



Kunnskapsdepartementet

Strategi

# Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata





# Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata



# Innhold

|          |   |    |
|----------|---|----|
| <b>1</b> | <b>Sammendrag</b> .....   | 7  |
| 1.1      | Forskningsdata bør deles og gjenbrukes mer .....  | 7  |
| 1.2      | Hva menes med offentlig finansierte forskningsdata? .....   | 8  |
| 1.3      | Hva menes med gjenbruk? .....   | 8  |
| 1.4      | Grunnprinsipper .....   | 9  |
| 1.5      | Forventninger og tiltak .....   | 9  |
| <b>2</b> | <b>Status</b> .....   | 15 |
| 2.1      | Overordnet IKT-politikk og åpne offentlige data internasjonalt og nasjonalt .....   | 15 |
| 2.2      | Åpen tilgang til forskningsdata internasjonalt og nasjonalt .....   | 16 |
| 2.3      | Personvern og rettigheter til data .....  | 18 |
| 2.4      | Nytt organ for høyere utdannings- og forskningstjenester .....  | 19 |
| 2.5      | Investeringer i forskningsdatainfrastrukturer .....   | 19 |
| 2.6      | Status ved forskningsinstitusjonene .....   | 21 |
| <b>3</b> | <b>Grunnprinsipper</b> .....  | 23 |
| 3.1      | Grunnprinsipp 1: Forskningsdata skal være så åpne som mulig, så lukkede som nødvendig .....                                 | 23 |
| 3.2      | Grunnprinsipp 2: Forskningsdata bør håndteres og tilrettelegges slik at verdiene i dataene kan utnyttes best mulig .....    | 25 |
| 3.3      | Grunnprinsipp 3: Beslutninger om arkivering og tilrettelegging av forskningsdata må tas i forskerfellesskapene .....        | 26 |
| <b>4</b> | <b>Kulturendring, insentiver, kompetanse, datahåndteringsplaner, teknisk tilrettelegging og finansieringsmodeller</b> ..... | 29 |
| 4.1      | Kulturendring, insentiver og kompetanse .....   | 29 |
| 4.2      | Krav om datahåndteringsplaner .....   | 31 |
| 4.3      | Bedre teknisk tilrettelegging og bedre faglig samordning nasjonalt .....  | 34 |
| 4.4      | Bærekraftig finansiering og drift av nasjonale forskningsdatainfrastrukturer og forskningsdataarkiver .....                 | 35 |
| <b>5</b> | <b>Offentlige data som er viktige datakilder for forskning</b> .....  | 41 |
| 5.1      | Bedre tilrettelegging for bruk av offentlige data til forskning .....   | 41 |
| 5.2      | Bedre samspill mellom offentlige data og forskningsdata .....   | 42 |
| 5.3      | Enklere tilgang til data fra Statistisk sentralbyrå for forskningsformål .....  | 43 |
| 5.4      | Enklere og sikrere tilgang til helsedata .....  | 45 |
| <b>6</b> | <b>Økonomiske og administrative konsekvenser</b> .....  | 47 |



# 1 Sammendrag

---

## 1.1 Forskningsdata bør deles og gjenbrukes mer

Digitaliseringen bringer med seg forandringer i forskningen. Arbeidsprosesser endres og det åpner seg nye muligheter for å bedre forskningskvaliteten og øke den kunnskapsintensive verdiskapingen. Bedre tilgjengeliggjøring og mer gjenbruk av forskningsdata kan bidra til vitenskapelige fremskritt ved at hver enkelt forsker får et større datagrunnlag å bygge på, til økt etterprøvnbarhet og bedre kvalitetssikring av tidligere forskningsfunn, og til å unngå at samme type datainnsamling må finansieres flere ganger. En sterkere delingskultur gjør det mulig for studenter å arbeide med interessante og aktuelle data og bidra til forskningen. Bedre tilgang til forskningsdata kan bidra til innovasjon og verdiskaping ved at andre aktører enn forskere kan se nye bruksområder. Det kan bidra til smartere tjenesteutvikling i offentlig sektor, muligheter for ny næringsvirksomhet og flere arbeidsplasser. Et selvstendig og viktig poeng er at større åpenhet og bedre innsyn i forskningen kan bidra til å styrke tilliten til forskere og til forskningsresultater.

For å bedre tilgjengeligheten og øke gjenbruken av forskningsdata trenger forskere kompetanse og verktøy til å håndtere data på en god og sikker måte i alle ledd av forskningsprosessen. De må kunne finne og få tilgang til relevante data blant eksisterende kilder. De må ha infrastruktur både for innsamling, analyse, arkivering og deling av data, og tilgang til oversiktlig informasjon om denne. Infrastrukturen må legge til rette for samarbeid og kunnskapsdeling på tvers av land og sektorer. Norske datasett bør være enkle å finne for internasjonale forskere. Brukergrupper i næringslivet eller i forvaltningen vil ha en del behov som sammenfaller med forskernes. For at en sterk åpenhetskultur skal kunne kombineres med en trygg ivaretagelse av viktige skjermingshensyn, for eksempel knyttet til personvern, må ressurser og oppmerksomhet rettes både mot tekniske aspekter ved databehandling og mot etiske, juridiske og administrative forhold.

Denne strategien skal bidra til å øke forskningskvaliteten og den kunnskapsintensive verdiskapingen gjennom å fastslå grunnprinsipper for håndtering og tilrettelegging av offentlig finansierte forskningsdata, klargjøre forventninger til forskningsinstitusjonene og presentere tiltak som legger bedre til rette for å bygge på tidligere arbeider og sammenstille forskningsdata på nye måter.

Strategiens del 1 handler om å øke tilgjengeligheten til og gjenbruken av offentlig finansierte forskningsdata. Her beskrives policystatus og status ved forskningsinstitusjonene. Tre grunnprinsipper for hvordan forskningsdata skal tilrettelegges og deles fastslås. I tillegg beskrives de viktigste barrierene mot tilgjengeliggjøring og gjenbruk av forskningsdata med tilhørende tiltak. Det pekes på behov for kulturendring, økt kompetanse, datahåndteringsplaner, bedre teknisk tilrettelegging, bedre faglig samordning nasjonalt og bærekraftige finansieringsmodeller.

Strategiens del 2 handler om forskeres tilgang til offentlige data som utgjør viktige datakilder for forskning. Problemstillingen omtales generelt, og to viktige områder behandles spesielt. Det første dreier seg om forskeres tilgang til registerdata hos Statistisk sentralbyrå. Det andre dreier seg om deres tilgang til data fra helseregistre, biobanker og helseundersøkelser.

Strategien omfatter ikke forskningsdata fra privat finansierte forskning og utviklingsarbeid. I tilfeller der private aktører mottar offentlig støtte til forskning eller samarbeider med offentlige forskningsinstitutter, universiteter, høyskoler og helseforetak i forsknings- og innovasjonsprosjekter som er offentlig finansierte, åpner grunnprinsippene (se 1.4) for at data kan skjermes for å beskytte forretningshemmeligheter eller når det er nødvendig for å kommersialisere resultatene. Det kan likevel finnes tilfeller der private aktører finner det ønskelig å dele sine data. Effektivitets- og soliditetshensynene vil også gjøre seg gjeldende for private, som kan finne det hensiktsmessig å spre risiko og gå sammen om investeringer i data som deles, og konkurrere på bruken av data heller enn på tilgangen til data. Dette vil det

være opp til private aktører å vurdere fra tilfelle til tilfelle. Strategiens prinsipper, virkemidler og oppfølging kan følgelig være relevante også for private aktører. Virkemidler som informasjon, standarder og verktøy for åpne data kan gjøres tilgjengelig på fritt grunnlag og også finnes relevante for private. Sist, men ikke minst, er næringslivet tjent med at offentlig finansierte forskningsdata gjøres mer åpent tilgjengelige, at forskningsresultater er så robuste som mulig, og at offentlige forskningsmidler utnyttes effektivt.

Strategien er en oppfølging av Meld. St. 27 (2015–2016) *Digital agenda for Norge – IKT for en enklere hverdag og økt produktivitet*. For å øke tilgjengeligheten til åpne offentlige data sluttet Stortinget seg til at det skal utarbeides strategier eller handlingsplaner for fem sektorer som er funnet spesielt samfunnsøkonomisk verdifulle. Disse sektorene er kultur, statlige utgifter, transport og samferdsel, kart og eiendom (geodata) i tillegg til forskning.

Utviklingen går fort, og det er mye vi ikke vet om hvordan økt digitalisering og mer datadeling vil påvirke forskningen og dens rammebetingelser. Strategien gjelder derfor for perioden 2018-2021. Deretter vil den bli vurdert på egnet måte, før den eventuelt rulleres i tråd med nye erfaringer og økt kunnskap.

Innledningsvis klargjøres definisjoner, og det gis en sammenfatning av strategiens grunnprinsipper, forventninger og tiltak, før nærmere begrunnelser og utdyping gis i de påfølgende delene.

## 1.2 Hva menes med offentlig finansierte forskningsdata?

I denne strategien refererer *offentlig finansierte forskningsdata* til:

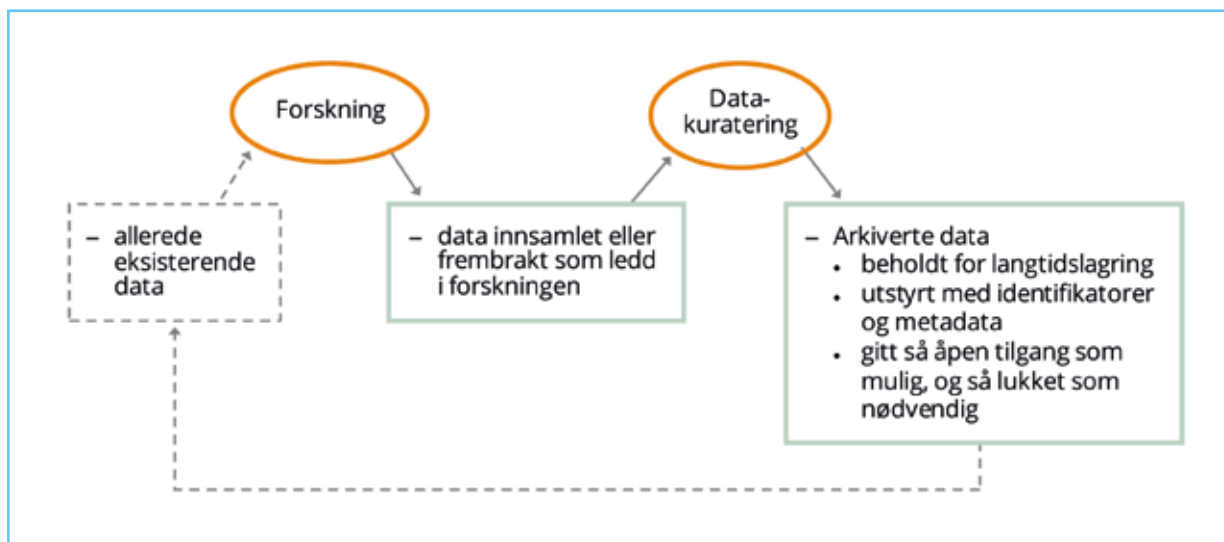
- (i) data som er samlet inn eller frembrakt til bruk for eller som et resultat av offentlig finansierte forskning, og
- (ii) data som utgjør grunnlaget for publikasjoner som er et resultat av offentlig finansierte forskning, uavhengig av hvilken kilde dataene kommer fra.

Med offentlige data menes all informasjon som offentlige organer produserer, samler inn eller betaler for. Slike data kan være både åpne og lukkede. Mange av disse dataene er ikke samlet inn primært for forskning, men de kan utgjøre viktige datakilder for forskning.

## 1.3 Hva menes med gjenbruk?

Politikken for tilgjengeliggjøring av offentlige data i Norge er beskrevet i Meld. St. 27 (2015–2016) *Digital agenda for Norge – IKT for en enklere hverdag og økt produktivitet*. Her brukes uttrykkene *gjenbruk* og *viderebruk*. *Gjenbruk* innebærer at offentlige virksomheter skal gjøre bruk av de dataene som det offentlige alt har, i stedet for å spørre brukerne på nytt om forhold de allerede har opplyst om. Dette omtales som prinsippet om «kun én gang» (kapittel 7). *Viderebruk* handler om å gi forskere, næringsliv og sivilsamfunnet tilgang til offentlige data på en måte som gjør at de kan

Figur 1.1 Datakretsløpet i forskning





brukes i nye sammenhenger også utenfor offentlig sektor (kapittel 14.2). All bruk vil være underlagt de generelle normene og reglene for kreditering, personvern, taushetsplikt mv. For denne strategiens formål er ikke skillet mellom gjenbruk og viderebruk og mellom offentlig og privat sektor viktig, og ordet gjenbruk betegner all slags videre bruk av forskningsdata, uavhengig av sektor.

## 1.4 Grunnprinsipper

Denne strategien slår fast følgende tre grunnprinsipper for offentlig finansierte forskningsdata i Norge:

**Grunnprinsipp 1: Forskningsdata skal være så åpne som mulig, så lukkede som nødvendig.** Dersom ingen legitime hensyn står i veien, skal forskningsdata tilrettelegges for åpen tilgang. Hensynet til sikkerhet, personvern, immaterielle rettigheter, forretningshemmeligheter o.l. tilsier i en del tilfeller at forskningsdata ikke kan gjøres helt åpent tilgjengelige. En del slike forskningsdata kan likevel gjøres tilgjengelige for bestemte brukere etter bestemte kriterier for tilgang. Behovet for skjerming kan også endre seg over tid, slik at dataene kan gjøres tilgjengelige på et senere tidspunkt.

**Grunnprinsipp 2: Forskningsdata bør håndteres og tilrettelegges slik at verdiene i dataene kan utnyttes best mulig.** Forskningsdata må håndteres slik at de er søkbare og gjenfinnbare og, der det er aktuelt, tilrettelegges slik at de er reelt gjenbrukbare. Det vil blant annet si at de må utstyres med gode metadata og med en lisens som gjør det klart hvordan dataene kan eller eventuelt ikke kan brukes. Forskningsdata skal gjøres tilgjengelig til lavest mulig pris og på like vilkår for alle brukergrupper som oppfyller kriterier for tilgang.

**Grunnprinsipp 3: Beslutninger om arkivering og tilrettelegging av forskningsdata må tas i forskerfellesskapene.** Forskere inngår i fellesskap, lokalt sammen med sine institusjonsledere, nasjonalt og internasjonalt, der god forsknings-skikk, herunder datadeling, utvikles kontinuerlig. Fordi det er ressurskrevende å tilrettelegge forskningsdata for gjenbruk, er det viktig med gode kost-nytte-vurderinger av hvilke data det gir merverdi for samfunnet å bruke ressurser på å tilrettelegge. Myndighetene kan ikke gi retningslinjer for hvilke data som skal arkiveres, og i hvilken grad

de skal tilrettelegges for gjenbruk. Slike vurderinger og beslutninger må tas av dem som har det nærmeste ansvaret for dataene. Det er liten tvil om at også forvaltningen og næringslivet vil kunne ha stor nytte av at forskningsdata gjøres tilgjengelige for gjenbruk. Forskningsmiljøene må også ta hensyn til denne nytten når de vurderer kostnad opp mot nytte ved å tilrettelegge for gjenbruk.

## 1.5 Forventninger og tiltak

### **Kulturendring, insentiver og kompetanse**

For at forskere skal kunne tilgjengeliggjøre og dele forskningsdata mer, må barrierer reduseres i form av gode støttetjenester, informasjon, opplæring og insentiver. For å understøtte en ønsket kulturendring der datatilretteleggingsarbeid i større grad anerkjennes og krediteres ved søknader om forskningsfinansiering og stillinger, må det legges til rette for at datasett kan siteres, og for at arbeid med forskningsdata synliggjøres bedre.

