

Investering i infrastrukturer for FAIR forskningsdata og særlig relevante forvaltningsdata for forskning

Organisering og finansiering av datainfrastruktur for best mulig
utnyttelse

Anbefalinger fra datainfrastrukturutvalget mai 2022

Mai 2022

Norges forskningsråd
Besøksadresse: Drammensveien 288
Postboks 564 , 1327 Lysaker
Telefon: 22 03 70 00
Telefaks: 22 03 70 01

post@forskningsradet.no / www.forskningsradet.no

ISBN nr. 978-82-12-03933-9 (PDF)

Publikasjonen kan bestilles via internett: www.forskningsradet.no/publikasjoner



Dette arbeidet er lisensiert med en [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Innhold

Forord	5
Sammendrag	6
Bakgrunn	8
<i>Oppdraget fra Kunnskapsdepartementet</i>	8
<i>Arbeidets avgrensning</i>	9
<i>Rapportens oppbygning</i>	10
DEL 1 Infrastrukturbehov i et framtidsrettet og datadrevet forsknings- og innovasjonssystem	11
<i>Om bruk av begreper i rapporten</i>	12
1. <i>FAIR data</i>	12
1.1 FAIR data krever godt organiserte datainfrastrukturer	13
1.2 FAIR data stiller nye krav til kompetanse	13
1.3 FAIR data forutsetter tydelig ansvarsdeling	14
1.4 FAIR data forutsetter internasjonalt samarbeid	14
1.5 FAIR data for bedre samhandling mellom offentlig og privat sektor	14
2. <i>Hva er behovet framover?</i>	15
DEL 2 Ambisjoner for datainfrastrukturer i Norge	17
<i>Utvalgets forslag til ambisjonsnivå for datainfrastrukturer i Norge</i>	18
3. <i>Ambisjonsnivå for bedre utnyttelse av data for forskning og samfunnsutvikling i Norge</i>	18
4. <i>Gvinster</i>	18
4.1 Gvinster for forskningen.....	19
4.2 Gvinster for forvaltningen, samfunnet og befolkningen	19
4.3 Gvinster for næringslivet	20
4.4 Internasjonalt samarbeid er avgjørende for å ta ut gevinstene.....	20
5. <i>Kostnader ved å investere i infrastruktur for FAIR forskningsdata og forvaltningsdata</i>	21
5.1 Kostnader til e-infrastruktur som er en forutsetning for databasert forskning og forvaltning	21
5.2 Kostnader til datainfrastruktur for FAIR data.....	22
6. <i>Trinnvis realisering av FAIR data</i>	24
7. <i>Vurdering av ambisjoner og risiko</i>	25
8. <i>Friske midler eller investering innenfor dagens rammer?</i>	26
Tabell over scenarier	27
DEL 3 Bærekraftig organisering og finansiering av datainfrastruktur	28
<i>Utvalgets anbefalinger om finansiering og organisering av datainfrastrukturer</i>	29
9. <i>Finansieringskilder og pengestrømmer</i>	30

9.1	Kunnskapsdepartementets rolle	30
9.2	Øvrige departementers rolle	30
10.	<i>Ansvarsdeling og koordinering i et uoversiktlig landskap av datainfrastrukturer.....</i>	<i>31</i>
10.1	Forskerfellesskapenes ansvar	32
10.2	Institusjonenes ansvar	32
10.3	Stimulere til nettverksvirksomhet og kompetansedeling	33
10.4	Investering i nasjonale datainfrastrukturer.....	33
10.5	Hvorfor internasjonal deltagelse er viktig for god koordinering og ansvarsdeling.....	34
10.6	Ansvar for å gjøre forvaltningsdata tilgjengelige for forskning.....	35
10.7	Dimensjonering og prioritering gjennom nettverk og samarbeid.....	36
	DEL 4 Forslag til rammeverk for finansiering av datainfrastruktur.....	37
	<i>Utvalgets overordnede anbefalinger.....</i>	<i>38</i>
11.	<i>Forslag til rammeverk for framtidig investering i og finansiering av datainfrastrukturer ...</i>	<i>38</i>
	<i>Utvalgets forslag til bærekraftig rammeverk for finansiering av datainfrastrukturer.....</i>	<i>40</i>

Forord

Data har alltid vært sentralt for forskningen, men har ofte vært den enkelte forskers ressurs og ansvar. Den teknologiske utviklingen har ført til langt flere og mer avanserte kilder til data, ofte fra store infrastrukturplattformer med mange brukere, noe som øker kostnadene ved å samle inn data. Den hurtige digitaliseringstakten har gitt nye muligheter for å dele data, noe som også har ført til en økt erkjennelse av verdien av forskningsdata. Dette ligger til grunn for oppmerksomheten om at offentlig finansierte forskningsdata og særlig relevante forvaltningsdata for forskning bør deles, slik at offentlige investeringer kommer alle til gode, slik at det gis gode muligheter for etterprøving av forskningen, og slik at samfunnet kan høste av de verdiene som offentlig finansierte forsknings- og forvaltningsdata representerer.

Dette arbeidet må sees i sammenheng med andre utredninger som på ulike områder har foreslått løsninger for å oppnå mer deling og gjenbruk av data. Det er i stor grad sammenfall i anbefalinger, og samlet vil det være et godt grunnlag for å følge opp med handling, slik at verdien av dataene blir utnyttet bedre framover.

Norge har et godt utgangspunkt for å ligge i front på dette området, med de gevinster det vil gi for norsk forskning, forvaltning og næringsutvikling, men det vil kreve politisk vilje og gjennomføringskraft på tvers av sektorer og god samhandling mellom de ulike aktørene.

Jeg vil takke medlemmene av utvalget, som har bidratt med kompetanse og erfaring fra ulike fag- og forvaltningsområder. Takk også for godt forberedte og poengterte innspill fra eksterne ressurspersoner underveis i prosessen. Mandatet har vært omfattende, og med relativt kort tidshorison og restriksjoner for fysiske møter har diskusjonene vært løftet opp på et generelt og prinsipielt nivå. Det har resultert i et foreslått rammeverk for organisering og finansiering av datainfrastrukturer i Norge.

På vegne av utvalget vil jeg takke Norges forskningsråd og Kunnskapsdepartementet for et spennende oppdrag. Jeg vil også rette en stor takk til sekretariatet, Rita Bergersen, Siri Lader Bruhn og Katrine Weisteen Bjerde, for deres uvurderlige støtte til utvalgets arbeid. Til sist en takk til de mange som har delt sin kunnskap og kompetanse med utvalget i løpet av vårt arbeid.

Kenneth Ruud
utvalgsleder

10. mai 2022

Sammendrag

For å oppnå forskning av høy kvalitet er det viktig at dataene som er brukt, er kvalitetssikret og kan brukes av andre. På den måten får vi forskning som er til å stole på, og som folk har tillit til. Dessverre er det for mange eksempler på at data som produseres av offentlig finansiert forskning, ikke blir delt og gjenbrukt. Det til tross for at Norge har sluttet seg til de internasjonale forpliktelsene om at data skal være FAIR (Findable, Accessible, Interoperable og Reusable). Å sørge for at data kan deles og brukes, kommer imidlertid ikke av seg selv. For at data skal være gjenfinnbare, tilgjengelige, kompatible eller samhandlende og gjenbrukbare (FAIR), kreves det både fysisk datainfrastruktur og kompetanse. Å vurdere kvaliteten på data krever fagkunnskap, og å sikre at data kan gjenbrukes av både mennesker og maskiner, kan forutsette spesialistkompetanse som ikke alle forskere har. Det er behov for å videreutvikle dagens løsninger til et økosystem som ivaretar alle disse forholdene.

I denne rapporten foreslås det hvordan et slikt økosystem kan organiseres og finansieres. Det pekes samtidig på hvilke gevinster som kan realiseres hvis Norge setter seg ambisiøse mål på området. Utvalget foreslår et rammeverk som ivaretar behovet for langsiktighet, koordinering, kvalitetssikring, etablering og videreutvikling av datainfrastrukturer for FAIR-data. Det framheves i tillegg hvilke aktører som må få ansvar for ulike deler av økosystemet.

Mer deling og gjenbruk av forskningsdata av god kvalitet åpner for helt nye muligheter, og evnen til å utnytte disse mulighetene vil være en kritisk faktor for deltakelse i forskningsfronten i årene som kommer. Det betyr at forskningssystemet må rigges på en annen måte enn tidligere. Det krever internasjonal standardisering, nasjonal koordinering og prioritering og endringer på institusjonelt nivå. Internasjonalt samarbeid er viktigere enn noen gang og nødvendig for at forvaltning av forskningsdata skal skape merverdi for forskning og samfunn. Gode infrastrukturer for tilrettelegging, tilgang til og gjenbruk av forskningsdata vil bli avgjørende for forskning på nær sagt alle områder i framtiden. Dette forutsetter igjen at forskere har grunnleggende kompetanse om datahåndtering og verktøy til å håndtere data på en god og sikker måte i alle ledd av forskningsprosessen. I tillegg til et generelt kompetanseløft blant forskere vil det bli mer behov for spesialistkompetanse innenfor data og dataforvaltning i og rundt forskningsmiljøene.

I rapporten blir det redegjort for viktige utviklingstrekk og hvilke krav som stilles til forsknings- og innovasjonssystemet for å innfri ambisjonsnivået som utvalget foreslår:

- Forskere ved norske universiteter og institutter skal ha tilgang til datainfrastrukturer som muliggjør forskning og utdanning i verdensklasse.
- Innen 2030 skal alle fagområder i Norge ha tilgang til kompetanse, veiledning og kuratering av forskningsdata, enten i form av nasjonale løsninger eller helt eller delvis gjennom deltakelse i europeisk eller internasjonale infrastruktursamarbeid.
- Norge skal på utvalgte områder ha datainfrastrukturer i verdensklasse som blir foretrukket av internasjonale brukere.
- Norge må ha en opptrappingsplan for organisering og finansiering av datainfrastrukturer som gjør det mulig å hente ut gevinstene av de store datamengdene som vil genereres med offentlig finansiering i årene som kommer.

For at det foreslåtte ambisjonsnivået skal kunne realiseres, må forskningsbudsjettet være stort nok til å romme nødvendige investeringer i datainfrastruktur og tilhørende støttefunksjoner.

Utvalget foreslår videre hvordan datainfrastrukturer kan organiseres og finansieres. Det blir blant annet pekt på at sektordepartementene i større grad enn i dag må ta ansvar og bidra med

finansiering av datainfrastrukturer. Det er dessuten nødvendig med samarbeid og en bedre koordinering og arbeidsdeling mellom aktører med ansvar for ulike datatjenester. Til slutt presenterer utvalget et rammeverk som kan bidra til en mer bærekraftig finansiering av etablering og drift av datainfrastrukturer. Rammeverket er ment som et verktøy som både departementene, virkemiddelapparatet og infrastrukturer kan bruke for å utvikle egnede finansieringsmodeller.

Bakgrunn

Høsten 2022 skal Kunnskapsdepartementet legge fram en revidert langtidsplan for forskning og høyere utdanning. Målet med revisjonen er å gi forsknings- og innovasjonssektoren et bedre verktøy for å ta innover seg endringene i kunnskapslandskapet og utvikle politikken i tråd med viktige nasjonale og internasjonale utviklingstrekk. Det grønne og det digitale skiftet er slike trekk, og et framtidig forsknings- og utdanningsystem må være rustet til å understøtte datadrevet forskning, forvaltning, tjenesteutvikling og verdiskaping på en bærekraftig og etisk forsvarlig måte. Åpen forskning er en tydelig prioritering både nasjonalt og internasjonalt hvor tilgjengeliggjøring av resultater fra forskning er et sentralt element.

Oppdraget fra Kunnskapsdepartementet

Fra Kunnskapsdepartementet (KD) fikk Forskningsrådet i oppdrag å nedsette et utvalg som skulle utforme et grunnlagsdokument med anbefalinger om investeringer i infrastrukturer for FAIR-forskningsdata og særlig relevante forvaltningsdata for forskning. Anbefalingene og vurderingene som presenteres i denne rapporten, skal brukes som underlag for Kunnskapsdepartementets arbeid med den nye langtidsplanen.

Utvalgets overordnede innfallsvinkel har vært å vurdere en mulig organisering og finansiering av relevant datainfrastruktur for FAIR-forskningsdata i Norge. Aktuelle momenter i arbeidet har vært

- å identifisere hvilke datatjenester det vil være behov for
- å foreslå hvem som bør ha ansvar for ulike typer data og håndteringen av dem
- å anbefale hva som bør løses lokalt, hva som bør organiseres som nasjonale fellestjenester, og hva som bør løses gjennom internasjonalt samarbeid
- å foreslå ulike investerings- og driftsmodeller
- å vise gevinstpotensial og gevinstrealisering som en godt utbygd infrastruktur for FAIR-forskningsdata vil ha for produsenter og brukere av data
- å peke på områder der det er nødvendig å sette i gang nye tiltak
- å anbefale ambisjonsnivå, inkludert hvilke ambisjoner Norge bør ha i det europeiske samarbeidet på området

Utvalget hadde sitt første møte i september 2021 og har i alt hatt ni møter. På flere av møtene har inviterte innledere fra inn- og utland belyst tematikken fra ulike sider. Det har i tillegg vært to innspillmøter, ett med inviterte deltakere og ett åpent for alle, i tillegg til dialog med aktuelle aktører underveis i arbeidet. I januar ble et notat lagt ut på høring for skriftlige innspill, og utvalget fikk inn i alt fjorten svar. Alle innspill er vurdert og tatt hensyn til i denne endelige rapporten fra utvalget.

Medlemmene i utvalget har kompetanse og erfaring fra ulike typer datainfrastrukturer og representerer både UH-sektoren, instituttsektoren, offentlig forvaltning og næringslivet:

Navn	Stilling	Arbeidsgiver
Kenneth Ruud, leder	Administrerende direktør	Forsvarets forskningsinstitutt
Ingrid Heggland, nestleder	Førstebibliotekar, prosjektleder	NTNU Universitetsbiblioteket
Ingunn Limstrand	Seksjonsleder	Miljødirektoratet

Klas Henning Pettersen	Leder	NORA, Norwegian Artificial Intelligence Research Consortium
Gunnar Bøe	Daglig leder	Sigma2 AS
Tanja Storsul	Direktør	Institutt for samfunnsforskning
Knut Kalgraff Skjåk	Avdelingsdirektør	Sikt – Kunnskapssektorens tjenesteleverandør
Christine Stansberg	Seniorrådgiver	UiB, ELIXIR Norge
Pål Richard Romundstad	Professor	NTNU, Institutt for samfunnsmedisin
Alexander Refsum Jensenius	Professor	UiO, Institutt for musikkvitenskap
Ann-Kristin Brændvang	Fagdirektør	SSB
Carina Hundhammer	Leder, høyere utdanning og forskning	Abelia
Geir Huse	Forskningsdirektør	Havforskningsinstituttet

Sekretariatet har bestått av Siri Lader Bruhn og Rita Bergersen (leder) fra Forskningsrådet og Katrine Weisteen Bjerde fra Direktoratet for høyere utdanning og kompetanse (HK-dir).

Arbeidets avgrensning

Et utvalg på denne størrelsen vil ikke ha erfaring og kompetanse fra alle forsknings- og forvaltningsområder. Det har derfor hele tiden vært viktig å løfte diskusjonene opp på et generelt og prinsipielt nivå. Dette har resultert i overordnede anbefalinger framfor konkrete anbefalinger rettet mot spesifikke områder.

