplever at dette er krevende, enten fordi de mangler kunnskap om hvordan de skal få det til, eller fordi de ikke har tilgang til det som trengs for å gjøre det.

## 2.5. DEFINISJONER

**Forskningsdata:** Med forskningsdata mener vi i denne utredningen registreringer/nedtegnelser/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, sammenstilling av eksisterende data, eller helt nye data generert gjennom ny datainnsamling. Forskningsdata er alltid et direkte resultat av forskningen, uavhengig av om dataene baserer seg på eksisterende data eller om de er samlet inn på nytt  
[Kilde: Forskningsrådet]



**Åpne data:** Åpne data er digitale data som er gjort tilgjengelige med de tekniske og juridiske egenskaper som er nødvendige for at de skal kunne brukes fritt, og bli distribuert av hvem som helst, når som helst og hvor som helst. [Kilde: <https://data.norge.no/concepts/b4f0dedb-9cde-489d-9a50-59d225f407ad> ]

**Metadata** er *data om data*, eller informasjon som beskriver annen informasjon. Det er strukturert informasjon som lokaliserer, beskriver, forklarer, eller på andre måter gjør det enklere å hente, (gjen)bruke eller forvalte en informasjonsressurs/et digitalt objekt.<sup>4</sup> Det er vanlig å skille mellom *deskriptive* metadata og *administrative* metadata.

*Deskriptive* metadata inneholder identifiserende egenskaper som tittel, sammendrag, nøkkelord, metode, o.l. Noen av disse opplysningene gjøres søkbare for gjenfinning. I deskriptive data inngår også opplysninger om hvor en kan finne ressursen/det digitale objektet.

*Administrative* metadata er informasjon som kan brukes til å administrere/forvalte en (digital) ressurs. Eksempler på administrative metadata er informasjon om når og hvordan ressursen ble opprettet, teknisk informasjon (datatyper, format, programvare, o.l.), og informasjon om hvem som har tilgang til den. Administrative metadata kan også inneholde informasjon om immaterielle rettigheter (opphavsrettigheter, lisenser, o.l.) og informasjon som trengs for å arkivere og bevare en ressurs ('bevaringsmetadata').

I en del sammenhenger gir det mening å skille mellom *søkemetadata* og *bruksmetadata*. Søkemetadata er den informasjon som er til hjelp for å finne datasett som er relevante for den aktuelle bruken. Det vil normalt være informasjon som kan forstås uavhengig av disiplin. Bruksmetadata på den andre siden er informasjon du ikke egentlig trenger før du faktisk skal bruke eller behandle dataene, men da er de helt nødvendige. Dette er oftest fagspesifikk informasjon f.eks. om metode for å samle inn dataene. Informasjonen er ofte uttrykt gjennom fagspesifikke standarder.

Begrepet metadata er helt sentralt i håndtering av både forskningsdata og andre typer data. Dette er maskinlesbare beskrivende data som bidrar til forståelse av hva slags data det er snakk om og hvordan de skal tolkes. Hvis selve datasettet for eksempel er en rekke tall, kan metadata beskrive at tallene dreier seg om høyde i cm på norske rekrutter gjennom et antall år. Dersom dataene er en serie røntgenbilder, kan metadata angi hvilken kroppsdel bildene viser, fødsels- og personnummer på personene det er bilde av osv.

**FAIR-prinsippene:** er et helt sentralt element i åpen forskning. FAIR står for Findable, Accessible, Interoperable og Re-usable og handler i all hovedsak om å håndtere og beskrive data på en måte som setter andre i stand til å finne, forstå og bruke dem. På norsk kan man si at dataene

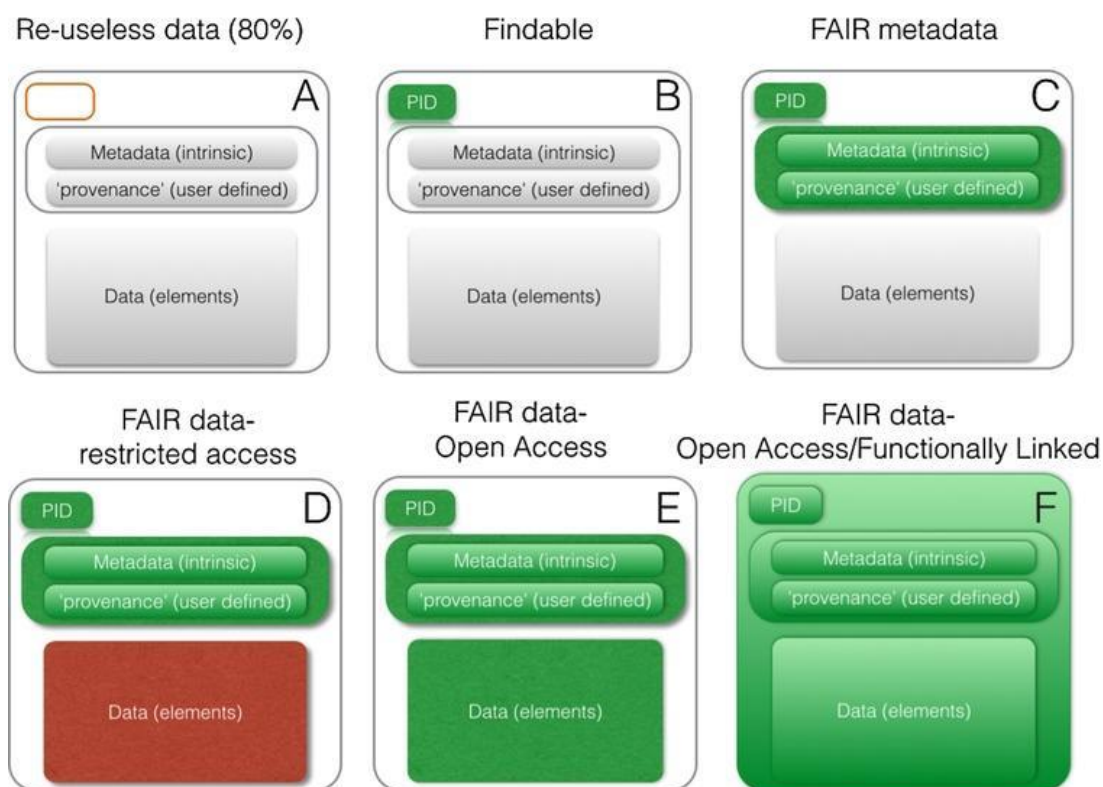
---

<sup>4</sup> Mer informasjon om metadata kan man finne f.eks. på <http://www.niso.org/publications/understanding-metadata-2017> og <https://data.research.cornell.edu/content/readme>

skal være gjenfinnbare, tilgjengelige, samhandlende og gjenbrukbare. Mer informasjon om hva FAIR data innebærer er gitt på [opendscience.no](https://opendscience.no) og på OG FAIR<sup>5</sup>.

For at data skal være fullt ut FAIR, må de ikke bare være gjenfinnbare og gjenbrukbare for et menneske, de må også være lesbare og gjenbrukbare for en datamaskin. Samtidig er FAIR ikke et absolutt begrep. Data kan være mer eller mindre FAIR. Figuren under er hentet fra en artikkel av Barend Mons et al i IOS Press<sup>6</sup> og viser at man kan definere mange grader eller trinn fra fullstendig lukkede data til helt FAIR data – og hvordan små grep kan gjøre at lukkede data mer tilgjengelige, selv om de fortsatt ikke er helt FAIR. I figuren er dataene i tilfelle A helt lukket og utilgjengelige. I tilfelle D finner man beskrivelsen av dataene, mens tilgangen til selve dataene er begrenset. I tilfelle F er dataene gjenfinnbare, tilgjengelige og funksjonelt mulig å utnytte.

## Data as increasingly FAIR Digital Objects



FIGUR 2: GRADER AV FAIR DATA. KILDE: BAREND MONS ET AL. (2017)

<sup>5</sup> <https://www.go-fair.org/fair-principles/>

<sup>6</sup> [Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. \(2017\). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use, 37\(1\), 49-56, DOI: 10.3233/ISU-170824](#)



**FAIR data:** [Fra lisensutvalget:] Mens åpne data hovedsakelig betyr at data kan brukes av enhver uten restriksjoner, legger betegnelsen FAIR data vekt på teknisk tilgjengelighet, noe som forutsetter gode, maskinlesbare metadata, inkludert lisensinformasjon. FAIR data skal være gjenfinnbare, tilgjengelige, compatible og gjenbrukbare uten at dette nødvendigvis betyr fri tilgang for alle.

**Infrastruktur:** Med infrastruktur mener vi i denne sammenheng *forskningsinfrastruktur*.

Forskningsinfrastrukturer er fasiliteter som tilbyr ressurser, verktøy og tjenester som forskningsmiljøene kan benytte seg av. Dette kan for eksempel være data(behandlings)systemer, kommunikasjonsnettverk, utstyr eller instrumenter, lagringsløsninger, arkiver, o.l.

[se: [https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/research-infrastructures\\_en](https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/research-infrastructures_en) ]

### Lagring vs arkivering

I et forskningsprosjekt er det viktig å ha klart for seg forskjellen mellom *lagring* og *arkivering* av data. Lagring handler gjerne om oppbevaring av aktive («varme») data *underveis* i et prosjekt. Aktive data kan prosesseres, analyseres, endres, deles eller på andre måter brukes operativt underveis i prosjektet. Aktive data kan i så måte bestå av både rådata og prosesserte data. En slik form for aktiv 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.

Arkivering handler om *langtidsbevaring* av data *etter* at prosjektet er avsluttet eller ved publisering av artikkel. Når data arkiveres så lages det *statiske* («kalde») versjoner av data, gjerne basert på et utvalg - for eksempel de dataene som danner grunnlaget for en forskningspublikasjon eller lignende. Sagt på en annen måte: de dataene som velges ut for arkivering anses å ha en viss *verdi* for (videre) forskning og/eller som grunnlag for validering av forskningsresultater. Data som arkiveres berikes med metadata og annen dokumentasjon, slik at de kan forstås, langtidsbevares og (eventuelt) gjøres tilgjengelige også for andre enn de som deltok i innsamlingen/genereringen av dataene.

### Publisering av data

Publisering eller deling av data innebærer det å gjøre data åpent tilgjengelige for gjenbruk. Dette kan for eksempel gjøres ved å deponere og publisere data i et arkiv for forskningsdata. Et viktig aspekt av datadeling handler om i hvilken grad publiserte data kan bli funnet, vurdert, forstått og gjenbrukt av andre enn den/de som opprinnelig samlet inn eller genererte dataene.

### Klassifisering av data

UH-sektorens sekretariat for informasjonssikkerhet har i samarbeid med UH-sektoren utarbeidet en [veileder for klassifisering av informasjon](#). Veilederen definerer fire klasser for konfidensialitet (klasser som beskriver grad av beskyttelse som kreves for informasjon), nemlig grønne, gule, røde og sorte data. *Åpen* (grønn) kategori betyr at data kan være tilgjengelig for alle uten særskilte tilgangsrettigheter; *intern* (gul) betyr at data må ha en viss beskyttelse, og at



tilgang kontrolleres med tilgangsrettigheter; *fortrolig* (rød) betyr at data skal ha høy beskyttelse og skal ha strenge tilgangsrettigheter; mens *strengt fortrolig* (sort) betyr at data skal ha de strengeste tilgangsrettigheter. Når det gjelder forskningsdata så kan disse kategoriene anvendes for å avgjøre hvor og hvordan data skal håndteres, lagres, prosesseres og arkiveres. Det som gjerne avgjør om forskningsdata havner i den ene eller andre kategorien, er om data inneholder personopplysninger (altså enhver opplysning om en identifisert/identifiserbar enkeltperson) eller ikke. Forskningsdata som ikke inneholder personopplysninger og/eller som er anonyme klassifiseres gjerne som grønne data, mens data med personopplysninger gjerne klassifiseres som gule, røde eller sorte, avhengig av grad av sensitivitet og datamengde.

Det finnes også andre grunner til at data krever ekstra beskyttelse. Eksempler på tilfeller der data ikke kan deles (eller kun kan deles med begrensninger) kan være kontraktsforpliktelser og taushetsavtaler med oppdragsgiver/finansiør, interesser knyttet til opphavsrettigheter og patenter, kommersielle interesser, sensitiv informasjon om bedrifter eller organisasjoner, og lignende.