### **Regjeringen forventer**

- at forskningsinstitusjonene bidrar til å heve kompetansen hos ansatte og studenter gjennom opplæring i datahåndtering og gjenbruk av data
- at utdanningsinstitusjonene vurderer nasjonalt og nordisk samarbeid med henblikk på behovet for å etablere utdanningstilbud for håndtering av forskningsdata (datarøktere o.l.)

### **Regjeringen vil**

- be det nye organet for høyere utdannings- og forskningstjenester om å formidle informasjon om gjeldende krav, regelverk og ressurser knyttet til håndtering av forskningsdata
- gi det nye organet for høyere utdannings- og forskningstjenester koordineringsansvar for å vurdere hvordan nasjonale forskningsdataarkiver best kan tilby opplæring av forskere og studenter om egne tjenester
- be det nye organet for høyere utdannings- og forskningstjenester om å tilrettelegge for sitering av datasett og kreditering av dataarbeid gjennom utstedelse av permanente identifikatorer (DOI-nummer) og digital forsker-ID (ORCID)
- be det nye organet for høyere utdannings- og forskningstjenester om å utarbeide statistikk for å synliggjøre etablering og gjenbruk av forskningsdata

### **Krav om datahåndteringsplaner**

Forskere møter økte krav og forventninger knyttet til håndteringen av forskningsdata fra tidsskrifter, forskningsfinansierer og egne institusjoner. Fra 2017 er åpen tilgang til forskningsdata og krav om datahåndteringsplaner standard i EUs forsknings- og innovasjonsprogram Horisont 2020. I sin oppdaterte policy for tilgjengeliggjøring av forskningsdata krever Norges forskningsråd at godkjente datahåndteringsplaner skal foreligge ved kontraktsinngåelse for prosjekter som genererer data. Dette krever gode prosedyrer og retningslinjer ved den enkelte forskningsinstitusjon.

### **Regjeringen forventer**

- at forskningsinstitusjonene utvikler prosedyrer for (i) godkjenning av datahåndteringsplaner og (ii) beslutninger om at et forskningsprosjekt eventuelt er av en slik art at egen datahåndteringsplan ikke er nødvendig eller hensiktsmessig

### **Regjeringen vil**

- be Norges forskningsråd gjøre godt kjent sitt krav om at en datahåndteringsplan godkjent av forskningsinstitusjonen skal foreligge ved kontraktsinngåelse, der dette er relevant
- be Norges forskningsråd fortsette å være en pådriver nasjonalt og internasjonalt for bedre tilgjengeliggjøring og gjenbruk av forskningsdata
- be forskningsinstitusjonene om å utvikle retningslinjer for forskningsdata som skal lagres, tilgjengeliggjøres og tilrettelegges, med angivelse av hvilke løsninger som bør benyttes for håndtering av ulike datasett
- be det nye organet for høyere utdannings- og forskningstjenester bidra til at tjenestetilbydere av nasjonale forskningsdataarkiver samordner seg og, i samarbeid med forskningsinstitusjonene, bidrar til at løsninger som utvikles for datahåndteringsplaner er mest mulig kompatible, og følger internasjonale standarder

### **Bedre teknisk tilrettelegging og bedre faglig samordning nasjonalt**

Det er behov for større grad av samordning mellom tilbydere av nasjonale forskningsdataarkiver, slik at tjenestene fremstår oversiktlige for brukerne. Det bør finnes en samlet oversikt med informasjon om norske forskeres publikasjoner og datasett. For at forskningsdata skal være gjenfinnbare og gjenbrukbare på tvers av ulike IKT-systemer, kreves bedre teknisk tilrettelegging og gode metadata som følger internasjonale standarder.

### **Regjeringen forventer**

- at forskningsinstitusjoner, forvaltere av forskningsdatainfrastrukturer og fagmiljøer bidrar til standardisering og harmonisering som muliggjør deling og gjenbruk av data i henhold til internasjonale standarder og god skikk innenfor ulike fagområder (for eksempel gjennom å etablere fagområdespesifikke nasjonale kommunikasjonsarenaer)

### **Regjeringen vil**

- be det nye organet for høyere utdannings- og forskningstjenester om å bidra til informasjon om nasjonale forskningsdataarkiver og -tjenester, slik at tjenestene fremstår samlet og lett tilgjengelige for alle forskere og forskningsutførende institusjoner
- be det nye organet for høyere utdannings- og forskningstjenester ta ansvar for å utrede hvordan et nasjonalt vitenarkiv skal realiseres

### **Bærekraftig finansiering og drift av nasjonale forskningsdatainfrastrukturer og forskningsdataarkiver**

Det finnes en rekke mulige finansieringskilder for infrastrukturer for lagring og tilgjengeliggjøring av forskningsdata. Det er imidlertid ikke alltid klart hvilke finansieringsmodeller som vil fungere mest effektivt for å ivareta langsiktig og bærekraftig drift i hvert enkelt tilfelle. I tillegg til særskilte finansieringsvirkemidler for etablering av infrastruktur er det behov for kunnskap om hvordan ulike inntekter og kostnader knyttet til arbeidet med forskningsdata kan kombineres i effektive og bærekraftige forretningsmodeller.

### **Regjeringen forventer**

- at det der det er hensiktsmessig utvikles selvbetjeningsløsninger for å redusere kostnader ved drift av forskningsdatainfrastrukturer gjennom å forenkle prosesser for deponering av og tilgang til forskningsdata og metadata

### **Regjeringen vil**

- videreføre Norges forskningsråds ordning *Nasjonal satsing på forskningsinfrastruktur*
- be Norges forskningsråd, i samarbeid med det nye organet for høyere utdannings- og forskningstjenester og med representanter fra forskningsinstitusjonene, om å utrede og gi råd om bærekraftige finansieringsmodeller for drift av forskningsdatainfrastrukturer og forskningsdataarkiver

### **Bedre tilrettelegging for bruk av offentlige data til forskning**

Datasett generert av ulike offentlige aktører kan ha stor relevans for forskning, blant annet ulike typer registerdata. Noen av disse datasettene er åpne og lett tilgjengelige for forskere, mens tilgang til andre datasett kan være begrenset av en rekke ulike årsaker. Det kan skyldes tekniske forhold, at dataene inneholder opplysninger som tilsier hel eller delvis skjerming eller kostnader knyttet til bearbeiding før de kan gjøres tilgjengelige. Det er et mål at det skal bli enklere for forskere å få tilgang til denne type data.

#### ***Regjeringen vil***

- be offentlige aktører legge til rette for gjenbruk av data til forskning, innovasjon og verdiskaping ved oppgradering og utvikling av systemer for datalagring
- legge vekt på tilgang til data for forskning i arbeid med relevante lov- eller forskriftsendringer

### **Bedre samspill mellom offentlige data og forskningsdata**

Arkiverte og gjenfinnbare forskningsdata representerer verdier som kan styrke den offentlige forvaltningen og dermed gi bedre tjenester tilbake til samfunnet. Det er behov for større grad av nasjonal koordinering for at arkivverdige forskningsdata lagres og håndteres på en slik måte at de i størst mulig grad kommer offentlig forvaltning til gode. Dette krever dialog mellom forskersamfunnene, representanter fra offentlig forvaltning og de som eier og/eller drifter nasjonale infrastrukturer for lagring av både forsknings- og forvaltningsdata.

#### ***Regjeringen vil***

- vurdere å etablere et nasjonalt forum for gjenbruk av forskningsdata med representasjon fra relevante offentlige organer for å få råd i spørsmål om samordning, datakvalitet, tilgangsstyring, rutiner og finansiering av nasjonale datainfrastrukturer som skal betjene både forskere, forvaltere, næringsliv og innbyggere

### **Enklere tilgang til data fra Statistisk sentralbyrå for forskningsformål**

En av Statistisk sentralbyrås (SSBs) oppgaver er å gi opplysninger til statistisk bruk for forskningsformål, innenfor rammene av blant annet personvern og statistisk konfidensialitet. Data fra SSB inngår i en rekke forskningsprosjekter. Det er et mål at denne strategien skal bidra til at det blir enklere å få tilgang til SSB-data til forskningsformål.

#### ***Regjeringen forventer***

- at forskningsinstitusjonene bidrar til bedre bestillerkompetanse hos egne forskere og støttepersonell

#### ***Regjeringen vil***

- utrede en finansieringsmodell for tilgang til data fra Statistisk sentralbyrå til forskning som ivaretar både hensynet til en effektiv bruk av ressurser på datatilgangen og samfunnsmessige gevinster av forskningen
- støtte opplæringstiltak for forskere og støttepersonell i regi av Statistisk sentralbyrå
- vurdere forslag fra Statistikklovutvalget og innspill fra høringen med sikte på å bedre forskeres tilgang til data fra Statistisk sentralbyrå

### **Enklere og sikrere tilgang til helsedata**

Norge har et omfattende kunnskapsgrunnlag i ulike typer helsedata bygget opp over lang tid, blant annet gjennom nasjonale helsetjenester. I regjeringens handlingsplan for oppfølging av HelseOmsorg21-strategien legges det vekt på å gjøre helsedata lettere tilgjengelig for forskning. Helsedatautvalget ble oppnevnt for å gi anbefalinger om et bedre og mer effektivt system for behandling av helsedata for statistikk, planlegging, helseanalyser, kvalitetsforbedring, forskning, innovasjon og næringsutvikling. Utvalget har anbefalt tiltak for å gjøre tilgangen til helsedata enklere og sikrere.

#### ***Regjeringen forventer***

- at databehandlingsansvarlige for helseregistre og Statistisk sentralbyrå gjennom Helsedataprogrammet bidrar til at data i større grad skal kunne analyseres sikkert og effektivt på tvers av datakilder

#### ***Regjeringen vil***

- vurdere forslagene fra Helsedatautvalget og innspill fra høringen med sikte på å etablere et enklere og sikrere system for tilgang til helsedata
- vurdere etablering av en plattform for helseanalyser, jf. oppdrag til Direktoratet for e-helse om å utrede etablering av en slik plattform



# Del 1

Data som er samlet inn  
eller frembrakt som et resultat av  
offentlig finansiert forskning





## 2 Status

### 2.1 Overordnet IKT-politikk og åpne offentlige data internasjonalt og nasjonalt

Økt tilgang til offentlige data, inkludert forskningsdata, er viktig for effektivisering, innovasjon, næringsutvikling og et åpent og demokratisk samfunn.

Europakommisjonen definerer offentlige data som all informasjon som offentlige organer produserer, samler inn eller betaler for. Eksempler er geografisk informasjon, statistikk, værdata, data fra offentlig finansierte forskningsprosjekter eller digitaliserte bøker fra biblioteker.

EUs viderebruksdirektiv 2003/98/EC ("PSI-direktivet") gir det lovmessige rammeverket for tilgang til offentlige data. Direktivet bygger på to av grunnpillarene i EUs indre marked: transparens og rettferdig konkurranse. I tillegg ønsker Kommisjonen å sikre mer effektiv utnyttelse av denne typen data over landegrensene. Direktivet ble revidert i 2013 (2013/37/EU) og omfatter nå også arkiver, biblioteker og museer, samt rett til å få utlevert data i maskinlesbare formater (dersom dette eksisterer). Direktivet gjelder tekstdokumenter, databaser, lydfiler og filmklipp, men omfatter ikke områdene utdanning, forskning og kringkasting. Direktivet inneholder bestemmelser om ikke-diskriminering, gebyrer, avtaler om enerett, gjennomsiktighet, lisenser og praktiske redskaper som kan lette registrering og viderebruk av offentlige data.<sup>1</sup> EUs viderebruksdirektiv er implementert i viderebruksbestemmelsene i offentlighetsloven.

Norsk politikk for tilgjengeliggjøring av offentlige data er beskrevet i Meld. St. 27 (2015–2016) *Digital agenda for Norge — IKT for en enklere hverdag og økt*

*produktivitet*. Regjeringen har videre gitt føringer for gjenbruk og viderebruk av offentlige data i *Retningslinjer ved tilgjengeliggjøring av offentlige data*.<sup>2</sup> En sammenstilling av pålegg og anbefalinger finnes i *Digitaliseringsrundskrivnet*.<sup>3</sup> Dette gjelder for departementene, statens ordinære forvaltningsorganer, forvaltningsorganer med særskilte fullmakter og forvaltningsbedrifter.

*Digital agenda* fastsetter at den enkelte virksomhet skal ha oversikt over hvilke data den håndterer, hva dataene betyr, hva de brukes til, hvilke prosesser de inngår i og hvem som kan bruke dem. Dette innebærer å ta stilling til hvilke data som kan gjøres tilgjengelig for gjen- og viderebruk. Rundskrivnet slår også fast at virksomhetene skal gjøre egnet informasjon tilgjengelig i maskinlesbare og helst standardiserte formater, fortrinnsvis gjennom maskinelle grensesnitt (API-er).

Regjeringens *Retningslinjer ved tilgjengeliggjøring av offentlige data* består av 15 punkter, og skal sikre at tekniske, organisatoriske og juridiske forhold er lagt til rette for en best mulig utnyttelse av offentlige data. Retningslinjene innebærer at bruksvilkårene skal åpne for så mange bruksområder som mulig (gjennom åpne standardlisenser), at data som hovedregel skal være gratis, og at datasettene og datakvaliteten bør dokumenteres.

Allmennheten har innsynsrett i offentlige dokumenter og registre, inkludert data i maskinlesbare formater, innenfor de unntakene som følger av offentlighetsloven. Dette kan beskrives som en reaktiv åpenhet, der innsyn forutsetter at allmennheten har tilstrekkelig kunnskap til å stille relevante innsynskrav. Digitaliseringsrundskrivnet legger opp til en proaktiv åpenhet, der egnet informasjon skal gjøres enkelt tilgjengelig.

<sup>1</sup> Europaparlamentets og rådets direktiv 2013/37/EU av 26. juni 2013 om viderebruk av offentlig informasjon. <http://data.europa.eu/eli/dir/2013/98/2013-07-17>

<sup>2</sup> <https://www.regjeringen.no/no/dokumenter/retningslinjer-ved-tilgjengeliggjoring-av-offentlige-data/id2536870/>

<sup>3</sup> [www.regjeringa.no/digitaliseringsrundskrivnet](http://www.regjeringa.no/digitaliseringsrundskrivnet)

## 2.2 Åpen tilgang til forskningsdata internasjonalt og nasjonalt

På overordnet nivå har det vært enighet om prinsipper og retningslinjer for åpen tilgang til offentlig finansierte forskningsdata i over ti år. I 2006 anbefalte Rådet i OECD prinsipper og retningslinjer for tilgang til offentlig finansierte forskningsdata for å bidra til økt kvalitet og effektivitet i forskning, der dette balanseres mot behov for å skjerme data for å beskytte sosiale eller økonomiske interesser.<sup>4</sup> Prinsippene omfatter blant annet at data bør gjøres tilgjengelig på like vilkår og til lavest mulig kostnad, og at systemer og betingelser for tilgang må ivareta alle legitime interesser, inkludert lovmessig beskyttelse av nasjonal sikkerhet, personvern, immaterielle eiendeler og forretningshemmeligheter. Videre omtaler prinsippene krav til datahåndteringen. Den må følge relevante tekniske standarder og kvalitetskrav, systemene må fungere kostnadseffektivt, de må være mulig å opprettholde over tid og fleksibelt kunne tilpasses teknologiske og andre endringer. Norge sluttet seg til OECDs anbefalinger og retningslinjer for tilgang til forskningsdata gjennom Stortingets behandling av St.meld. nr. 30 (2008–2009) *Klima for forskning*.