Gitt oppdragets relativt korte varighet har det vært nødvendig å gjøre noen avgrensninger i det omfattende mandatet. Utvalget har derfor ikke gjennomført egne kartlegginger eller analyser i arbeidet. Det er imidlertid gjennomført flere andre prosesser den siste tiden som har skjæringspunkter mot utvalgets oppdrag. Vurderinger og anbefalinger som er gitt i disse prosessene, har vært en del av kunnskapsgrunnlaget og er tatt med i utvalgets diskusjoner der det har vært relevant. Følgende prosesser har vært spesielt relevante:

- *Behov og finansieringsstrategi for nasjonal e-infrastruktur for forskning for perioden 2020–2030.* Arbeidsgruppen anbefalte blant annet å øke de nasjonale investeringer i e-infrastrukturer til 250 millioner kroner årlig og en modell for framtidig justering av dette investeringsbehovet.
- Anbefaling om å øke investeringene til Forskningsrådet, både til de generiske og til de fag- og temaspesifikke nasjonale plattformene, til minst 400 millioner kroner årlig og at 5 prosent av nasjonal FoU bør investeres i tilgjengeliggjøring av forskningsdata (brev fra Forskningsrådet til KD 30.9.2021).
- *Datadeling i næringslivet* (2020). Rapport fra ekspertgruppe med anbefalinger til Stortinget om hvordan data fra næringslivet kan deles: [rapport-fra-datadelingsutvalget2.pdf \(regjeringen.no\)](https://www.regjeringen.no/dokument/rapp-fra-datadelingsutvalget2).

- *Hvordan skal vi dele forskningsdata?* Utredning og anbefaling om lisensiering og tilgjengeliggjøring (Forskningsrådet 2021): [rapport rettighets--og-lisenssporsmal-ved-tilgjengeliggjoring-av-forskningsdata.pdf \(forskningsradet.no\)](#).
- *Evaluation of the INFRASTRUKTUR initiative as a funding instrument* (Forskningsrådet 2021). Evalueringskomiteen anbefaler at INFRASTRUKTUR-ordningen videreføres og blir den viktigste mekanismen for investeringer i nasjonale forskningsinfrastrukturer og norske noder i internasjonale forskningsinfrastrukturer: [infrastruktur-evalueringsrapport.pdf \(forskningsradet.no\)](#).
- *Evaluation of UNINETT Sigma2* (2019) fastslår at infrastrukturen er viktig for norske forskere og nødvendig å øke aktivitetene, tjenestene og dens finansiering for å kunne ivareta de voksende behovene fra stadig flere brukere: [Evaluation of UNINETT Sigma2 - DFØ | kudos \(dfo.no\)](#).
- *Felles infrastruktur og tjenester for FAIR forskningsdata* (FAIR-utredningen, 2022). Utredningsgruppen påpeker i sin sluttrapport at det er store mangler i dagens videreutvikling av både infrastrukturer, kompetanse og samarbeid om FAIR-forskningsdata: [Open Science | Felles infrastruktur og tjenester for FAIR forskningsdata](#).

I tillegg vil utvalget nevne fire påbegynte prosesser som er relevante for utvalgets arbeid, men som foreløpig ikke har levert sine forslag eller rapporter:

- Lovutvalgsarbeid: [Helhetlig regulering av videre bruk av offentlig informasjon \(data\)](#)
- Ekspertgruppe, ledet av Simen Markussen: [Effektiv og sikker infrastruktur for deling og bruk av relevant statistikk og data i kriser](#)
- Ekspertgruppe, ledet av Mari Rege: [Juridiske og etiske problemstillinger knyttet til innsamling, deling og bruk av randomiserte forsøk i kriser](#)
- Ekspertgruppe, ledet av Heri Ramampiaro: [Deling av industridata](#)

Rapportens oppbygning

I del 1 av rapporten beskrives konteksten som anbefalingene til utvalget må sees i lys av, og de infrastrukturbehovene som må imøtekommes dersom Norge ønsker å legge til rette for at norske forskningsdata skal utnyttes bedre. I del 2 går utvalget gjennom de gevinster og kostnader som er knyttet til investeringer i datainfrastrukturer, og gir en anbefaling om hvilket ambisjonsnivå Norge bør ha. I del 3 gir utvalget anbefalinger om hvordan norsk datainfrastruktur bør organiseres og finansieres på en mest mulig kostnadseffektiv og bærekraftig måte. Rapporten avsluttes med del 4, som presenterer et rammeverk for finansiering av datainfrastruktur for forskningsdata og særlig relevante forskningsdata for forskning.

DEL 1

Infrastrukturbehov i et framtidsrettet og datadrevet forsknings- og innovasjonssystem

Forskningssystemet gjennomgår nå to transformasjoner parallelt. Forskingen blir både mer dataintensiv og datadrevet som resultat av digitalisering og ny teknologi, mens det samtidig er en sterk dreining mot åpen forskning, herunder deling og gjenbruk av data. Dette fører til at grensene mellom ulike forskningsdisipliner blir mindre skarpe, og mulighetene for tverrfaglighet blir større. Slik blir det mulig å oppnå mer effektiv og kvalitetssikret flyt av data på tvers av forskning, forvaltning og næringsliv. I denne delen av rapporten gir utvalget en vurdering av hvilke endringer som må til for at Norge skal hevde seg i forskningsfronten og bidra til presis og tilgjengelig kunnskap på viktige samfunnsområder. Dette vil danne et bakteppe for anbefalingene som utvalget gir i del 2, 3 og 4 i rapporten.

Om bruk av begreper i rapporten

Forskningsdata:

Data som genereres eller oppstår underveis i forskningsprosessen.

Særlig relevante forvaltningsdata for forskning:

Data fra offentlig forvaltning som er interessante for bruk i forskning.

Datainfrastruktur:

De verktøyene, tjenestene og systemene, inkludert kompetanse, som er nødvendige for å innhente, analysere, lagre, organisere, dokumentere og tilgjengeliggjøre data.

Videre i denne rapporten vil vi for enkelhets skyld

- bruke ordet **datainfrastruktur** som en forkortelse for *infrastrukturer for FAIR-forskningsdata og særlig relevante forvaltningsdata for forskning*
- bruke ordet **data** når vi omtaler både *forskningsdata og særlig relevante forvaltningsdata for forskning* samlet, men presisere når det spesielt dreier seg om forskningsdata eller forvaltningsdata

1. FAIR data

Datadeling er ikke noe nytt verken innenfor forskningen eller forvaltningen, men omfattende digitalisering innenfor de fleste samfunnsområder har gitt muligheter for slik deling i mye større utstrekning og på helt andre måter enn tidligere. For å utnytte disse mulighetene må forskningsprosessene endres og investeringer i datainfrastrukturer og kompetanse som muliggjør deling og gjenbruk, må prioriteres.

At data fra det offentlige skal være åpne, har vært en politisk føring i mange år. At data er *åpne*, betyr veldig forenklet at de deles uten begrensninger på hvem som kan få tilgang til dem, og hva de kan brukes til. I regjeringens gjeldende strategi [Én digital offentlig sektor](#) (2019–2025) er det et grunnleggende prinsipp at offentlig sektor skal dele data når den kan, og skjerme data når den må. Dette er et flott og viktig prinsipp, men en forutsetning for at verdien i data skal kunne utnyttes utenfor sammenhengen de har oppstått i, er at de håndteres på en måte som setter andre, både mennesker og maskiner, i stand til å gjenfinne, forstå og bruke dem.

For å optimalisere bruk og gjenbruk av forskningsdata ble FAIR-prinsippene presentert i artikkelen [The FAIR Guiding Principles for scientific data management and stewardship](#) i Nature i 2016. Akronymet FAIR står for Findable (gjenfinnbare), Accessible (tilgjengelige), Interoperable (kompatible eller samhandlende) og Reusable (gjenbrukbare). Disse prinsippene har blitt retningsgivende innenfor de fleste forskningsfelt i store deler av verden, og de senere årene har også offentlig sektor og private aktører begynt å forholde seg til dem. OECD, UNESCO og EU har alle gitt sin tilslutning til at forskningsdata bør være FAIR, og dette er også et grunnprinsipp i [Nasjonal strategi for tilgjengeliggjøring og gjenbruk av forskningsdata](#). Alle utlysninger i Horisont Europa krever at prosjektene har en datahåndteringsplan som beskriver hvordan datahåndteringen er i tråd med FAIR-prinsippene, og hovedmålet til partnerskapet og stiftelsen European Open Science Cloud (EOSC) er å utvikle "a Web of FAIR Data and services". Norges forskningsråd har også innført krav om datahåndteringsplaner i forskningsprosjekter.

Det er med andre ord stor enighet i og tilslutning til at data bør deles etter FAIR-prinsippene. Hvis vi skal oppnå dette i praksis, må imidlertid enigheten følges opp med nødvendige tiltak. Det er viktig å

merke seg at å gjøre data FAIR i de aller fleste tilfeller vil være mer ressurskrevende enn å gjøre data åpne. Datainfrastrukturer som legger til rette for slik deling, blir helt avgjørende, men det må også inkludere økt kompetanse og god samordning og ansvarsdeling, både nasjonalt og internasjonalt.

1.1 FAIR data krever godt organiserte datainfrastrukturer

FAIR-data krever datainfrastrukturer som legger til rette for god datahåndtering. Med datainfrastruktur mener vi her verktøyene, tjenestene og systemene som er nødvendige for å innhente, analysere, lagre, organisere, dokumentere og tilgjengeliggjøre data. Å realisere FAIR-data vil kreve generiske plattformer for lagring, arkivering og tilgjengeliggjøring av store datamengder, og fag- og domenespesifikke datainfrastrukturer for ulike formål.

Det er behov for datainfrastrukturer som gir rask og enkel tilgang til data som er utstyrt med gode beskrivende metadata i tråd med internasjonale standarder, metoder for behandling av dataene, systemer for håndtering av personvern og opphavsrett og lisenser for videre bruk, slik at det er enkelt å sammenstille datasett fra mange ulike kilder. Datainfrastrukturene må fungere godt hver for seg, men ulike infrastrukturer leverer funksjonalitet for ulike faser i livsløpet til dataene. De må derfor også kunne utveksle informasjon og samvirke, slik at brukerne opplever (tilnærmet) sømløs og god funksjonalitet gjennom alle faser.

Dette vil kreve datainfrastrukturer som er tilrettelagt for samarbeid og kunnskapsdeling på tvers av land, fagområder og sektorer. For datainfrastrukturer for deling og gjenbruk av klausulerte/konfidensielle data må det stilles tilleggskrav til informasjonssikkerhet og trygg datahåndtering. Det gjelder blant annet for data som nasjonen har et ansvar for å beskytte, og som vil kreve juridisk kompetanse og teknologi for å ivareta slike særskilte behov.

1.2 FAIR data stiller nye krav til kompetanse

Når data blir en stadig mer sentral del av forskningsprosessen, må forskningssystemet rigges om. Det vil stille nye krav til kompetanse på alle nivåer, fra basiskompetanse om FAIR-data og datahåndtering hos forskerne til mer spesialisert kompetanse i støttetjenestene og ved infrastrukturene som tilbyr datatjenester til forskningsmiljøene. Det er også nødvendig at ledelsen ved forskningsinstitusjonene har en grunnleggende forståelse av nødvendige endringer som må gjøres. I [sluttrapporten fra FAIR-utredningen](#) blir behovet for kompetanseheving i forskermiljøene framhevet og foreslått fulgt opp med tiltak.

For å kunne betjene en datainfrastruktur må man ha kompetent personell med flere former for spesialisert ekspertise. Dataforvaltere (data stewards) er ansatte med spesiell kompetanse på datahåndtering. *En dataforvalter bistår forskere eller andre som håndterer data, med å sikre god datahåndtering gjennom alle fasene av datalivssyklusen*¹. Dette er en relativt ny kompetansegruppe i kunnskapssystemet, og vi må forvente at behovet vil øke i takt med økende krav om gjenbruk av data. Det er vanskelig å forutsi hvor stort behovet vil bli for denne typen kompetanse, men det har blitt foreslått at vi må regne med at opptil fem prosent av alle midler til forskning vil måtte brukes til dataforvaltning i framtiden (Mons, 2020). Personalet i en domenespesifikk datainfrastruktur trenger ofte inngående kunnskap om fagområdet de skal støtte, om hvordan data har oppstått, om hvilke analyser som er meningsfylt å gjøre på ulike typer data, og om hvilke typer metadata og standarder som er relevante. Samtidig må personalet ha teknisk kompetanse på det datafaglige området.

¹ Beskrivelsen er hentet fra [Open Science | Termliste for forskningsdatahåndtering](#)

En del oppgaver innenfor dataforvaltning er så grunnleggende at alle forskningsinstitusjoner må sørge for å ha et visst nivå av kompetanse. Andre oppgaver, som å vurdere juridisk hvorvidt et konkret datasett kan deles åpent eller ikke, krever spisskompetanse som det kan være hensiktsmessig å samle på ett eller noen få miljøer. For å gi forskerne god støtte og utnytte ressursene best mulig er det viktig med dialog for å sikre at alle forskningsmiljøer har tilgang på nødvendig kompetanse, samtidig som spisskompetanse ikke dubleres unødige.

1.3 FAIR data forutsetter tydelig ansvarsdeling

Investering i datainfrastrukturer og kompetanseheving er kostbart. Det blir derfor viktig å bruke ressurser på en effektiv og fornuftig måte. En avgjørende faktor for å få til dette er at det etableres et bedre samarbeid på tvers av infrastrukturer og institusjoner, og at roller og ansvar defineres tydelig. Det blir viktig å sortere hvilke datainfrastrukturtenester som bør tilbys på hver institusjon, og hvilke det er behov for å etablere som nasjonal datainfrastruktur eller en fellestjeneste. Med en tydelig rolle- og ansvarsdeling er det mulig å unngå overlappende og unødvendige investeringer, og det gir dessuten en bedre utnyttelse av kompetanse på tvers av aktørene i forskningssystemet.

Med FAIR-data vil dataflyten innad i forskningssystemet og mellom forskningen og andre sektorer bli bedre. Dette er i seg selv positivt, men det kan by på utfordringer med tanke på hvor ansvaret for datainfrastrukturer og forvaltning av data skal ligge. Det vil derfor også bli særlig viktig å tydeliggjøre ansvarsdelingen mellom Kunnskapsdepartementet og andre sektordepartementer som har et stort behov for forskningsbasert kunnskap som grunnlag for politikkutforming og forvaltning.

1.4 FAIR data forutsetter internasjonalt samarbeid

Deltakelse i internasjonalt samarbeid sikrer kvalitet og fornyelse av forskningen og er spesielt viktig for å oppnå FAIR-data. Slikt samarbeid gjør at Norge kan påvirke utviklingen av standarder for datahåndtering, metadata, metoder og teknologi, men gir også norske forskningsmiljøer tilgang til relevante datainfrastrukturer i Norge og i utlandet. Norge har gjennom mange år deltatt i internasjonalt samarbeid om datainfrastruktur og i internasjonale kompetansenettverk. I Europa vil slikt samarbeid i stor grad skje gjennom deltakelse i European Open Science Cloud (EOSC) i tiden som kommer. Norge er nå medlem av partnerskapet, noe som medfører nye muligheter, men også forpliktelser. Gjennom deltakelsen i EOSC får europeiske forskere tilgang til FAIR-data og relevante tjenester i virtuelle miljøer på tvers av vitenskapelige disipliner og landegrensler. EOSC utvikler noe nytt, men bygger i stor grad på eksisterende data- og e-infrastrukturer (også norske) for gjenfinning, tilgang til og gjenbruk av data. Deltakelse i internasjonalt samarbeid innebærer at risikoen og kostnadene deles mellom aktørene som er med. Dette er i tråd med ansvaret Norge har for å bidra til den internasjonale kunnskapsutviklingen og bidra til å løse felles problemer og utfordringer, som beskrevet i [Internasjonalt forskningssamarbeid – regjeringen.no](https://www.regjeringen.no).