### **Dataturering**

Dataturering innebærer å innordne (tilfang) i vitenskapelige eller andre faglige system. Det er ofte en oppgave man gjør 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

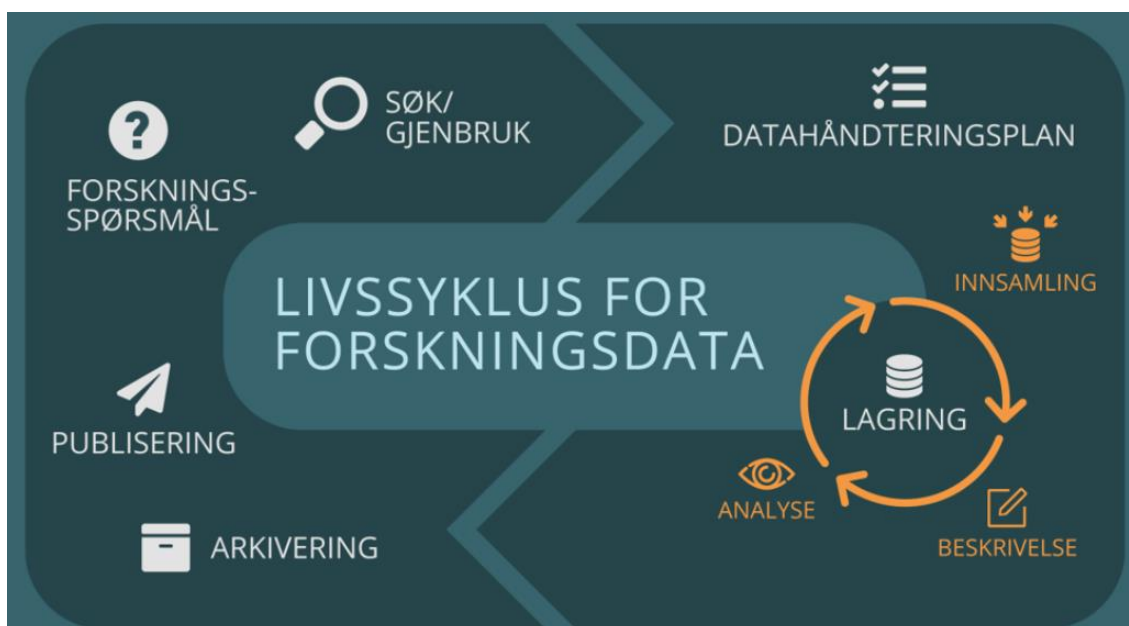


### 3. HVILKE TJENESTER OG VERKTØY BRUKES I DAG?

I dette kapitlet beskrives tjenester og verktøy for FAIR forskningsdata som brukes av forskerne og institusjonene i dag i henhold til kartleggingen. Kapitlet er delt opp etter de ulike stegene i livssyklusen til forskningsdata. Tilgang til kompetanse og bistand er skilt ut da dette ofte går på tvers av stegene.

#### 3.1. INTRODUKSJON - LIVSSYKLUSEN TIL FORSKNINGSDATA

Forskning er et svært heterogent område og det er ingen standardisert måte å gjennomføre et forskningsprosjekt på. I denne rapporten bruker vi livssyklusen til forskningsdata for å illustrere hvordan datahåndtering treffer ulike deler av forskningsprosessen fra innledende idé til avsluttet prosjekt.



FIGUR 3: LIVSSYKLUS FOR FORSKNINGSDATA. CC-BY NTNU

Det er en forenklet fremstilling, og i prosjektene beveger forskerne seg mellom de ulike fasene mer smidig enn det kan fremstå i modellen. Det er åpenbare ulikheter mellom disiplinene når det gjelder hvordan man jobber i de ulike fasene. Ett eksempel er forskjellen mellom et forskningsprosjekt der man primært baserer seg på eksisterende data og ett datainnsamling en stor del av jobben. Til tross for åpenbare forskjeller er det såpass store likheter, at fasene i modellen kan fungere som knagger for de ulike aktivitetene knyttet til datahåndtering.

## 3.2. INNLEDENDE FASE

Den innledende fasen består av flere delprosesser etter at forskningsidéen er etablert. Et ledd i dette arbeidet er å gjøre seg kjent med eksisterende forskning og eventuell tilgjengelig data på området. Videre planlegges datainnsamling og håndtering av data underveis og etter endt prosjekt. En datahåndteringsplan kan strukturere denne prosessen og bidra til at det tas valg innledningsvis som gjør det enklere å gjøre data FAIR på sikt. I denne fasen har vi etterspurt hva forskerne bruker av søketjenester for å finne relevante data og verktøy for å lage datahåndteringsplan.

### 3.2.1. Søke etter eksisterende data

For å avklare behovet for ny datainnsamling er det nyttig å søke etter eksisterende data. Det er mulig å søke direkte i arkivene eller via søketjenester som muliggjør søk på tvers av arkiver.

Det finnes både fagspesifikke og generelle søketjenester. Et eksempel på en fagspesifikk tjeneste er Open Polar, beskrevet i Boks 1. For søk i arkivene viser vi til oversikten over arkiver som brukes av forskerne i dag.

#### Eksempler på søketjenester:

- [OpenAIRE Explore](#)
- [Re3data.org](#)
- [Zenodo](#)
- NSDs [arkiver](#)
- [NIRD Research data archive](#) fra Uninett Sigma2
- [DataverseNO](#)
- [DataCite Search](#)
- [DataCite Commons](#)
- [Google Dataset Search](#)
- Bielefeld Academic Search Engine ([BASE](#))

#### Andre ressurser:

- CESSDAs [informasjonsside](#) med veiledning om hvordan man kan finne eksisterende data

#### CASE: OPEN POLAR

Open polar er et samarbeid mellom Universitetet i Tromsø – Norges arktiske universitet og Norsk Polarinstitutt. Portalen samler metadata om datasett fra polarforskning fra ulike kilder for å gi en samlet oversikt over tilgjengelig data på området. Per i dag finnes det metadata fra over 4600 kilder i portalen. Kilder inkluderer blant annet norske og internasjonale arkiver.

Kilder:

<https://site.uit.no/open-polar/about-2/>  
<https://doi.org/10.7557/5.6209>

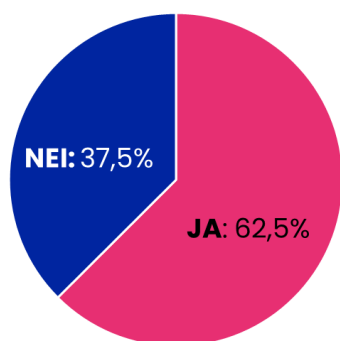
Boks 1

### 3.2.2. Lage datahåndteringsplan

En datahåndteringsplan er et dokument som beskriver hvordan data planlegges håndtert fra start til slutt i et forskningsprosjekt<sup>7</sup>. Dette er det sjelden mulig å forutsi ved prosjektstart, så planen er ment å være et levende dokument som oppdateres underveis i prosjektet. I dag lages planene direkte som tekstfiler eller ved hjelp av nettbaserte verktøy. Ved bruk av nettbaserte verktøy kan datahåndteringsplanen lagres i et maskinlesbart format og forskerne kan legge på innledende metadata. En FAIR tjeneste må kunne tilby andre å gjenbruke disse metadataene.

I dag brukes både nasjonale og internasjonale verktøy. Det er flere av institusjonene som har laget egne tekstmalere for datahåndteringsplan, og i flere av de aktuelle verktøyene kan man lage digitale maler for ulike formål, f.eks. for å tilpasse til en spesiell fagdisiplin.

Har brukt verktøy/tjenester for å lage datahåndteringsplan



I spørreundersøkelsen til forskerne var det 62,5 % som oppga at de hadde erfaring med tjenester for å planlegge datahåndtering. Av de som hadde erfaring hadde flest brukt mal fra egen institusjon og NSDs DMP var det mest brukte verktøyet. Av de som ikke hadde erfaring oppga forskerne manglende kjennskap til tjenestene som hovedårsak.

FIGUR 4: ANDEL FORSKERE SOM HADDE ERFARING MED TJENESTER FOR DATAHÅNTERINGSPLAN

I spørreundersøkelsen til institusjonene var også NSDs verktøy for datahåndteringsplan det verktøyet flest anbefalte sine forskere per i dag. Noen av institusjonene har laget egne veiledere for hvilke verktøy som kan brukes til ulike prosjekter og som inneholder informasjon om hvordan lage en datahåndteringsplan. Se eksempel på veiledere fra institusjonene ved å følge lenkene i Boks 2.

#### Eksempler på guider fra institusjonene:

- [NTNUs informasjonsside om datahåndteringsplaner](#)
- [UiTs informasjonsside om datahåndteringsplaner](#)
- [UiOs informasjonsside om datahåndteringsplaner](#)

#### BOKS 2

Eksempler på datahåndteringsplanverktøy (i tilfeldig rekkefølge) som brukes av forskerne i dag:

- [NSDs DMP](#)
- [easyDMP](#) fra Uninett Sigma 2
- [DMPonline](#) fra Digital Curation Centre (DCC)
- [DMP](#) fra Data Stewardship Wizard
- <https://dmptool.org/> fra California Digital Library (CDL)

<sup>7</sup> <https://innsida.ntnu.no/wiki/-/wiki/Norsk/Datah%C3%A5ndteringsplan>

- ELIXIR [Research Data Management Kit](#) (RDMkit)
- [Argos](#) fra OpenAIRE

#### Andre tjenester og ressurser:

- [CEDAR](#) – plattform for metadata rettet mot biomedisin
- [Veileder](#) i hvordan klassifisere informasjon fra Unit
- [Ressurside](#) om datahåndteringsplan på openscience.no

### 3.3. HÅNDBERE FORSKNINGSDATA I PÅGÅENDE PROSJEKTER

Den aktive fasen av et forskningsprosjekt består av flere delprosesser; innhenting og/eller innsamling av data, lagring og analyse. For innsamling og analyse brukes det svært mange ulike verktøy spesifikt for de ulike disiplinene, og vi har derfor avgrenset vårt hovedfokus til tjenester for lagring av aktive data.

#### Innsamling og analyse

I undersøkelsen til forskerne og institusjonene nevnes det eksempler på flere ulike typer løsninger for innsamling og analyse av data. For innsamling finnes det flere eksempler på tjenester der forskerne både kan samle inn data, f.eks. gjennom spørreundersøkelser, og lagre dataene. Et eksempel på en slik tjeneste er Nettskjema fra UiO<sup>8</sup> der man kan lage spørreskjemaer og samle inn og lagre opptil sorte (sensitive) data. For analyse er det tilsvarende ulike løsninger der man kan lagre og analysere data, for eksempel NIRD Service Platform<sup>9</sup> fra Uninett Sigma2.

#### Eksempler på tjenester for innsamling og behandling av data:

- [UiOs nettskjema](#)
- [eFORSK](#) fra HEMIT (Helse Midt-Norge IT)

#### Lagring og deling i pågående prosjekter

Det er flere gode eksempler på forskningsmiljøer og fagfelt som jobber aktivt med deling av data underveis i prosjektene. Et eksempel er arbeidet til EBRAINS beskrevet i Boks 3.

---

<sup>8</sup> <https://nettskjema.no/>

<sup>9</sup> <https://www.sigma2.no/nird-service-plattform>

### CASE: EBRAINS

EBRAINS er en digital forskningsinfrastruktur laget i forbindelse med EU-prosjektet Human Brain Project. Den organiseres som et internasjonalt samarbeid og består av nasjonale noder. Infrastrukturen samler data og verktøy for hjerneforskning. De tilbyr også forskerstøtte og bistand til datakuratering.