Europakommisjonen har under overskriften «Det indre digitale marked» satt blant annet åpen forskning tydelig på den politiske dagsordenen (*Open Science, Open Innovation, Open to the World*).<sup>5</sup> Det europeiske rådet vedtok ambisiøse rådskonklusjoner om åpen forskning i mai 2016. Her uttrykte Europas forskningsministre støtte til Kommisjonens mål om at forskningsdata som er fremkommet gjennom finansiering fra Horisont 2020, som hovedregel skal være åpne med

mindre sikkerhetshensyn, personvern hensyn, kommersielle hensyn eller andre legitime hensyn tilsier noe annet. Rådet ber Kommisjonen arbeide for å fremme god dataforvaltning, inkludert opplæring og økt bevissthet, og om å implementere datahåndteringsplaner som en integrert del av forskningen. Rådet oppfordrer også landene om å utarbeide nasjonale datadelingsstrategier og om å implementere datahåndteringsplaner som standard vitenskapelig praksis i nasjonale forskningsprogrammer. Rådet understreker videre at data må være gjenfinnbare og gjenbrukbare, jf. punkt 3.2 om de såkalte FAIR-prinsippene, og minner om betydningen av langsiktig arkivering og kuratering av forskningsdata, og om at metadata må være basert på internasjonale standarder.<sup>6</sup> Kostnader til datahåndtering og klargjøring av data for deling er støtteberettiget i Horisont 2020. Det er etablert flere rådgivende ekspertgrupper, blant annet *Open Science Policy Platform* (OSPP) og ekspertgruppen for den europeiske forskningsskyen *European Open Science Cloud* (EOSC).<sup>7</sup>

EOSC er en del av det europeiske initiativet som ble presentert av Europakommisjonen i april 2016. Forskningsskyen er tenkt som et interessentdrevet virtuelt miljø av tjenester for lagring, håndtering, analyse og gjenbruk av forskningsdata på tvers av landegrenser og fagområder. Forskningsskyen er tenkt bygget på:

- integrering og konsolidering av e-infrastrukturplattformer (som GÉANT, EGI, PRACE o.l., se tekstboks 2.1)
- en føderasjon av eksisterende forskningsinfrastruktur og forskningsskyer (som dataintensive ESFRI<sup>8</sup> forskningsinfrastrukturer)
- utvikling av skybaserte programvaretjenester for åpen forskning

<sup>4</sup> «Recommendation of the Council concerning Access to Research Data from Public Funding», Organisasjonen for økonomisk samarbeid og utvikling, 14. desember 2006.

<sup>5</sup> <http://ec.europa.eu/research/openscience/index.cfm?pg=home>

<sup>6</sup> «Council conclusions on the transition towards an Open Science system», Rådet for Den europeiske union, 27. mai 2016, dokument 9526/16

<sup>7</sup> <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform>, norsk deltaker: Ernst Kristiansen Sintef, <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

<sup>8</sup> European Strategy Forum on Research Infrastructure (ESFRI): [https://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri](https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri)



## Boks 2.1 Viktige europeiske e-infrastrukturplattformer

E-infrastruktur for forskning er IKT-baserte infrastrukturer som muliggjør avansert og samarbeidsorientert forskning. E-infrastruktur omfatter regneressurser for store beregninger (tungregning), lagringsressurser, høykapasitets datanettverk og tilhørende tjenester, som autentisering og autorisering, verktøy for effektiv arbeidsflyt og programvare for simulering og analyse av data. Begrepet e-infrastruktur benyttes også om digitale registre og databaser, samt verktøy og tjenester for å sikre og gjøre disse tilgjengelige.

E-vitenskap betegner nye modeller, metoder, algoritmer og programvare rettet mot beregnings- og dataintensive vitenskapelige problemstillinger.

GÉANT (Gigabit European Academic Network): leverer høykvalitets nettverkstjenester utenfor «det vanlige internettet» for internasjonale brukere fra FoU-miljøer. Tjenester inkluderer for eksempel IP-nettverk med høy hastighet, skytjenester, VPN-tjenester, virtuelle testområder og et stort Wi-Fi som gir 50 millioner studenter i over 70 land tilgang gjennom en enkel tilgangsprotokoll. UNINETT Sigma2 representerer Norge som et assosiert nettverk.

EGI (European Grid Initiative): forening av mer enn 300 leverandører av data- og skytjenester i Europa (delvis også fra ikke-europeiske land). EGI tilbyr tjenester på forskjellige anvendelsesområder for forskning og innovasjon, som for eksempel skyregning, høyhastighetsberegning, datalagringstjenester, operasjonell løsningsutvikling og støttefunksjoner. Man kan få tilgang til ressursene på ulike måter: gratis søknadsbasert tilgang ut fra prosjektkvalitet, kjøp av medlemskap eller betaling av engangspris for ønsket tjeneste.

PRACE (Partnership for Advanced Computing in Europe): europeisk forskningsinfrastruktur for tungregningsressurser distribuert i Frankrike, Tyskland, Italia, Spania og Sveits. Infrastrukturen har 25 medlemsland, som også tilbyr lokale tungregningstjenester i sine land. Norge er medlem i PRACE, der UNINETT Sigma2 har ansvar for Norges deltagelse. Forskere kan få tilgang til PRACE tungregningsressurser etter søknad (frist to ganger i året), der det kun er kvaliteten på den planlagte forskningen som vurderes.

I 2016 beregnet Europakommisjonen kostnaden for implementering av forskningsskyen til 6,7 milliarder euro. Kommisjonen har antydnet at den vil investere omkring 2 milliarder euro i forskningsskyen, hovedsakelig gjennom Horisont 2020 og EUs neste rammeprogram for forskning og innovasjon. Kommisjonen forventer at resten finansieres gjennom nasjonale offentlige og private investeringer.

På nordisk nivå har landene særlig samarbeidet om tilgang til forskningsdata gjennom Nordforsk. NordForsk har utviklet og støttet samarbeid om såkalt e-infrastruktur og e-vitenskap (se boks 2.1) gjennom initiativene Nordic e-Infrastructure Collaboration (NeIC) og The Nordic eScience Globalisation Initiative (NeGI). NordForsk har også bidratt til samarbeid om data mellom nordiske statistikk-

myndigheter og mellom de nordiske meteorologiske instituttene.

Norges forskningsråd har fra 2000 stilt krav gjennom sine «Generelle vilkår for FoU-prosjekter» om at forskningsdata skal arkiveres på forsvarlig måte i minimum ti år. I 2014 vedtok Forskningsrådet en policy for tilgjengeliggjøring av forskningsdata. Der slås det fast at Forskningsrådet vil sikre at søknader inneholder planer for datahåndtering der det er relevant, og sørge for at disse blir fulgt opp som en del av ordinær prosjektoppfølgning. Forskningsrådet oppdaterte sin policy høsten 2017, se nærmere omtale i punkt 4.2.

Bedre tilgang til forskningsdata må ses i sammenheng med utviklingen når det gjelder åpen tilgang til forskningsresultater. Regjeringen har som mål at alle norske vitenskapelige artikler finansiert

av offentlige midler skal være åpent tilgjengelige innen 2024, og la i august 2017 frem retningslinjer og tiltak som skal å bidra til at målet nås.<sup>9</sup>

## 2.3 Personvern og rettigheter til data

All gjenbruk av forskningsdata og andre offentlige data som er viktige datakilder for forskning vil være underlagt de generelle reglene om personvern, taushetsplikt osv. Hensynet til sikkerhet, immaterielle rettigheter, forretningshemmeligheter o.l. må også ivaretas. Gjenbruk endrer ikke på ansvaret for de som utleverer data, bruker data og/eller lagrer data. I tillegg kommer ansvaret for å følge forskningsetiske normer. Etter forskningsetikkloven av 2017 ligger dette ansvaret både på den enkelte forsker og på forskningsinstitusjonene.

### Personvern

En del forskningsdata inneholder personopplysninger. Tilgjengeliggjøring og gjenbruk av slike data vil derfor, på samme måte som ved ordinær bruk i forskning, utløse spørsmål om ivaretagelse av personvernet. Personopplysningsloven gir generelle regler om behandling av personopplysninger. I tillegg finnes en rekke sektorspesifikke lover og bestemmelser om personvern, slik som helseforskningsloven og helseregisterloven. Mange særlover har dessuten enkelte personvernrelaterte bestemmelser, som taushets- og innsynsbestemmelser. Videre finnes det en rekke veiledere, forskningsetiske retningslinjer med mer som forskere også må kjenne og ta hensyn til.

EUs nye forordning om behandling av personopplysninger (kjent som *GDPR*), trer i kraft i mai 2018. Formålet med forordningen er blant annet å harmonisere medlemslandenes personvernregler, modernisere reglene og styrke personvernet. Justis- og beredskapsdepartementet har foreslått å gjennomføre forordningen i norsk rett gjennom en ny personopplysningslov.<sup>10</sup> Det tas sikte på at den nye loven skal tre i kraft i mai 2018. Forordningen åpner for nasjonale regler på en del områder, blant annet når det gjelder forskning.

På mange områder innebærer forordningen en videreføring av gjeldende norsk regelverk. På andre områder vil vi få nye regler. For virksomheter som behandler personopplysninger, erstattes den tidligere melde- og konsesjonsplikten med en plikt til å vurdere personvernkonsekvenser, til å gjennomføre eventuelle risikoreduserende tiltak og til å inngå forhåndsdrøftinger med tilsynsmyndighetene. Tilsynsmyndighetene skal legge mer vekt på virkemidler som forhåndskonsultering, etterkontroll (tilsyn og enkeltsaksbehandling) og normarbeid. Dette innebærer at de som behandler personopplysninger, må ta mye større ansvar for at behandling skjer i samsvar med regelverket. Forordningen medfører i tillegg at flere virksomheter må ha personvernombud (personvernrådgiver), og at datasystemer skal bygges etter prinsippene for innebygd personvern.

### Rettigheter til data

Mange forskningsinstitusjoner har utarbeidet egne retningslinjer for lagring og for rettigheter til forskningsdata og forskningsresultater. Universiteter og høyskoler har som del av sitt samfunnsansvar å sikre samfunnsnyttene av sin forskning. Dette innebærer blant annet formidling, herunder offentliggjøring og utnyttelse. Etter universitets- og høyskoleloven skal de også bidra til innovasjon og verdiskaping basert på forskningsresultater. Dette er overordnede mål for all offentlig finansiert forskning.

Forskningsinstitusjoner må ha retningslinjer internt som regulerer disse forholdene, og sikre at de har tilsvarende avtaler med ansatte og samarbeidspartnere. Mange har dette alt. For eksempel sier *Reglement om sikring og forvaltning av arbeidsresultater ved UiT* at universitetet som hovedregel har rett til å overta rettigheter til forskningsdata skapt av ansatte i tilknytning til stillingen ved universitetet.<sup>11</sup> *Prinsipper og retningslinjer for forvaltning av forskningsdata ved UiT* sier videre at rettigheter til data skal avtales når forskningsdata genereres i prosjekter finansiert av en tredjepart.<sup>12</sup> Universitetet i Oslo har *Politikk for håndtering av immaterielle rettigheter ved Universitetet i Oslo*.<sup>13</sup>

<sup>9</sup> <https://www.regjeringen.no/no/tema/forskning/artikler/open-tilgang-til-vitskapelege-artiklar/id2567744/>

<sup>10</sup> Høringsnotat av 6. juli 2017

<sup>11</sup> *Reglement om sikring og forvaltning av arbeidsresultater ved UiT* gjeldende fra 19.11.2009, sist endret 15.02.2012. Det vises videre til at reglementet gjelder med forbehold for de ufravelige bestemmelsene i arbeidstakeroppfinningsloven (arbeidstakers rett til godtgjøring mv.) og åndsverkloven (respektretten og opphavsmannens rett til å bli navngitt).

<sup>12</sup> *Prinsipper og retningslinjer for forvaltning av forskningsdata ved UiT* datert 9. mars 2017

<sup>13</sup> *Politikk for håndtering av immaterielle rettigheter ved Universitetet i Oslo*.

Her berøres blant annet forholdet mellom forsker, institusjon og akademiske tidsskrifter.

Forskningsinstitusjonene bør gjennomgå sine retningslinjer og arbeidsavtaler for å sikre at de er utformet slik at de tilrettelegger for institusjonenes gjennomføring av denne strategien og for gjennomføring av de nye retningslinjene for åpen tilgang til vitenskapelige artikler.

## 2.4 Nytt organ for høyere utdannings- og forskningstjenester

I juni 2017 presenterte regjeringen sin plan for omorganisering av kunnskapssektoren og gjorde det i den anledning kjent at oppgaver fra flere ulike virksomheter innenfor høyere utdanning og forskning skal samles i et forvaltningsorgan som skal levere tjenester til høyskoler, universiteter, forskningsinstitutter og helseforetak. CERES blir slått sammen med BIBSYS og enkelte oppgaver fra UNINETT. Disse virksomhetene jobber blant annet med studieadministrative systemer og tjenester, samordning av opptak ved universiteter og høyskoler og andre nasjonale IKT-oppgaver knyttet til høyere utdanning. De jobber også med forskningstjenester for alle de forskningsutførende institusjonene i Norge. Viktige komponenter er forskningsinformasjonssystemet Cristin, arbeid for åpen publisering, kontraktsforhandlinger med tidsskriftseiere og utgivere om tilgang til vitenskapelige artikler og bøker og arbeidet med *BIBSYS Infrastructure for Research Data* (BIRD, se nærmere omtale under punkt 2.5).

Det nye organet for høyere utdannings- og forskningstjenester blir viktig for oppfølging av regjeringens digitaliseringspolitikk, herunder arbeidet med åpen tilgang til forskningsresulta-

ter og forskningsdata. God brukermedvirkning blir avgjørende i dette arbeidet. Det er nødvendig å videreføre og styrke dagens medvirkning knyttet til tjenesteleveransene, og det er behov for god medvirkning på overordnet strategisk nivå. Organets oppgaver skal ivaretas sammen med Norsk senter for forskningsdata (NSD) og UNINETT, som i større grad enn i dag vil bli spisset inn mot leveranse av infrastruktur og støttetjenester.

Kunnskapsdepartementet har fastsatt en *Digitaliseringsstrategi for universitets- og høyskolesektoren for 2017–2021*, som på en del områder også omfatter infrastruktur og støttetjenester til alle forskningsutførende institusjoner.<sup>14</sup> Strategien inneholder målbilder som angir hvilken retning departementet ønsker at utviklingen skal ta, og beskriver tiltak som anses som de viktigste i arbeidet med digitalisering av høyere utdanning og forskning. Strategien stiller tydelige forventninger og krav til etablering av lokale strategier, planer og tiltak for digitalisering og kompetanseutvikling ved den enkelte institusjon. Digitaliseringsstrategien legger også føringer for videre rolle- og arbeidsdeling mellom tjenesteorganet, UNINETT og Norsk senter for forskningsdata (NSD). Denne vil bli endelig avklart i løpet av 2018.

## 2.5 Investeringer i forskningsdatainfrastrukturer

Det finnes en rekke forskjellige infrastrukturer for forskningsdata i Norge, både fagområdespesifikke og generiske, som betjener flere fagmiljøer. De fagområdespesifikke datainfrastrukturene drives som regel enten med støtte fra en vertsinstitusjon eller gjennom prosjektmidler fra Norges forskningsråd eller Horisont 2020 som dekker investeringer og i noen grad drift i etableringsfasen.

<sup>14</sup> <https://www.regjeringen.no/contentassets/779c0783ffee461b88451b9ab71d5f51/no/pdfs/digitaliseringsstrategi-for-universitets-og-hoysk.pdf>

## Boks 2.2 Nasjonal satsning på forskningsinfrastruktur

Siden 2009 har Forskningsrådet gjennom ordningen *Nasjonal satsing på forskningsinfrastruktur* inngått kontrakter for omkring fem milliarder kroner til etablering og fornyelse av norsk forskningsinfrastruktur. Av dette har ca én milliard i perioden 2009–2015 gått til etablering og/eller drift av infrastruktur der hele eller en betydelig del av formålet med infrastrukturen er datahåndtering.<sup>15</sup> Dette omfatter både generisk og fagspesifikk e-infrastruktur, vitenskapelige databaser og samlinger, og tjenester for metodeutvikling og dataanalyse. Noen av infrastrukturene er tverrgående og faller innenfor flere fagområder.