1.5 FAIR data for bedre samhandling mellom offentlig og privat sektor

Stadig mer forskning gjennomføres i partnerskap mellom offentlig sektor og næringsliv. Dette er en villet politisk utvikling, og å tilrettelegge for samarbeid blir viktig dersom Norge skal nå sine mål om verdiskaping med data som ressurs i den nye dataøkonomien ([Data som ressurs – Datadrevet økonomi og innovasjon](#)). Datainfrastrukturer som tilrettelegger for enklere og smidigere tilgang til forskningsdata og forvaltningsdata, gir privat sektor mulighet til å utnytte disse i næringsutvikling og innovasjon. [Datafabrikken](#) som er under utvikling, skal hjelpe bedrifter og offentlige aktører med verdiskaping på grunnlag av data. Dette er et initiativ som skal bygge nettverks- og

kompetansemiljøer for å styrke samarbeid på tvers av sektorer. Ambisjonen er å redusere tekniske, juridiske og forretningsmessige barrierer knyttet til deling og gjenbruk av data. [Gaia-X](#) er et europeisk initiativ med representanter fra næringsliv, politikk og forskning fra Europa og verden for øvrig som samarbeider om å bygge en distribuert og sikker datainfrastruktur. Denne infrastrukturen gir bedrifter og innbyggere mulighet for å utveksle og dele data og samtidig beholde kontroll på egne data.

2. Hva er behovet framover?

For å lykkes med FAIR data må tilbudet av løsninger langs hele livsløpet til dataene forbedres. Det trengs mer brukervennlig støtte til planlegging av datahåndtering og bedre støtte i prosjektfasen for å gjøre dataene klare for å bli FAIR helt fra starten. Ikke minst trengs det gode løsninger for å gjøre arkiverte data enklere tilgjengelig for både mennesker og maskiner. For at data faktisk skal tas i bruk, trengs også gode søketjenester som setter mulige brukere i stand til å oppdage hva som finnes av tilgjengelige data fra både forskning og forvaltning. På dette området er det mye godt arbeid i gang på europeisk nivå som det er viktig at Norge deltar i og støtter opp om.

Det er investert i mange og varierte infrastrukturer i Norge i de senere årene, både gjennom INFRASTRUKTUR-ordningen i Forskningsrådet, gjennom bevilgninger direkte fra sektordepartementene og gjennom lokale initiativ ved forskningsinstitusjonene. Internasjonalt er det investert i forskningsnett og datainfrastrukturer som norske forskningsmiljøer får tilgang til gjennom samarbeid eller medlemskap. I tillegg har mange norske forskningsinstitusjoner gjort mye for å legge til rette for FAIR-data for sine forskere ved at de har laget tydelige retningslinjer for datahåndtering, og gjennom satsing på styrket forskningsstøtte, lokale datatjenester og kompetanseheving.

Det er også viktig å påpeke at det finnes mye god datainfrastruktur i forvaltningen som legger til rette for bruk av forvaltningsdata i forskning. Forvaltningsdata er innsamlet av offentlig sektor og finnes i registre over helse, skatt og arbeidsmarked, Brønnpøysundregistrene, Nav og Arkivverket for å nevne noe. I tillegg har ABM-sektoren (ABM: arkiv, bibliotek og museum) en rekke samlinger og databaser som er tilrettelagt for gjenbruk. Det vil bli avgjørende for framtidig utnyttelse av data fra offentlig forvaltning at de etablerte strukturene får gode vilkår for å videreutvikle sine tilbud og tjenester, samtidig som det også legges til rette for å etablere infrastrukturer på områder hvor tilbudet ikke er godt nok i dag. For å utnytte verdipotensialet i dataene bedre vil det kreve et vesentlig arbeid med å klargjøre data med riktig kvalitet og å gjøre dem tilgjengelige for forskere og andre aktører.

Samtidig er det fortsatt en lang vei å gå før vi har på plass infrastrukturer for FAIR-forskningsdata og særlig relevante forvaltningsdata for forskning innenfor alle områder. Av FAIR-utredningen framgår det at det er stor variasjon i organisering av infrastrukturene, og at bare et fåtall av datainfrastrukturene er organisert på en måte som legger til rette for samhandling og dataflyt mellom infrastrukturene. Blant annet opplever flere av forskerne at tilbudet av verktøy for håndtering av forskningsdata er fragmentert, og de etterlyser at forskningsinfrastrukturer og datahåndteringstjenester for de ulike delene av forskningsprosessen henger bedre sammen.

Rapporten fra ekspertgruppen for [datadeling i næringslivet](#) peker på mange av de samme utfordringene som FAIR-utredningen, og det anbefales bedre samordning (spesielt på statlig nivå) og kompetanseheving: "Staten bør påse at etatene bidrar til sektorovergripende samarbeid om dataflyt der dette kan skape samfunnsøkonomisk verdi, uavhengig av om tiltaket gagnar den enkelte etats kjerneoppgaver." De peker også på at "manglende kompetanse og manglende søkelys på behovet for å bearbeide data for å kunne benytte avanserte metoder, er viktige barrierer for datadrevet

verdiskaping i næringsliv og i offentlig sektor". Det blir i den sammenheng også påpekt at manglende tillit mellom aktørene er en barriere for deling av data – og at det derfor er viktig å styrke kompetanse og kapasitet.

Det finnes ingen fullstendig oversikt over hvilke datainfrastrukturer som er i bruk i norsk forskning og forvaltning, og det finnes mange ulike løsninger knyttet til organisering og finansiering. Innenfor noen områder har man kommet langt når det gjelder infrastrukturer og kompetanse og funnet en god ansvarsdeling med hensyn til organisering og finansiering, mens det på andre områder gjenstår mye. For Norge blir det viktig framover å bygge videre på det som fungerer, og finne bærekraftige løsninger på organisering og finansiering av datainfrastrukturer der det i dag mangler. Utvalget kommer nærmere inn på løsninger i del 3 av rapporten.

DEL 2

Ambisjoner for datainfrastrukturer i Norge

I denne delen av rapporten foreslår utvalget ambisjonsnivået Norge bør ha for datainfrastrukturer. Gevinstene som er mulig å nå med et høyt ambisjonsnivå, og hvilken risiko nasjonen løper ved å velge et for lavt ambisjonsnivå, beskrives. Til slutt presenteres hvilke utfall det vil kunne få for forskning og for samfunnet å velge ulike ambisjonsnivåer.

Utvalgets forslag til ambisjonsnivå for datainfrastrukturer i Norge

- Forskere ved norske universiteter og institutter skal ha tilgang til datainfrastrukturer som muliggjør forskning og utdanning i verdensklasse.
- Innen 2030 skal alle fagområder i Norge tilbys kompetanse, veiledning og kuratering av forskningsdata, enten i form av nasjonale løsninger eller helt eller delvis gjennom deltakelse i europeisk eller internasjonalt infrastruktursamarbeid.
- Norge skal på utvalgte områder ha datainfrastrukturer i verdensklasse som blir foretrukket av internasjonale brukere.
- Norge må ha en opptrappingsplan for organisering og finansiering av datainfrastrukturer som gjør det mulig å hente ut gevinstene av de store datamengdene som vil genereres med offentlig finansiering i årene som kommer.

For at det foreslåtte ambisjonsnivået skal kunne realiseres, må forskningsbudsjettet være stort nok til å romme nødvendige investeringer i datainfrastruktur og tilhørende støttefunksjoner.

3. Ambisjonsnivå for bedre utnyttelse av data for forskning og samfunnsutvikling i Norge

Utvalget mener at Norge må ha et ambisjonsnivå som svarer til de overordnede målene i gjeldende [Langtidsplan for forskning](#). Det vil si å styrke Norges konkurransekraft og innovasjonsevne, møte store samfunnsutfordringer og utvikle fagmiljøer av fremragende kvalitet. Muligheten til å samarbeide med og måle seg med de beste miljøene internasjonalt er avgjørende for å utvikle fremragende fagmiljøer. Uten tilgang til data og muligheten til å koble data begrenses mulighetene for nye gjennombrudd i forskningen, og med det mulighetene til å utvikle forskning som hevder seg internasjonalt. Ambisjonsnivået vil ha konsekvenser for hvor mye framtidig datainfrastruktur vil koste, men det vil også påvirke hvilke gevinster det kan gi Norge.

4. Gevinster

Det er mye å vinne på god forvaltning og deling av data. Norge har et godt utgangspunkt for å ligge i front på dette området, med de gevinster det vil gi for norsk forskning, forvaltning og næringsutvikling. Det vil kreve ressurser og innsats å dele data, men gevinsten er økt verdiskaping, flere arbeidsplasser og en effektiv offentlig sektor som kan tilby innbyggerne bedre og mer sammenhengende tjenester. Europakommisjonen har gjennomført studier som viser at den økonomiske verdien av data fra offentlig sektor vil øke fra 52 milliarder euro i 2018 til 194 milliarder

euro i 2030. Det betyr at offentlige data har en stor verdi om de forvaltes og deles på riktig måte og brukes til verdiøkende tjenester og produkter.

Data for bedre folkehelse

HUNT har vokst fra det som startet som én enkelt medisinsk studie, til å bli forskningscenter og kompetansemiljø som inkluderer biobank, databank og infrastruktur for sensitive data. Den første Helseundersøkelsen i Nord-Trøndelag, HUNT 1, ble gjennomført i 1984–1986. Siden har over 240 000 trøndere deltatt i undersøkelsene. HUNT forskningscenter har en unik samling av helsedata og biologisk materiale.

Langsiktig forvaltning i forskningscenteret legger til rette for viktig forskning. Hvert år publiseres over ett hundre forskningsartikler basert på materialet, både fra interne og eksterne forskere. Det omfattende datasettet bidrar til innsikt og ny kunnskap om folkehelse og sykdom, genetikk, livsstil og miljøpåvirkning.

4.1 Gevinster for forskningen

En åpenbar gevinst med FAIR-forskningsdata er at de kan gjøres tilgjengelige for gjenbruk av andre forskere, uten at de som gjenbraker dem, unødige trenger å bruke tid på ny datainnsamling. Dette er effektiviserende og frigjør tid til blant annet dataanalyse og annen forskningsaktivitet. I tråd med prinsippene for åpen forskning generelt vil FAIR-forskningsdata også ha gode metadata og metodebeskrivelser, som tilrettelegger for reproduksjon og etterprøvnbarhet av dataene. På denne måten blir det mulig å rette opp feil, utfylle eksisterende data, forbedre metoder og forklare at forskjellige studier gir ulike resultater. Tilgjengelige forskningsdata vil med andre ord både effektivisere forskning og bidra til å styrke tilliten til forskningen. Også kvalitativ forskning, som ikke nødvendigvis er hypotesedrevet, vil ha nytte av å kunne vise til strukturert og tilgjengelig datagrunnlag. Dette kan gjelde humaniora og samfunnsvitenskapelige fag, hvor deling av blant annet kvalitative data kan bidra til å bygge opp under resonnementer og konklusjoner.

Innebygd personvern for effektivitet og lavere kostnader

Gjennom microdata.no får forskere rimelig og umiddelbar tilgang til deler av SSBs datagrunnlag uten å søke og uten å knytte bruken til definerte prosjekter. De kan bruke tjenesten innovativt og eksplorativt for å utvikle et prosjekt og som arbeidsbank og datalager når de gjennomfører prosjektet.

Metadata er en integrert del av plattformen, og forskerne kan fritt dele datasett og analyseopplegg. Det pågående prosjektet [National Microdata Platform for Norwegian and International Research and Analysis](https://microdata.no) videreutvikler microdata.no til en generisk distribusjonsplattform for data fra en hvilken som helt dataeier. Kreftregisteret er pilotpartner i prosjektet. Dette er mulig fordi microdata.no har innebygd personvern. Brukerne arbeider metadatatrevet uten å se individdata, og dataene forlater aldri SSB. Alle resultater går gjennom et sett av konfidensialitetsfiltre som gjør dem anonyme. Eventuelle forsøk på å omgå konfidensialiseringen blir stoppet av det innebygde personvernet.

Fafo-rapporten [Seniorer i skolen \(2022:05\)](https://www.fafo.no/rapporter/seniorer-i-skolen-2022-05) undersøker alder, kjønn, arbeidstid, skoleslag og pensjonsalder for lærere i perioden 2010–2019. Dataene er hentet fra microdata.no. SSB har beregnet at et tilsvarende tradisjonelt datautlån ville kostet rundt 30 000–40 000 kroner og hatt en produksjonstid på 2–4 måneder, avhengig av pågang. I tillegg kommer forskernes kostnader og tidsbruk for å sette opp en søknad.

Det vil være store muligheter i gjenbruk av data fra offentlig forvaltning i forskning på områder som befolkning, helse, transport, naturressurser, klima og språk m.m. Norske registerdata er for eksempel i særklasse, og det ligger mye uutnyttet potensial i bruk av disse i forskning. Bedre tilrettelegging for gjenbruk av forvaltningsdata i forskning vil gi kunnskap om komplekse samfunnsforhold som går på tvers av fag og tema. Resultater fra forskningen vil i sin tur gi verdifull innsikt og kunnskap som kan gi bedre og mer kunnskapsbasert grunnlag for utvikling og beslutninger i offentlig forvaltning.

4.2 Gevinster for forvaltningen, samfunnet og befolkningen

Muligheten til å hente inn og bearbeide data fra ulike kilder gir et bedre beslutningsgrunnlag og sikrere prediksjoner innenfor viktige forvaltningsområder. Data som offentlig sektor samler inn selv,

eller som brukes som grunnlag for analyse i forskning, gir et viktig kunnskapsgrunnlag for planlegging og beslutninger som har innvirkning på samfunnsutviklingen på viktige områder. Med FAIR-data blir også sektorgrensene mindre skarpe, og det blir mulig å tilby mer helhetlige og brukervennlige tjenester til innbyggerne, jf. overordnet strategi for departementsfellesskapet [Gode hver for oss. Best sammen 2021–2025](#).

Det er viktig at forskningsresultater brukes i samfunnet, som grunnlag for mediasaker, tjenesteutvikling, politikkutvikling eller næringsvirksomhet. Åpenhet om datagrunnlaget for resultatene er nødvendig for at folk skal ha tillit til forskningen, og for å kunne diskutere på tvers av interesser og verdier. Indirekte bidrar dette til bedre forskning og dermed gode beslutningsgrunnlag.

Nyttige datatjenester for folk flest

- [yr.no](#) – en datatjeneste fra Meteorologisk institutt og NRK
- [Artsdatabanken – Kunnskapsbank for naturmangfold](#)
- [Ordbøker og andre språkressurser gjennom CLARINO](#)
- [Biblioteker, museer og arkiver](#)
- [Helsenorge.no](#)
- [Statistikk, forskning og analyse fra SSB](#)
- [Leksikon, slik som \[Store norske leksikon\]\(#\)](#)

Et godt dokumentert kunnskapsgrunnlag basert på systematisk bruk av forskningsdata på områder som helse, økonomi, arbeidsmarked, vær, klima, trafikk og luftkvalitet gjør at folk kan vurdere egen risiko og ta informerte valg. På den måten bidrar det til et åpent og opplyst demokrati.