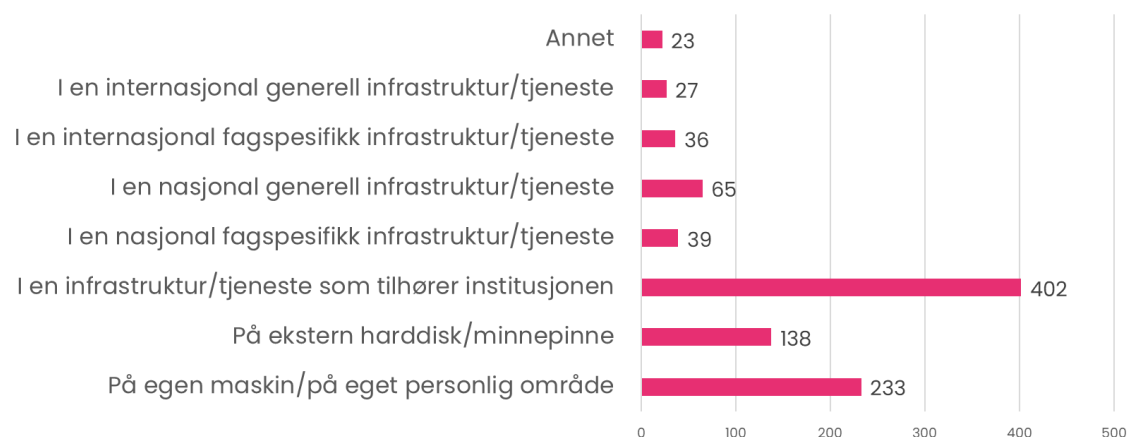
Prosjektet jobber aktivt for at forskerne skal dele data under pågående prosjekt, og har blant annet lagt vekt på å lage enkle, brukervennlige metadataskjemaer. Forskerne kan også velge mellom en «standard» og «avansert» datakurateringsprosess. Det er noden ved Universitetet i Oslo som koordinerer kuratering av metadata.

Mer informasjon: <https://ebrains.eu/>

#### Boks 3

I spørreundersøkelsen til forskerne var et av spørsmålene hvor de primært lagrer data i pågående prosjekter. Figur 5 viser en oversikt over hvor forskerne som deltok i spørreundersøkelsen primært lagrer data. De fleste lagret data i tjenester og infrastrukturer som tilhører institusjonen og en del lagrer på personlig område eller eksterne medier som harddisk og minnepinne.

### Hvor lagres primært data i pågående prosjekter?



FIGUR 5: OVERSIKT OVER HVOR FORSKERNE I UNDERSØKELSEN PRIMÆRT LAGRET DATA (FLERE SVAR MULIG)

Videre stilte vi spørsmål til institusjonene hvor de anbefalte sine forskere å lagre data i pågående prosjekter. Mange har egne forskningsservere eller bruker kommersielle tjenester som Microsoft 365, Sharepoint, Onedrive og lignende. Det er forskjeller mellom institusjoner og fagfelt. Flere har egne lagringsguider og anbefalinger basert på klassifisering av dataene. For behandling av sensitive data er *Tjenester for sensitive data* (TSD) fra UiO en tjeneste som flere anbefaler.



### Eksempler på generiske tjenester for lagring av data under pågående prosjekt:

- [NIRD Data Storage](#) fra Uninett Sigma 2
- [Tjenester for Sensitive Data](#) (TSD) fra UiO
- [Educloud Research](#) fra UiO
- [SAFE](#) (Sikker Adgang til Forskningsdata og E-infrastruktur) fra UiB

### Eksempler på fagspesifikke tjenester for lagring av data under pågående prosjekt:

- [HUNT Cloud](#) driftes av NTNU – plattform for sensitive helsedata
- [EBRAINS](#) – digital infrastruktur for hjerneforskning
- [ELIXIR](#) – ressurser og plattformer for livsvitenskapsdata

Det finnes også tjenester der forskerne både kan lagre data underveis i pågående prosjekter og arkivere dataene etter prosjektslutt. Et eksempel på en helhetlig tjeneste for forskningsdata er Yoda<sup>10</sup> fra Universitetet i Utrecht som kan brukes til både lagring, arkivering og publisering av opptil sensitive data.

### Metadata

En viktig del av prosessen for å gjøre data FAIR er å legge på gode metadata. For å lage strukturerte metadata kan forskerne bruke verktøy og/eller en metadatastandard. Det finnes både generelle og disiplinspesifikke metadatastandarder. Et eksempel på en generell metadatastandard er Dublin Core som består av 15 datafelter for å beskrive datasett og andre ressurser som er tilgjengelig via nettet<sup>11</sup>.

Arbeidet med gode metadata ved publisering av datasett forenkles dersom metadata legges på underveis i prosjektene. Gjennom utredningen har vi snakket med forskere fra ulike disipliner, og elektroniske lab-bøker med standardiserte metadata trekkes frem som et godt hjelpemiddel for forskere som gjennomfører forsøk. Et eksempel på et prosjekt som har arbeidet mye med standardisering av metadata er Arven etter Nansen som beskrives i Boks 4 Et annet eksempel er UiOs prosjekt Fair@UiO beskrevet i Boks 5.

#### **CASE: Arven etter Nansen og deres arbeid med metadata**

Arven etter Nansen er et forskningsprosjekt der forskere fra ti norske institusjoner undersøker klima og økosystem i det nordlige Barentshavet. Forskningsskipet og isbryteren «Kronprins Haakon» er hovedplattform for forskningen. I prosjektet har de jobbet aktivt med å få forskerne til å bruke tid på å legge på gode metadata i forkant av innsamlingen. De bruker også åpne, internasjonale metadatastandarder.

Kilder: <https://arvenetternansen.com/nb/om-oss/>

### Boks 4

<sup>10</sup> <https://www.uu.nl/en/research/yoda/what-is-yoda>

<sup>11</sup> <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/Documentation-and-metadata>

**CASE:****Fair@UiO**

UiO har et pågående prosjekt der de arbeider med å etablere en sentral oversikt over metadata hentet fra ulike kilde-systemer. En viktig del av arbeidet er å etablere et minimumssett av metadata som alle prosjekter må beskrive.

Kilde og mer informasjon: <https://www.usit.uio.no/prosjekter/fair-uio/>

**Boks 5****Andre ressurser:**

- CESSDAs [informasjonsside](#) om prosjektdokumentasjon og metadata
- RDAs [oversikt](#) over metadatastandarder og verktøy for metadata

### 3.4. ARKIVERE OG PUBLISERE FORSKNINGSDATA

Arkivering av data etter prosjektslutt er viktig for å sikre langtidsbevaring av dataene. Publisering av data gjør forskningen etterprøvable og gjør det enklest mulig for andre forskere å bruke dataene i nye prosjekter. Data kan arkiveres og publiseres åpent eller med tilgangsbegrensninger, avhengig av hva slags data det dreier seg om. Dersom det er personidentifiserbare opplysninger det dreier seg om, vil måten data arkiveres og eventuelt publiseres på avgjøres av hva det er gitt samtykke til.

Det er flere fagområder og miljøer som har gode eksempler på publisering av data. Et eksempel er arbeidet med åpne data til Metrologisk institutt beskrevet i Boks 6. Et annet eksempel er infrastrukturen CLARINO der et eksempel på en samling er NTNUs arbeid med et norsk tegnspråkkorpus beskrevet i **Error! Reference source not found.** NTNUs arbeid med tegnspråkkorpus er også et eksempel på hvordan man kan innhentesamtykke til åpen deling av persondata. Ungdata-undersøkelsene i regi av NOVA/OsloMet, beskrevet i Boks 8, er et eksempel på publisering av data om ungdoms livskvalitet og psykiske helse, fritidsbruk, rus, sosiale ulikheter etc., i en nasjonal database som brukes til mange ulike anvendelser.

### CASE: Meteorologisk institutt

Meteorologisk institutt (MET) tilgjengeliggjør store deler av sine data åpent via selvbetjente nedlastningstjenester. Instituttets dataprodukt består av:

- Alle data som er bearbeidet ved at det er beregnet statistiske elementer eller gjort transformasjoner, for eksempel av typen utplukk, formatering, visualisering, sammenstilling og tolkning utgjør METs produkter. Inkluderer tilhørende metadata.
- Produktene er et resultat av en bearbeiding med en verdiøkning ut over det datagrunnlaget gir og er basert på METs offisielle data.

Mer informasjon:

Om Met.no: <https://www.met.no/om-oss/om-meteorologisk-institutt>

Om Met.no sine datatjenester: <https://www.met.no/frie-meteorologiske-data/data-og-produkter-tilgjengelig-fra-met>

#### BOKS 8

### CASE: CLARINO og NTNUs korpus for tegnspråk

CLARINO er et norsk infrastrukturprosjekt som skal implementere den norske versjonen av Common Language Resources and Technology Infrastructure (CLARIN). Målet er å gjøre språkressurser enkelt tilgjengelig.

Et eksempel på en samling som er publisert i CLARINO er NTNUs korpus for tegnspråk. I prosjektet dokumenterer de blant annet bruk av norsk tegnspråk ved bruk av videoopptak. Forskerne har innhentet samtykke fra deltakerne til åpen deling av videoene og prosjektet er derfor også et eksempel på bruk av samtykke for å dele ikke-anonymisert persondata.

Mer informasjon:

Om CLARINO: <https://clarin.w.uib.no/about/>

Om korpus for tegnspråk: <https://www.ntnu.no/blogger/humsam/2019/01/10/skal-dokumentere-norsk-tegnsprak/>

#### BOKS 6

### CASE: UNGDATA

Ungdata er en nasjonal spørreskjema-undersøkelse blant elever på ungdoms- og i videregående skole. Formålet med undersøkelsen er å samle inn ny kunnskap om ungdoms oppvekstvilkår i Norge generelt. Dataene samles i en database, og er blant annet tilgjengelig via NSDs arkiver. De lånes ut til undervisnings- og forskningsformål etter søknad.

Velferdsforskningsinstituttet NOVA ved OsloMet – storbyuniversitetet er ansvarlig for gjennomføringen sammen med de regionale kompetansesentrene innen rusfeltet. På Ungdatas nettsider publiseres det også en [oversikt](#) over forskningsbaserte publikasjoner som bruker dataene.

Mer informasjon: <https://www.ungdata.no/ungdataforskning/>

#### BOKS 7





## Tjenester for arkivering og/eller publisering

Av forskerne som deltok i vår spørreundersøkelse var det 43,5 % som oppga at de hadde brukt tjenester for å arkivere eller dele data i avslutningsfasen av et prosjekt. Forskerne oppgir å bruke både nasjonale og internasjonale fagspesifikke og generiske arkiver. Det er også forskere som arkiverer data hos tidsskriftene i forbindelse med publisering av forskningsartikler.

Institusjonene anbefaler både fagspesifikke og generiske arkiver til forskerne. Flere av institusjonene har egne institusjonsarkiv tilknyttet DataverseNO, et nasjonalt generisk arkiv for åpne data som eies og driftes av UiT. Mange anbefaler også NSDs arkivtjenester. Hvilke arkiv som er egnet avhenger også av eventuelle behov for å beskytte dataene noe som påvirker anbefalingene til institusjonene.

### Eksempler på generiske tjenester for arkivering og/eller publisering:

- [DataverseNO](#) fra UiT
- NSDs [arkivtjenester](#)
- [NIRD Research Data Archive](#) – Uninett Sigma 2
- [Zenodo](#) – driftes av CERN/EU
- [Dryad](#) Digital Repository
- [Open Science Framework](#) (OSF) fra Center for Open Science, Inc

### Eksempler på fagspesifikke tjenester for arkivering og/eller publisering:

- [HUNT Cloud](#) driftes av NTNU
- [TROLLing](#) - The Tromsø Repository of Language and Linguistics driftes av UiT
- Norwegian eInfrastructure for Life Science ([NeLS](#))
- [CLARINO](#) - Common Language Resources and Technology Infrastructure Norway
- [EBRAINS](#)
- [Norwegian Marine Data Centre](#) (NMDC)
- [Bjerknes Climate Data Centre](#) (BCDC)
- [Norwegian Polar Data Centre](#) – fra Norsk Polarinstitut
- ELIXIR [Core Resources](#) and Recommended Deposition Databases
- CESSDA
- Flere eksempler på <https://www.re3data.org/>

Forskerne arkiverer også data i andre arkiver som inkluderer data fra andre sektorer og i forskningsregistrene. Det er blant annet flere registre der forskere som bruker registerdata leverer bearbejdede data tilbake til registeret. Andre eksempler på arkiver og oversikter som også brukes av forvaltningen er Digitalarkivet (med det tilhørende Digitalpensjonatet), data.norge.no og Geonorge.

### 3.5. TILGANG TIL KOMPETANSE OG BISTAND

En viktig forutsetning for god datahåndtering i forskningsprosjektene er kompetanse både hos forskerne selv og i ulike typer støttefunksjoner i institusjonene. I undersøkelsen til institusjonene stilte vi spørsmål om hvordan de arbeidet med kompetanseheving for forskerne, administrativt ansatte og hvordan de opplevde egen kapasitet på bistand.