<sup>15</sup> Se oversikt i Forskningsrådets rapport «Tilgjengeliggjøring av forskningsdata og data for forskning. Kunnskapsgrunnlag til KDs arbeid med en nasjonal strategi for tilgjengeliggjøring av data», januar 2017, s 10-19.

Det er fem dataarkiver/infrastrukturer som kan betegnes som generiske, dvs. at de tilbyr tjenester på tvers av de fleste fagområdene. UNINETT Sigma2 AS har etablert *National e-Infrastructure for Research Data* (NIRD), som tilbyr tjenester og kapasitet til alle fagområder som trenger tilgang til avanserte storskalaressurser for lagring, prosessering og publisering av forskningsdata eller søk i digitale databaser og samlinger.<sup>16</sup> Norsk senter for forskningsdata (NSD) er i ferd med å etablere *Norwegian Open Research Data Infrastructure* (NORDi), en ny løsning for opplasting, bevaring og deling av forskningsdata, som skal støtte åpen tilgang til og gjenbruk av data fra samfunnsvitenskapelig og humanistisk forskning og forskning innenfor medisin, helse, klima og miljø.<sup>17</sup> *Tjenester for sensitive data* (TSD) på Universitetet i Oslo (UiO) tilbyr et fullt sett med tjenester fra innsam-

ling av data til analyse, behandling og lagring, i sikrede omgivelser. TSD brukes foruten på UiO også av flere andre nasjonale forskningsinstitusjoner.<sup>18</sup> *UiT Open Research Data* er en generisk infrastrukturteneste for forskere ved UiT, som i tillegg tilbyr tjenesten DataverseNO også til andre norske forskningsinstitusjoner som ønsker et institusjonelt arkiv for forskningsdata. Tjenesten er også åpen for norske enkeltforskere som har behov for et åpent arkiv for arkivering, publisering og sitering av egne forskningsdata, spesielt for å gi et tilbud som oppfyller krav fra tidsskrifter om at bakgrunnsdata skal være tilgjengelig. Partnerinstitusjoner får også tilgang til opplæring, støtte til superbrukere og veiledning/manual for kuratering. BIBSYS BIRD er en teneste for forvaltning av forskningsdata som er utviklet gjennom et samarbeid mellom BIBSYS (nå en del av det nye organet for høyere utdannings- og forskningstjenester) og Handelshøyskolen BI. Infrastrukturen er et tilbud til utdannings- og forskningsinstitusjoner som ønsker å tilby sine forskere et felles verktøy for administrasjon av forskningsdata.

Det er stor variasjon i de fagspesifikke datainfrastrukturene for håndtering, lagring og tilgjengeliggjøring av forskningsdata. Noen er databaser der primærdata fra måleinstrumenter eller fra store undersøkelser lagres og tilgjengeliggjøres. Andre er tjenester der data fra forskningsprosjekter lagres og tilgjengeliggjøres. Noen tilbyr nettsider med innsynsløsninger for data, andre tjenester som tilbyr sikker mellomlagring og analyse av sensitive data, eller søke- og metadatatjenester som forbedrer tilgangen til eksisterende data.

I tillegg til forskningsdatainfrastrukturene over, finnes det også flere offentlig eide datainfrastrukturer/-samlinger som er opprettet for forvaltningsformål, men som i stor grad også benyttes til forskning. Vertsinstitusjonene har jevnt over strukturell finansiering fra staten for å dekke sine formålsbestemte oppgaver, og det varierer hvordan disse dekker kostnadene ved å gi tilgang til data for forskere utenfor etaten selv. Inndekning av

<sup>16</sup> For Uninett Sigma2 kommer en vesentlig andel av inntektene fra medlemsorganisasjonene (Universitetet i Oslo, Universitetet i Bergen, Norges teknisk-naturvitenskapelige universitet og Universitetet i Tromsø – Norges arktiske universitet).

<sup>17</sup> NSD har grunnfinansiering forvaltet gjennom Norges forskningsråd, inntekter gjennom en medlemskaps-/abonnementsmodell som dekker NSDs rolle som personvernombud for rundt 150 forsknings- og utdanningsinstitusjoner. Utover dette har NSD også inntekter knyttet til håndtering av offentlige databaser og registre.

<sup>18</sup> Oppbygging av TSD er finansiert gjennom prosjekter (Nasjonal satsning på forskningsinfrastruktur og diverse forsknings- og utviklingsprosjekter med støtte fra Forskningsrådet) og finansiering fra Universitetet i Oslo. Driften finansieres gjennom brukerbetaling, oppdrag og støtte fra UiO.

disse kostnadene gjøres vanligvis enten gjennom at vertsinstitusjonen finansierer forskerens tilgang til data gjennom egne midler, slik at det oppleves som gratis for forskerne, eller at man tar betalt for tilgang gjennom å ta betalt for de verdiøkende tjenestene knyttet til administrasjon og utlevering av data. Folkehelseinstituttet og Statistisk sentralbyrå (SSB) er eksempler på institusjoner som tar betaling for verdiøkende tjenester knyttet til utlevering av data. (Se nærmere omtale av SSB og helsedataproblematikken i kapittel 5). Tjenester som Nasjonalbiblioteket, Arkivverket og Meteorologisk institutt finansierer i stor grad forskernes tilgang til data gjennom egne midler slik at dette er gratis for forskerne. Enkelte av disse har utviklet selvbetjeningsløsninger for forskerne som sikrer forskerne tilgang til data på en kostnadseffektiv måte.

## 2.6 Status ved forskningsinstitusjonene

### Status ved universitetene og høyskolene

Det varierer hvor langt universitetene og høyskolene har kommet i sitt arbeid med policy og planer for håndtering av forskningsdata. Det generelle bildet er at institusjonene erkjenner at området krever både institusjonspolicy og lederoppmerksomhet. Flere institusjoner har pågående prosesser, men har i mindre grad etablert gjennomgående systemer for operativ støtte til enkeltforskere. Ifølge arbeidsgruppen for en IKT-strategi for universitets- og høyskolesektoren savner mange forskere mer tilrettelagte samhandlingsverktøy som er felles for sektoren, og som lett kan involvere deltakere utenfor sektoren, både nasjonalt og internasjonalt.<sup>19</sup> Det pekes på at forskere har behov for sikker lagring av data også mens forskningsprosjekter pågår, ikke bare etter at et prosjekt er avsluttet.

Flere av institusjonene har utarbeidet informasjon som viser til nasjonale og internasjonale mål og føringer, og til arkiver eller datainfrastrukturer der forskningsdata kan lagres. Universitetet i Tromsø – Norges arktiske universitet (UiT) har kanskje kommet lengst i arbeidet. I september 2017 ble Prinsipper og retningslinjer for forvaltning av forskningsdata ved UiT lansert og Forskningsdataportalen UiT,

en egen operativ tjeneste for lagring av forskningsdata, lokalt åpnet. Av andre eksempler kan nevnes at Universitetet i Oslo nylig har vedtatt Politikk og retningslinjer for forskningsdatahåndtering. Norges teknisk-naturvitenskapelige universitet (NTNU) er i ferd med å lansere en Handlingsplan for åpen tilgang til forskningsdata ved NTNU 2018–2020, og Universitetet i Bergen har etablert en operativ tjeneste for sikker adgang til forskningsdata og e-infrastruktur, SAFE, for å håndtere sensitive data.

Universitets- og høyskolerådet (UHR), som er et samarbeidsorgan for akkrediterte institusjoner under universitets- og høyskoleloven, har ikke utarbeidet noen felles politikk for medlemsinstitusjonene knyttet til økt tilgjengeliggjøring av forskningsdata.

### Status ved forskningsinstituttene

Instituttsektoren er svært mangfoldig, og det er stor variasjon både med hensyn til hvordan instituttene er organisert, hvordan de er finansiert (statlig basisfinansiering av forskningsinstitutter, direkte grunnfinansiering fra departementer, konkurranseutsatte offentlige og private forskningsmidler nasjonalt og internasjonalt), og hvilke data som frembringes og benyttes. Dette gir også et mangfold av behov og muligheter knyttet til lagring og tilgjengeliggjøring av data. For mange av instituttene kan ikke forskningsdata deles åpent umiddelbart, for eksempel fordi de inneholder identifiserbare personopplysninger, forretningshemmeligheter, eller er fremkommet gjennom samarbeid med private aktører med avtaler om konfidensialitet. Generelt er mange av de teknisk-industrielle instituttene og primærnæringsinstituttene opptatt av at deres forskning ofte gir resultater der forskningsdataene bidrar til verdiskaping og har stort kommersielt potensial.

Andre forskningsinstitutter, som hav- og miljøinstitutter som for eksempel forvalter tidsserier, er godt i gang med å få på plass strategier og opplæring for å dele data. Norsk institutt for naturforskning, Norsk institutt for luftforskning, Norsk polarinstitutt og Norsk institutt for bioøkonomi har både retningslinjer og opplæringsstrategier for håndtering av forskningsdata. Artsdatabanken samler ikke selv inn data, men praktiserer full åpen tilgang til de dataene de formidler.

<sup>19</sup> «IKT-strategi for forskning», <https://www.uninett.no/sites/default/files/ikt-strategi-uh-forskning-31012017.pdf> (delrapport 31. januar 2017 til samlerrapporten «IKT-strategi og helhetlige løsninger for norsk universitets- og høyskolesektor», overlevert Kunnskapsdepartementet fra en arbeidsgruppe ledet av Morten Dæhlen. Rapporten utgjorde et viktig grunnlag for regjeringens «Digitaliseringsstrategi for universitets- og høyskolesektoren 2017–2021».



### Boks 2.3 Norsk institutt for luftforskning

Norsk institutt for luftforskning (NILU) har praktisert åpen tilgang til data siden 1973 ved sitt datasenter for internasjonale måleprogrammer. Instituttet har en styrevedtatt IT-strategi med fokus på datasikkerhet, som fastslår prinsippet om åpen tilgang til alle instituttets data. Åpenheten understøtter legitimitet og fremmer bruken av data. Gode eksempler på slik datadeling er EBAS og EVDC.

EBAS er en database for målinger av atmosfærens kjemiske og fysiske egenskaper. Dette inkluderer i hovedsak konsentrasjoner og egenskaper av sporgasser og partikler, og målingene inkluderer parametere knyttet til luftkvalitet og sur nedbør, klimagasser, aerosoler, med mer. EBAS lagrer data på vegne av en rekke nasjonale og internasjonale programmer, alt fra langsiktig overvåking til tidsbegrensede forskningsprosjekter. De aller fleste av

dataene stammer fra programmer som oppfordrer til åpen og ubegrenset bruk av data for ikke-kommersiell bruk. For vitenskapelig bruk er disse dataene fritt tilgjengelig og kan gratis lastes ned fra databasens nettside.

EVDC (ESA Atmospheric Validation Data Centre) er den europeiske romorganisasjonens offisielle sentrale database for langtidslagring av såkalte korrelative atmosfæredata. Observasjoner fra fly, ballonger, bakkestasjoner med mer fra hele verden lagres i databasen og brukes deretter av forskere for validering av atmosfæremålinger fra satellitter. Databasen er tilgjengelig online og tilgang til data er gratis, forutsatt at brukerne godtar EVDCs datapolicy. Foruten tilgang til korrelative data tilbyr EVDC verktøy for ekstraksjon, konvertering og lagring av store mengder satellittdata.

De samfunnsvitenskapelige forskningsinstituttene samler inn store mengder data. Der det ikke er krav til sletting ut fra personvern hensyn finnes det gode prosedyrer for deponering av data hos Norsk senter for forskningsdata (NSD). Dette gjøres med store datasett som for eksempel valgforskningsdata.

Forskningsinstituttene følger utviklingen nøye, ikke minst internasjonalt, gjennom sin europeiske organisasjon European Association of Research and Technology Organisations (EARTO).

Forskningsinstituttens fellesarena (FFA) organiserer de instituttene som inngår i basisfinansierings-systemet for forskningsinstitutter. FFA har i innspill til Kunnskapsdepartementet uttrykt at det bør stimuleres til datadeling, gitt at behovet for å håndtere datamangfold ivaretas på en god måte.

FFA peker på at digitaliseringen av økonomien innebærer at mer deling av data vil komme, og noen institutter deltar i utvikling av internasjonale plattformer etablert av store næringsaktører for slik deling. Et eksempel er for eksempel SINTEFs deltakelse i *Industrial Data Space*.<sup>20</sup> Aktørene i slike internasjonale plattformer fremhever at disse vil kunne akselerere dataøkonomien dersom de bidrar til økt tillit, sikkerhet og utviklingen av forretningsmodeller, og samtidig tillater at bedriftene beholder kontrollen over egne data.

#### Status i helseforetakene

Helseforetakene har i dag ingen felles policy, prinsipper eller retningslinjer for forvaltning av forskningsdata. Det arbeides med retningslinjer for datahåndtering i kliniske studier i regi av NORCRIN, den nasjonale forskningsinfrastrukturen for gjennomføring av kliniske studier.

<sup>20</sup> *Industrial Data Space* er en sammenslutning av virksomheter, med en kjerne av store tyske bedrifter, som har gått sammen for å dele data. <http://www.industrialdataspace.org/en/the-association/#mission>. Et annet eksempel er *proDataMarket*, et samarbeid om eiendomsdata finansiert av Horisont 2020, <https://blog.prodatamarket.eu/about/>.

# 3 Grunnprinsipper

## 3.1 Grunnprinsipp 1: Forskningsdata skal være så åpne som mulig, så lukkede som nødvendig

Det første grunnprinsippet for tilgjengeliggjøring og deling av forskningsdata er at data skal være så åpne som mulig, og så lukkede som nødvendig. Åpenhet er ikke en enten-eller-kategori i denne sammenhengen. Det finnes grader av åpenhet langs flere dimensjoner som omhandler hvilke data som blir tilgjengeliggjort, hvem som får tilgang, når tilgang gis og på hvilke vilkår. At data skal være så lukkede som nødvendig, reflekterer legitime behov for ulike former for begrensning av tilgangen til enkelte typer data. I noen tilfeller kan lover og regelverk gjøre det nødvendig å skjermes data, eller kun gi tilgang på bestemte vilkår. I andre tilfeller kan skjerming være nødvendig for å ivareta bestemte interesser, selv om offentliggjøring eller åpen tilgang ville vært lovlig. Noen hensyn som gjør at data ikke *kan* gjøres åpent tilgjengelig, er blant annet:

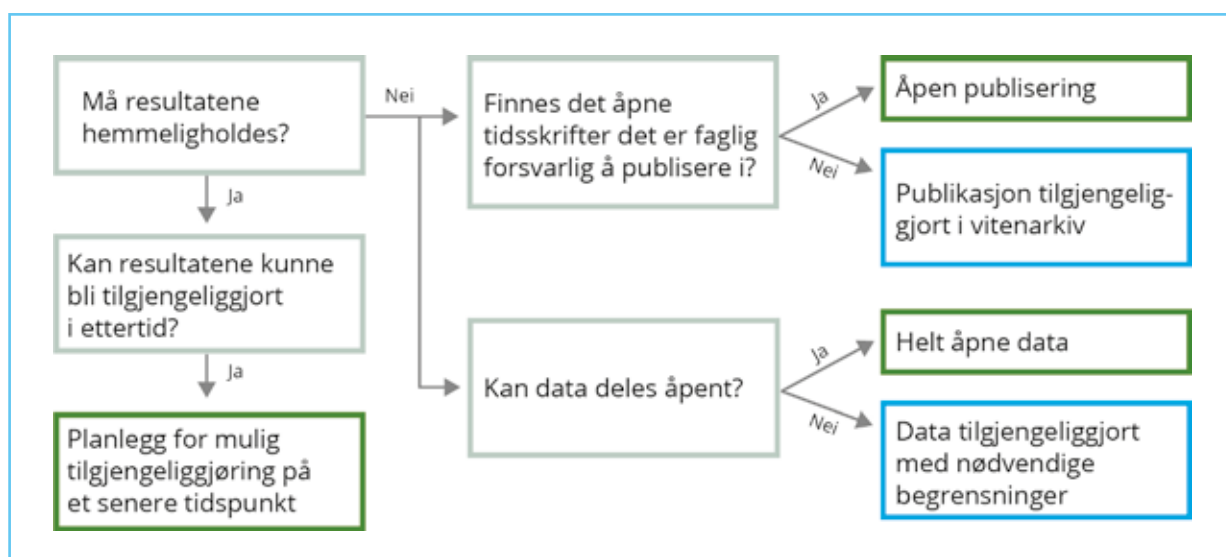
- **Sikkerhet:** dersom det vil kunne skade noens sikkerhet (inkludert for eksempel truede arter) eller nasjonale sikkerhetsinteresser
- **Personvern:** dersom det bryter med gjeldende personvernregler eller kan gå ut over noens rett til privatliv og personlig integritet

- **Immaterielle rettigheter:** dersom det bryter med opphavsretten til åndsverk eller andre immaterielle rettigheter
- **Forretningshemmeligheter o.l.:** dersom det vil avsløre noens forretningshemmeligheter eller lignende konfidensiell informasjon

Selv om noen av hensynene over gjør seg gjeldende i et forskningsprosjekt, trenger ikke det innebære at *alle* forskningsdata må skjermes. Personvern hensyn kan for eksempel ivaretas dersom data kan gjøres tilgjengelig uten å kunne knyttes til enkeltpersoner, eventuelt i form av aggregerte data.