4.3 Gevinster for næringslivet

Offentlige data gir muligheter for innovasjon og nyskaping i næringslivet, blant annet til utvikling av nye tjenester, produkter og forretningsmodeller. Samarbeid om infrastruktur og bruk av data effektiviserer verdikjeder i næringslivet. Det gir gevinst gjennom å redusere feil og fjerne arbeidsoppgaver og mellomledd. Store og helst åpne databaser er en forutsetning for å bruke kunstig intelligens som gir muligheter for avanserte analyser. Slike åpne databaser gir også store muligheter for å skape verdier for blant annet oppstartsbedrifter. I en stadig mer datadrevet økonomi vil smart sammenstilling og analyse av store datamengder kunne forbedre økonomisk konkurransevne og

Fra jorda til skyene

DigiFarm er en norsk oppstartsbedrift som hjelper bønder og landbruksbedrifter med å øke avlingene sine og redusere kostnader til gjødsel og plantevern. DigiFarm har ved hjelp av dype nevralt nettverk (AI) basert på høyoppløste satellittdata utviklet en helautomatisk og standardisert metode for å oppdage og oppdatere jordsmonnsgrensene og sådd areal for hver vekstsesong. Dette kan erstatte manuelle, unøyaktige og tidkrevende verktøy som bøndene bruker i dag for blant annet tilskuddssøknader, jordprøver og gjødselplaner. På myndighetsnivå automatiseres tidligere manuelle kontroller av produksjonstilskuddssøknadene.

innovasjon. Dette understøttes i [Meld. St. 22 \(2020–2021\) Data som ressurs. Datadrevet økonomi og innovasjon](#), som framhever gevinstene som ligger i å dele data for å blant annet styrke næringskonkurransekraft nasjonalt og internasjonalt.

4.4 Internasjonalt samarbeid er avgjørende for å ta ut gevinstene

Gjennom en rekke strategiske investeringsinitiativer og nye regelverk har EU-kommisjonen siden 2014 jobbet målrettet med å legge til rette for mer balansert og målrettet flyt og bruk av data på tvers av sektorer. Norges deltakelse i dette forsknings- og datasamarbeidet gir gevinster i form av enklere tilgang til data fra andre bidragsytere, bedre analytiske resultater og nye digitale tjenester. Initiativer som har fremmet utviklingen av en robust europeisk dataøkonomi, er for eksempel forordningen om [fri flyt av ikke-personlige data](#), [Cybersecurity Act](#) og [Open Data-direktivet](#). Nylig la kommisjonen fram et nytt forslag til forskrift om harmoniserte regler for rettferdig tilgang til og bruk av data ([Data Act](#)). Forskriften inneholder et rammeverk for tverrsektoriell forvaltning av data som

skal sikre at brukere beholder tilgang til egne data, noe som fremmer ikke-personlig datadeling og sikkerhetstiltak for å hindre tredjeland i å få tilgang til EU-data i fravær av en internasjonal avtale.

Selv om Norge ikke er med i EU, har vi tilgang til både EUs datamarked og europeiske finansieringsinitiativer som Horisont Europa og Digital Europe Programme ([DIGITAL](#)). Her kan Norge samkjøre nasjonale investeringer i datainfrastrukturer med europeiske investeringer for utvikling av tverrsektoriell datainfrastruktur. Deltakelse i internasjonale datainfrastrukturer gir norske aktører muligheten til å realisere gevinstene av at det er forbedret flyt og bruk av data på tvers av sektorer i Norge og andre europeiske land. Deltakelse i internasjonale datainfrastrukturer er en forutsetning for nyskapende forskning av høy internasjonal kvalitet. EU har også satset store summer på analyse- og beregningskapasitet for data gjennom [EuroHPC](#).

Kunstig intelligens bidrar til verdensledende forskning

Å forstå proteinfunksjon er essensielt for alt liv. Et proteins form avgjør dets funksjon. Eksperimenter for å finne strukturen til proteiner har tidligere vært svært vanskelige å utføre og dermed også kostbare. Det er ikke uvanlig at de avanserte metodene tar ett år å gjennomføre, og det kan normalt koste opptil én millioner kroner å finne strukturen eksperimentelt.

I 2020 skjedde noe helt spesielt innenfor dette feltet. Firmaet DeepMind utviklet AI-algoritmen AlphaFold, som basert på data fra Protein Data Bank raskt predikerer formen på helt nye proteiner. Etter kort tid hadde algoritmen kartlagt hele det menneskelige proteom (alle proteiner i menneskekroppen), og art for art blir nå kartlagt.

I 2021 ble AlphaFold kåret til "Method of the Year" av tidsskriftet Nature. DeepMind åpnet algoritmen, og på initiativ fra NORA, USIT ved UiO og Uninett Sigma2 ble algoritmene raskt installert på norsk infrastruktur og tilgjengelig for norske forskningsmiljøer. Programmet er tilgjengelig for hele forskningscommunityet og kan blant annet benyttes til å utvikle medisiner.

5. Kostnader ved å investere i infrastruktur for FAIR forskningsdata og forvaltningsdata

Utvalget beskriver her kostnadene som må beregnes inn i en datainfrastruktur som skal tilrettelegge for FAIR data. I slutten av kapittelet vurderes de økonomiske konsekvensene og gevinstene av å ha et høyt ambisjonsnivå. Utvalget har innenfor rammen av dette arbeidet ikke hatt mulighet til å frambringe pålitelige tall for kostnader som blir brukt på datainfrastrukturer i Norge i dag.

5.1 Kostnader til e-infrastruktur som er en forutsetning for databasert forskning og forvaltning

E-infrastruktur er IKT-baserte infrastrukturer som er viktige for datadrevet forskning, for eksempel tungregning og lagring av stordata underveis i forskningsprosessen. Slike infrastrukturer er ikke direkte relatert til FAIR-deling av data, men er like fullt en forutsetning for forskning, analyser og datautvinning, det vil si gjennomgang av store, eksisterende databaser for å generere ny informasjon. Denne typen infrastrukturer konkurrerer om de samme midlene som andre typer datainfrastrukturer. Det er uheldig fordi de binder opp store deler av de konkurranseutsatte midlene. Utvalget støtter derfor Forskningsrådets anbefaling til KD om at slik grunnleggende datainfrastruktur bør finansieres direkte og utenfor konkurranse. Anbefalingen fra evalueringen av INFRASTRUKTUR-ordningen konkluderer også med at dette er grunnleggende infrastrukturer som ikke bør konkurranseutsettes. Vi kommer nærmere tilbake til dette i kapittel del 4.

Forskningsinfrastrukturer som krever regelmessige, store investeringer og påvirker hele forskningssystemet, bør finansieres sentralt og ikke gjennom konkurranseutsatte INFRASTRUKTUR-

midler. Sentral finansiering vil sikre kontinuitet og begrense vanskelige prioriteringer mellom disse infrastrukturene og mer spesialiserte infrastrukturer (2021:8).

For e-infrastrukturer som tilbyr datatjenester for tungregning og lagring av store datamengder, er det høye kostnader til utstyr. Investeringsbehovet vises blant annet til i rapporten [Behov og nasjonal finansieringsstrategi for nasjonal e-infrastruktur for forskning for perioden 2020–2030](#) fra 2019, og utvalget støtter disse anbefalingene. Investeringsbehovet øker parallelt med etterspørselen etter tjenestene knyttet til lagring av data og tungregning. Det viser utviklingen i søknader til Forskningsrådet og anslått budsjetttramme til oppgradering av utstyr til Sigma2 som ligger på om lag 200 millioner kroner per år. Dette illustrerer behovet for å øke den økonomiske rammen for finansiering betraktelig. Det er viktig både for å videreutvikle teknologi og for å møte behovene som ulike fagmiljøer og sektorer har.

5.2 Kostnader til datainfrastruktur for FAIR data

Kostnadselementene i en datainfrastruktur er tjenester som håndterer data i hele forskningslivssyklusen – før, under og etter et forskningsprosjekt – og som påser at forskningen gjennomføres teknisk, metodisk, juridisk og etisk forsvarlig og holder høy kvalitet. Det er i tillegg nødvendig å finansiere arbeid som må pågå utover de enkelte forskningsprosjektene, for eksempel for å legge til rette for utvikling og vedlikehold av standarder og for forvaltning og langtidsbevaring av data etter prosjekters avslutning.

Store verdier går tapt ved å ikke dele data

Manglende reproduserbarhet har fått mye oppmerksomhet i helseforskningen. Studier gjort i USA av preklinisk forskning har for eksempel konkludert med at rundt halvparten av studiene ikke er reproduserbare. Av dette skyldes rundt en fjerdedel manglende rapportering av og tilgang til relevante forskningsdata. Konsekvensen er store økonomiske tap ved at forskningen har liten verdi for både samfunnet og annen forskning. Manglende reproduserbarhet skaper også problemer for kvalitetssikring av forskningen, og man risikerer i verste fall at resultater fra forskning som ikke er gyldig, tas i bruk.

Det vil som tidligere nevnt bli større behov for kompetent personell for å sikre oppdatert datainfrastruktur på lengre sikt. Blant annet må det påregnes økte kostnader knyttet til et kompetanseløft blant studenter og forskere gjennom grunnutdanning og etter- og videreutdanning og forskeropplæring. Dette er kostnader som utvalget mener

forskningsinstitusjonene i hovedsak selv må ta ansvaret for å dekke, og som må integreres i regulær forskerutdanning. Like fullt er det viktig å ta hensyn til dette i beregningen av kostnader. Dette understrekes også i FAIR-utredningen:

Utredningen har avdekket at manglende kompetanse om og kultur for deling og gjenbruk av forskningsdata sannsynligvis er det største hinderet for å komme videre. For å dekke behovet for kompetanseheving og veiledning på en effektiv måte for alle forskere i Norge, vil det være nødvendig med en betydelig ressursinnsats og god koordinering av kompetansehevingstilbud mellom lokale, institusjonelle og nasjonale miljøer. Generelt må kompetansen på alle nivåer i forskningssystemet heves, og tilgjengelig kapasitet for veiledning og operativ datahåndtering må økes betydelig (2022:4).

FAIR-utredningen dokumenterer at det er stor variasjon med hensyn til hvor langt norske forskningsinstitusjoner har kommet i arbeidet med å dele forskningsdata. Vetenskapsrådet i Sverige påpeker mye av det samme i [tilstandsrapporten om FAIR-data](#) og gir følgende anbefalinger om behovet for investeringer:

- En særskild finansiell satsning gjøres for å øke det infrastrukturelle stødet til det nasjonale arbeidet med åpen tilgang til forskningsdata ved lærosäten, myndigheter og infrastrukturer.
- Vetenskapsrådet ges særskilte medel til å støtte aktiviteter som fremjar åpen tilgang til forskningsdata nasjonalt og internasjonalt, til eksempel gjennom deltagande i det europeiske öppna forskningsmolnet.

Kostnader til internasjonal deltagelse for å delta i utviklingen av tilrettelegging av datainfrastrukturer er viktig å ta hensyn til. Norges partnerskap i EOSC innebærer en forpliktelse om å bidra med 500 millioner euro i in-kind-midler for perioden 2021–2030. Dette kan være bidrag til aktiviteter finansiert av kommisjonen eller til [aktiviteter](#) som er i tråd med den [strategiske agendaen](#) til EOSC.

Enklere innsikt i overvåkingsdata

Det nasjonale bakkesegmentet ([satellittdata.no](#)) er en nasjonal data-hub tilrettelagt for å møte norske brukeres behov for tilgang til data fra Copernicus Sentinel-satellittene. Dataene spenner fra overvåking av skipstrafikk, oljesøl, snøkartlegging, værvarsling, flomovervåking, deteksjon av snøskred og overvåking av barkbilleangrep på skog, for å nevne noen. Infrastrukturen tilbyr også verktøy for dataprosessering og tilgang til dataprodukter. Denne finansieres av Norsk Romsenter og driftes og utvikles av Meteorologisk institutt.

Den norske noden utgjør en del av et europeisk nettverk som raskt sprer data fra Sentinel-satellittene til aktører fra statlig forvaltning, næringslivet og forskningen. For noen av satellittene er data tilgjengelig bare noen titalls minutter etter passering. Dette gjør det mulig mye tidligere å sette inn tiltak som kan avverge eller avgrense katastrofer.

Behovet for tjenester for FAIR-forskningsdata blir stadig større. Hva behovet tilsvarer i kroner og øre, er vanskelig å tallfeste, men en interessant kilde til informasjon er det økende søknadsvolumet for slike infrastrukturer til blant annet INFRASTRUKTUR-ordningen i Forskningsrådet. Siden etableringen av ordningen i 2009 har andelen som søker om støtte til datainfrastrukturer, som helt eller delvis gjelder verktøy for datahåndtering, økt. Til den siste utlysningen i 2020 søkte om lag 40 prosent av prosjektene om midler til datahåndtering (3,3–3,7 milliarder kroner av totalt omsøkt beløp på 9,3 milliarder kroner). Dette viser at en stor andel av indikert behov for finansiering av datainfrastrukturer ikke blir dekket gjennom de konkurranseutsatte midlene.

Institusjonene har lagt en betydelig egeninnsats i utvikling og drift av datainfrastrukturene. En gjennomgang av bevilgede datainfrastrukturprosjekter gjennom INFRASTRUKTUR-ordningen viser at ordningen har gitt støtte til 66 prosent av de totale prosjektkostnadene, og at institusjonene selv i stor grad står for resten (se tabell).

Tabell 1 Oversikt over investeringer gjennom Forskningsrådets INFRASTRUKTUR-ordning i perioden 2009–2018.

Datainfrastrukturprosjekter per område	Sum totalt (mill. kr)	Kontraktsum med F.rådet (mill. kr)
Naturvitenskap	942,6	609,3
Helse- og persondata	532,1	488,8
Humaniora	316,9	214,4
Samfunnsvitenskap	145,5	107,3
Generiske	1321,0	784,3*
SUM	3258,1	2204,1

*Inkluderer 237 millioner kroner bevilget over post 53 til Sigma2 i 2019

Kartleggingen fra FAIR-utredningen, evalueringen av INFRASTRUKTUR-ordningen og vurderinger fra Vetenskapsrådet viser at det vil være ressurskrevende å videreutvikle og drifte datainfrastrukturer med tjenester som forskere og andre brukere trenger. I motsetning til investering i utstyr, der hoveddelen av kostnadene kommer på gitte tidspunkter i forbindelse med innkjøp og oppgradering, vil etablering av en datainfrastruktur binde opp betydelige personellressurser i hele infrastrukturens levetid. Det er med andre ord viktig å ha et langsiktig perspektiv i beregningen av kostnadene i en datainfrastruktur.

Som tidligere nevnt anslår estimater fra EU-kommisjonen at det er nødvendig å bruke 5 prosent av total nasjonal forskningsbevilgning til datahåndtering. For Norge vil det utgjøre om lag 2 milliarder kroner dersom det tas utgangspunkt i [NIFUs statsbudsjettanalyse for 2021](#). Her anslås de samlede bevilgningene i Prop. 1 S (2020–2021) til forskning og utviklingsarbeid (FoU) til 40,9 milliarder kroner. Anbefalingen om at 5 prosent av nasjonal FoU bør investeres i forskningsdata, bygger på en [beregning PwC EU Services foretok](#) i 2018 på oppdrag fra DG-Research and Innovation. PwC estimerte det faktiske tapet ved manglende FAIR-tilpasning i Europa til mellom 3,4 prosent og 8,7 prosent av samlet FoU-investering.

6. Trinnvis realisering av FAIR data

Som vist i det foregående er det kostbart å gjøre data FAIR, både fordi det innebærer spesialisert kompetanse og investering i datainfrastruktur. Når dette sees i sammenheng med de store mengdene data som samles inn, må det nødvendigvis gjøres en vurdering og prioritering av hvilke data som skal eller kan gjøres FAIR. Etter hvert vil forhåpentligvis gode infrastrukturer med stor kapasitet være på plass innenfor alle forskningsområder, og det som i dag ansees som spesialistkompetanse, vil være en mer grunnleggende kompetanse i alle ledd. Da vil muligens behovet for vurdering og prioritering bli mindre, og alle data kan i prinsippet gjøres FAIR. Før vi eventuelt kommer dit, er det fornuftig å tenke en trinnvis realisering av FAIR-data.