#### Institusjonenes oppfølging av forskerne

I undersøkelsen til institusjonene stilte vi spørsmål om hvordan de arbeider med kompetanseheving blant forskere innen temaet datahåndtering og FAIR-prinsippene. Her er det ulikheter mellom institusjonene. Noen forteller at de er helt i oppstartsfasen eller at de ikke arbeider systematisk med dette. Andre har mer systematisk opplæring, tilbyr jevnlig kurs og støtte.

#### Tilgang til kompetanse ved institusjonene

Det er store ulikheter mellom institusjonene når det gjelder tilgang på kompetanse om datahåndtering. Noen institusjoner har ansatte i spesialiserte forskerstøttestillinger knyttet til håndtering av forskningsdata. Flere av de store universitetene har egne miljøer som arbeider med forskningsdatahåndtering og forskerstøtte, ofte tilknyttet bibliotekene.

Det er også institusjoner har egne ressursider om datahåndtering, se eksempler i Boks 9. Hos noen er nettsidene åpent tilgjengelig for forskere og andre som ikke er tilknyttet institusjonen. I kartleggingen ser vi også eksempler på institusjoner som henviser til eller bruker materiell som andre institusjoner har gjort tilgjengelig.

Flere av institusjonene forteller at de arrangerer egne interne kurs for forskere og studenter. Hos noen er disse åpent tilgjengelig. UiT har blant annet digitale kurs i datahåndtering som er åpne også for deltakere utenfor institusjonen.

#### Eksempler på ressursider fra institusjonene:

- NMBUs [informasjonsside](#) om forskningsdata
- UiTs [forskningsdataportal](#)
- NTNUs [informasjonsside](#) og støttetjeneste for forskningsdata – Research Data @NTNU
- UiOs [informasjonsside](#) om forskningsdata

#### Boks 9

I undersøkelsen er det også eksempler på institusjoner med få ansatte i administrative stillinger. De beskriver begrenset kapasitet, og de ansatte dekker en stor bredde av problemstillinger. Samtidig beskriver flere at temaet vil få økende fokus fremover. I vår undersøkelse var det blant annet fem institusjoner som viste til nyopprettede stillinger eller nyansatt kompetanse på området.

Mange beskriver at de ser utfordringer med å finne finansiering til å etablere tilstrekkelig kapasitet på slike støttefunksjoner, også ved de store institusjonene.

#### Tilgang til kompetanse i forskermiljøene

Flere av miljøene som har kommet langt på området understreker viktigheten av tilgang til støttepersonell med domenekunnskap. Flere av de fremste forskningsmiljøene har egne ansatte

som jobber med datahåndtering, og som bidrar til å integrere den nødvendige kompetansen i forskningsarbeidet. I tillegg til flere miljøer nevnt i tidligere casebeskrivelser, er ELIXIR og deres bistand til datakuratering, beskrevet i Boks 10, et eksempel. Teaching Learning Video lab (TLVlab) fra Institutt for lærerutdanning og skoleforskning (ILS) ved UiO, beskrevet i Boks 11 er også et eksempel på et miljø som bistår forskere med praktisk støtte, også for forskere utenfor egen institusjon.

#### **CASE: ELIXIR**

ELIXIR er en organisasjon som samler ressurser (databaser, verktøy, opplæringsmaterieil, lagringskapasitet og maskiner for prosessering) for livsvitenskap på tvers av Europa. Målet er å koordinere de ulike ressursene slik at de til sammen utgjør en infrastruktur. Arbeidet er organisert med en sentral hub og ulike noder med egne ansvarsområder. Den norske noden forvalter blant annet Norwegian e-infrastructure for Life Sciences (NeLS). Hovedinnsatsen nasjonalt går på å sørge for at livsvitenskapsdata gjøres FAIR langs alle steg i datalivssyklusen, finansiering går i stor grad til personale som bistår ved alle disse stegene.

Kilde og mer informasjon: <https://elixir-europe.org/> og <https://elixir.no/>

#### **Boks 10**

#### **CASE: Teaching Learning Video lab (TLVlab) fra Institutt for lærerutdanning og skoleforskning (ILS) ved UiO**

TLVlab tilbyr rådgivning og praktisk støtte til forskere og studenter som bruker video i forskningen. De bistår i alle faser av forskningsprosessen fra innledende idé til avslutning. Eksempler på områder de bistår med er datahåndteringsplaner, samtykker, utlån av utstyr, råd om lagring, metadata og arkivering av dataene. Tjenestene tilbys også eksterne forskere (mot kostnad).

Kilde og mer informasjon: <https://www.uv.uio.no/ils/om/organisasjon/tlvlab/tjenester-arbeidsomrader/>

#### **Boks 11**

#### **Nasjonale ressursider og kurs**

God tilgang til ulike typer veiledningsmaterieil og kurs er viktig. Det finnes både internasjonale og nasjonale ressursider om datahåndtering og FAIR forskningsdata. Et eksempel på en nasjonal ressurside er Openscience.no som driftes av Unit. Gjennom prosjektet Open Science Toolbox, beskrevet i Boks 12, ble det utviklet et verktøy med ressurser og informasjon om datahåndtering. Dette er et eksempel på et samarbeid som senere har blitt til et nasjonalt verktøy.

### Open Science Toolbox

Prosjektet Open Science Toolbox var et samarbeid mellom bibliotekene ved Norges teknisk-naturvitenskapelige universitet (NTNU), Universitetet i Stavanger (UiS) og Høgskolen i Innlandet (INN). Prosjektet utviklet et verktøy med ressurser og informasjon om datahåndtering og åpen forskning. Verktøyet er nå overlevert til Unit og er publisert som del av openscience.no.

Kilde: <http://openscience.prototyp.io/om-open-science-toolbox>

#### Boks 12

Det tilbys også kurs i datahåndtering fra nasjonale aktører. Blant annet holder tjenesteleverandører som Uninett Sigma2 og NSD jevnlig kurs.

#### Informasjonssider:

- <https://www.openscience.no/> som driftes av Unit
- Informasjon om [FAIR-prinsippene](#) hos goFAIR
- Elixirs [informasjonsside](#) for forskere med informasjon om rolle og nyttige ressurser
- CESSDAs [Data Management Expert Guide](#)

#### Oversikt over arkiver

- <https://www.re3data.org/> - oversikt over fagspesifikke arkiver
- European Open Science Cloud (EOSC) og deres [portal](#) med oversikt over policies, tjenester og ressurser innen ulike domener

#### Informasjon om lisenser

- Lisensutvalgets lisensvettregler s.40 i [rapporten fra utvalget](#)
- Elixirs [informasjonsside](#) om lisenser.

## 3.6. MULIGHETER FOR FINANSIERING AV DATADELING

### Kostnader for å tilgjengeliggjøre data

Det er i dag mulig å legge inn kostnader for tilgjengeliggjøring av viktige data som en del av driftskostnadene i søknadene til flere av finansiererne. Forskningsrådet lister for eksempel opp «å gjøre forskningsdata som oppstår i prosjektet tilgjengelig i tråd med eventuell datahåndteringsplan» som eksempel på slike driftskostnader<sup>12</sup>.

Også i Horisont Europa åpnes det for å legge kostnader til datahåndtering inn som del av driftskostnadene.

<sup>12</sup> <https://www.forskningsradet.no/sok-om-finansiering/budsjett/budsjettet-skal-innehold/>

## 4. HVA MANGLER?

I dette kapittelet beskrives meldte behov og innspill knyttet til verktøy, tjenester, kompetanse og kultur/insentiver for deling som forskerne og institusjonene har trukket frem i spørreundersøkelsene og/eller i arbeidet med utredningen.

### 4.1. VERKTØY OG TJENESTER

Både enkeltforskere og institusjonene har beskrevet behov for videreutvikling i eksisterende verktøy og utvikling av nye verktøy og tjenester. Her følger noen av de viktigste innspillene i denne sammenheng.

#### Behov for FAIRifisering av tjenestene

Tjenestene må i enda større grad enn i dag bidra til å sikre at data blir FAIR. Det er behov for generell «FAIRifisering» av de tjenestene som er tilgjengelige i dag, slik at de sikrer noen minimumskrav.

For å sikre at data er FAIR må dataene som minimum ha<sup>13</sup>:

- En fast, vedvarende ID som beskriver hvor data er lagret og kan gjenfinnes
- Et tilstrekkelig og standardisert sett med metadata, som gjør dataene søkbare, gjenfinnbare og som i tilstrekkelig grad beskriver dataene
- En tydelig lisens som redegjør for hvordan dataene kan brukes

Det finnes mange tjenester i dag som muliggjør publisering av data og som gjør data gjenfinnbare og tilgjengelige (F og A). Tjenester som sikrer interoperabilitet og gjenbrukbarhet (I og R) mangler i mange sammenhenger. Dette skyldes at data er mangelfullt beskrevet (metadata) eller er samlet inn eller beskrevet med referanse til divergerende kodeverk, begrepsbruk (semantikk), formater eller lisensbestemmelser – noe som gjør det vanskelig eller umulig å sammanestille dataene med andre data eller gjenbruke dem i nye sammenhenger på en meningsfull måte. Manglende API-er i flere av tjenestene gjør også at det er krevende å gjenbruke dataene.

#### Sammenhengende tjenester og verktøy

Noen opplever at dagens tjenester og verktøy er fragmenterte. De savner at verktøy og tjenester for de ulike delene av forskningsprosessen henger bedre sammen. For eksempel nevnes tjenester der forskerne både kan lagre og behandle data underveis i pågående prosjekter og senere arkivere de samme dataene. Dette gjelder både rådata og metadata.

Et konkret eksempel som ble trukket frem av en forsker var det å sammenstille data som er hentet fra arkivet til NSD med andre datakilder i egen forskning. Når forskeren skal laste opp de sammenstilte dataene til DataverseNO kan ikke metadataene fra NSDs arkiv brukes direkte fordi tjenestene bruker ulike formater. Dermed må forskeren kopiere og lime inn informasjonen.

---

<sup>13</sup> <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/1.-Plan/FAIR-data>



Informantene har ulike tilnærminger til hva som er mulige løsninger på utfordringene. Noen etterlyser tjenester som dekker flere faser, f.eks. ved at en kan lage datahåndteringsplan og arkivere i samme tjeneste. Andre etterlyser API-er og standardisering som sikrer et mangfold av tjenester for spesialiserte områder samtidig som en sikrer interoperabilitet mellom tjenestene.

### **Enkle, fleksible tjenester**

Forskerne håndterer svært ulike mengder data. Det etterlyses tjenester tilsvarende Uninett Sigma2 for mindre prosjekter («den lange halen»). Dvs prosjekter som har behov for lagrings-/beregningsskapasitet som de ikke har tilgang til lokalt, men som ikke er store nok (Sigma2 er ment for lagring av datamengder over 10 TB) til å få tilgang til Sigma2.

Standardisering av formater er viktig for at data enkelt skal være gjenbrukbare og 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 tjenester uten slike krav, fordi det vil gjøre det enklere å publisere dataene. Et eksempel som ble trukket frem er datasett fra måleutstyr. Forskeren ønsket å publisere data fra forsøkene i det originale (men ikke godkjente) formatet og opplevde at dette ville kunne gi verdi for andre forskere. Dette kan tyde på et behov også for fleksible tjenester med minimale krav til formater og annet knyttet til datasettene.

### **Fleksibel tilgangskontroll**

Det etterlyses arkivløsninger med fleksibel tilgangskontroll. Det innebærer å gi ulik tilgang til data basert på for eksempel gruppetilhørighet eller rolle. Et eksempel på en tjeneste med fleksibel tilgangskontroll i dag er infrastrukturen CLARINO som har såkalte «akademiske lisenser» som kun gir tilgang med FEIDE-innlogging. Lignende løsninger etterspørres for flere av arkivene.

### **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 persondata og data som kan ligge åpent tilgjengelig.

Unit/HK-dir gjennomfører en årlig kartlegging av informasjonssikkerhet ved institusjonene. Et av funnene fra årets undersøkelse var at bruk av lokale servere (f.eks. på instituttnivå), som ikke er underlagt det normale driftsregimet til institusjonen, kan medføre økt risiko for brudd på informasjonssikkerheten.

### **Tjenester for lagring og arkivering av sensitive data**

For sensitive data beskrives behov for tjenester som muliggjør deling med samarbeidende forskere utenfor institusjonen i pågående prosjekter. Noen etterlyser også flere alternativer for lagring av sensitive data. Forskerne savner også gode tjenester og verktøy for å anonymisere eller pseudonymisere personopplysninger. Også syntetisering av datasett (lage kunstige, men realistiske datasett) er et tema som løftes. Informantene beskriver også et behov for oversikt over hvilke typer data som kan lagres hvor.