Figur 3.1. illustrerer prinsippet om at det i et forskningsprosjekt må tas ulike beslutninger om grad av åpenhet knyttet til forskningsresultatene. Et overordnet spørsmål er om resultatene skal tilgjengeliggjøres gjennom publikasjoner eller lignende, eller skjermes (i hvert fall for en viss tid), for eksempel av sikkerhetsgrunner eller kommersielle hensyn. Dersom forskningsresultatene i seg selv skal skjermes, er det ofte ikke naturlig med åpenhet om dataene heller. For forskningsresultater som skal publiseres, må det vurderes om også forskningsdataene kan gjøres tilgjengelige i en eller annen form. Dersom ingen legitime hensyn står i veien for offentliggjøring, skal forskningsda-

Figur 3.1 Skjematisk oversikt over sentrale beslutninger for åpen tilgang til offentlig finansierte forskningsdata og artikler



taene tilrettelegges for åpen tilgang, og ellers må det vurderes om det kan gis tilgang innenfor nærmere angitte rammer. Det må også vurderes om det kan være aktuelt å gi (økt) tilgang til resultatene på et senere tidspunkt. Denne typen beslutninger knyttet til håndtering av forskningsdata må inngå i prosjektenes datahåndteringsplaner, se 4.2.

Åpenhet innebærer ikke nødvendigvis fri tilgang for alle. I tilfeller der full åpenhet ikke er aktuelt, kan det gis tilgang på bestemte vilkår, for eksempel gjennom autentisering, autorisering eller med særskilte krav til datahåndteringen (som sikker lagring, kontroll av utdata o.l.). Ved slik tilgangsstyring er det viktig med likebehandling av alle aktører som oppfyller vilkårene for tilgang.

Behovet for å skjerme data vil ofte endre seg over tid, slik at forskningsdata kan gjøres tilgjengelige på et senere tidspunkt. Immaterielle rettigheter har for eksempel begrenset varighet, og behovet for å skjerme personopplysninger kan avta med tiden, jf. bestemte sperrefrister for når arkiver med persondata kan åpnes.<sup>21</sup> Videre kan den teknologiske utviklingen gjøre forretningshemmeligheter irrelevante etter relativt få år. Opplysninger om produkter eller prosesser kan være utdatert eller konkurrenter kan allerede ha tilsvarende opplysninger, slik at det ikke lenger er av konkurransemessig betydning å hemmeligholde dataene.

Et annet spørsmål knyttet til åpenhet dreier seg om vilkår for hva andre kan gjøre med data som er gjort tilgjengelige. Tidsskriftsutgivere og institusjoner kan ha ulike retningslinjer knyttet til hvilke lisenser for datalagring som skal eller bør brukes, og det er viktig med et bevisst forhold til hva lisensene innebærer. Her er det i stor grad snakk om å velge mellom ulike åpne standardlisenser, som Norsk lisens for offentlige data (NL0D) eller lisenser fra Open Data Commons eller Creative Commons.

Det er ønskelig at det benyttes lisenser som åpner for videst mulig gjenbruk av forskningsdataene. En viktig begrunnelse er at den som samler inn dataene, ikke nødvendigvis er den beste til å se potensialet for hvordan de kan brukes på nye måter, for eksempel til verdiskaping. Forskere må også være bevisst hva som kan utgjøre data for andre. For én forsker er data det som ligger til grunn for en publikasjon, mens publikasjonen i seg selv utgjør data for en annen. Det er derfor viktig å legge til rette for tekst- og datautvinning.

At lisensene bør være åpne, betyr imidlertid ikke at det ikke kan stilles vilkår til hvordan dataene kan og eventuelt ikke kan brukes. De fleste aktuelle lisensene krever for eksempel at lisensgiveren blir navngitt ved videre bruk, at det blir opplyst om endringer dersom det blir dannet derivater, og at lisensgiver ikke blir fremstilt som ansvarlig for disse. Lisensene kan også tydeliggjøre at de ikke innebærer tillatelse til gjenbruk som ellers ikke er tillatt (for eksempel av personvern hensyn).

Prinsippet om åpenhet om data innebærer ikke at hver forsker selv må arkivere kopier av alle data som er brukt og stå for tilgjengeliggjøringen av disse. I forskning som hovedsakelig er basert på åpent tilgjengelige data og hvor de aktuelle bearbeidingene lett kan beskrives og gjenskapes, vil det kunne være mer relevant å tilgjengeliggjøre disse prosedyrene med referanser til de opprinnelige kildene. I noen tilfeller vil forskeren ikke ha mulighet til å arkivere data over tid, for eksempel data som man bare kan få tilgang til under vilkår med krav om sletting eller behandling i sikrede omgivelser. I slike tilfeller vil prinsippet om åpenhet dreie seg om å dele informasjonen som er nødvendig for at andre forskere kan gjenskape eller få tilgang til dataene under samme vilkår. Når datasett er generert ut fra administrative datasystemer, kan det for eksempel være snakk om å arkivere og tilgjengeliggjøre koder for å gjenskape de samme uttrekkene.

<sup>21</sup> Personopplysningsloven gjelder ikke avdøde personer. Helseregisterloven og pasientjournalloven fastsetter at helseopplysninger om avdøde personer er taushetsbelagte og underlagt personvernreglene. Forskningsetiske retningslinjer fra den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora pkt. 17 omtaler behov for aktsomhet og respekt ved forskning på avdøde personer.



### Boks 3.1 Norsk lisens for offentlige data

Norsk lisens for offentlige data (NLOD) er en lisensavtale som er anbefalt brukt ved tilgjengeliggjøring av åpne data fra det offentlige.<sup>22</sup> Lisensen ble utarbeidet av det daværende Fornyings- og administrasjonsdepartementet i 2011 og er tilgjengelig på norsk og engelsk. NLOD er basert på prinsippene om navngivelse (krav om å oppgi kilde) og ansvarsfraskrivelse (utgiver gir ingen garantier for feil eller mangler). Fra og med januar 2017 er *Creative Commons Navngivelse 4.0* (CC-BY 4.0) sidestilt med NLOD som anbefalt lisens.<sup>23</sup>

<sup>22</sup> Se *Retningslinjer* ved tilgjengeliggjøring av offentlige data, pkt 1. <https://data.norge.no/nlod/no>

<sup>23</sup> <https://creativecommons.org/licenses/by/4.0/deed.no>

## 3.2 Grunnprinsipp 2: Forskningsdata bør håndteres og tilrettelegges slik at verdiene i dataene kan utnyttes best mulig

Det er ikke tilstrekkelig med åpen og sikker lagring av forskningsdata. Dataene utgjør bare en verdi i den grad de tas i bruk, og da må de holde høy kvalitet og være reelt gjenfinnbare og gjenbrukbare.

Å legge til rette for at verdiene i forskningsdata kan utnyttes handler for det første om å senke terskelen for å gjenbruke dem, og for det andre om å tilføre data kvaliteter som gjør at de kan komme til nytte på forskjellige bruksområder.

Et sett av prinsipper for god tilrettelegging for videre bruk av forskningsdata, er etablert i de

internasjonale FAIR-prinsippene, et akronym basert på ordene *findable*, *accessible*, *interoperable*, og *reusable*.<sup>24</sup> FAIR-prinsippene er rettet både mot å redusere tekniske og administrative hindre for tilgang og gjenbruk, og mot at data skal tilfredstille ulike kvalitetskrav. De legger også stor vekt på at data befinner seg i et digitalt økosystem, og at de må kunne håndteres maskinelt.

For at data skal være gjenfinnbare, bør de for eksempel være utstyrt med en entydig og bestandig identifikator, som gjør det mulig å referere og peke til dem (i praksis DOI-nummer, se nærmere omtale under punkt 4.1). De bør også være tilgjengelige i kataloger og søketjenester med beskrivende informasjon som gjør det mulig å oppdage dataene også ut fra deres kjennemerker.

For at data skal være tilgjengelige må de lagres et sted der andre kan finne dem, og det må legges til rette for at flest mulig kan få tilgang til dem. En potensiell terskel for videre bruk er dersom datatilgangen ikke er gratis. Prisen for tilgang bør derfor være lavest mulig, og helst ikke overstige grensekostnaden for brukerens tilgang. For offentlig finansierte forskningsdata er det sentrale prinsippet at det offentlige allerede har finansiert innsamlingen av data, og at disse dermed bør anses som et offentlig gode. Samtidig er det viktig at det er mulig å opprettholde tilgangen over tid, og utenom kostnadene til selve datainnsamlingen kan det også være lagt inn betydelige grunnlagskostnader for at det skal være mulig å gi tilgang. For å være gjenbrukbare bør dataene være lagret i et format som gjør informasjonen tilgjengelig og mulig å håndtere fleksibelt. Både data og metadata bør være så rikholdige som mulig. Dette innebærer at detaljer knyttet til innsamling, avgrensninger, definisjoner og annet som er relevant for videre bruk må være beskrevet. Datasettene bør følge relevante standarder, med strukturerte vokabularer, identifikatorer og referanser med mer for at de enklere kan forstås i forskerfellesskapene, håndteres maskinelt og brukes sammen med andre datasett. I tillegg bør datasettene være utstyrt med en (standardisert) lisens som gjør det klart hva dataene kan brukes til.

<sup>24</sup> Wilkinson, Mark D. mfl. (2016) «The FAIR Guiding Principles for scientific data management and stewardship». *Scientific Data*. 3(160018). <https://doi.org/10.1038/sdata.2016.18>. <https://www.force11.org/group/fairgroup/fairprinciples>

### 3.3 Grunnprinsipp 3: Beslutninger om arkivering og tilrettelegging av forskningsdata må tas i forskerfellesskapene

Det er krevende å gjøre forskningsdata tilgjengelige og verdifulle for gjenbruk på en god måte. Å tilrettelegge dataene og systemene for tilgang fullt ut i tråd med beste praksis for prinsippene nevnt over er svært kostnadskrevenne. Selv om kostnader knyttet til digital lagring og håndtering av data stadig synker, genereres det til gjengjeld stadig større datamengder som skal lagres. Det kreves også betydelige investeringer for at dataenes sikkerhet og bruksverdi kan opprettholdes over lang tid, blant annet i form av oppdatering av formater o.l.. Disse kostnadene, kombinert med det enorme og økende omfanget av forskningsdata, gjør at ikke alle data kan arkiveres og vedlikeholdes like lenge, og ikke alle data kan eller bør tilrettelegges for gjenbruk med samme innsats og kvalitetskrav.

Alle forskningsdata kan ikke behandles på samme måte. Hvor mye ressurser det gir mening å investere for å tilrettelegge for gjenbruk vil variere, og disse investeringene må ses i forhold til den potensielle verdien dataene utgjør i dag og i fremtiden. Det er ikke alle forskningsdata som kan lagres utover det konkrete prosjektet, selv om det er lovlig å gjøre det. Dette gjelder særlig noen forskningsområder der det skapes enorme mengder rådata. Av alt som lagres, er det bare noe som utstyres med identifikatorer, noe som kan offentliggjøres, og noe som kan tilrettelegges særskilt og vedlikeholdes for permanent lagring. Ikke alle data og metadata er like rikholdige, og investeringene i utvikling av standarder og detaljerte koblinger vil ikke være like store på alle områder. Noen, men ikke alle, datasett vil gjennomgå fagfellevurderinger, autoriseringer, eller lignende.

Selv om regjeringen forventer adskillig bedre tilgjengeliggjøring og større gjenbruk av forskningsdata enn i dag, så kan ikke myndighetene vurdere hvilke data det er verdt å arkivere og gjøre gjenbrukbare. Dette er spørsmål som forskere selv må ta stilling til gjennom diskusjoner i de institusjonelle, nasjonale og internasjonale forskerfellesskapene om hva som er hensiktsmessig og hva som vurderes som god skikk innenfor forskjellige fagområder. På noen fagområder har dette arbeidet kommet langt, og det er

stor grad av enighet og etablert praksis for hvordan forskningsdata skal håndteres, og hvilke datainfrastrukturer eller dataarkiver som bør brukes. Dette kan være infrastrukturer eller arkiver nasjonalt eller internasjonalt. På mange fagområder har vi for eksempel tilgang til internasjonale arkiver gjennom Norges medlemskap i paneuropeiske forskningsdatainfrastrukturer (som ESFRI-infrastrukturer). På andre fagområder har arbeidet kommet kortere, og fagmiljøene har først i de senere årene begynt diskusjonene om datadeling nasjonalt og internasjonalt.

Kostnadene ved å gjøre datasett reelt gjenbrukbare (og ikke bare gjenfinnbare) må veies opp mot nytten for forskerfellesskapene og samfunnet ved at dataene kan gjenbrukes. Fordi det kan være ressurskrevende for forskningsinstitusjonene å gjøre forskningsdata gjenbrukbare, vil forskernes naturlige hovedhensyn i vurderingene av hvilke data det er viktig å arkivere og kuratere være nytten for forskerfellesskapene. Mange av kjennetegnene ved data som er velegnet for gjenbruk, er ikke bare til nytte når data deles med andre, men kan også øke kvaliteten eller effektiviteten på databruken til den som har produsert dem. Samtidig kan tilrettelegging for åpen tilgang kreve merarbeid av både teknisk og administrativ art som dataprodusentene selv ikke har noen direkte glede av. Det er derfor viktig at man tar hensyn til den potensielle nytten for andre forskere i beslutningene om deling og datahåndtering. Det er også viktig at forskere som gjør data gjenbrukbare, belønnes for nytten disse dataene har for andre gjennom anerkjennelse i forskningens systemer for verdsetting. Det er viktig at data er gjenfinnbare og siterbare for at bruken av data skal kunne registreres, og at anerkjennelsen tilfaller opphavsmiljøene, se 4.1.