Noen data er enklere enn andre å gjøre FAIR fordi de krever mindre tilpasning for å kunne deles og gjenbrukes, mens andre krever mye bearbeiding og/eller beskyttelse. Noen fagområder har i tillegg kommet langt i arbeidet med å gjøre data FAIR ved for eksempel standardiserte metadata, formater, ontologier og prosessering, mens andre bare er i startgropen. I stedet for å ha som ambisjon at alle data skal bli FAIR på én gang, kan det være fornuftig å legge til rette for at de "lavhengende fruktene" høstes først, og at man deretter legger vekt på de mer kompliserte datasettene og fagområdene som har kommet kortere. På denne måten vil man kunne oppnå stor overføringsverdi både når det gjelder tekniske løsninger og kompetanse.

Bedre helseeffekter i befolkningen ved bedre datautnyttelse

Samfunnskostnadene knyttet til sykdom i form av tapte kvalitetsjusterte leveår ble i 2013 beregnet av Helsedirektoratet til å være over 1400 milliarder kroner. En reduksjon i tapte kvalitetsjusterte leveår på 1 prosent som følge av bedre bruk av helsedata gir en samfunnsøkonomisk effekt på 14 milliarder kroner. Bedre bruk av helsedata kunne føre til mer effektiv drift av helsevesenet, som utgjør en vesentlig del av offentlige utgifter. Realisering av gevinstene vil imidlertid kreve en helt annen innovasjonstakt enn vi har hatt fram til i dag. Det krever etablering av et sterkt helseøkosystem: Man er avhengig av å jobbe tett med både akademia, helsenæringen og IKT-næringen dersom man skal lykkes (Menon Economics, 2018:3).

Videre har noen data potensielt stor verdi for videre utnyttelse, mens andre antakeligvis har begrenset verdi utover forskningen de ligger til grunn for. Data med potensielt stor verdi bør da prioriteres før data med antatt begrenset verdi. Det kan være vanskelig å vurdere og forutse framtidig verdi av data fordi verdiene kan tas ut i helt andre fagområder, bransjer eller samfunnsområder enn dem de ble innhentet i. Et hovedprinsipp bør likevel være at forskerfelleskapene selv gjør

denne vurderingen, og da helst i dialog med andre relevante fagområder og andre aktører som kan være aktuelle brukere.

En annen måte å realisere FAIR trinnvis på er å stille ulike krav til bokstavene i FAIR-prinsippene. Det er ofte mye mer ressurskrevende å gjøre data samhandlende/interoperable og gjenbrukbare (I og R) enn å gjøre dem gjenfinnbare og tilgjengelige (F og A). Bare det at dataene er gjenfinnbare, kan gi stor gevinst. Man kan derfor i første omgang legge til rette for at alle datasett skal være gjenfinnbare gjennom maskinlesbare datahåndteringsplaner og registrering i datakataloger og registre, slik at potensielle brukere vet hvor dataene finnes, og hvordan de kan få tilgang til dem. Så kan man prioritere "lavhengende frukter" og datasett med potensielt stor verdi ved gjenbruk for I og R i første omgang.

Uansett hvilken måte man velger å realisere FAIR-data trinnvis, forutsetter det koordinert nasjonal og internasjonal innsats, teknisk tilrettelegging og standardisering samt kompetanseheving. Utvalget kommer nærmere inn på anbefalinger om hvordan dette kan løses, i del 3 av denne rapporten.

7. Vurdering av ambisjoner og risiko

Utvalget mener nivået på nasjonale investeringer i infrastruktur for data må være i samsvar med forventningene som EU-kommisjonen og Horisont Europa har om krav til åpen forskning. Et viktig hensyn for investeringene er at det skal gi forskere ved norske universiteter og institutter tilgang til

Raske svar i kriser

Da verden tidlig i 2020 ble konfrontert med et nytt og farlig virus, tok det ikke lang tid før forskerne kom til enighet om hvilken type virus det var snakk om. Dette var mulig takket være førti år med kartlegging, lagring og tilgjengeliggjøring av ulik biologisk informasjon om mange typer virus. Hadde man ikke hatt slike databaser, ville trolig langt flere menneskeliv gått tapt i de første månedene av koronapandemien.

I Europa utvikles og vedlikeholdes alle disse databasene av ELIXIR, den europeiske ESFRI-infrastrukturen for livsvitenskapelige data. Da pandemien var et faktum, kunne man i Europa slå mynt på de årelange investeringene i ELIXIRs åpne databaser og raskt opprette en covid-19-dataplattform. Ved hjelp av denne ekstraordinære innsatsen kan man i dag overvåke mutasjoner og spredning av ulike stammer av koronaviruset, noe som har vært og er helt avgjørende for at samfunnet skal kunne håndtere pandemien videre.

infrastruktur som gjør det mulig med forskning og utdanning i verdensklasse. På samme måte mener utvalget at brukere i offentlig og privat sektor skal ha tilgang til datainfrastruktur som gir enkel tilgang til forskningsdata av høy kvalitet i planlegging og beslutninger om samfunns- og næringslivsutvikling der det er behov for det. Fordi gevinstene i mange tilfeller vil bli realisert i andre deler av økosystemet enn der innsamling og tilrettelegging for gjenbruk av dataene skjer, er det helt avgjørende at kostnadene til dataforvaltningen blir dekket. Utvalget kommer nærmere tilbake til dette i del 3.

Norges konkurransefortrinn når det gjelder digitalisering og dataregistre med mulighet til å koble data fra ulike kilder, står i fare for å forsvinne dersom de teknologiske mulighetene ikke utnyttes. Langsiktig forvaltning av data krever både datafaglig kompetanse og kapasitet og tilsvarende ressurser til oppgradering og modernisering av teknisk datainfrastruktur og datainnhold. Dersom slike kostnader ikke dekkes, vil oppgradering og utvikling hindres, og risikoen er stor for at løsningsene går ut på dato, at infrastruktur forvirrer, og at de langsiktige investeringene går tapt. Oppdatering og videreutvikling av systemer er en forutsetning for at norske og internasjonale forskningsmiljøer skal ha mulighet til å utnytte datasamlinger til nytte for samfunnet. Risikoen ved ikke å investere i framtidsrettet datainfrastruktur er at norske forskningsmiljøer ikke får delta i kunnskapsutviklingen fra internasjonalt samarbeid, og at Norge blir hengende etter som kunnskapsnasjon. På alle

samfunnsområder er det viktig å ha et pålitelig datagrunnlag når vurderinger og beslutninger som har konsekvenser for samfunnsutviklingen, skal tas. I det siste har vi også sett eksempler på hvor viktig det er å raskt få tilgang til data for avgjørelser som påvirker folks liv og helse.

- Konsekvensene av et for lavt ambisjonsnivå vil være at norske forskningsmiljøer risikerer å miste muligheten til å delta i internasjonal forskningsfront dersom det ikke investeres i gode og framtidrettede infrastruktur for FAIR-forskningsdata og forvaltningsdata med særlig relevans for forskning.
- Det risikeres at gevinstene i datainfrastrukturtenestene ikke kan høstes fullt ut dersom kostnader til personell og kompetanseheving ikke tas hensyn til.

EUs publikasjon [*Cost-benefit analysis for FAIR research data: Cost of not having FAIR research data 2019*](#) viser også hvilke verdier man risikerer ved å ikke gjøre data FAIR:

FAIR data has a direct and quantifiable benefit on research and the knowledge economy. A recent study by the European Commission and PwC noted the cost of not making Europe's research data FAIR to over €10 billion per year.

8. Friske midler eller investering innenfor dagens rammer?

Som vist over medfører investering i etablering og drift av datainfrastrukturer store kostnader, men vil på sikt gi store gevinster. Investeringene bør derfor vurderes i et kost–nytte-perspektiv. Nyttens vil ofte være stor, og gevinsten på kort og lang sikt ved å gjøre data FAIR må sannsynliggjøres når investeringer skal foretas. Illustrasjonen under (tabell over scenarioer på side 27) viser gevinster og risikoer ved å tilføre friske midler og øke investeringene (scenario 1), ved å omprioritere innenfor dagens rammer (scenario 2) og ved å verken tilføre friske midler eller omprioritere (scenario 3). Kostnadene som må dekkes for å gjøre data FAIR, krever en omprioritering av midler innenfor nåværende økonomiske ramme eller en økonomisk ramme som ligger høyere enn dagens nivå.

Risikoen ved ikke å investere tilstrekkelig er at kostnadene påløper, og at gevinstene uteblir. Dersom Norge skal fortsette å delta i internasjonalt forskningssamarbeid, må vi også være kvalifiserte for dette og prioritere midler deretter. Det er nødvendig med større investeringer i datainfrastrukturer i Norge. Uten å satse på gode rammebetingelser for datadeling og bruk går vi glipp av kunnskaps- og teknologiutvikling på områder som er viktige for å utvikle løsninger for samfunns- og næringsliv. På noen utvalgte områder bør Norge også ha som ambisjon å drive utviklingen.

Utvalgets vurdering

Utvalget mener at det bare er scenario 1 som vil innfri ambisjonsnivået som foreslås innledningsvis i denne delen. Scenario 2 vil medføre risiko for at andre deler av forskningssektoren må nedprioriteres, eventuelt organiseres på andre måter som kan forringe kvaliteten på forskingen. Scenario 3 er, slik utvalget ser det, ikke et alternativ dersom Norge skal ha høye ambisjoner på norsk forsknings vegne.

Tabell over scenarier

Scenarier	Investeringsmuligheter	Forskerårsverk	Forskningsaktivitet	Kvalitet i forskningen	Verdiskaping utenfor forskningssektor
Scenario 1: Friske midler til datainfrastruktur	Norge får mulighet til å investere i nødvendig framtidrettet datainfrastruktur.	Investering i datainfrastruktur påvirker ikke andre deler av forskningsbudsjettet, og antall forskere kan forbli som i dag.	Framtidrettet datainfrastruktur vil gi forskere tilgang til en helt annen mengde data og på en helt annen måte enn i dag og gi mulighet for å øke forskningsaktiviteten.	Norske forskere får tilgang til datainfrastruktur som gjør at de kan ta del i den datadrevne forskningen og dermed være en del av forskningsfronten.	Gode datainfrastrukturer vil gi andre sektorer og næringsliv store muligheter til å skape verdier basert på forskningsdata.
Scenario 2: Omprioritering – mer av forskningsbudsjettet settes av til datainfrastruktur	Norge får mulighet til å investere i nødvendig framtidrettet datainfrastruktur.	Mer penger til investering i datainfrastruktur vil gå på bekostning av andre deler av forskningsbudsjettet, og antall forskere må kanskje reduseres.	Framtidrettet datainfrastruktur gir forskerne tilgang til data på en helt annen måte. Dette vil øke forskningsaktiviteten på viktige områder. Færre forskerårsverk og mindre ressurser vil imidlertid gi mindre forskningsaktivitet på andre områder.	Norske forskere får tilgang til datainfrastruktur som gjør at de kan ta del i den datadrevne forskningen og dermed være en del av forskningsfronten – men reduksjon i andre deler av forskningsbudsjettet kan likevel føre til lavere kvalitet.	Gode datainfrastrukturer vil gi andre sektorer og næringsliv mulighet til å skape verdier basert på forskningsdata – men lavere kvalitet kan føre til mindre verdiskaping.
Scenario 3: Ingen omprioritering – potten til datainfrastruktur økes ikke	Norge får ikke mulighet til å investere i nødvendig framtidrettet datainfrastruktur.	Investeringer i infrastrukturer for datainfrastruktur påvirker ikke andre deler av forskningsbudsjettet, og antall forskere kan forbli som i dag.	Norske forskere vil ikke få tilgang til framtidrettet datainfrastruktur og kan derfor ikke ta i bruk data på nye og mer effektive måter. Dette vil etter hvert føre til mindre forskningsaktivitet sammenlignet med andre land som har tilgjengelig datainfrastruktur.	Norske forskere vil raskt falle bakpå uten tilgang til datainfrastruktur som internasjonale kollegaer har, og vil ikke være i stand til å være en del av forskningsfronten.	Mangel på gode datainfrastrukturer vil gjøre det vanskelig for andre å gjenbruke disse dataene til verdiskaping utenfor sektoren.

Grønne felter: Situasjonen forbedres fra i dag ✕ Gule felter: Det er risiko for forverring av situasjonen ✕ Røde felter: Situasjonen forverres fra i dag

DEL 3

Bærekraftig organisering og finansiering av datainfrastruktur

For at Norge skal få mest mulig ut av investeringer i datainfrastrukturer, må løsningene være langsiktige og bærekraftige. Det betyr at man må tydeliggjøre ansvar, foreta prioriteringer, samordne og koordinere overlappende datainfrastrukturer og etablere gode modeller for finansiering av etablering og drift. I denne delen av rapporten gir utvalget innspill til hvordan dette kan løses, og dette kan sees som et grunnlag for rammeverket for finansiering av datainfrastrukturer som foreslås i del 4.

Utvalgets anbefalinger om finansiering og organisering av datainfrastrukturer

Finansiering og pengestrømmer

- 1 Kunnskapsdepartementet må ta det overordnede koordineringsansvaret for finansiering av nasjonal datainfrastruktur og sette av tilstrekkelige midler til denne budsjettposten.
- 2 Sektordepartementenes forskningsbudsjetter må inkludere både direkte og indirekte kostnader til datainfrastruktur på de områdene hvor de har forsknings- og kunnskapsbehov.

Koordinering og ansvarsdeling

- 3 Forskerfellesskapene må ta et særlig ansvar for å identifisere og kommunisere behov for datainfrastrukturer. I dette ligger også å synliggjøre gevinstpotensialet.
- 4 Forskningsinstitusjonene har selv ansvar for å tilrettelegge for FAIR deling av data der dette er påkrevet. Dette kan skje gjennom tilpassede institusjonelle tjenester eller tilgang til nasjonale eller internasjonale datainfrastrukturer. Statlige institusjoner har et særlig ansvar på dette området.
- 5 Det bør etableres stimuleringsordninger for infrastrukturens samarbeid og kompetanseoverføring mellom relevante miljøer i form av nettverk, møteplasser og/eller sentre.
- 6 Beslutning om etablering og videreutvikling av nasjonale datainfrastrukturer bør gjøres sentralt, enten via Forskningsrådets INFRASTRUKTUR-ordning eller gjennom andre mekanismer som sørger for en vurdering av infrastrukturens innplassering i det nasjonale og internasjonale datainfrastrukturlandskapet.
- 7 Det bør legges til rette for deltakelse i internasjonalt datainfrastrukturens samarbeid for å oppnå effektiv bruk av ressurser og videreutvikling av norske noder og tilhørende kompetansemiljøer.
- 8 Sektordepartementene må ta ansvar for å gjøre relevante forvaltningsdata tilgjengelige for forskning.
- 9 Det må etableres arenaer for prioritering av hvilke datasamlinger og datainfrastrukturer det skal brukes ressurser på å gjøre FAIR. Dette må gjøres i fellesskap mellom forskningsmiljøene, offentlig forvaltning, næringslivet og andre relevante samfunnsaktører.

9. Finansieringskilder og pengestrømmer

KD har ansvaret for et velfungerende forskningssystem, og i dette ligger også ansvaret for infrastrukturer for forskning. Som vist i del 1 vil mer datadrevet og dataintensiv forskning medføre at behovet for datainfrastrukturer vil øke. Som vist i del 2 vil investeringer i FAIR-forskningsdata gi store gevinster for alle deler av samfunnet. Det er utvalgets oppfatning at det er helt nødvendig å se helhetlig på finansieringskilder og pengestrømmer. Gevinstene som vi høster, er et kollektivt gode for Norge, og det er derfor nødvendig at også fellesskapet sikrer nødvendige investeringer. For å oppnå ambisjonene for datainfrastrukturer vil det være nødvendig med en klar ansvarsfordeling også når det gjelder finansiering. I *Veileder for sektoransvaret for forskning* (2017) understrekes det at hvert departement har ansvar for forskning innenfor sine ansvarsområder, og at de gjennom dette skal bidra til gjennomføringen av Norges felles forskningspolitikk. I veilederen står det ikke spesifikt at dette gjelder forskningsinfrastruktur, men utvalget mener dette er gitt i og med at slik infrastruktur i de fleste tilfeller er en forutsetning for gjennomføring av god forskning.