Flere institusjoner oppgir UiOs Tjenester for Sensitive Data (TSD) som anbefalt tjeneste for arkivering av sensitive data. Det er også flere forskere som oppgir at de har arkivert data i TSD. TSD er per i dag ikke et arkiv og er ikke ment for langtidslagring og arkivering av data etter prosjektslutt. Den nåværende bruken kan indikere at det faktisk mangler tjenester for arkivering av sensitive forskningsdata i Norge. Det kan dermed være behov for å videreutvikle TSD og/eller utvikle andre sikre lagringsløsninger for sensitive data.

Flere av helseforetakene har lokale eller regionale løsninger under utvikling, og Direktoratet for e-helse har sin store satsing gjennom Helsedataprogrammet. Selv om Helsedataprogrammet i Helseanalyseplattformen i første omgang tar sikte på å tilgjengeliggjøre data fra helseregistre, har de ambisjoner om på sikt å også kunne lagre data fra forskningsprosjekter. Denne uken økte usikkerheten rundt tidsplanene for en slik løsning, da Direktoratet for e-helse meldte at utvikling av Helseanalyseplattformen settes på vent på grunn av juridiske utfordringer.

#### **Etablering av nye infrastrukturer – udekkede behov?**

Flere av institusjonene arbeider for tiden med å etablere egne infrastrukturer for deling av forskningsdata. Dette kan tyde på at eksisterende tjenester ikke dekker institusjonens behov.

Det kan samtidig se ut som behovet for nye infrastrukturer ofte 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 utfordring som kan oppstå når det etableres infrastrukturer i forbindelse med nye forskningsprosjektene er den manglende finansiering etter prosjektslutt. For noen infrastrukturer vil dette føre til manglende videreutvikling og drift.

#### **Behov for forutsigbare kostnader**

Noen av tjenestene har betalingsmodeller der forskerne betaler for faktisk bruk av tjenester og kapasitet etterskuddsvis. 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 rimelige, helst gratis infrastrukturer for arkivering/publisering av data.

## **4.2. KUNNSKAPSBEHOV OG KOMPETANSE**

I undersøkelsen etterlyses kompetanse om ulike aspekter ved datahåndtering i forskningsprosjektene. Det gis også klare indikasjoner på behov for økt kunnskap om tjenester og verktøy for FAIR forskningsdata, først og fremst blant forskere, men også på institusjonsnivå.

#### **Kunnskap om verdien av å dele data**

Det er flere eksempler på fagfelt og forskningsmiljøer der deling av forskningsdata allerede er godt etablert og hvor det er tradisjon på å tilgjengeliggjøre data for gjenbruk. Samtidig er det også eksempler på forskere og institusjoner som mangler gode eksempler på hvordan dataene

de håndterer kan være til nytte for andre. Det etterlyses informasjonsmateriell og gode beskrivelser av case fra ulike fagfelt.

### **Grunnleggende kunnskap om FAIR forskningsdata**

I undersøkelsene kommer det frem at noen forskere og institusjoner omtaler FAIR forskningsdata og åpne data som det samme. Dette viser at det er behov for å øke kunnskapen på området. Blant annet er det viktig å få større oppmerksomhet om mulighetene til å lagre og arkivere data med tilgangsbegrensninger, og om mulighetene for å dele metadata uten nødvendigvis å dele selve datasettene.

### **Viktigheten av tidlig planlegging og innsats**

For å sikre at data fra prosjektene kan gjøres FAIR er det viktig at forskerne tenker på dette allerede under planleggingen av sine forskningsprosjekter. De må tenke på 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 for deling. Det finnes eksempler på forskere som ønsker å dele data fra pågående eller avsluttede prosjekter, men som mangler de rette samtykke eller tillatelser til dette. Både institusjonene og forskerne peker på et behov for gode eksempler på samtykkeskjemaer og personvernerklæringer som beskriver mulig deling på en god måte for deltakerne.

Gode metadata er en forutsetning for at datasett skal være gjenfinnbare og gjenbrukbare. Dersom dataene skal deles kan det også stille krav til formatene data lagres i underveis i prosjektet. Ved å arbeide med maskinlesbare metadata i alle fasene forenkles senere arkivering og eventuell publisering. Dette kan i dag være krevende for små miljøer uten tilgang til etablerte løsninger for dette.

### **Ulik modenhet i retningslinjer og politikk for FAIR datahåndtering hos institusjonene**

Forskerne forholder seg til ulike retningslinjer når de planlegger datahåndtering i sine prosjekter. I spørreundersøkelsen oppga mange at de forholdt seg til retningslinjer fra de spesifikke infrastrukturene de brukte, fra finansiører (som Forskningsrådet), relevante nasjonale retningslinjer og sist men ikke minst retningslinjer fra de institusjonene de var tilknyttet. Institusjonelle retningslinjer er det som oftest refereres til i kartleggingen, noe som viser viktigheten av institusjonenes arbeid på området.

I spørreundersøkelsen til institusjonene ble de bedt om å vurdere egen modenhet når det gjelder arbeid med politikk og retningslinjer for å følge FAIR-prinsippene for datahåndtering. Svarene oppsummeres i Figur 6. Fritekstsvarene viste at noen av institusjonene er helt i oppstartsfasen, mens andre har jobbet med dette over lengre tid. I hvilken grad institusjonenes retningslinjer er til hjelp for forskerne vil med andre ord variere.



Hvordan vurderer dere institusjonens modenhet når det gjelder arbeid med politikk og retningslinjer for å følge FAIR-prinsippene for datahåndtering?



FIGUR 6: INSTITUSJONENES VURDERING AV EGEN MODENHET PÅ ARBEID MED POLITIKK OG RETNINGSLINJER FOR FAIR FORSKNINGSDATA (OPPDATERT ETTER 58 SVAR)

### Manglende kunnskap om behovet for og nytten av datahåndteringsplaner

Datahåndteringsplaner er ment som et hjelpemiddel for forskere, i håndteringen av alle typer data. I undersøkelsene finner vi imidlertid eksempler på forskere og institusjoner som knytter datahåndteringsplaner tett til håndteringen av personopplysninger. En institusjon beskrev at de ikke anbefalte tjenester for datahåndteringsplan til sine forskere fordi de ikke hadde prosjekter som behandlet data med behov for beskyttelse («røde og sorte data»). Dette viser at det kan mangle kunnskap om hva en datahåndteringsplan er for noe, og hvorfor det er viktig å planlegge all forskningsrelatert datainnsamling med tanke på tilgjengeliggjøring og gjenbruk av data.

### Behov for mer kunnskap om metadata

Utredningen viser at både forskerne og institusjonene har behov for mer kunnskap om metadata og verdien av gode metadata. Flere ønsker også bistand og opplæring på dette området.

Det er eksempler på både forskere og institusjoner i kartleggingen som beskriver at de hverken deler forskningsdata eller metadata fordi de håndterer datasett med behov for beskyttelse. Dette tyder på et behov for mer kunnskap om muligheter for å dele kun metadata, i tillegg til informasjon om hvordan data med behov for beskyttelse kan deles. Selv om data ikke kan deles åpent er det fortsatt mulig å dele metadata slik at andre får informasjon om hvilke data som finnes. Her er det viktig å benytte metadata (metadataskjema) som i størst mulig grad er standardiserte og kjent innen fagdomenet. På denne måten kan forskere som mener å ha legitim grunn til å få tilgang, enkelt finne fram til datasettene og søke om tilgang.

For forskere som gjennomfører laboratorieforsøk er elektroniske lab-journaler et nyttig hjelpemiddel. Det finnes i dag slike lab-journaler der metadata kan legges på automatisk. Denne type automatiserte metadata-generatorer er nyttige verktøy, og som antagelig kunne gitt verdi for flere – i det jobben med å legge på metadata blir betydelig enklere.

### **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 på å bistå forskerne innen juridiske, etiske, fagområdespesifikke og innen tekniske problemstillinger og informasjonssikkerhet.

Juridiske spørsmål skiller seg ut om et område som mange institusjoner opplever begrenset kapasitet på, med hensyn til å bistå forskere. 50 % (29 av 58) vurderer egen kapasitet som begrenset eller svært begrenset på dette området. Informasjonssikkerhet virker å være et område mange institusjoner har fokusert på, siden det i vår kartlegging ikke er noen som beskriver svært begrenset kapasitet på dette området.

For å støtte arbeidet med forskningsdata er det behov for ulik kompetanse. Det pågår nå et arbeid for å utvikle et nasjonalt kompetanserammeverk for forskningsstøtte beskrevet i Boks 13.

#### **CASE:**

##### **Nasjonalt kompetanserammeverk for forskningsdatastøtte**

«For å få på plass en helhetlig forvaltning av forskningsdata i Norge er det nødvendig at ansatte i støttetjenester har relevant kompetanse. Denne typen støtte ved fagbibliotek i Norge er fortsatt under utvikling. Nye roller og stillinger stiller også nye krav til samarbeid og koordinering både internt og eksternt. Målet med prosjektet er å utarbeide et nasjonalt kompetanserammeverk for forskningsdatastøtte»

Kilde: <https://bibliotekutvikling.no/prosjektbank/prosjekt/kompetanserammeverk-for-forskningsdatastotte/>

#### **Boks 13**

### **Behov for oversikt over hva som finnes av verktøy og tjenester**

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

## **4.3. HÅNDTERING AV DATA MED BEHOV FOR BESKYTTELSE**

Flere beskriver utfordringer og uavklarte problemstillinger knyttet til data som krever beskyttelse. Eksempler på slike data er persondata, spesielt særlige kategorier, data som er vernet gjennom åndsverksloven og data som er beskyttet av immaterielle rettigheter (forretningshemmeligheter).

### **Persondata**

Sykehusene og andre som forsker på helsedata beskriver gjeldende regelverk som en barriere for å dele data med forskere utenfor institusjonen. Både Kreftforeningen og Kreftregisteret

trekker frem at mange pasienter ønsker at deres data skal deles, men at gjeldende taushetsbestemmelser og lovverk gjør det svært krevende. Deling av data med forskere utenfor EU/EØS beskrives av noen som krevende på nåværende tidspunkt pga lovverk som hindrer. For tiden er det også umulig (iht. Schrems II) å dele data med forskere og forskningsinfrastrukturer under USAs jurisdiksjon.

Noen opplever også usikkerhet knyttet til om dataene kan deles og beskriver en frykt for å gjøre feil som kan føre til at man velger å ikke dele for å være på den sikre siden. En av institusjonene beskriver blant annet et ønske om «en fellestjeneste som kan gjennomgå datasett før publisering for å vurdere om åpen arkivering av data er OK i forhold til personvern/sensitivitet, f.eks. som NSDs meldeskjema fungerer i dag ved prosjektstart».

Basert på kartleggingen kan det se ut til å være et behov for mer informasjon om hvordan data som inneholder personopplysninger kan arkiveres og eller publiseres. Som NSD skriver på egne nettsider «later det til å ha spredt seg en misforståelse om at personvern og deling av data står i konflikt»<sup>14</sup>.

Det foreslås også å utarbeide eksempler på samtykkeskjemaer som viser hvordan man kan innhente samtykke til åpen deling eller deling av aidentifiserte data.

#### **Utfordringer knyttet til håndtering av registerdata**

Selv om dette kanskje handler om utfordringer med tilgang til data mer enn om utfordringer med å håndtere (nye) data, er registerforskningen et tema i denne sammenheng. Mange opplever at det tar tid å søke om – og få, tilgang til å bruke data fra helseregistre, SSBs registre og andre relevante personregistre som kilde 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.

En kunne tenke seg at institusjonene som gjør registerne tilgjengelig for forskere (som SSB) også har ansvar for å dokumentere FAIR metadata for leverte datasett. Ideelt sett burde dette gjøres på nivået “data som faktisk blir levert” i stedet for universet av data som potensielt er tilgjengelig. I dag er det ofte vanskelig for interesserte tredjeparter å vite hva det skal søkes om for å kunne reprodusere publisert forskning. Siden registerne kan være data i bevegelse kan en også tenke seg at for eksempel SSB gjør en intern langsiktig lagring av alle leverte datasett.