Data som før ble brukt bare av forskere, blir i økende grad anvendt også av andre, i næringsliv, forvaltning, undervisning og brukergrupper som man tidligere ikke så for seg ville utvikle interesse for disse dataene eller kompetanse til å ta dem i bruk. Det er liten tvil om at også forvaltningen og næringslivet vil kunne ha stor nytte av at forskningsdata gjøres tilgjengelige for gjenbruk. Forskningsdatasett er også et læremiddel, og forskningsdatainfrastrukturene bør også legge til rette for at forskningsdata kan brukes i undervisning. Forskningsmiljøene bør ta hensyn til denne nytten også og vurdere den opp mot kostnadene ved å tilrettelegge for gjenbruk. Det er ikke rimelig

å forvente at forskere vil kunne se for seg detaljene i hvordan data kan blir gjenbrukt utenfor forskersamfunnet, og de kan derfor ikke forventes å tilrettelegge spesifikt for dette. Men i all hoved-

sak kan man forvente at det som skal til for å gjøre data gjenbrukbare for forskning, også vil gjøre data verdifulle for andre formål.

### Boks 3.2

#### Eksempel på forvaltningens behov for forskningsdata: Mattilsynet

Mattilsynet er et statlig, landsdekkende forvaltningsorgan som skal ivareta trygg mat og trygt drikkevann, fremme dyre- og plantehelse og miljøvennlig produksjon og etisk forsvarlig hold av fisk og dyr på tvers av flere sektorer. Mattilsynet har også et internasjonalt ansvar for å delta i utviklingen av regelverk og standarder på sitt område (EU). Mattilsynet mottar videre vitenskapelige vurderinger av nytte- og risiko og kunnskap fra Vitenskapskomiteen for mat og miljø (VKM). Mattilsynet og VKM er avhengig av vitenskapelig dokumentasjon og forskningsdata for å utføre sine oppgaver nasjonalt og internasjonalt. Jo bedre tilgang til forskningsdata disse får, jo bedre vurderinger kan de gjøre. For Mattilsynet og VKM er det derfor viktig med kompatibilitet og effektive systemer som åpner for datadeling og gjenbruk av forskningsdata.

Det er imidlertid ikke sikkert at resultatet av forskerfellesskapenes kost-nytte-vurderinger gir det ønskede resultatet også for forvaltningen eller næringslivet. Der forvaltningen eller andre brukergrupper ser behov for at data fra forskning gjøres gjenbrukbare utover hva det kan forventes at forskningsmiljøene tar ansvar for, må det utvikles bærekraftige finansieringsmodeller som kan dekke den ekstrakostnaden det innebærer å gjøre dataene tilgjengelige og av en slik kvalitet at de kan gjenbrukes for andre formål. Dette gjøres allerede i dag på flere områder, som for eksempel i miljøforvaltningen, i havforvaltningen eller i helseregistrene. Flere miljø- og havforskningsinstitutter har et dobbelt formål og samler inn data for både forskning og forvaltning, som gjøres åpent tilgjengelige for alle. Behovet for tilsvarende ordninger også på andre områder er et viktig spørsmål som antagelig vil måtte vurderes nærmere innenfor flere sektorer fremover.

### Boks 3.3 Hvilke typer data skal tilgjengeliggjøres?

I denne strategien er begrepet «forskningsdata» ment å omfatte nedtegnelser i ethvert format som brukes som grunnlag for argumentasjon, tolkninger eller beregninger i forskning. Data kan være numeriske eller tekstlige eller registreringer av lyd, bilde osv. Strategien og prinsippene om tilgjengeliggjøring og tilrettelegging av data gjelder med andre ord både kvantitative og kvalitative data, og både data fra simuleringer eller empiriske data i form av målinger eller nedtegnede observasjoner. Data kan være enkeltvis og direkte nedtegnelser («rådata»), eller bearbejdede data som kan ha gjennomgått flere ledd av transformasjoner og sammenstillinger. De kan finnes i ustrukturerte samlinger eller være tilrettelagt i databaser.

Utover de generelle grunnprinsippene gir strategien ingen detaljerte retningslinjer for avgrensningen av akkurat hvor mye eller hvilke av forskningsdataene knyttet til en gitt aktivitet som skal arkiveres eller tilgjengeliggjøres. Dette må avgjøres av forskerne og de som håndterer dataene, innenfor de retningslinjer som er satt av de relevante institusjonene, finansieringsorganene, dataarkivene eller andre.

Et sentralt minimumskriterium for hva som bør inkluderes, er det som er nødvendig for å etterprøve forskningen. Andre relevante hensyn kan være i hvilken grad datasettene lar seg gjenskape, og den potensielle nytteverdien man vurderer at disse kan ha for andre. For data fra simuleringer vil det i mange tilfeller kunne være mest hensiktsmessig å tilgjengeliggjøre selve modellen og programvaren som genererer data, mens terskelen for å arkivere data er lavere når det er snakk om unike observasjonelle data. For datasett som kun innebærer trivielle bearbejdingar av data som allerede er tilgjengeliggjort på en hensiktsmessig måte, kan det være mer relevant å referere originalkildene og forklare hvordan de er brukt, enn å arkivere bearbejdsen. På

hvilket bearbejdningsnivå det er mest relevant å tilgjengeliggjøre data, må også bero på en pragmatisk vurdering ut fra de samme hensynene. I mange tilfeller er det gunstig å inkludere så detaljerte og ubearbejdede data som mulig, selv om et høyere bearbejdningsnivå kunne vært tilstrekkelig for å etterprøve selve funnene. Det kan for eksempel gi større rom for nye bruksmåter som forskeren selv ikke hadde sett for seg. Dersom bearbejdsen ikke fjerner så mye informasjon eller det vil være svært kostbart å inkludere rådata, kan et mer bearbejdet datasett være mer relevant og hensiktsmessig. På mange områder kan det være aktuelt å avlede flere datasett, der tilgangen til de mer detaljerte dataene må kontrolleres strengt, mens et mer bearbejdet datasett kan tilgjengeliggjøres fritt.

Kvalitative data kan reise noen særegne problemstillinger. Når data primært er ment å utgjøre grunnlaget for analytiske tolkninger fremfor maskinell behandling, kan strukturen i dataene lett bli mer idiosynkratisk og avhengig av intuitive tolkninger. At data kan være uryddige og potensielt vanskelige å forstå for andre, er imidlertid ingen god grunn for å unnlate å gjøre dem tilgjengelige. I det minste vil den tilretteleggingen av dataene som har gjort dem brukbare i det aktuelle forskningsarbeidet, gjøre dem relevante for å forstå og etterprøve dette konkrete arbeidet for fagfeller som skal vurdere kvaliteten.

For både kvantitative og kvalitative data er det sentralt at datahåndteringen fra starten av gjøres ut fra en plan som ivaretar grunnprinsippene om tilgjengeliggjøring og tilrettelegging. Det som gjør data gode for gjenbruk, gjør dem som regel også gode for bruk generelt, og forskeren selv er den eneste som er garantert å dra nytte av at dataene er strukturert og dokumentert godt fra første stund. Forskerne som har produsert dataene, kan også selv fort kjenne seg igjen i rollen som «gjenbruker» når de ser tilbake på sine egne data etter noe tid.

# 4 Kulturendring, insentiver, kompetanse, datahåndteringsplaner, teknisk tilrettelegging og finansieringsmodeller

Til tross for en generell positiv holdning til deling av data blant mange forskere viser DAMVADs rapport om deling og arkivering av offentlig finansierte forskningsdata i Norge at det likevel er mange grunner til at deling ikke skjer. De viktigste er mangel på infrastruktur og standarder for data og metadata, manglende kompetanse og manglende kultur for tilgjengeliggjøring av egne data for gjenbruk.<sup>25</sup> Tillit ser ut til å være et nøkkelbegrep. Data-deling forutsetter tillit til at dataene blir håndtert på riktig måte av dem man deler med, tillit til at man selv vil få anledning til å publisere resultater og høste akademiske frukter av datainnsamlingen, og kanskje viktigst, tillit til at datatilretteleggingsarbeidet vil bli anerkjent og kreditert av andre forskere ved søknader om forskningsfinansiering og stillinger.

For å øke tilgjengeliggjøring og gjenbruk av forskningsdata er det behov for å understøtte den kulturendringen som er i gang. Det er behov for endringer i akademiske insentiver, økt kompetanse, datahåndteringsplaner, bedre teknisk tilrettelegging og bærekraftige finansieringsmodeller.

## 4.1 Kulturendring, insentiver og kompetanse

For forskere på noen fagområder vil deling og tilgjengeliggjøring av forskningsdata være noe relativt nytt og ukjent. For forskere på andre områder er datadeling en selvfølgelig del av forskningsprosessen, og arbeidet med harmonisering og standarder har kommet mye lenger. Når det gjelder arkivering av forskningsdata, var det i DAMVADs undersøkelse blant norske forskere fra 2014 så mye som 85 prosent av respondentene som ikke plasserte

data i dedikerte dataarkiver etter at prosjektet var ferdig, men som lagret egne data utelukkende på personlige enheter eller på institusjonens ordinære servere. Andelen varierte lite på tvers av sektorer, forskningsfelter og forskererfaring.

### Insentiver for lagring og deling av data

Sett fra forskeres synspunkt er det i en del fag fremdeles slik at det er få insentiver knyttet til å dele data, mens arbeidet med å gjøre data tilgjengelig for andre forskere kan være omfattende, og gå på bekostning av andre aktiviteter som i større grad fremmer den enkeltes vitenskapelige karriere. Det er en viss interessekonflikt mellom forskere og samfunnet på dette punktet. Mer datadeling og gjenbruk er åpenbart i samfunnets interesse, mens dagens vurderingsregimer i akademia kan trekke i motsatt retning og forskere kan tenkes å la være å gjøre data gjenbrukbare av konkurransegrunner.

Forskere bør ikke tape karrieremessig på å bruke tid på å gjøre kvalitetssikrede, verdifulle datasett tilgjengelige for andre forskningsmiljøer og øvrige aktuelle brukergrupper. Etter hvert som det blir enklere og vanligere å dele data, erfarer stadig flere forskere at god kunnskap om forskningsdata som er tilrettelagt for gjenbruk, gjør dem attraktive som samarbeidspartnere i gode, internasjonale forskningssamarbeid.

Det er forskerfellesskapene selv som må bidra til at det blir attraktivt å tilrettelegge viktige data for gjenbruk. Forskere kan blant annet velge å legge vekt på og premiere gode datasett som viktige bidrag til forskningen ved søknader om forskningsfinansiering og stillinger.

<sup>25</sup> DAMVAD, «Sharing and archiving of publicly funded research data: Report to the Research Council of Norway», april 2014.

Myndighetene på sin side har ansvar for å legge til rette for at forskere skal kunne premieres for publisering av datasett. Datasettene må kunne identifiseres og siteres, og forskerne må kunne knyttes entydig til de relevante datasettene.

Norge er medlem av DataCite, en internasjonal ikke-kommersiell organisasjon som gir medlemsorganisasjonene mulighet til å utstede DOI-nummer til forskningsdata. I Norge vil medlemsskapet og utstedelsen av DOI-nummer til forskningsdata ivaretas av det nye organet for høyere utdannings- og forskningstjenester. DOI står for digital object identifier, og er en internasjonalt standardisert permanent identifikator, som blant annet gjør det enklere og mer treffsikkert å både sitere og gjenfinne forskningsdata og publikasjoner. Den som tildeler DOI, forplikter seg til langtidslagring av objektet, minimum i ti år, og til å vedlikeholde korrekt URL til presentasjonssiden og korrekte metadata til objektet.

Norge er også medlem av ORCID, en internasjonal ikke-kommersiell organisasjon som tilbyr permanente identifikatorer for forskere. ORCID gir forskere større kontroll over opplysninger om hvilke publikasjoner og data de har medvirket til, og gjør det enklere å dele og gjenbruke disse opplysningene, for eksempel i CV-er eller rapporteringer til finansieringsorganer. Identifikatoren gjør det også mulig å knytte informasjonen om den enkelte publikasjon eller datasett til den riktige forskeren, slik at man både kan gå fra en enkelt forsker eller publikasjon og se datasettene som vedkommende har frembrakt eller brukt, og gå fra et datasett til å se forskere eller publikasjoner som har en forbindelse til dette.

Gjennom medlemskapene i DataCite og ORCID har myndighetene lagt til rette for at norske forskere kan kobles til datasettene de publiserer, og at datasettene kan siteres på en effektiv måte. Viktige rammebetingelser for at forskere skal kunne krediteres og premieres for publisering av datasett er dermed i ferd med å komme på plass.

En del forskere og institusjoner har foreslått at det bør innføres et økonomisk insentiv i forskningsinstitusjonenes finansieringsmodeller for publisering av datasett, på linje med insentivet for publisering av vitenskapelige artikler og bøker. Per dags dato har ikke myndighetene den kunnskapen og den datakvaliteten i institusjonenes rapportering som er nødvendig for å kunne vurdere et slikt

insentiv. På sikt kan dette være en mulighet som kan bidra som insentiv for datadeling.

Det vil være behov for å overvåke hva slags type data som faktisk blir gjenbrukt og har verdi over tid. Oversikt over gjenbruk kan også gi et viktig bidrag til forskningsinstitusjonenes kost-nytte-vurderinger av hvilke forskningsdata det er hensiktsmessig å tilgjengeliggjøre og tilrettelegge for gjenbruk.

### **Sperrefrister før forskningsdata tilgjengeliggjøres**

For mange forskere forutsetter datadeling at de først får anledning til å publisere resultater selv og høste akademiske frukter av datainnsamlingen. Mange er derfor opptatt av spørsmålet om sperrefrister før forskningsdata må tilgjengeliggjøres.

Generelt gjelder målsettingen om at forskningsdata skal gjøres tilgjengelig på et så tidlig tidspunkt som mulig, men med henvisning til grunnprinsipp 3: Hva som er en rimelig sperrefrist innenfor forskjellige fagområder må vurderes og besluttes i forskerfellesskapene selv. Forskningsrådet anbefaler i sin policy for tilgjengeliggjøring av forskningsdata at dataene som ligger til grunn for publikasjoner gjøres tilgjengelig så tidlig som mulig, og aldri senere enn ved publiseringstidspunktet. Andre data som kan være av interesse for annen forskning, bør gjøres tilgjengelige innen rimelig tid, og aldri senere enn tre år etter endt prosjekt. Flere av forskningsinstitusjonene i Norge planlegger, eller har allerede vedtatt, tilsvarende formuleringer i sine retningslinjer for forvaltning av forskningsdata.

### **Kompetansebehov**

Økt kompetanse blant forskere og studenter vil på sikt effektivisere tidsbruk og kostnader knyttet til lagring, tilrettelegging av og tilgang til datasett.

Tydligere krav til lagring og tilgjengeliggjøring av forskningsdata medfører behov for bedre kunnskap om datahåndtering på alle nivåer. Forskere bør tilbys opplæring i hvordan man klargjør forskningsdata for lagring og tilgjengeliggjøring og hvordan man gjenbraker og siterer datasett. De nye forskningsmulighetene som den digitale utviklingen muliggjør, krever nye typer kompetanse og kombinasjoner av kompetanser i fagmiljøene. Studentene må forberedes til et digitalisert og mer data-intensivt forskningslandskap, og vil spille en viktig rolle i overgangen til en sterkere og mer innarbeidet



delingskultur. På master- og ph.d.-nivå kan det være hensiktsmessig å inkludere opplæring spesifikt rettet mot tilgjengeliggjøring av data og relevante krav og prinsipper for tilretteleggingen. Data er et område som er velegnet for digitale undervisningsformer og læringsressurser og for internasjonale utdanninger, og det finnes et økende tilbud av nettkurs o.l.