9.1 Kunnskapsdepartementets rolle

Kunnskapsdepartementets overordnede ansvar for å sikre et velfungerende forskningssystem innebærer også et særlig ansvar for å koordinere, samordne og finansiere forskningsinfrastruktur. Midler på post 54 i KDs budsjett skal brukes til "relevant og oppdatert forskningsinfrastruktur av nasjonal, strategisk interesse som understøtter forskning av høy kvalitet for eit innovativt og berekraftig samfunn". Disse midlene finansierer Forskningsrådets INFRASTRUKTUR-ordning i tillegg til norske bidrag i utvalgte internasjonale infrastrukturer og basisbevilgningen til Sigma 2 AS. Bevilgningene fra denne posten har økt jevnt de siste årene for å dekke de store behovene for forskningsinfrastrukturer, og en stadig større andel av posten bevilges til datainfrastrukturer. Det er avgjørende at denne budsjettposten ikke svekkes dersom ambisjonsnivået for gjenbruk og videre bruk av data i Norge skal være mulig å oppnå.

Utvalget anbefaler:

Kunnskapsdepartementet må ta det overordnede koordineringsansvaret for finansiering av nasjonal datainfrastruktur og sette av tilstrekkelige midler til denne budsjettposten.

Rammebevilgningene til universiteter og høyskoler, som også går over Kunnskapsdepartementets budsjett, gir institusjonene et strategisk handlingsrom til å bruke midlene i tråd med mål for sektoren og egne virksomhetsmål. Utvalget mener at det er viktig at KD i styringsdialogen med institusjonene understreker ansvaret de har for å bidra til målet om FAIR deling av data og med det sørge for at forskere får nødvendig støtte til å gjøre data FAIR.

9.2 Øvrige departementers rolle

Departementene har gjennom sitt sektoransvar for forskning ansvar for blant annet å bidra til forskning og kompetansebygging for sektoren, sørge for forskning for politikktutforming og forvaltning og legge til rette for høy vitenskapelig kvalitet og relevans i forskningen. Det er viktig å se dette ansvaret i sammenheng med behovet for å investere i forskningsinfrastruktur som er sektorovergripende.

Mange datainfrastrukturer er i dag finansiert av staten gjennom departementer og direktorater. Dette gjøres når sektordepartementet har behov for etablering og drift av en datainfrastruktur på et bestemt område. Finansieringen foregår som regel enten som forvaltningsoppdrag til et forskningsinstitutt eller som tilskudd til et forvaltningsinstitutt. Ved forvaltningsoppdrag finansierer departementet eller direktoratet utvikling og drift av datainfrastruktur som ligger til grunn for at

forskningsinstituttet skal forvalte data for staten. Infrastrukturer drevet av NILU er et eksempel på slik finansiering i form av forvaltningsoppdrag. Ved tilskudd til forvaltningsinstitutter kommer derimot gjerne bevilgingen til utvikling og drift av infrastruktur sammen med tildelingsbrevet fra det aktuelle departementet. Et eksempel på dette er Havforskningsinstituttet, som får slike tildelinger fra sitt eierdepartement.

Noen sektordepartementer bevilger store summer til forskning og genererer gjennom dette mye data som ofte trenger regnekraft og datainfrastruktur for lagring, tilgjengeliggjøring og arkivering. I noen tilfeller innebærer bevilgingen også midler til datainfrastruktur, som beskrevet over. I andre tilfeller bevilges det kun midler til selve forskningsaktiviteten, for eksempel ved bevilgninger gjennom Forskningsrådet og andre deler av virkemiddelapparatet. I sistnevnte tilfelle blir det ofte en utfordring at sektordepartementene ser på selve forskningsaktiviteten som sitt ansvar, men ikke tar hensyn til konsekvensene dette har for de nasjonale datainfrastrukturene. Dette er utfordrende fordi forskningen binder opp kapasitet ved infrastrukturene uten at infrastrukturene får tilført ekstra midler. Ut fra sektoransvaret for forskning som hvert departement har, er det rimelig å forvente at de bidrar med finansiering av datainfrastruktur også utover de direkte investeringene de gjør i forbindelse med forvaltningsoppdrag og tilskuddsbevilgninger til institutter. På de områdene sektordepartementene finansierer dataintensiv forskning, må de også bidra med midler til datainfrastruktur som kreves for at de skal få dekket kunnskapsbehovene, som ofte overskrider fag- og sektorgrenser.

Det finnes også eksempler på infrastrukturer som ligger i grenseland mellom forskning, forvaltning og formidling, som innenfor ABM-sektoren (arkiv, bibliotek og museum). I slike infrastrukturer er ofte primæraktiviteten formidling, noe som kan gi utfordringer med å legge til rette for god forskningsutnyttelse. Her kan man tenke seg betydelige gevinster ved bedre samhandling med de bibliotekene og museene som er drevet av universitetene.

Utvalget anbefaler:

Sektordepartementenes forskningsbudsjetter må inkludere både direkte og indirekte kostnader til datainfrastruktur på de områdene hvor de har forsknings- og kunnskapsbehov.

10. Ansvarsdeling og koordinering i et uoversiktlig landskap av datainfrastrukturer

Som beskrevet i FAIR-utredningen og i del 1 av denne rapporten er status i dag at det finnes en rekke datainfrastrukturer (og tilhørende tjenester) av ulik karakter og med ulik størrelse og nedslagsfelt. Noen infrastrukturer er lokale, noen er nasjonale, mens andre er internasjonale. Noen datainfrastrukturer er generiske og tilbys for bruk på en rekke fagområder, mens andre er tilpasset bruk innenfor spesielle fagområder. Andre igjen kan ha brede formidlingsoppdrag (som biblioteker og museer). Datainfrastrukturer etableres som oftest for å dekke et eksisterende behov, og slike behov har ofte oppstått på flere steder og innenfor flere fagområder samtidig. Resultatet er til dels overlappende datainfrastrukturer på noen områder. Samtidig er det på andre fagområder mangel på datainfrastrukturer, fordi disse tradisjonelt ikke har håndtert så store mengder data, eller fordi det ikke er insentiver eller kultur for å dele data.

I kjølvannet av at det stilles krav fra myndigheter og finansører om at data skal gjøres FAIR, har det oppstått behov for datainfrastrukturer også innenfor disse områdene. Det finnes ingen fullstendig oversikt over landskapet av datainfrastrukturer som norske miljøer bruker, og det er heller ikke kartlagt i detalj hva behovet vil være framover. Som del av FAIR-utredningen ble det påbegynt en

kartlegging av forskningsinfrastrukturer og tjenester som er i vanlig bruk blant norske forskere i dag. Denne gir god oversikt over generiske datainfrastrukturer, men går ikke inn på de ulike fagområdene. Resultatet av kartleggingen er publisert på nettstedet opendscience.no². Nettstedet videreutvikles løpende av Sikt i samarbeid med Forskningsrådet, UHR og institusjonene.

FAIR-utredningen gir en pekepinn om hva som fungerer, og på hvilke områder det er spesielt nødvendig å følge opp, blant annet at det vil kreve betydelig finansiering å oppnå bedre funksjonalitet for tjenester som krever kapasitet og kompetanse (2022:39). Dette gjelder særlig data som ikke kan deles åpent og må beskyttes.

For å komme fram til kostnadseffektive investeringer som dekker behovene for datainfrastrukturer, er det viktig med en tydelig ansvarsdeling og koordinering. Fordi det finnes et stort mangfold av data, med ulike egenskaper, kompleksitet og håndteringsbehov, mener utvalget at det ikke er fornuftig å foreslå én "superinfrastruktur" som skal dekke alle behov. Det er likevel viktig å fremme en organisering som i større grad enn i dag bidrar til at datainfrastrukturene er samordnet, og at de samlet dekker det totale behovet og støtter tverrfaglig forskning og bedre flyt av data. Det er behov for en bedre arbeidsdeling og koordinering enn tilfellet er i dag, og en oversikt over ansvar og roller som ulike aktører har for å gjøre data FAIR.

Ansvar i denne sammenheng impliserer ofte også et ansvar for finansiering. Dette behøver imidlertid ikke bety at den ansvarlige alene skal bære kostnadene ved etablering og drift. Vi kommer nærmere inn på ulike finansieringsmåter i del 4, hvor vi foreslår et rammeverk for finansiering av datainfrastruktur.

10.1 Forskerfellesskapenes ansvar

Det er forskerfellesskapene selv som vet hvilke behov de har for datainfrastrukturer, både for å kunne drive forskning av høy kvalitet, men også for å innfri krav og retningslinjer om å gjøre data fra forskningen FAIR. For at institusjoner, virkemiddelapparatet og myndighetene skal kunne ta riktige valg for investering i datainfrastruktur, er det viktig at behovene blir identifisert og tydelig kommunisert fra forskerfellesskapene. Som et grunnlag for prioritering mellom investeringer vil det være nyttig at gevinstpotensialet blir synliggjort, og at innspill til beslutningstakere kommer samlet fra miljøer med sammenfallende behov. Dette kan blant annet skje gjennom godt koordinerte søknader om konkurranseutsatte midler.

Utvalget anbefaler:

Forskerfellesskapene må ta et særlig ansvar for å identifisere og kommunisere behov for datainfrastrukturer. I dette ligger også å synliggjøre gevinstpotensialet.

10.2 Institusjonenes ansvar

I løpet av forskningsdataenes livssyklus vil forskeren ha behov for ulike datainfrastrukturer og tilhørende tjenester. Noen av disse datatjenestene må oppfattes som "basis"-datainfrastrukturer, fordi det er tjenester som det må forventes at alle institusjoner tilbyr sine ansatte. I den grad institusjoner ikke tilbyr slike tjenester selv, må det forventes at de inngår avtaler med andre som kan gi et tilbud. Mange forskningsinstitusjoner utvikler egne, mer spesialiserte datainfrastrukturer, og flere av disse tilbys også forskere ved andre institusjoner. I slike tilfeller er det viktig at institusjonene er bevisste på eventuelt overlapp med andre institusjoners tjenester og hva som finnes av nasjonal

² Se artikler på opendscience.no om de ulike typene infrastrukturer og verktøy, [Open Science | Åpen forskning \(open science\)](#)

datainfrastruktur, og ikke unødig duplerer disse. Ledelsen ved forskningsinstitusjonene må også ta ansvar for at de ansatte får tilgang til nasjonal datainfrastruktur. Det innebærer blant annet å sette av penger til eventuell brukerbetaling i forskningsprosjekter som institusjonene finansierer over ordinært budsjett. Det er i tillegg viktig å tenke livstidssyklus når man fatter beslutninger. Mange forskere lokkes i dag til å benytte enkle og gratis tilgjengelige kommersielle tjenester med høyst uklart tidshorison og manglende nasjonal forankring.

Utvalget anbefaler:

Forskningsinstitusjonene har selv ansvar for å tilrettelegge for FAIR deling av data der dette er påkrevet. Dette kan skje gjennom tilpassede institusjonelle tjenester eller tilgang til nasjonale eller internasjonale datainfrastrukturer. Statlige institusjoner har et særlig ansvar på dette området.

10.3 Stimulere til nettverksvirksomhet og kompetansedeling

For at forskere skal ta i bruk datainfrastruktur og tilhørende tjenester på en god måte, er lokal kompetanse en viktig faktor. Forskningsadministrasjon, forskningsbibliotekene, IT-avdelingene og lokal forskerstøtte har en viktig rolle som førstelinje for forskere som er usikre på hvordan de kan og bør håndtere data. Det er imidlertid lite hensiktsmessig og kostnadseffektivt at hver enkelt forskningsinstitusjon skal ha spesialistkompetanse på alle områder i datahåndtering og dataforvaltning. I FAIR-utredningen ble det avdekket at det er behov for bedre tilgang til kunnskap og veiledning innenfor spesialiserte områder som flere av institusjonene ikke har forutsetninger for å tilby. Det kan derfor være fornuftig å tilrettelegge for å dele på ansvaret for ulike kompetanseområder.

Spesielt for mindre institusjoner, som ikke har godt utbygde støttefunksjoner, for eksempel egne bibliotekstjenester, er det relevant at tjenesten hentes utenfor egen institusjon. Her er det muligheter for å oppnå gode effekter av samarbeid mellom større forskningsinstitusjoner og nasjonale aktører gjennom å stimulere til styrking av eksisterende nasjonale og internasjonale nettverk og eventuelt etablere nye. Det kan også være aktuelt å vurdere å etablere kompetansesentre og/eller kompetanse-huber for å utvikle og utnytte kompetansen til å gjøre data FAIR. Der det trengs juridisk, etisk, organisatorisk, semantisk og teknisk kompetanse, kan slike kompetansemiljøer eller kompetansesentre innenfor sentrale domener bidra med fagutvikling, veiledning og råd. Disse må arbeide i nær kontakt med infrastrukturene og forskermiljøene. Samlet vil dette utnytte kompetansen på en god måte, fremme effektiv utnyttelse av data og være kostnadseffektivt. I Nederland lyste for eksempel regjeringen ut midler til utvalgte universitets- og sykehusmiljøer for å bygge opp teknologisk og faglig kompetanse i å gjøre data FAIR. Det var lagt opp til at søkermiljøene måtte dekke 70 prosent av investeringskostnadene.

Utvalget anbefaler:

Det bør etableres stimuleringsordninger for infrastrukturensamarbeid og kompetanseoverføring mellom relevante miljøer i form av nettverk, møteplasser og/eller sentre.

10.4 Investering i nasjonale datainfrastrukturer

I denne sammenheng bruker vi betegnelsen nasjonal datainfrastruktur om infrastruktur som ikke bare er tilknyttet én institusjon eller ett miljø, men som tilbys for bruk til alle relevante miljøer i Norge. Slik datainfrastruktur finansieres i dag enten direkte eller gjennom konkurranseutsatte midler. I begge tilfeller er det viktig at investeringene sees i sammenheng med eksisterende datainfrastrukturer, både nasjonalt og internasjonalt.

Forskningsrådet er gjennom INFRASTRUKTUR-ordningen ansvarlig for store deler av investeringene som gjøres i nasjonal datainfrastruktur. Ordningen dekker nyetablering av datainfrastruktur eller oppgradering av eksisterende. Mange av de domenespesifikke datainfrastrukturene er i en implementerings- og driftsfase, og disse faller strengt tatt utenfor ordningen. Felles for flere av infrastrukturene er at de er medlem av [ESFRI – the European Strategy Forum on Research Infrastructures](#), der de oppfyller nasjonale forpliktelser i et internasjonalt samarbeid på vegne av Norge og Forskningsrådet. Som nevnt er dette kompetanseintensive datainfrastrukturer som krever jevnlig oppdateringer på tekniske, datafaglige og vitenskapelige områder. For å sikre at datainfrastrukturene kan koordinere og levere sine tjenester på et høyt nasjonalt nivå og samtidig oppfylle sine forpliktelser over sine ESFRI-medlemskap, er det behov for stabil og forutsigbar finansiering. Svakheten ved konkurranseutsatt finansiering i dette tilfellet er tidsbegrensningen.

Evalueringen av INFRASTRUKTUR-ordningen oppsummerer at satsningen har hatt en strukturerende effekt og forbedret samarbeid og synergier mellom aktører. I tillegg har den bidratt til å styrke det internasjonale samarbeidet og tydeliggjøre betydningen av FAIR-data. Generelt gjenstår imidlertid mye for å oppnå kompatible løsninger for datainfrastrukturer. Et eksempel på dette kommer tydelig fram i rapporten om [Fremtidens miljødata](#) (Menon, 2021), hvor det blir det pekt på behovet for å samordne miljødata bedre. Situasjonen for miljødata er tungvinte løsninger for å finne og dele data, og ulike data er registrert i ulike formater og på ulike måter. FAIR-utredningen påpeker noe av det samme. For å unngå overlappende ansvarsområder og duplisering av datainfrastruktur og tjenester er det viktig at arbeidet samordnes sentralt.