## **4.4. 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. Kulturelle elementer påvirker også omfanget av deling. Forskerne må ha en vilje til å dele, og det å sørge for at data kan gjenbrukes må oppfattes som en naturlig del av det å være forsker. Det handler til viss grad om kunnskaper, men mest om at «systemet» forventer av dem at de deler, og at de har insentiver til å gjøre det.

<sup>14</sup> Tekstboks nederst på siden: <https://www.nsd.no/arkiver-data/>

### **Manglende tradisjoner for deling**

I undersøkelsen 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.

### **Behov for insentiver for datadeling**

Flere av forskerne peker på et behov for insentiver for datadeling. Det kan ikke påregnes å være selvsagt at forskerne prioriterer å legge ressurser i å tilrettelegge for deling og gjenbruk av egne data. Det bør utvikles ulike insentiver for mer deling. Det å sikre at det er mulig å sitere datasett, og oppfordre forskere til å gjøre det, vil være et virkemiddel som vil bidra til å gjøre det mer attraktivt å publisere datasett. Sitering av forskningsdatasett i et arkiv er fullt mulig, men fordrer at lagringsinfrastrukturen eller arkivet tilbyr mulighetene for fast lenking (DOI eller tilsvarende) til datasettet.

Et annet insentiv er å etablere ulike mekanismer for å anerkjenne arbeid med publisering av datasett, for eksempel i redaksjonelle omtaler i vitenskapelige journaler, i oversikter institusjonene utarbeider, etc. Det å rette oppmerksomheten mot nye datasett som er publisert vil gjøre det mer lukrativt for forskere å publisere. Man kunne også se for seg at publisering av datasett i anerkjente kanaler ga institusjonene poeng i det nasjonale tellekantsystemet (norsk publiseringsindikator) og at dette ville kunne bidra til høyere prioritering og bedre tilrettelegging for deling fra institusjonenes side.

## **4.5. REGULERENDE VIRKEMIDLER**

I tillegg til tilgang til verktøy og tjenester, og til kompetanse, kunnskap og vilje til deling blant forskere og institusjoner, kan det være behov for ulike regulerende virkemidler og mekanismer for å «tvinge» igjennom mer deling og gjenbruk av forskningsdata.

### **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. American Economic Association krever for eksempel nå pre-aksept «computational reproducibility» og FAIR data (i alle fall til en viss grad) av alle artikler<sup>15</sup>.

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

---

<sup>15</sup> <https://www.aeaweb.org/articles?id=10.1257/pandp.111.808>

### **Krav fra finansiørene**

Tydligere krav og oppfølging fra finansiørene 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 åpent tilgjengelig (med noen unntak). Krav fra de som styrer publiseringskanalene vil også krav fra de nasjonale eller internasjonale aktører som finansierer forskningen være et effektivt virkemiddel for å stimulere til mer publisering av data.

I undersøkelsen til institusjonene nevnes det at forskerne gjør det de er kontraktsforpliktet til å gjøre. Strengere kontraktsfestede krav fra finansiørene og en oppfølging av disse i etterkant foreslås derfor som mulige tiltak.

## **4.6. OPPSUMMERING AV DAGENS UTFORDRINGSBILDE**

I dette kapittelet oppsummerer vi dagens utfordringsbilde og behov, på basis av den kartlegging vi har gjort.

### **Behov for verktøy og tjenester for FAIR forskningsdata**

- Flere etterlyser nasjonale maler og verktøy for datahåndteringsplaner med mulighet til å gjøre individuelle tilpasninger, f.eks. en egen versjon for institusjonen der man kan vise til lokale retningslinjer, rutiner og eventuelle verktøy.
- Forskerne har behov for infrastruktur som muliggjør sikker deling med/tilgang for forskere ved andre institusjoner, spesielt for forskningssamarbeid om sensitive data.
- Metadata:
  - Flere etterlyser tjenester som bistår og veileder forskerne i det å legge på gode metadata på datasettene.
  - Det er behov for tjenester der man kan publisere søkbare metadata uten å publisere datasett og behov for verktøy som hjelper forskerne i å legge på og vedlikeholde/ha kontroll på metadata fra oppstart til avsluttet prosjekt.
  - Det etterlyses også standardisering av metadata, og veiledning om hvilke standarder som anbefales. I dag finnes det mange ulike standarder som vanskeliggjør interoperabilitet på tvers av fag.
- Noen opplever dagens tjenester og verktøy som fragmenterte og savner at de ulike delene av prosessen henger bedre sammen, f.eks. ved å kunne lagre underveis og senere kunne arkivere data uten å måtte laste opp alt på nytt i en annen infrastruktur.
- Noen etterlyser små, enkle tjenester for analyse som ikke krever mye arbeid for å opprette brukerprofiler o.l. og tjenester som enkelt lar deg dele data i ulike (også ikke standardiserte) formater.



- Det etterlyses tjenester tilsvarende Sigma2 for små, kortvarige prosjekter («den lange halen»). Dvs. prosjekter som har behov for lagring/beregningskapasitet som de ikke disponerer lokalt, men som ikke er store nok til å få tilgang til Sigma2
- I noen prosjekter opprettes egne infrastrukturer og databaser for lagring av data underveis. Det kan være utfordrende med finansiering til videre drift etter prosjektslutt. Forslag om å kunne overføre hele databaser til infrastrukturene.
- Det etterlyses nasjonale infrastrukturer og arkiver som kan brukes uten store kostnader.

### Behov for informasjon og kompetanse

- Det er behov for å synliggjøre hvilke muligheter deling av data kan gi i forskningen
- Det er ulikheter mellom disiplinene og det er ikke alle forskere som forstår hvordan data de har fremskaffet kan ha verdi for andre (også på tvers av disipliner). Det er behov for å vise mangfoldet og vise eksempler på hvordan data fra ulike disipliner kan være til nytte.
- For noen er det et utydelig skille mellom «åpne data» og FAIR data. Det er viktig at informasjonen som formidles fremhever at det å gjøre data FAIR ikke betyr at alle skal få tilgang til dataene. Data kan deles under kontrollerte forhold med tilgangsbegrensning. Det er også mulig å dele metadata uten å dele selve datasettene.
- Institusjonene har svært ulik tilgang på kompetanse om håndtering av forskningsdata internt. Det gjør at behovet for kompetanse og bistand varierer mellom institusjonene. Noen er helt i oppstartsfasen mens andre har kommet langt på området.
- Miljøene som har kommet langt på området fremhever behovet for datarøkttere/støtte med domenekompetanse som jobber tett på forskerne.
- Det er ressurskrevende å bygge opp kompetanse på datahåndtering og å lære opp forskerne i dette. Institusjonene etterlyser finansiering og bistand.
- Alle fagfelt produserer data som kan være viktige å dele for gjenbruk. Flere forskere kjenner imidlertid ikke til verktøy for datahåndteringsplan, lagring og arkivering av data. Noen etterlyser mer informasjon om de ulike verktøyene og tjenestene og hva som kan brukes til hva.
- Kapasiteten på juridisk bistand er begrenset hos mange av institusjonene. Det etterlyses juridisk kompetanse og bistand i å vurdere hvilke data som kan deles. Noen beskriver en kultur der man er redd for å gjøre feil, spesielt etter innføringen av GDPR.
- Det etterspørres eksempler på samtykkeskjemaer og veiledning rundt hvordan man kan utforme samtykkeskjemaer på en måte som legger til rette for senere deling.
- Det er behov for juridiske avklaringer knyttet til deling av data med behov for beskyttelse, som data beskyttet av åndsverkloven og registerdata fra offentlig

forvaltning.

### **Behov for incentiver – skape kultur for deling**

- Det pekes på at det er behov for sterkere incentiver for datadeling. Det oppleves som tidkrevende å tilrettelegge data for deling og publisering, spesielt i de tilfellene dataene må bearbeides først. Et eksempel på dette er anonymisering av data. Incentiver som foreslås er:
  - Tilrettelegge for og oppfordre til at det i større omfang enn i dag blir vanlig å sitere datasett som er publisert
  - Aktiviteter knyttet til deling og publisering av datasett må bli meritterende i ulike vurderingssituasjoner. Aktiviteter som på sikt bør vektlegges:
    - Publisere datasett
    - Publisere metadata dersom selve datasettet ikke kan deles
    - Bidra med nye data til et etablert/anerkjent register
    - Bidra til å etablere/forbedre internasjonale standarder knyttet til FAIR data

### **Behov for regulerende virkemidler**

- Det er behov for tydeligere krav fra ulike aktører, både fra de som finansierer forskningen, og fra de som publiserer forskningsresultater. Krav fra tidsskrifter og andre om å tilgjengeliggjøre forskningsdataene FAIR vil være et effektivt virkemiddel.

## 5. DISKUSJON

I denne utredningen har formålet vært primært å kartlegge hvilke tjenester for håndtering av forskningsdata som per i dag er tilgjengelige for norske forskere, og komme med anbefalinger om veien videre for å komme frem til tjenester som gir muligheter for tilgjengeliggjøring og deling av forskningsdata som er FAIR. Økt deling av forskningsdata er et viktig og riktig fokus i norsk og internasjonal forskningspolitikk. På mange måter har kunnskapsdeling i samfunnet både nasjonalt og internasjonalt i stor grad blitt dreid mot deling av data. Blant annet er global økonomi et eksempel på hvordan datadeling ikke bare endrer hvordan samfunnet fungerer gjennom økt tilgang til data, men også på hvordan datadelingen skaper nye muligheter for å utnytte data (f.eks. gjennom stordata og kunstig intelligens).

Samfunnet står overfor store utfordringer som krever både rask(ere) kunnskapsutvikling og bedre utnyttelse av kunnskapen – både i og utenfor forskningssystemet. Åpen forskning og datadeling er grunnpilarer for å sikre dette. Datadeling kan bidra både til å sikre kvaliteten til forskningsresultater og tilliten til disse i befolkningen. For å motvirke polariserte samfunnsdebatter om for eksempel klima eller vaksineeffekter er mulighetene for å få tilgang til de underliggende dataene sentralt. Svake muligheter til å reprodusere forskningsresultater er fortsatt en bekymring innen mange fag.

I denne utredningen har vi kartlagt eksisterende tjenester som norske forskere og forskningsinstitusjoner kan bruke til datadeling. Vi har identifisert barrierer mot datadeling, og behov for tjenester, infrastruktur og kompetanse. Vi kommer med anbefalinger til begrepsapparat rundt datadeling og gir anbefalinger til kriterier som tjenester knyttet til forskningsdata bør oppfylle. Disse vil publiseres på samme nettside som en oversikt over tilgjengelige tjenester. Vi gir også anbefalinger om prosjekter som kan bidra til å komplettere de tjenester vi har kartlagt, med spesiell vekt på langtidsbevaring av forskningsdata. Denne utredningen er en av flere relativt samtidige offentlige arbeider som skal bidra til å sikre effektiv og helhetlig datadeling i forskningssystemet. Sammen med utredningene til Datalisensutvalget og Datainfrastruktur-utvalget gjennomgås ulike perspektiver av datadeling. Det er et naturlig mål at anbefalingene i disse parallelle utredningene konvergerer.

Vi opplever at denne utredningen dokumenterer stor heterogenitet med hensyn til hvor langt norske forskningsinstitusjoner har kommet i arbeidet med å dele forskningsdata, og at forskjellen på store og etablerte forskningsinstitusjoner versus de mindre miljøene kan være et viktig fokus for videre arbeid. Samtidig indikerer den at det er store interne variasjoner på institusjonene og mellom ulike fagdomener og disipliner i arbeidet med datadeling. Vår undersøkelse indikerer at insentiver, tekniske forhold, kompetanse og juridiske forhold alle er utfordringer som representerer barrierer mot datadeling.

En sentral leveranse i dette prosjektet er å levere en oversikt over de tjenester som er tilgjengelige for norske forskningsinstitusjoner, og samtidig komme med konkrete anbefalinger til nye tjenester eller forbedring av de eksisterende – altså identifisere mangler i dagens tilbud. På bakgrunn av diskusjoner rundt funnene blant prosjektets deltakere (sekretariat



utredningsgruppe, referansegruppe og andre) kommer vi i denne rapporten med noen konkrete anbefalinger og råd. Disse handler både om hvordan norske forskere og forskningsinstitusjoner bør gå fram for å velge tjenester, hvilke tjenester som kan brukes for arkivering av forskningsdata, håndtering av store datasett og håndtering av sensitive persondata. Vi håper dette kan være en viktig praktisk hjelp i hverdagen for forskningsinstitusjonene og for enkeltforskere, og håper også det kan være starten på en verdifull tjeneste i den brede forskningssektoren – og som kan videreutvikles i takt med sektorens læring om hvordan disse fungerer og strukturene rundt datadeling modnes og videreutvikles.