Ekspertgruppen for den europeiske forskningskyen (EOSC) argumenterer for at kulturkollisjon og motsetninger mellom forskere og IKT-eksperter er en barriere det er viktig å overstige for å få utnyttet potensialet som ligger i åpen tilgang til forskningsdata.<sup>26</sup> De peker på behovet for å utdanne flere personer som kombinerer teknisk dataekspertise med innsikt i fagområdene de jobber på, som kan fungere som en brobygger mellom de tunge disiplinforskere og IKT-eksperter. Slike datarøktere (*core data experts*) vil være involvert i planleggingen av datahåndteringen i prosjekter fra begynnelsen av, og bidra til at dataene innsamles, tilrettelegges og tilgjengeliggjøres på en måte som er hensiktsmessig for det enkelte fagfelt, for eksempel når det gjelder valg av formater, metadatastandarder, og arkiv- og publiseringssteder. Utdanningsinstitusjonene bør vurdere behovet for å opprette nye undervisningsstilbud for å utdanne slike datarøktere, som kan videreutvikle datahåndteringstjenester og veilede forskere.

I tillegg er det et stort behov for opplæring av forskere om ulike nasjonale tjenester, og det er behov for veiledning om lisenser og rettigheter, forskrifter og lovverk, ikke minst knyttet til data-sikkerhet og personvern. EUs nye personvernforordning legger som nevnt et større ansvar på institusjonene som behandler personopplysninger.

Nasjonale dataarkiver og forskningsdatainfrastrukturer må vurdere behovet for kurs og opplæring av forskere om krav, regelverk og egne tjenester knyttet til håndtering av forskningsdata. Forskere har for eksempel bruk for støtte i form av standardformuleringer de kan bruke når de skal redegjøre overfor internasjonale tidsskriftseiere eller utgivere om hva institusjonenes retningslinjer og norsk regelverk sier om anledning til å utlevere eller gi tilgang til visse typer data.

### **Regjeringen forventer**

- at forskningsinstitusjonene bidrar til å heve kompetansen hos ansatte og studenter gjennom opplæring i datahåndtering og gjenbruk av data
- at utdanningsinstitusjonene vurderer nasjonalt og nordisk samarbeid med henblikk på behovet for å etablere utdanningstilbud for håndtering av forskningsdata (datarøktere o.l.)

### **Regjeringen vil**

- be det nye organet for høyere utdannings- og forskningstjenester om å formidle informasjon om gjeldende krav, regelverk og ressurser knyttet til håndtering av forskningsdata
- gi det nye organet for høyere utdannings- og forskningstjenester koordineringsansvar for å vurdere hvordan nasjonale forskningsdataarkiver best kan tilby opplæring av forskere og studenter om egne tjenester
- be det nye organet for høyere utdannings- og forskningstjenester om å tilrettelegge for sitering av datasett og kreditering av dataarbeid gjennom utstedelse av permanente identifikatorer (DOI-nummer) og digital forsker-ID (ORCID)
- be det nye organet for høyere utdannings- og forskningstjenester om å utarbeide statistikk for å synliggjøre etablering og gjenbruk av forskningsdata

## **4.2 Krav om datahåndteringsplaner**

Forskningsfinansierende myndigheter er ikke i stand til å sette detaljerte krav til hvordan data skal håndteres i et prosjekt, og akkurat hvilke forskningsdata som skal tilgjengeliggjøres på hvilken måte, jf. grunnprinsipp 3. Men for å understøtte en god datakultur, der prinsippene for god tilrettelegging og tilgjengeliggjøring av data blir integrert i forskningsprosessen, er det behov for tydeligere krav om at forskere lager planer for datahåndteringen.

En datahåndteringsplan er et dokument som beskriver hvordan forskere vil håndtere data gjennom forskningsprosjektets og dataenes livsløp. Planen er først og fremst et verktøy for de aktuelle forskerne selv. Den er et hjelpemiddel for å sikre kvaliteten på data og at man etterlever de ulike reglene og forventningene som databruken faller inn under. Planen gir også

<sup>26</sup> Realising the European Open Science Cloud – First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud. 2016

eksterne aktører mulighet for innsyn i datahåndteringen. Dette er til nytte for forskningsinstitusjonene, Forskningsrådet eller andre som har ansvar for å følge opp ulike krav eller retningslinjer. Det er også nyttig for andre forskere, som kan få innsikt i ulike sider av datahåndteringen og lære av god praksis. For å være til nytte som verktøy bør datahåndteringsplanen inngå i prosjektplanleggingen fra begynnelsen av, oppdateres jevnlig og følge prosjektet til det er ferdig. Forskere som har planlagt godt gjennom hele prosjektet, står bedre rustet til å møte eventuelle krav om datainnsyn fra utgivere eller andre i forbindelse med publisering.

En datahåndteringsplan bør omfatte alle beslutninger knyttet til data som forskere ser nytte av å planlegge, enten de er faglige, administrative, tekniske, juridiske eller etiske. Noen aktuelle spørsmål er blant annet:

- Hvilke data vil prosjektet samle inn, generere eller behandle?
- Hvilken nytte ser man for seg at dataene vil kunne ha for andre, både forskere og andre grupper i forvaltning og næringsliv?
- Hvilke juridiske og etiske problemstillinger er involvert? Hvilke avtaler og informasjon er det behov for i kontakten med forskningsdeltakere, dataleverandører og andre partnere for å ivareta de øvrige delene av planen?
- Finnes det grunner til at dataene ikke kan tilgjengeliggjøres helt åpent, og hvordan kan de da gjøres så åpne som mulig innenfor de nødvendige rammene (avledede datasett med mindre skjermingsbehov, mulighet for tilgang for noen brukergrupper på bestemte vilkår, osv.)?
- Hvordan vil datasikkerheten ivaretas?
- Hvordan vil dataene behandles underveis i prosjektet for å gjøre dem mest mulig gjenfinnbare og gjenbrukbare (rikholdighet og kvalitet, dokumentasjon av data og tilhørende programvare, bruk av standarder eller prinsipper for identifikatorer og variabler, osv.)?
- Hvor vil dataene arkiveres og tilgjengeliggjøres? Er det behov for å planlegge eventuell fremtidig sletting?

Gode datahåndteringsplaner vil kunne bidra til at grunnprinsippene i denne strategien blir etterlevd på en god måte. På sikt bør derfor de fleste forskere

utarbeide datahåndteringsplaner som en selvfølgelig del av forskningsprosessen. Krav fra forskningsfinansierende myndigheter om at det etableres slike planer vil kunne bidra til at flere skaffer seg et gjennomtenkt forhold til datahåndteringen sin, og hvordan data kan tilrettelegges og tilgjengeliggjøres for gjenbruk.

Fra 2017 er krav om en datahåndteringsplan som innebærer åpen tilgang til forskningsdata standard i Horisont 2020.<sup>27</sup> Datahåndteringsplanen må leveres senest seks måneder etter prosjektstart, og kravet til åpenhet følger samme grunnprinsipp som i denne strategien om «så åpent som mulig, så lukket som nødvendig». Grunner som gjør at data ikke kan være helt åpne må inkluderes i datahåndteringsplanen. Kravet gjelder de dataene som er nødvendige for å etterprøve resultatene, og de dataene man selv velger i datahåndteringsplanen utover det. Kravet om åpen tilgang og datahåndteringsplaner er standardalternativet i Horisont 2020, men man kan både før og etter kontraktsinngåelsen velge å unnta prosjektet fra disse kravene når det foreligger legitime grunner. Likeledes har Forskningsrådets policy for tilgjengeliggjøring av forskningsdata inntil høsten 2017 omtalt datahåndteringsplaner som en anbefalt fremgangsmåte uten å fremsette et krav om slike planer for prosjekter som finansieres av Forskningsrådet. Til nå har det med andre ord i stor grad vært opp til forskerne selv om de vil lage datahåndteringsplaner og gjøre datasettene sine tilgjengelige i tråd med dem.

Nå er situasjonen endret. Som det fremgår av statusbeskrivelsen i kapittel 2, er flere av de store forskningsinstitusjonene i gang med å få på plass retningslinjer og praktiske løsninger for datahåndtering. Nasjonale forskningsdataarkiver er tilgjengelige og en rekke fagspesifikke infrastrukturer for lagring og tilgjengeliggjøring av data er etablert eller under etablering både i Norge og internasjonalt. Viktige rammebetingelser for å kunne kreve at forskere lager datahåndteringsplaner for prosjektene sine er dermed i ferd med å komme på plass.

### **Forskningsrådets rolle**

I sin oppdaterte policy for tilgjengeliggjøring av forskningsdata forutsetter Norges forskningsråd

<sup>27</sup> «H2020 Programme: AGA – Annotated Model Grant Agreement». Europakommisjonen, 2017. Versjon 4.1 hentet fra [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/amga/h2020-amga\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf). «H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020». Europakommisjonen, 2016. Versjon 3.0 hentet fra [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).



at det skal være en datahåndteringsplan i prosjekter med finansiering fra Forskningsrådet dersom prosjektet genererer data. I tråd med grunnprinsipp 3 om at beslutninger om tilrettelegging og arkivering av data må tas i forskerfellesskapene skal den forskningsutførende institusjonen eller foretaket selv ha ansvar for å vurdere behovet for og kvaliteten på datahåndteringsplanen i et prosjekt som institusjonen eller foretaket er ansvarlig for. Spørsmålet om en datahåndteringsplan tas opp med Forskningsrådet først ved kontraktsinngåelse. Det er prosjektansvarlig institusjon som avgjør om det er behov for en slik plan, og det er prosjektansvarlig institusjon som i så fall skal godkjenne at planen tilfredsstillers institusjonens krav til datahåndteringsplaner. Forskningsrådet vil på sin side være tydelige i sine forventninger om at data skal lagres i arkiver som følger de internasjonale FAIR-prinsippene.

Denne ansvarsfordelingen innebærer at det ikke skal være noen vurdering av kvaliteten på et prosjekts datahåndteringsplan i forbindelse med fagfelle vurderingen som gjøres i Forskningsrådet. Den medfører med andre ord ikke flere kriterier i Forskningsrådets søknadsvurdering.

Ansvaret for datahåndteringen og godkjenningen av datahåndteringsplaner legges til institusjonene, og det vil være opptil institusjonene selv å velge relevante arkiveringsløsninger for ulike forskningsdata. Selv om dette er hovedregelen, skal Forskningsrådet fortsatt kunne kreve at bestemte prosjekter benytter spesifikke nasjonale datainfrastrukturer for lagring av data. Begrunnelsen for dette er at det innenfor enkelte fagområder finnes gode, nasjonale eller internasjonale infrastrukturer for lagring og tilgjengeliggjøring av data som helt klart bør brukes. Et slikt krav om at et spesifikt arkiv benyttes bør signaliseres i utlysningen eller gjøres klart etter nærmere dialog med de institusjonene Forskningsrådet inngår kontrakt med.

Institusjonelle rutiner for godkjenning av datahåndteringsplaner i forbindelse med søknader om forskningsfinansiering fra Norges forskningsråd, EU eller andre kan være til god hjelp for å øke forskernes kompetanse når det gjelder datahåndtering, lagring og deling. Men det er som nevnt ikke alle forskningsprosjekter som genererer data i nevneverdig grad. For en del forskningsprosjek-

ter kan det derfor være at egen datahåndteringsplan ikke er nødvendig eller hensiktsmessig.

### **Samordning av systemene for utvikling av datahåndteringsplaner**

For at det skal oppleves enklest mulig for den enkelte forsker, bør det tilbys IKT-verktøy for å utarbeide datahåndteringsplaner som er i tråd med hva som kreves av forskningsfinansierer internasjonalt. Slike verktøy bør være generiske, med mulighet for å tilpasses ulike fag. I dag har norske leverandører som tilbyr arkivering av forskningsdata, ulike retningslinjer og systemer for utvikling av datahåndteringsplaner. Det er ønskelig at tjenestetilbyderne samordner seg og i samarbeid med forskningsinstitusjonene bidrar til at løsningene som utvikles er mest mulig kompatible, og følger internasjonal standard.<sup>28</sup>

#### ***Regjeringen forventer***

- at forskningsinstitusjonene utvikler prosedyrer for (i) godkjenning av datahåndteringsplaner og (ii) beslutninger om at et forskningsprosjekt eventuelt er av en slik art at egen datahåndteringsplan ikke er nødvendig eller hensiktsmessig

#### ***Regjeringen vil***

- be Norges forskningsråd gjøre godt kjent sitt krav om at en datahåndteringsplan godkjent av forskningsinstitusjonen skal foreligge ved kontraktsinngåelse, der dette er relevant
- be Norges forskningsråd fortsette å være en pådriver nasjonalt og internasjonalt for bedre tilgjengeliggjøring og gjenbruk av forskningsdata
- be forskningsinstitusjonene om å utvikle retningslinjer for forskningsdata som skal lagres, tilgjengeliggjøres og tilrettelegges, med angivelse av hvilke løsninger som bør benyttes for håndtering av ulike datasett
- be det nye organet for høyere utdannings- og forskningstjenester bidra til at tjenestetilbydere av nasjonale forskningsdataarkiver samordner seg og, i samarbeid med forskningsinstitusjonene, bidrar til at løsninger som utvikles for datahåndteringsplaner er mest mulig kompatible og følger internasjonale standarder

<sup>28</sup> Jf. anbefaling i Morten Dæhlen m.fl. 2017, "Forslag til IKT-strategi for UH-sektoren, delstrategi for forskning."

### 4.3 Bedre teknisk tilrettelegging og bedre faglig samordning nasjonalt

I tillegg til den tekniske infrastrukturen som er nødvendig for å lagre og analysere data, må det tilrettelegges for at arkiveringen og tilgjengeliggjøringen av forskningsdata med identifikatorer og gode metadata ikke innebærer en for stor administrativ byrde for forskerne.

Hva som er riktig arkiveringssted for et forskningsdatasett, kan variere fra en lokal arkivløsning ved den enkelte forskningsinstitusjon til felles nasjonale eller internasjonale databaser. Utvalget av relevante steder – og spørsmål om det er behov for å bygge opp nye – vil også variere fra fagfelt til fagfelt. For noen finnes det allerede definerte internasjonale databaser der data skal deponeres, mens for andre finnes det ikke lagringssteder i dag.

Fagmiljøer på flere områder har meldt om at det er behov for bedre tekniske muligheter for lagring og tilgjengeliggjøring av forskningsdata, og behov for gode selvbetjeningsløsninger for lagring, gjenfinning og gjenbruk av datasett.<sup>29</sup>

Det er også et stort behov for gode standarder for data og metadata. Dette er av avgjørende betydning for at dataene skal være interoperative, det vil si kunne brukes i andre datasystemer og av andre personer enn dem som lagde dem. Standarder for identifikatorer er også nødvendig for å kunne koble sammen ulike datasett, ved at man kan fastholde identiteten til objektene som data beskriver, på tvers av ulike datakilder. Spesifikke metadatastandarder for de enkelte fagfeltene må være godt forankret i fagfelleskapene for å få den nødvendige oppslutningen. Dette vil i de fleste tilfeller innebære internasjonale standarder, som bør benyttes i den grad det er mulig. I tillegg må man finne standarder som fungerer på tvers av fag. Standardene påvirker også i hvilken grad løsninger kan fungere på tvers av sektorgrenser, for både forskning, forvaltning og næringsliv, se avsnitt 5.2.

Et eksempel på en infrastruktur som er beregnet å betjene både forskning og næringsliv, er Språkbanken ved Nasjonalbiblioteket. Denne ressurskatalogen tilgjengeliggjør digitale språkressurser fra både forskning og forvaltning, som kan utgjøre viktige grunnlag for utviklingen av språkteknologiske produkter. En slik samling av relevante data i kataloger og datainfrastrukturer er nyttig for å gjøre data bedre gjenfinnbare, og kan være særlig viktig for brukergrupper som ikke deltar i de vitenskapelige kommunikasjonskanalene.