Utvalget anbefaler:

Beslutning om etablering og videreutvikling av nasjonale datainfrastrukturer bør gjøres sentralt, enten via Forskningsrådets INFRASTRUKTUR-ordning eller andre mekanismer som sørger for en vurdering av infrastrukturenes innplassering i det nasjonale og internasjonale datainfrastrukturlandskapet.

10.5 Hvorfor internasjonal deltakelse er viktig for god koordinering og ansvarsdeling

Norsk deltakelse i internasjonalt samarbeid om datainfrastrukturer er viktig for norsk forskning fordi det gir forskere tilgang til avansert infrastruktur som ikke Norge har, det gir tilgang til bedre og større datamengder, det legger til rette for samarbeid mellom land, og det bidrar til å utvikle og forbedre norske infrastrukturer og tilhørende kompetansemiljøer. Finansiering gjennom internasjonale ordninger skjer først og fremst gjennom Horisont Europa og DIGITAL. Gjennom deltakelsen i europeiske nettverk og initiativer har Norge en mulighet til å samkjøre nasjonale investeringer i datainfrastrukturer med europeiske investeringer for utvikling av tverrsektoriell infrastruktur. Mye av utviklingen og implementeringen av EOSC har foregått i EU-finansierte EOSC-prosjekter i Horisont 2020-rammeprogrammet og fortsetter i Horisont Europa i arbeidsprogrammet for forskningsinfrastrukturer, hvor EOSC er en egen destinasjon.

De aller fleste datainfrastrukturene på [ESFRIs veikart](#) er distribuerte. Det vil si at forskjellige land bidrar med komplementære delinfrastrukturer (kalt noder) i en felles infrastruktur. Gjennom EUs programmer er det mulig å søke støtte til infrastrukturer som bidrar til å oppfylle målene i ESFRI-veikartet. Eksempler på internasjonale infrastrukturer med støtte fra Forskningsrådet og EUs programmer er

- ELIXIR Norge er den norske noden i den paneuropeiske ESFRI-infrastrukturen ELIXIR. ELIXIR Norge tilbyr e-infrastruktur for analyse, håndtering og deling av molekyllære

livsvitenskapsdata. Norwegian e-infrastructure for Life Sciences (NeLS) for ikke-sensitive data bygger på Sigma2, løsninger for sensitive data er etablert i samarbeid med USIT-TSD, SAFE og HUNT Cloud. ELIXIR Norge utvikler og benytter ontologier og metadataskjemaer som er brukt over hele Europa. Videre tilbys generelle databehandlingstjenester og rådgivning.

- CLARINO er den norske noden av ESFRI-infrastrukturen CLARIN ERIC (Common Language Resources and Technology Infrastructure). CLARINO bygger opp historiske og nåtidige elektroniske språkressurser og kobler sammen språk- og tekstdatabaser. Infrastrukturen kobler norske databaser sammen med utenlandske, og forskere kan enkelt laste ned data som har åpne lisenser (Creative Commons). En node i CLARINO er CLARINO Bergen Centre, som omfatter en lagrings- og kurateringsressurs.

På det operative nivået finnes en rekke internasjonale organisasjoner, initiativer og prosjekter som utvikler standarder, løsninger og infrastrukturer for mer deling og gjenbruk av forskningsdata. Blant annet er ESFRI, EOSC, RDA, OpenAire, GO FAIR og CODATA viktige aktører. I 2019 ble det etablert Fem [ESFRI-klyngeprosjekter](#) for å gi ulike ESFRI-prosjekter og landemerker et samlingspunkt for å koble seg til EOSC. Flere norske ESFRI-noder er tilknyttet klyngeprosjektene, deriblant [ELIXIR Norge](#) gjennom [EOSC-Life](#) og [CESSDA](#) gjennom [SSHOC](#). Noen norske forskningsinfrastrukturer er dermed allerede involvert og viktige i etableringen og implementeringen av EOSC via deltakelse i internasjonale konsortium og nettverk. Deres bidrag inn mot EOSC er med på å synliggjøre og gjøre egen datainfrastruktur og tjenester tilgjengelige og attraktive blant europeiske forskningsmiljøer. Ved selv å bidra med FAIR-data og forskningsresultater kan norske forskere også synliggjøre sin egen forskning og skape nye muligheter for tverrfaglige synergier og samarbeid.

Utvalget anbefaler:

Det bør legges til rette for deltakelse i internasjonalt datainfrastruktursamarbeid for å oppnå effektiv bruk av ressurser og videreutvikling av norske noder og tilhørende kompetansemiljøer.

10.6 Ansvar for å gjøre forvaltningsdata tilgjengelige for forskning

Mange departementer og direktorater samler inn data til forvaltningsformål. Dette er data som kan være svært verdifulle for forskere, men det kan være tungvint og tidkrevende å få tilgang til dem. I noen tilfeller kan dette skyldes at dataene ikke rett fram kan deles slik de er, og da må de behandles manuelt i hvert enkelt tilfelle. Det kan også være at dataene ikke er utstyrt med gode nok metadata, eller at de har formater som ikke er kompatible med dataene forskerne ønsker å kombinere forvaltningsdataene med. I en del tilfeller er dataene kun tilgjengelige for overlevering til mennesker, ikke direkte fra maskin til maskin. Det er med andre ord behov for større grad av FAIR-deling av forvaltningsdata til forskning enn hva som er tilfellet i dag.

Strategien [Én digital offentlig sektor](#) (2019–2025) setter som mål at alle offentlige institusjoner skal dele sine data der dette er mulig. To av innsatsområdene i strategien er: "Offentlig sektor skal samhandle bedre om digitale tjenester og effektivisere ressursbruken gjennom styrket samordning på tvers av forvaltningsnivåer og sektorer, og systematisk uthenting av gevinster fra digitalisering" og at "Data skal i større grad deles og gjenbrukes i offentlig sektor, og åpne data skal publiseres for innovasjon og verdiskaping i næringslivet."

Vi har mange eksempler på forvaltningsorganer som på ulike måter tilgjengeliggjør data til forskning, for eksempel Nasjonalbiblioteket, Arkivverket, Meteorologisk institutt, Folkehelseinstituttet og Statistisk sentralbyrå. Noen har kommet langt i å gjøre data FAIR, mens andre har mer tidkrevende og kostbare prosesser knyttet til å gjøre dataene tilgjengelige for forskning. Skillet mellom

datainfrastrukturer for forskningsdata og datainfrastrukturer for forvaltningsdata er ikke alltid tydelig, noe som gjør det hensiktsmessig å se dem i sammenheng. Flere av infrastrukturene som i hovedsak tilbyr offentlige forvaltningsdata til bruk i forskning, tilbyr også arkivering av forskningsdata i den samme infrastrukturen. Dette gjelder for flere av infrastrukturene for miljødata, for eksempel Norwegian Marine Data Center. Et annet eksempel er at microdata.no nå også tilbyr forskerne å legge inn sine egne forskningsdata i infrastrukturen for å analysere disse sammen med data fra SSB.

Det kan i noen tilfeller være nødvendig med ekstra tilrettelegging for å gjøre data anvendbare for forskning. Det kan være tvil om hvem som er ansvarlig for å finansiere det ekstra arbeidet. Det kan argumenteres for at KD bør ha et hovedansvar, og enkelte datainfrastrukturer har også fått støtte gjennom Forskningsrådets INFRASTRUKTUR-ordning. Det er utvalgets oppfatning at også sektordepartementene har et ansvar for å sikre finansielle og organisatoriske rammer som tilrettelegger forvaltningsdata for forskningen innenfor sitt sektoransvar.

Utvalget anbefaler:

Sektordepartementene må ta ansvar for å gjøre relevante forvaltningsdata tilgjengelige for forskning.

10.7 Dimensjonering og prioritering gjennom nettverk og samarbeid

Som beskrevet i del 2 er det kostbart å gjøre data FAIR, og utvalget diskuterte også hvorvidt det skal være en ambisjon at alle forskningsdata skal være FAIR. I overskuelig framtid vil ikke alle data kunne gjøres FAIR, både på grunn av penger, teknologi og kompetanse. Innenfor de enkelte forskningsprosjektene og forskningsgruppene skal forskermiljøene selv beslutte hvilke data som skal tas vare på. På et mer overordnet nivå vil det være behov for å beslutte hvilke datasamlinger og datainfrastrukturer som skal prioriteres i første omgang. Slike beslutninger bør tas av forskningsmiljøene i samarbeid med andre aktører som har interesse av tilgang til dataene. I noen sammenhenger ser vi også at ulike fagområder baserer seg på samme type data, men har gjort dette uavhengig av hverandre. Når nye tverrfaglige samarbeid oppstår, kan det bli nødvendig med koordinering for å sikre felles forståelse. Tverrfaglig samarbeid vil også bidra til kompetanseheving og læring på tvers, som det blir stort behov for framover.

Behovet for kunnskap som kan bidra til å belyse samfunns- og naturfenomener på tvers, forutsetter samarbeid, tverrsektoriell styring og finansiering. Ofte er det enkelte departement opptatt av å fremme interesser i sin sektor og avgrenser dermed den helhetlige tilnærmingen som ofte trengs i utforskningen av komplekse problemstillinger. Det bidrar ikke til tverrfaglig samarbeid og til å finne nye løsninger på komplekse problemer. Nå er det imidlertid satt ned flere ulike tverrdepartementale grupper som har som mål å oppnå bedre koordinering mellom departementene. Utvalget synes [forslaget fra Skate](#) (Styring og koordinering av tjenester i e-forvaltning) til KMD om tverrsektoriell styring er fornuftig. For å lykkes anbefaler Skate blant annet å samordne tildelingsbrev, å ha felles budsjettposter og at et departement bør ta koordineringsansvar for utvikling som omfatter flere departementer.

Utvalget anbefaler:

Det må etableres arenaer for prioritering av hvilke datasamlinger og datainfrastrukturer det skal brukes ressurser på å gjøre FAIR. Dette må gjøres i fellesskap mellom forskningsmiljøene, offentlig forvaltning, næringslivet og andre relevante samfunnsaktører.

DEL 4

Forslag til rammeverk for finansiering av datainfrastruktur

Utvalget har gjennom sitt arbeid kommet fram til et forslag til rammeverk som vil kunne bidra til en mer bærekraftig finansiering av etablering og drift av datainfrastrukturer. I denne delen av rapporten presenteres først noen overordnede anbefalinger før selve rammeverket beskrives mer i detalj. Rammeverket er ment som et verktøy som både departementene, virkemiddelapparatet og infrastrukturer kan bruke for å finne egnede finansieringsmodeller. Utvalget presiserer at dette ikke er tenkt som en standard, men heller som mulige løsninger som kan kombineres og må tilpasses i hvert tilfelle. For infrastrukturer som i dag har en velfungerende finansieringsmodell, ser vi ingen grunn til å gjøre endringer.

Utvalgets overordnede anbefalinger

- For finansiering av etablering, videreutvikling og drift av nasjonal datainfrastruktur anbefaler utvalget å bruke rammeverket for langsiktig og bærekraftig finansiering som presenteres under.
- Finansiering av etablering, videreutvikling og drift av datainfrastrukturer som det forutsettes at alle forskningsinstitusjoner skal tilby sine forskningsmiljøer (basisinfrastruktur), bør skje innenfor institusjonenes ordinære budsjetter.
- Nasjonale datainfrastrukturer som mottar langsiktig finansiering, må ha klare mål for hvilke tjenester de skal levere, og jevnlig evalueres for å sikre at de leverer i tråd med målene som er satt for dem.
- Investeringer i nasjonale datainfrastrukturer bør samordnes og sees i sammenheng med norsk deltakelse i internasjonale datainfrastrukturer, både i form av norske noder og medlemskap i internasjonalt infrastruktursamarbeid.

I tilfeller hvor det allerede finnes velfungerende mekanismer for organisering, styring og finansiering, vil utvalgets generelle anbefaling være at dette videreføres.

11. Forslag til rammeverk for framtidig investering i og finansiering av datainfrastrukturer

I det anbefalte rammeverket tar utvalget hensyn til at det er stor variasjon i eksisterende datainfrastrukturer, for eksempel hvilke behov de dekker, hvor godt utviklet de er, om de tilbyr generiske eller domenespesifikke datatjenester, og om de tilbyr kun lokale tjenester eller er åpne for nasjonale brukere utenfor vertsinstitusjonen. Fordi behovene datainfrastrukturene dekker er så forskjellige, mener utvalget at det vil være urealistisk og lite hensiktsmessig å etablere en "superinfrastruktur" som skal dekke alle behov, men at det utvikles løsninger som gjør det mulig å dele og bruke data på tvers av infrastrukturer som betjener ulike områder. Infrastrukturenes funksjon, organisasjonsmodell, kostnadsnivå, kapasitet, brukermasse m.m. vil også påvirke hvordan kostnadene knyttet til driften kan og bør dekkes. Av denne grunn vil heller ikke utvalget foreslå én måte å finansiere datainfrastrukturer på, men et rammeverk for finansiering med flere komponenter som kan kombineres på ulike måter.

Det er viktig å understreke at det allerede brukes mye ressurser på investering og drift av infrastrukture- og datatjenester. For noen infrastrukturer fungerer dagens situasjon godt. På andre områder er dagens finansieringsmodell lite bærekraftig, spesielt sett i lys av behovet for videreutvikling mot å følge FAIR-prinsippene. Et av formålene med utvalgets anbefalinger er å sørge for kostnadseffektiv bruk av ressursene uten at det går på bekostning av ulike brukerbehov eller begrenser forskernes tilgang til gode og relevante infrastrukturer og tjenester. Et viktig tiltak for å oppnå mer optimal bruk av ressurser er å koordinere investeringer i infrastruktur på et nasjonalt nivå for å unngå at det etableres *unødvendige* parallelle infrastrukturer og/eller tjenester ved institusjonene. Det bør også på flere områder vurderes om norske miljøer bør delta i internasjonale infrastruktursamarbeid, enten i kombinasjon med eller som en erstatning for investering i nasjonal infrastruktur.

Det er kostbart å etablere, utvikle og drifte datainfrastrukturer dersom de skal imøtekomme kunnskapsbehovene og også utnytte de teknologiske mulighetene. Datainfrastrukturer må være i kontinuerlig utvikling for å møte samfunnets og forskningsmiljøenes behov. Kravet i FAIR-prinsippene om at data skal være maskinlesbare, øker for eksempel kompleksiteten i infrastrukturene. Derfor er det behov for regelmessige investeringer og oppgraderinger, også i de etablerte infrastrukturene.

Langsiktig finansiering av en datainfrastruktur vil sikre forutsigbar drift, slik at basisfunksjoner, inkludert nødvendig kompetanse, kan opprettholdes. En slik garantert finansiering må imidlertid ikke dyrke fram en monopolsituasjon uten å være oppmerksom på tjenestekvalitet og ROI (return of investment). Jevnlig evaluering må derfor gjennomføres, og utfallet må gi grunnlag for å endre innretning i organisering, finansiering og oppgaver. Kritisk brukerevaluering basert på kvalitet og kostnad er viktig for enhver datainfrastruktur, og institusjonene må kunne velge hvilke tjenester de finansierer og bruker. I tillegg må det stilles krav til årlige rapporteringer og formelle kontaktflater med brukermiljøene, som gir disse reell medvirkning i videreutviklingen.