Etablering av norske begreper for bruk knyttet til forskningsdata er kritisk for å kunne videreutvikle og sikre konvergens i de tjenester som etableres, og i samfunnets videre arbeid med deling av forskningsdata. Som en del av denne utredningen vil den norske noden av Research Data Alliance (RDA) foreslå et slik begrepsapparat. Forslaget kan være et tiltak for å standardisere arbeidet med deling av forskningsdata både internt i norsk forskningssektor, men også å sikre gode strukturer for interaksjon med internasjonale enheter.

Håndteringen av forskningsdata ses av mange som en lineær, stegvis modell, hvor dataenes livssyklus starter med planlegging av et forskningsprosjekt og avsluttes med arkivering av dataene. Vi tror dette er en litt begrensende modell. De infrastrukturene som etableres bør ha et livsløpsperspektiv for forskningsdata, hvor det fra forskerens ståsted er en sømløs brukeropplevelse fra datahåndteringsplaner til arkivering av data – og gjenbruk av disse dataene. Institusjonene bør anbefale funksjonalitet i datahåndteringsplaner som gjør det mulig å gjenbruke og dele informasjonen senere i prosessen, både institusjoner og infrastrukturer bør legge til rette for aktiv bruk av internasjonale standarder og eventuell videreutvikling av disse der de mangler. I tråd med livsløpsperspektivet må også tjenester som har funksjonalitet for å legge på metadata, både sikre at dette kan gjøres tidlig i prosessen og at de kan oppdateres og vedlikeholdes underveis i prosessen og etter hvert som selve forskningsprosessen går framover.

«Livsløpsmodellen» med linearitet knyttet til kortvarige forskningsprosjekter bør i dette perspektivet utfordres, siden behovet for økt datadeling innebærer økt fokus på data og mindre fokus på publikasjoner. Dette vil i mange disipliner kunne endre måten det forskes på. Forskningsinfrastrukturene i seg selv vil være determinanter i forskningsprosessene – på samme måte som tidsskriftene og forlagene i dag er det i publikasjonsorienterte disipliner.

Insentiver styrer i stor grad norsk og internasjonal forskningsaktivitet både på individnivå og institusjonelt nivå. Tidsskriftenes store innflytelse på forskningsadferd har så langt preget bildet. Vi også gjort en vurdering av infrastrukturenes samspill med insentiver i forskningssektoren i denne utredningen, da vi opplever nettopp insentiver som et helt sentralt virkemiddel for å øke datadeling i forskningssystemet. For eksempel vil et viktig spørsmål være hvordan Forskningsrådet og andre aktører som finansierer forskning kan følge opp de prosjektkontrakter som inngås ved tildeling av midler. Et annet sentralt spørsmål er hvordan omfanget av deling og gjenbruk av forskningsdata fra infrastrukturene kan måles, som bruk både til å evaluere nytten av infrastrukturene, men også bruken av enkeltdatasett og mulige koblinger til insentivene for forskerne.



En sentral del av utredningens mandat har vært å vurdere hvilke kriterier tjenester som er knyttet til forskningsdata bør oppfylle. På bakgrunn av våre funn og diskusjoner i prosjektet mener vi at tjenestene blant annet bør sikre forutsigbar langtidsbevaring, gi forskere sømløse brukeropplevelser og sikre at eventuelle restriksjoner ved deling og gjenbruk av data ivaretas. Et viktig prinsipp bør være at grad av FAIR («FAIRness») må ses på som et kontinuum, der hver av de fire prinsippene i FAIR kan være oppfylt i større eller mindre grad. Hvordan infrastrukturer kan legge til rette for FAIR-data på en effektiv måte bør være et viktig mål i det videre arbeid på området.

En viktig utfordring med åpen deling av data, som dette arbeidet har fokusert spesielt på, er hvordan data med beskyttelsesbehov bør lagres. Slike data dekker et stort spekter, fra personidentifiserbare helsedata til data med beskyttelsesbehov av helt andre grunner, som industrihemmeligheter. Våre funn bekrefter at det er stor usikkerhet i sektoren om hvordan slike data bør lagres for å både sikre juridiske og etiske aspekter – samtidig som behovet for deling og gjenbruk av slike data ivaretas. Vi har i løpet av arbeidet sett svært gode eksempler på hvordan merverdi i samfunnet kan oppnås gjennom god håndtering og FAIRifisering av data der selve dataene som følge av blant annet personvern hensyn ikke kan deles åpent, men der enten metadata eller aidentifiserte versjoner av datene kan deles.

Vårt inntrykk er at manglende kompetanse knyttet til juridiske forhold i mange tilfeller er en barriere for å få til deling av FAIR-kompatible, sensitive data. I mange tilfeller er det lite empiri å støtte seg på i fagmiljøene. For eksempel er spørsmålet om hvilken grad av anonymisering av persondata som er nødvendig for å unngå bakveis identifisering i fremtiden, en krevende utfordring for mange. Det er bekymringsfullt om forskningsmiljøer tenderer til å ikke dele data gjennom FAIR-prinsipper som følge av opplevde, men ikke nødvendigvis reelle, juridiske barrierer mot deling av sensitive forskningsdata. Kompetanseheving innen juridiske forhold er derfor en helt sentralt særlig internt i forskningsmiljøer og institusjoner som baserer forskning på sensitive data.

Samfunnsmessig vil økt deling av forskningsdata representere ekstra kostnader for forskningen, og gitt typisk lav betalingsvilje fra brukere av forskningsdata er finansiering og økonomiske modeller for infrastrukturer som bruker forskningsdata en sentral debatt fremover. Vi oppfatter dette som utenfor prosjektets mandat, men vil her peke understreke at det vil være nødvendig med betydelige investeringer for å utvikle et forskningssystem som kjennetegnes av datadeling som standard.

Vi har gjennom utredningsarbeidet også diskutert den delvis uløste samfunnsutfordringen med hvordan data med kommersiell verdi bør håndteres. Både den store kommersielle verdien av forskningsdata i seg selv (på tross av lav betalingsvilje som nevnt over), forskningsinstitusjonenes oppgave med å legge til rette for innovasjon inkludert kommersialisering av forskningsresultater, samt samspillet mellom offentlige forskningsinstitusjoner og næringsliv er alle faktorer hvor det er knyttet utfordringer. Det datadrevne samfunnet vil innebære nye spilleregler og en endret dynamikk mellom aktørene. Eksisterende utfordringer vil antagelig forsterkes. Vi følger her opp med anbefaling om å se dypere på disse utfordringene i egne utredninger.



Deling og gjenbruk av forskningsdata og de samfunnsmessige merverdiene av dette krever standardisering, herunder at det utvikles gode standarder for metadata. Innen mange fagfelt finnes det gode internasjonale standarder, og vi opplever det som en kritisk faktor at norske fagmiljøer og institusjoner sammen jobber for å utvikle funksjonelle standarder for metadata. Forskningsinfrastrukturene bør være en sentral node i utvikling og konvergens av metadata både internt og på tvers av fagdisipliner. Gode forvaltningsrutiner rundt metadata bør være et sentralt fokus for infrastrukturene.

Mange av problemstillingene reist i denne utredningen er ikke unike for forskningen. Norsk offentlig forvaltning forvalter enorme dataressurser, som resultat av sin tjenesteproduksjon og gjennom egen datainnsamling og samfunnsrolle. Aktører som Kartverket, NVE og Miljødirektoratet forvalter store mengder data. I et livsløpsperspektiv er det viktig å se sammenhengene mellom forsknings- og forvaltningsdata. Ofte er de sammenvevd i brukssituasjonene, og grensene er utflytende. Innen en del domener er dataene de samme, eller deles mellom forskningsmiljøer og forvaltningsorganer. Et eksempel er sosiodemografiske bakgrunnsdata fra SSB – som benyttes mye innen samfunnsfaglig og medisinsk forskning. Forvaltningens utfordringer med metadata og gjenfinning, strukturer for deling, langtidslagring og gjen- og viderebruk, og personvernproblematikk er nøyaktig de samme som for forskningen. Noe av det som skiller er det sterke internasjonale preg forskningen innen flere domener har, og de hensyn dette innebærer for norske forskningsmiljøer. Vi tror likevel at mange av de tiltak som er relevante i det videre arbeidet, for eksempel innen kompetanse om personvern hensyn og andre juridiske forhold, er tiltak som forvaltningen og forskningen med fordel kan se på sammen. For eksempel er det naturlig med et samarbeid med Digitaliseringsdirektoratet som jobber med mange av de aktuelle temaene på et mer generisk nivå, for forvaltningen.

Vi håper denne utredningen og dens leveranser kan være et viktig hjelpemiddel og verktøy som på kort sikt understøtter overgangen mot deling av forskningsdata som standard i norsk forskning, men som samtidig bidrar i det mer langsiktige arbeidet med å sikre hensiktsmessige og samfunnsnyttige strukturer rundt den store transisjonen som både norsk og internasjonal forskning nå går gjennom.

## 6. ANBEFALINGER

### 6.1. FORSKNINGENS BRUK AV VERKTØY OG TJENESTER

En gjennomgående anbefaling er å bruke disiplinspesifikke internasjonale tjenester der det finnes. (Dette omfatter norske tjenester som betjener en internasjonal gruppe og norske noder av internasjonale nettverk.) Internasjonale tjenester har ofte et stort utvikler- og brukermiljø og denne kompetansen kan man dra nytte av. Det er også større sannsynlighet for at de internasjonale tjenestene utvikles etter eller bidrar til å etablere og forbedre internasjonale standarder og at det lages integrasjoner mot andre tjenester og verktøy. Det vil normalt også være betydelig mer kostnadseffektivt å samarbeide over landegrensene enn at hvert nasjonale fagmiljø skal lage sin egen løsning.

#### Verktøy for datahåndteringsplan

- Institusjonene bør anbefale og forskerne bør bruke verktøy for datahåndteringsplaner som inneholder strukturert informasjon som gjør det mulig å søke maskinelt på og gjenbruke innholdet. Det fins både internasjonale og nasjonale alternativer. Rene Word-maler bør man etter hvert bevege seg bort fra.
- Institusjonene bør anbefale og forskerne bør bruke verktøy for datahåndteringsplaner som har funksjonalitet (APIer) som gjør det mulig å gjenbruke og dele informasjonen når forskningsdata senere i prosjektet skal lagres og arkiveres.

#### Metadata

- Forskere bør dokumentere data underveis i prosjektet, både med tanke på senere bruk i eget miljø og med tanke på bruk av andre etter prosjektavslutning. Dette vil gjøre deling enklere. Det finnes flere gode guider for bruk av metadata.
- Forskere bør bruke internasjonale metadatastandarder der det finnes. De internasjonale fagmiljøene har ofte utviklet gode metadatastandarder for data innen sin disiplin, noe som gir best mulighet for gjenbruk. RDA har laget en [oversikt](#) over relevante standarder innen ulike disipliner.
- Institusjonene og forskningslederne må arbeide mot at stadig flere norske forskere bruker internasjonale standarder for beskrivende data/metadata. Dette gjelder både på generisk nivå og på disiplinspesifikt nivå. Der det mangler gode standarder (eller er for mange) bør Norge bidra til å etablere slike.
- Institusjonene bør anbefale bruk av tjenester som har god funksjonalitet for å hjelpe forskerne til å legge på metadata tidlig i prosessen og forbedre/utvide disse løpende gjennom hele forskningsprosessen.

#### Håndtering av aktive data

Åpne data/data som kan deles fritt:

- Institusjonene bør anbefale og forskerne bør velge løsninger med god informasjonssikkerhet selv om dataene kan deles åpent. Dette for å sikre integritet i dataene (unngå tilsiktet eller utilsiktet endring av andre) og sikre tilgjengelighet (at ikke data slettes).
- Lagre data i disiplinspesifikke ( gjerne internasjonale) infrastrukturer styrt av forskersamfunnet dersom dette finnes. Dette vil normalt være infrastrukturer som mange aktører har interesse av å sikre at opprettholdes over tid og det vil oftest finnes et miljø som kan gi god hjelp til bruk, så man sikrer at dataene faktisk vil være mulig å finne og gjenbruke.
- Bruk generiske nasjonale infrastrukturer (noen av disse er drevet av en institusjon) der det ikke finnes gode disiplinspesifikke infrastrukturer.
  - Vi anbefaler NIRD fra Uninett Sigma 2 for behandling av store datasett
- Bruk lokale infrastrukturer der heller ikke nasjonale infrastrukturer dekker behovet

#### Persondata:

- Sensitive data skal lagres i nasjonale eller internasjonale tjenester som er godkjent for dette. Det bør utvises forsiktighet med å lagre sensitive data på lokale filservere.
  - Vi anbefaler bruk av TSD eller andre infrastrukturer godkjent for gule/røde/sorte data dersom det ikke fins fagspesifikk godkjent infrastruktur.

#### Arkivering av data

##### Generell anbefaling

- Bruk disiplinspesifikke ( gjerne internasjonale) infrastrukturer der disse finnes og følg deres metadatastandarder, for å sikre at dataene faktisk vil være mulig å søke opp / gjenfinne.
- Bruk generiske nasjonale infrastrukturer (noen av disse er drevet av en institusjon) der det ikke finnes gode disiplinspesifikke infrastrukturer
- Bruk lokale infrastrukturer der heller ikke nasjonale infrastrukturer dekker behovet
- Det anbefales å ikke 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.

#### Åpne data/data som kan deles fritt:

- Bruk NSD, DataverseNO eller andre nasjonale infrastrukturer (noen av disse er drevet av en institusjon) der det ikke finnes gode disiplinspesifikke infrastrukturer

Persondata:

- I en del tilfeller kan persondata (men ikke sensitive data) arkiveres hos NSD, men dette forutsetter samtykke.
- Det mangler gode nasjonale generiske infrastrukturer for arkivering og langtidsbevaring av persondata generelt og sensitive data spesielt.
- Dersom det ikke er mulig å publisere dataene åpent, vurder om det er mulig å publisere metadata, slik at personer med legitim grunn til å få tilgang, kan oppdage dataene og søke om slik tilgang.

## 6.2. FORBEDRING OG UTVIKLING AV VERKTØY OG TJENESTER

### Verktøy for datahåndteringsplan

- Leverandører må videreutvikle verktøyene med funksjonalitet til å gi forskerne best mulig hjelp til å planlegge datahåndtering.
- Leverandører bør tilby funksjonalitet som gjør det mulig for forskerne å legge på prosjektrelaterte metadata allerede i planleggingsfasen.
- Leverandører bør utvikle funksjonalitet (API-er) for å la andre aktører gjenbruke informasjonen senere i prosjektet.
- Forskningsrådet og andre finansiører bør utvikle sine søknadssystemer til å kunne lese strukturerte datahåndteringsplaner, slik at forskerne ikke trenger å skrive dette inn på nytt.
- Leverandører bør gi mulighet for disiplinspesifikke og institusjonsspesifikke tilpasninger av maler for datahåndteringsplaner.
- Verktøy for datahåndteringsplaner og leverandører av samtykkeløsninger må gi veiledning i hvordan samtykker kan utformes på en måte som legger til rette for senere deling for informasjon.

### Generelle anbefalinger for videreutvikling av tjenester og verktøy for håndtering av forskningsdata:

- Leverandører av eksisterende tjenester og verktøy må utvikle dem i retning FAIR. Første trinn bør være å tilby funksjonalitet for å utstyre dataene med FAIR metadata. (Her vil utredningsgruppen jobbe videre med anbefalinger frem til endelig leveranse gjennom å prioritere enkelte av FAIR-prinsippene)
- Leverandører må sikre at datasett som deles påføres en unik og persistent identifikator (PID/DOI), slik at de kan siteres og forskerne kan få uttelling og anerkjennelse for arbeidet



- Generiske tjenester for lagring og arkivering av forskningsdata bør støtte et sett av internasjonale standarder for generiske metadata, men også tillate bruk av fagspesifikke standarder.
- Det bør defineres nivåer av FAIR (flere internasjonale forslag finnes), slik at miljøer som er i startgropen kan få noe oppnåelig å strekke seg etter samtidig som de miljøene som ligger langt fremme kan synliggjøre at dette er tilfelle.
- Det må etableres (basert på internasjonale eksempler) metoder og verktøy for å vurdere graden av FAIR i en datasamling.
- Noen leverandører bør også fortsatt legge til rette for å kunne lagre data som ikke er fullt ut FAIR, men som f.eks. gjør at dataene er gjenfinnbare/søkbare og tilgjengelige. Det å tilgjengeliggjøre data fortsatt nytt for mange forskere, og det er i en overgangsfase fortsatt behov for verktøy og tjenester som ikke krever FAIR data fullt ut.
- Det bør etableres en eller flere nasjonale tjenester for behandling og analyse ala NIRD fra Uninett Sigma2, men for mindre datamengder – Den lange halen.
- Det bør vurderes å etablere (evt utvide eksisterende) tjenester nasjonalt for arkivering og langtidsbevaring av persondata/sensitive data.
- Leverandørene må både kunne hente informasjon fra andre via APIer og kunne tilby datadeling til andre på samme måte. Dette vil bidra til å redusere fragmentering og silosystemer, spare forskerne for dobbeltregistrering og gjøre det mulig for datamaskiner å automatisk gjenfinne og sammenstille.
- Det bør gjennomføres en større gjennomgang av hvilke infrastrukturer som finansieres per i dag. Etablere incentiver for å videreutvikle de infrastrukturer som finnes i «FAIR-retning» og samarbeide med andre. (Vi er usikre på hva som er det rette tiltaket her. Forskningsrådet har strammet inn en del og krever nå sterkere begrunnelse for å få lov til å etablere en ny infrastruktur. Bør dette styrkes ytterligere?)
- Finansiørene bør unngå å gi støtte til etablering av helt nye infrastrukturer der det allerede finnes eksisterende infrastrukturer som kan brukes. Mange av de eksisterende gode datasamlingene er etablert som overbygginger til generiske lagringsløsninger.
- Leverandører av IT-løsninger knyttet til håndtering av forskningsdata må gjøre seg kjent med konseptvalgutredningen Kunnskapssektorens datafelleskap ([KUDAF](#)).
- Det kan eventuelt vurderes større grad av samarbeid mellom institusjonene om (anskaffelse av) elektronisk labjournal.

### Nasjonalt vitenarkiv

- Nasjonalt vitenarkiv er planlagt å bli en sammensmelting og videreutvikling av funksjonaliteten i dagens Cristin, vitenarkivet Brage og datalagringsløsningen BIRD.
- Hovedvekten nå i første fase ligger på funksjonalitet knyttet til publikasjoner (inkludert kunstnerisk utviklingsarbeid), både det å få inn metadata om og lenke til publikasjoner, eventuelt kopi av fulltekst hvis den ikke ligger trygt et annet sted. Det er imidlertid også mulig å legge på henvisninger til datasett brukt i publikasjonene.
- I dagens BIRD er det mulig å lagre datasett med tilkoblet unik identifikator (DOI el.l.) og med noen enkle metadata. Det er i øyeblikket ikke planlagt mer avansert funksjonalitet, så denne planen opprettholdes, blir BIRD-delen av et «hvis du ikke har et annet sted som egner seg, så ligger i alle fall dataene dine trygt her». BIRD er ikke formelt godkjent for lagring av sensitive data, men kan for andre typer data brukes i alle faser av et forskningsprosjekt.

## 6.3. KOMPETANSE

- Nasjonale aktører bør, i samarbeid med institusjonene etablere oversikt over opplæringstilbud for datarøktene. Det må totalt sett (nasjonalt eller på institusjonene) etableres tilstrekkelig kapasitet for opplæring av datarøktene.
- Nasjonale aktører bør i samarbeid med de institusjonene som har kommet lengst, utarbeide veiledningsmateriell til institusjonene som kan brukes i deres opplærings- og informasjonsarbeid.
- Kapasiteten på nasjonalt nivå på bistand til juridiske og personvernrelaterte problemstillinger i forskning bør økes, og det må avklares hvordan dette best løses. Spesielt de mindre institusjonene melder behov for støtte på dette området. Slik kompetanse kan f.eks. tilbys enten ved å øke kapasiteten i NSD eller ved å kjøpe av tjenester mellom institusjonene.
- Det anbefales å etablere regulatoriske sandkasser, dvs tjenester der man kan ta med en case, og så teste ut om det er mulig/lovlig å gjøre det slik man har tenkt. Dette kan være et alternativ for å utforske problemstillinger knyttet til deling av data med behov for beskyttelse. Brukes i dag en del i forbindelse med AI, f.eks. av Datatilsynet og Arkivverket.
- Nasjonale aktører bør ta initiativ til en nasjonal kampanje for å øke den grunnleggende kjennskapen til FAIR blant alle som skaper og bestiller data. Det mangler basiskunnskap og kjennskap til FAIR blant veldig mange, ikke bare i forskningen, men også i hele forvaltningen (som bl.a. «bestiller» data fra forskningen og fra konsulenter og andre). En



nasjonal kampanje på nivået over forskningsdata-tematikken, men hvor også UH-/forskningssektoren slutter opp. Adressat for anbefalingen: KMD/Digdir. Det kan jo f.eks. oppleves underlig (og forvirrende) at Digitaliseringsdirektoratet ikke promoterer FAIR i sine tilnærminger. Ref. også at regjeringen nå igangsetter flere utredninger om dataforvaltning for gjenbruk og viderebruk.

- Institusjonene må ha informasjon om datahåndtering og FAIR forskningsdata lett tilgjengelig i sine informasjonskanaler. Institusjonenes egne nettsider og retningslinjer er en viktig informasjonskilde for forskerne. Institusjonene spiller derfor en viktig rolle for å informere om god datahåndtering og verdien av å dele forskningsdata.
- Opplæring i datahåndtering må inn tidlig og inngå som en del av utdanningsløpet for studenter og Phd-kandidater. De som veileder dem må også få god opplæring.
- Den enkelte institusjon må prioritere å skaffe seg enten egen kompetanse eller tilgang til kompetanse for å jobbe med datahåndtering. Sette som mål at det skal finnes (tilgang til) 3 årsverk for datahåndtering pr 100 forskere. De som er for små til å ha egne ansatte må samarbeide om felles tjenester til dette, eventuelt i samarbeid med større institusjoner og nasjonale aktører

#### 6.4. KULTUR FOR DELING

- Forskerne må oppleve at de får noe igjen for å dele datasett. Man trenger ikke nødvendigvis ha finansiering knyttet til det, men det må gi merittering og telle i ulike typer vurderinger. Det pågår arbeid både på EU-nivå<sup>16</sup> og nasjonalt i regi av UHR (NOR-CAM<sup>17</sup>) for å etablere et bredere sett av kriterier for å vurdere akademiske prestasjoner. Det vil være disiplinforskjeller på hvordan en bredere «palett» av vurderingskriterier skal brukes, men dagens ensidige fokus på publikasjoner vil for mange disipliner være medvirkende i å hindre datadeling.
- Det bør utredes hvordan standardkontrakter som sektormyndighetene benytter, f.eks. i forbindelse med utsetting av kartleggingsoppdrag og ulike former for oppdragsforskning har formuleringer om FAIR (f.eks. krav til metadata, lisenser etc.) for de data som samles inn. Likeledes kan det i slike utlysninger etterspørres tidligere bidrag til FAIR data.
- Det bør settes av midler til å støtte opp om og initiere tiltak for å fremme delingskultur, både deling av egne og gjenbruk av andres forskningsdata.
- Norske forskningsmiljøer og myndigheter bør fortsette å støtte deltakelse i internasjonalt samarbeid om alle aspekter ved håndtering av forskningsdata, som f.eks. EOSC og RDA.

<sup>16</sup> Rapport fra EU: <https://op.europa.eu/en/publication-detail/-/publication/36ebb96c-50c5-11ec-91ac-01aa75ed71a1/language-en/>

<sup>17</sup> Veileder fra NOR-CAM: <https://www.uhr.no/temasider/karrierepolitikk-og-merittering/nor-cam-veileder-for-vurdering-i-akademiske-karrierelop/>



- Sikt har planer om å følge opp punktet i strategi for tilgjengeliggjøring og deling av data om indikatorer/statistikk over åpne data. Det er viktig at dette tiltaket gjennomføres.



## **7. VEDLEGG**

Se vedlegg med spørsmål fra spørreundersøkelsene i egne PDF-er.

# UNIT

DIREKTORATET FOR IKT OG FELLESTJENESTER  
I HØYERE UTDANNING OG FORSKNING