I noen deler av modellen anbefales brukerbetaling, enten i form av stykkpris eller abonnement. For å få brukerbetaling til å fungere er det viktig at forskningsmiljøene gis mulighet til å budsjettere med reelle kostnadsestimater for datahåndtering i planleggingen av et forskningsprosjekt. Det er derfor viktig at datainfrastrukturene har prislister for de ulike tjenester de tilbyr som inngår i finansieringsmodellen. Det kan være vanskelig å lage slike prisestimater, fordi det ikke er mulig å forutse framtidig interesse og bruk av infrastrukturen og tjenester knyttet til den, men det er likevel helt nødvendig.

Utvalgets forslag til bærekraftig rammeverk for finansiering av datainfrastrukturer

Etablering, videreutvikling og drift av datainfrastrukturer kan omfatte mange ulike kostnadselementer, for eksempel anskaffelse og oppgradering av fysisk infrastruktur, drift av fysisk infrastruktur, utvikling og drift av digitale tjenester, brukerstøtte, utvikling av relevant kompetanse for brukere og tilbydere, løpende nødvendige oppgraderinger og eventuelt videreutvikling av infrastrukturene. En bærekraftig finansieringsmodell for slike infrastrukturer må ta opp i seg at alle disse kostnadselementene må dekkes inn.

Inndeling i tre tjenestekategorier

Forskningsdata vil typisk gå gjennom et dynamisk livsløp i et forskningsprosjekt, fra planlegging og innsamling, via aktiv bruk med analyser og bearbeiding, til arkivering og eventuell deling, og underveis i denne prosessen vil det oppstå ulike infrastruktur- og tjenestebehov. Utvalgets anbefalinger baserer seg i det følgende på en grov, overordnet inndeling i tre kategorier med ulike typer verktøy, tjenester eller funksjonalitet (heretter "tjeneste") som kan tilbys av datainfrastrukturer underveis i denne livssyklusen. Hvilken kategori tjenesten faller inn under, vil ha innvirkning på hvilken type finansiering som er hensiktsmessig og realistisk.

En enkelt infrastruktur kan ha tilbud innenfor flere av kategoriene, og beskrivelsen av disse kategoriene er ikke uttømmende og vil ikke i detalj dekke alle typer funksjonalitet og varianter av tjenester tilknyttet datainfrastrukturer.

- 1 Grunnleggende, generisk e-infrastruktur som er en forutsetning for datadrevet forskning. Dette vil inkludere fysisk infrastruktur (for eksempel høyhastighetsnett og datalagring), tungregning samt generisk (interdisiplinær) programvare for analyse og prosessering av stordata.
- 2 Verktøy og tjenester knyttet til aktiv bruk, deling og gjenbruk av data. Dette kan være tilknyttet både generelle og fagspesifikke infrastrukturer, men vil i betydelig grad være fagspesifikke og disiplin tilpassede verktøy og tjenester.
 - 2.1 Tjenester som tilgjengeliggjør data for gjenbruk i forskning. Dette inkluderer publisering av data, databaser, datakataloger for søk, plattformer og tjenester for utlevering/tilgang til data osv.
 - 2.2 Tjenester og verktøy for aktiv bruk, håndtering, analyse og prosessering av data i forskningsprosjekter. Dette inkluderer samhandlingsplattformer, verktøy for metadata, ontologier og annotasjoner, "data management systems" og elektronisk labjournal, programmer og tjenester (pipelines/workflows) for reproduserbar analyse, visualisering og prosessering.
- 3 Tjenester for langtidsbevaring og langsiktig forvaltning av data. Dette gjøres gjerne i konteksten av et dataarkiv (repository) som ofte tilbyr mange ulike tjenester, men i denne kategoriseringen skilles selve den langsiktige forvaltningen ut som en egen, grunnleggende tjeneste separat fra tilgjengeliggjøringen av data (kategori 2.1). Infrastruktur som tilbyr slik langtidsforvaltning, kan både være generelle, tverrfaglige eller fagspesifikke.

I tillegg er vellykket datadrevet og delingsbasert forskning avhengig av gode datatjenester og systemer som sørger for blant annet gjenfinnbarhet og kreditering (PID-utstedelse, CRIS-systemer, siteringsverktøy,

lisenser m.m.) som er godt integrert i datainfrastrukturene som brukes. Mange av disse tjenestene dekkes gjennom nasjonale og internasjonale fellestjenester kombinert med kompetanse og bistand i støttesystemene ved institusjonene og er derfor ikke like relevante for utvalgets anbefalinger om investering i infrastruktur.

Tre hovedformer for finansiering

Infrastrukturer som ligger til grunn for tjenestene som er beskrevet over, kan finansieres på flere ulike måter. Utvalget har delt inn finansieringen i tre hovedformer, som vi vil bruke i anbefalingene under. I mange sammenhenger vil finansieringen være en kombinasjon av disse typene.

- 1 **Langsiktig finansiering** – stabil og forutsigbar finansiering. Finansieringen kan for eksempel komme i form av grunnfinansiering fra ett relevant departement eller fra flere departementer etter en fordelingsnøkkel basert på hvilke forvaltningsområder infrastrukturen leverer forskning og forvaltning til. Slik finansiering kan også komme i form av spesifikke forvaltningsoppdrag eller gjennom forpliktende medlemskap eller partnerskap fra institusjoner som bruker infrastrukturen.
- 2 **Konkurransetsatt finansiering** – finansiering som et resultat av konkurranse mellom flere aktører som vurderes etter gitte kriterier, for eksempel Forskningsrådets infrastrukturutlysning eller EUs rammeprogrammer.
- 3 **Brukerfinansiering eller brukerbetaling** – finansiering som kommer fra de som bruker tjenestene. Dette kan enten være i form av abonnement, stykkpris eller forbruksfinansiering, eller en kombinasjon av disse.

Utvalgets anbefalinger

Basert på tidligere utredninger³, diskusjoner i utvalget, innspill som utvalget har fått, og funn som er presentert i [FAIR-utredningens rapport](#), har utvalget kommet fram til følgende hovedanbefalinger om finansiering av ulike datainfrastrukturer som er viktige for å oppnå FAIR-forskningsdata:

- 1 **Finansieringen av grunnleggende, generisk e-infrastruktur må økes og være forutsigbar i årene som kommer for å imøtekomme behovene for slik infrastruktur på tvers av hele forskningssektoren.** Fra flere hold, blant annet fra Forskningsrådet og den nylige [evalueringen](#) av Forskningsrådets INFRASTRUKTUR-ordning, er det slått fast at det er behov for å øke størrelsen på finansieringen av denne typen infrastruktur, og at dette ikke bør finansieres gjennom konkurransetsatte midler. Dette er etablerte infrastrukturer som tilbyr tjenester det er behov for nå og i framtiden.
 - a) **En større andel av finansieringen av grunnleggende, generisk infrastruktur bør komme gjennom langsiktig finansiering.** Det er uheldig at en stor andel av finansieringen av slik infrastruktur er konkurransetsatt. Det skaper usikkerhet og risiko for tap av viktig kompetanse og en reell fare for at viktige forsknings- og forvaltningsoppgaver ikke kan gjennomføres. Denne typen infrastruktur betjener alle fagområder og prosjekter innenfor både UH-sektoren, helsesektoren og instituttsektoren. I takt med at forskningen innenfor alle fag blir mer og mer dataintensiv og datadrevet, vil behovet øke. For å oppfylle nasjonale føringer om langsiktig lagring av forsknings- og forvaltningsoppgaver vil det kreve en langsiktig og forutsigbar finansiering.
 - b) **Generiske e-infrastrukturtenester som spesielt tilbyr lagringsressurser underveis i et forskningsprosjekt kan helt eller delvis brukerfinansieres.** Midler til betaling for bruk må da

³ Blant annet [Behov og finansieringsstrategi for nasjonal e-infrastruktur for forskning for perioden 2020–2030](#) og [Tilgjengeliggjøring av forskningsdata og data for forskning](#)

legges inn som del av budsjettene til forskningsprosjekter, både i prosjekter med ekstern finansiering og i prosjekter som institusjonene selv dekker. For større institusjoner som har mange prosjekter som genererer store mengder data, kan betaling med fordel skje gjennom institusjonsbidrag til infrastrukturene. Dette vil bidra til mindre administrasjon og mer stabil og forutsigbar finansiering av nasjonale e-infrastrukturer som tilbyr slike tjenester.

2 Datainfrastrukturer for aktiv bruk, deling og gjenbruk av forskningsdata anbefales finansiert i en kombinasjon av langsiktig finansiering, konkurranse- og brukerbetaling.

- a) **Dersom infrastrukturen dekker et nasjonalt behov, kan det vurderes om deler av driften skal dekkes gjennom en langsiktig finansiering.** I dag får også de fleste av disse infrastrukturene konkurranseutsatt finansiering. Som i pkt. 1a vil dette skape usikkerhet og medfølgende risiko for tap av viktig kompetanse. For noen av dem er dette spesielt uheldig, for eksempel hvis de er del av en internasjonal infrastruktur der Norge har forpliktet seg til deltakelse. For å få grunnfinansiering må datainfrastrukturer oppfylle bestemte krav til hvilke tjenester de skal levere. Dette kan være krav om å gjøre data FAIR, gjøre data komplementære til andre data, at data må gi faglig merverdi ut over eksisterende grunnfinansierte tjenester og godkjente internasjonale datainfrastrukturer, og at det må stilles tekniske krav til tjenesten. Dette gjør at vi kan bygge på dagens eksisterende tjenester, og med det ivareta domene- og institusjonsheterogenitet samtidig som vi sørger for at ikke alle institusjoner eller enkeltprosjekter lager egne datainfrastrukturer og -tjenester. Utvalget tar ikke stilling til hvem som skal vurdere om en tjeneste oppfyller disse kravene, men vil peke på at dette kan sammenlignes med jobben Forskningsrådet i dag gjør med å vurdere om forskningsinstitutter oppfyller kriteriene for å få grunnfinansiering, og delvis også vurderingen av hvorvidt en forskningsinfrastruktur fyller kravet om å være nasjonal. Det finnes også flere internasjonale sertifiseringsstandarder, for eksempel gjennom Core Trust Seal, GoFair og EOSC som vil være relevante i denne sammenhengen.
- b) **Nyetablering av datainfrastrukturer eller omfattende utvidelse eller videreutvikling av tilbudet til en datainfrastruktur for deling og gjenbruk av forskningsdata konkurranseutsettes.** Dette kan gjennomføres gjennom dagens INFRASTRUKTUR-ordning i Forskningsrådet eller gjennom andre konkurransearenaer. Fordelen med å tildele midler gjennom konkurranse er at infrastrukturene blir vurdert etter tydelige krav i utlysningen. Dette vil identifisere og fremme infrastrukturer som svarer på nasjonale behov innenfor bestemte fagområder eller bidrar til tekniske løsninger tilpasset behovene i forskning, forvaltning, næringsliv og frivillig sektor. Det er viktig å få en kvalifisert vurdering av hvilken merverdi det vil ha å utvikle tjenestetilbudet, og derfor vil konkurranse, inkludert faglige og strategiske vurderinger, egne seg.
- For å sikre varig drift av en nyetablert infrastruktur kan finansieringen komme fra ulike kilder avhengig av hvilket behov infrastrukturen skal imøtekomme. Modell for framtidig drift må vurderes før det fattes beslutning om å tildele midler til nyetablering. Etablerte datainfrastrukturer kan, når de er godt etablert, kvalifisere til finansiering i kategori 2a.
- c) **Tjenester som er rettet spesielt mot enkeltprosjekter eller er tett knyttet til fysiske infrastrukturer, kan brukerfinansieres.** Dette kan for eksempel være tilpassede tjenester som gir forskerne tilgang til behandlede eller sammenstilte data, og som det vil være naturlig at prosjektet betaler for. Det kan også være tjenester knyttet til avlevering og publisering av forskningsdata i tråd med FAIR-prinsippene. Ved ekstern finansiering må slike kostnader legges inn som en del av prosjektets budsjett. Institusjoner som stiller krav om FAIR-forskningsdata

fra sine forskere, bør da også forventes å legge dette inn i sine budsjetter, enten gjennom å "belaste" enkeltprosjekter for kostnadene, eller gjennom institusjonsbidrag til relevante datainfrastrukturer. Det vil være naturlig at infrastrukturene, i den grad den kommende [gjennomføringen i norsk rett av direktivet om åpne data tillater det](#), legger en overhead på tjenestene for å dekke inn deler av driftskostnadene gjennom slik brukerbetaling.

- 3 Finansiering av dataarkiver for langtidsbevaring og kuratering av verdifulle forskningsdata og forvaltningsdata som er relevante for forskning, må være et nasjonalt ansvar og gis langsiktig finansiering**, slik man til en viss grad gjør innenfor ABM-sektoren i dag. Langtidsbevaring kan ikke dekkes av enkeltprosjekter med begrenset budsjettperiode, og mange typer data kan ha verdi utover et enkelt forskningsprosjekt. For å sikre langsiktige arkivtjenester for forskning, næringsliv og forvaltning er det også nødvendig med forutsigbar finansiering. Det betyr at slike arkivtjenester i all hovedsak bør ha en langsiktig finansiering. Om denne skal komme fra én eller flere kilder, må vurderes. Aktuelle finansieringskilder er alle departementer som har et sektoransvar for forskning. Kunnskapsdepartementet, som har et hovedansvar for forskning og utdanning (herunder infrastruktur), må ta ansvaret for å samordne og koordinere departementene på en måte som understøtter samordning av den langsiktige finansieringen. Dette kan også kombineres med et krav om at alle datagenererende institusjoner selv bidrar med midler til infrastrukturen.

Et eksempel

For å vise hvordan en slik modell vil slå ut i praksis, har vi tatt utgangspunkt i Statistisk sentralbyrås funksjon som dataleverandør for forskere. I tabellen under har vi kategorisert ulike tjenester som SSB har i den forbindelse, og beskrevet hvordan de ulike tjenestene kan finansieres ut fra den foreslåtte modellen.

Tjenestekategori/aktivitet	Innsamling og forvaltning av statistikkdata	Tradisjonell tilgjengeliggjøring av tilrettelagte data for forskere	Etablering og videreutvikling av microdata.no	Drift av microdata.no
Finansieringsform	Langsiktig finansiering	Brukerfinansiering, stykkpris	Konkurransebasert finansiering	Planlegges delvis brukerfinansiert

Konkurransarenaer som sikrer utvikling og modernisering av teknologi, tjenester og kompetanse

I tillegg til investering i selve infrastrukturene (utstyr og personell) som skal bidra til mer FAIR-forskningsdata og relevante forvaltningsdata, ser utvalget et stort behov for å øke investeringene til utvikling av ny teknologi og nye tjenester for å gjøre infrastrukturene relevante og internasjonalt konkurransedyktige for framtidige forskningsbehov. Finansiering av slike utviklingsprosjekter må sees i tett sammenheng med investering i infrastrukturene. Disse prosjektene faller gjerne utenfor etablerte finansieringsordninger fordi vi ikke har gode nok mekanismer til å identifisere, vurdere og kvalitetssikre dem. De kan være vanskelige å definere inn under etablerte kategorier fordi de ofte befinner seg i skjæringspunktet mellom forskning, metodeutvikling, tjenesteutvikling og infrastrukturetablering. Det er ofte forbundet mye usikkerhet og risiko med slike prosjekter fordi de søker å finne løsninger på delvis ukjente, framtidige behov og kan være eksperimentelle i sin natur. De taper da ofte i konkurranse med andre, mindre risikofylte prosjekter. Eksempler på slike prosjekter kan være utvikling av nye beregningsteknologier, som neuromorphic computing, analysemetoder for store og komplekse datamengder som baserer seg på maskinlæring og kunstig intelligens, nettverkstjenester i takt med for eksempel 5G/6G nettverksutvikling eller nye FAIR-løsninger for arkiver. Utvalget mener derfor at det er

behov for å utvide etablerte konkurransearenaer til også å omfatte utviklingsprosjekter, eller at det etableres egne arenaer med kriterier som gjør det mulig for risikoprosjekter å nå opp i konkurransen.