Det er ønskelig med kommunikasjonsarenaer der relevante datainfrastrukturer, ulike forskningsinstitusjoner og fagmiljøer kan samles for å samordne faglige og tekniske løsninger for å bedre mulighetene for gjenbruk av data.

Gode tekniske løsninger kombinert med god data-sikkerhet er viktige tiltak både for effektivisering for forskere og for dataeiere, og ikke minst for å håndtere personvernutfordringer på en god måte.<sup>30</sup>

Regjeringen har lagt frem nasjonale mål og retningslinjer for åpen tilgang til vitenskapelige artikler, med tilhørende tiltak. Innen 2024 skal alle norske vitenskapelige artikler finansiert av offentlige midler være åpent tilgjengelige. For å bidra til at målene nås, vil regjeringen forbedre funksjonaliteten for deponering av artikler via forskningsinformasjonssystemet *Cristin*, kreve deponering i vitenarkiv lokalt eller nasjonalt som en forutsetning for at artiklene skal telle i den resultatbaserte finansieringen (kravet forutsetter at vitenarkiv og annen nødvendig infrastruktur for deponering er tilgjengelig for alle forskningsinstitusjonene), bidra til videreutvikling av nye og bærekraftige modeller for finansiering av åpen publisering nasjonalt og internasjonalt, utvikle indikatorer og statistikk for åpen tilgang, samt utrede hvordan et nasjonalt vitenarkiv skal realiseres. Tiltak som krever økte budsjetttrammer, vil måtte vurderes i årlige budsjettprosesser.<sup>31</sup> Punktet om et vitenarkiv er også relevant med hensyn til forskningdata. Det bør finnes en samlet oversikt med informasjon om norske forskeres artikler, bøker, forskningsrapporter, avhandlinger

<sup>29</sup> Morten Dæhlen m.fl. 2017, "Delstrategi for forskning" (forslag til IKT-strategi for UH-sektoren). Selvbetjeningsløsninger vil kunne forenkle forskernes tilgang til- og deponering av data/metadata og redusere kostnadene knyttet til drift av datainfrastrukturer.

<sup>30</sup> Datatilsynet: Veileder om programvareutvikling med innebygd personvern. 2017

<sup>31</sup> <https://www.regjeringen.no/contentassets/ae7f1c4b97d34806b37dc767be1fce76/nasjonale-mal-og-retningslinjer-for-åpen-tilgang-til-vitenskapelige-artikler.pdf>

osv., sammen med informasjon om de tilhørende rettighetsklarte datasettene.

Det vil også kunne være behov for en eller flere nasjonale løsninger for sikker behandling av sensitive forskningsdata. Slike bør imidlertid sees i sammenheng med andre løsninger for håndtering av sensitive data, som for eksempel Helseanalyseplattformen. (Se nærmere omtale kapittel 5)

#### **Regjeringen forventer**

- at forskningsinstitusjoner, forvaltere av forskningsdatainfrastrukturer og fagmiljøer bidrar til standardisering og harmonisering som muliggjør deling og gjenbruk av data i henhold til internasjonale standarder og god skikk innenfor ulike fagområder (for eksempel gjennom å etablere fagområdespesifikke nasjonale kommunikasjonsarenaer)

#### **Regjeringen vil**

- be det nye organet for høyere utdannings- og forskningstjenester om å bidra til informasjon om nasjonale forskningsdataarkiver og tjenester, slik at tjenestene fremstår samlet og lett tilgjengelige for alle forskere og forskningsutførende institusjoner
- be det nye organet for høyere utdannings- og forskningstjenester ta ansvar for å utrede hvordan et nasjonalt vitenarkiv skal realiseres

## **4.4 Bærekraftig finansiering og drift av nasjonale forskningsdatainfrastrukturer og forskningsdataarkiver**

Arkivene for forskningsdata må ha langsiktighet for at de skal være attraktive både for dataeierne som skal lagre data, og for arkivene som skal bygge opp tjenestene. Derfor er det nødvendig å videreutvikle bærekraftige finansieringsmodeller både for etablering og drift av forskningsdatainfrastrukturer og forskningsdataarkiver.

Det er generelt en utfordring at forskningen ofte er organisert i prosjekter med begrenset varighet, mens forskningsdataene i mange tilfeller skal forvaltes i mange år etter prosjektslutt. Når det først er bestemt at forskningsdataene skal langtidslagres og dataene er deponert i et arkiv, må dataene være tilgjengelige over tid på en forutsigbar måte.

Kostnader til datahåndtering må dermed ivaretas ut over prosjektperioden, og datainfrastrukturer har behov for å ha et lengre tidsperspektiv enn de fleste andre forskningsinfrastrukturer. I noen tilfeller vil dataene som arkiveres i infrastrukturene være lange og nasjonalt viktige tidsserier eller referansedatasett for forskningen. Denne typen data skal i prinsippet være tilgjengelig til evig tid. Det normale vil likevel være at dataene har en begrenset levetid, men for å sikre gjenbruk og mulighet for etterprøving av forskningen, må brukerne – både de som deponerer og de som henter ut data – ha en forsikring om at dataene er tilgjengelige og brukbare i lang tid. Minimumsforpliktelsen ved tildeling av DOI-nummer er ti år, men ofte bør forskningsdata være tilgjengelige lenger enn det. Dette krever at det finnes et element av langsiktighet i finansieringsmodellene, et element som i mange tilfeller ikke er der i dag.

Dagens finansieringsmodeller for datainfrastrukturer kan inndeles i ulike typer. Finansieringen av en infrastruktur vil normalt være sammensatt av flere av disse kildene.

1. *Strukturell finansiering*: langsiktig finansiering uten krav om direkte gjenytelser. Den vil normalt komme som støtte fra et departement eller fra Forskningsrådet
2. *Forsknings- og utviklingsprosjekt*: konkrete og tidsavgrensede prosjekter og aktiviteter, der formålet normalt vil være å etablere eller videreutvikle datainfrastrukturen. Støtten kan komme direkte fra forskningsfinansiører eller via en forskningsutførende organisasjon som er ansvarlig for forsknings- og utviklingsprosjektet. Støtte fra Nasjonal ordning for forskningsinfrastruktur inngår her
3. *Engangskostnad for deponering (Data Processing Charge – DPC)*
4. *Kostnad for tilgang (Data Access Charge – DAC)*: betaling for kostnader ved å gi tilgang til data, for eksempel uttrekk av sensitive data. Det kan også være kostnader ved verdikjende tjenester, som for eksempel generering av metadata, kobling av data, visualisering og analyse av data
5. *Støtte fra vertsorganisasjon*: direkte eller indirekte støtte fra vertsorganisasjonen for dataarkivet, i form av frie midler, personale, lokaler, datautstyr osv.
6. *Medlemskap/abonnement*: betaling fra institusjoner for deponering og/eller tilgang.
7. I *medlemskapsmodellen* går flere organisasjoner sammen om å dekke kostnader knyttet til deponering og/eller tilgang til data.

8. *Abonnementsmodellen* skiller seg fra medlemskapsmodellen ved at kostnaden ved abonnement avregnes direkte etter bruk, mens den i medlemskapsmodellen også bestemmes av andre parametere og den vil ofte være fast over en periode
9. *Oppdrag*: knyttet til deponering, tilgang og/eller verdiøkende tjenester

Det er fordeler og ulemper ved de forskjellige finansieringsmodellene. Strukturell finansiering gir stabilitet og mulighet for langsiktig planlegging. Datainfrastrukturene vil trolig kunne tiltrekke seg høyt kompetent arbeidskraft om de har stor andel strukturell finansiering, og kostnadene med å skaffe finansiering vil være lave. På den andre siden kan stor strukturell finansiering over tid føre til at datainfrastrukturene blir lite effektive, med få insentiver for forbedringer og innovasjon. Stor andel strukturell finansiering bør derfor følges av jevnlig evalueringer, som også kan resultere i at finansieringen endres.

Betaling for deponering av data vil fungere godt sammen med normale mekanismer for finansiering av forskning, forutsatt at slike kostnader regnes som legitime. Denne finansieringskilden er også kompatibel med praksis som er etablert i forbindelse med åpen publisering, der en del åpne tidsskrifter finansieres gjennom artikkelavgifter. En utfordring vil være å avklare hva som er en riktig kostnad, ettersom den faktiske kostnaden for arkivering og kuratering over en lang tidsperiode vil være ukjent. Videre kan datainfrastrukturer som tar betalt for arkivering tape i konkurransen med datainfrastrukturer som ikke benytter denne mekanismen. Erfaring tilsier at betaling for deponering for eksempel kan være egnet for prosjekter med store behov for lagringskapasitet, mens arkivering av data som skal tilgjengeliggjøres i størst mulig grad bør være gratis for prosjektene.

Betaling for tilgang til data er i tråd med prinsippet om at brukeren betaler for de tjenestene hun eller han ønsker. Betaling for selve dataene, kan være i strid med prinsippet om åpen tilgang til offentlig finansierte data. Det kan imidlertid tas betaling for uttrekk av data og for verdiøkende tjenester som generering av metadata, kobling av data, visualisering og analyse av data. Videre kan et offentlig dataarkiv som har krav om selvfinansiering etter

gjeldende offentlighetsforskrift også ta betaling for informasjon (data).<sup>32</sup> Vektige innvendinger kan være at brukerbetaling for tilgang til data reduserer den faktiske bruken, og at det kan være mer hensiktsmessig at det er forskningsinstitusjonene og ikke enkeltprosjekter som betaler. Brukerbetaling kan også føre til at data generert i Norge blir lagret ved internasjonale datasentre som ikke tar betaling for tilgang isteden.

Finansiering av drift gjennom medlemskap/abonnement vil gi stabil og langsiktig finansiering, og de største brukergruppene vil få eierskap til datainfrastrukturen. Brukergruppene vil føle økt lojalitet til datainfrastrukturen og vil kunne få innflytelse på styringen av datainfrastrukturen. En ulempe dersom medlemskap dekker kostnad ved tilgang til data, vil være at brukere utenfor medlemsorganisasjonene må betale for tilgang, mens brukere innenfor normalt vil ha fri tilgang. Medlemskap for å dekke kostnader til deponering kan være enklere å forsvare utad, fordi det da vil være valgfritt for de som står utenfor å benytte datainfrastrukturen.

Finansiering fra vertsorganisasjonen gir god langsiktighet og god forankring hos eieren av datainfrastrukturen. Samtidig kan datainfrastrukturer som i hovedsak baserer finansiering av drift på denne kilden, få et stort innadrettet fokus. Det er naturlig at vertsorganisasjonen søker å ivareta egne behov. Konsekvensen kan bli at det bygges opp mange institusjonsarkiver, også der brukerne ville vært tjent med felles datainfrastrukturer, og den totale kostnaden ville være lavere om institusjonene gikk sammen.

Oppdrag kan være en god tilleggsfinansiering for datainfrastrukturene. Fordi både etterspørsel etter tjenestene og prisen det er mulig å oppnå for en gitt tjeneste vil være markedsbasert, er dette en usikker inntektskilde. Statsstøttereguleringen setter dessuten grenser for hvor stor økonomisk aktivitet en datainfrastruktur kan ha.

I helt spesielle tilfeller kan det søkes om støtte til langsiktig grunnfinansiering av infrastrukturene gjennom *Nasjonal satsing på forskningsinfrastruktur*. Det er i dag fare for at kriteriene for å kvalifisere til langsiktig grunnfinansiering ikke i tilstrekkelig grad fanger opp utfordringer som spesifikt er knyttet til drift av nasjonale forskningsdata-

<sup>32</sup> Forskrift 17. oktober 2008 til offentliglova

infrastrukturer. Det er ikke alltid klart hvilke finansieringsmodeller som vil fungere mest effektivt for å ivareta langsiktig og bærekraftig drift i hvert enkelt tilfelle. I tillegg til særskilte finansieringsvirkemidler for etablering av infrastruktur er det derfor behov for kunnskap om hvordan ulike inntekter og kostnader knyttet til arbeidet med forskningsdata kan kombineres i effektive og bærekraftige forretningsmodeller.

Det er også utfordringer knyttet til hvordan datainfrastrukturer som skal fungere på tvers av sektorer og betjene både forskere, innbyggere, forvaltning og næringsliv best finansieres. Dette er nærmere omtalt i kapittel 5.2.

### ***Regjeringen forventer***

- at der det er hensiktsmessig utvikles selvbetjeningsløsninger som reduserer kostnader ved drift av forskningsdatainfrastrukturer gjennom å forenkle prosesser for deponering av og tilgang til forskningsdata og metadata

### ***Regjeringen vil***

- videreføre Norges forskningsråds ordning Nasjonal satsing på forskningsinfrastruktur
- be Norges forskningsråd, i samarbeid med det nye organet for høyere utdannings- og forskningstjenester og med representanter fra forskningsinstitusjonene, om å utrede og gi råd om bærekraftige finansieringsmodeller for drift av datainfrastrukturer og dataarkiver



# Del 2:

## Offentlige data som er viktige datakilder for forskning







# 5 Offentlige data som er viktige datakilder for forskning

## 5.1 Bedre tilrettelegging for bruk av offentlige data til forskning

Statens arbeid med å digitalisere offentlig sektor og legge til rette for videre bruk av offentlige data foregår primært gjennom regjeringens arbeid med oppfølging av Meld. St. 27 (2015–2016) *Digital agenda for Norge – IKT for en enklere hverdag og økt produktivitet*.

I meldingen gjør regjeringen det klart at den ønsker å styrke arbeidet med å gjøre offentlige data tilgjengelig for viderebruk og gjenbruk. Tiltak i meldingen omfatter blant annet at det, i tillegg til bedre tilrettelegging for gjenbruk av forskningsdata, skal utarbeides strategier og handlingsplaner for økt tilgjengeliggjøring av data innenfor områdene kultur, geodata, statlige utgifter og transport. Økt tilgjengelighet av offentlige data vil bidra til å styrke forskningen, gjennom enklere og raskere tilgang til datasett for forskere. Lett tilgjengelige offentlige data kan også ha stor verdi for høyere utdanning, for eksempel som grunnlag for masteroppgaver. Kulturdepartementets strategi for åpne data ble lansert i oktober 2017.<sup>33</sup> For de øvrige områdene pågår det arbeid med utarbeiding av strategier eller handlingsplaner.

I noen tilfeller kan det være flytende overganger mellom forskningsdata og forvaltningsdata (eks. klima- og miljøovervåkningsdata), og det er ingen

tvil om at offentlige data som er samlet inn for andre formål utgjør viktige datakilder for forskning. Mange forskningsspørsmål krever at man kobler opplysninger fra flere ulike datakilder. Disse prosessene bør skje så effektivt som mulig. Ikke bare for å unngå sløsing med ressurser, men også fordi mye av den aktuelle forskningen skal ha konkrete anvendelser som krever at kunnskapen blir produsert i tide for å være til nytte. Det bør utarbeides felles regler, standarder og formater på data og gode metadata der det er naturlig. Arbeidet bør skje i nært samarbeid med sentrale brukergrupper. For at forskere skal kunne planlegge sine prosjekter både med hensyn til forskningsdesign, tidsplaner og budsjetter, er det viktig med åpenhet og forutsigbarhet om vilkårene for tilgang, datakvalitet, og tidsbruk og kostnader for datatilgang. Derfor er det viktig at offentlige data tilrettelegges godt for videre bruk, slik at også forskningen kan dra nytte av dem.

Generelt er det slik at en del offentlige data allerede er åpent tilgjengelig for alle aktører. Det varierer imidlertid hvor godt de er tilrettelagt for gjenbruk. At data som er samlet inn for andre formål, tilrettelegges godt slik at de også er enkle å bruke som grunnlag for forskning, kan være et viktig bidrag til å øke kvaliteten og relevansen i forskningen. God tilrettelegging for gjenbruk på en måte som også tjener forskningen, kan derfor anses som en viktig del av departementenes sektoransvar for forskning.

<sup>33</sup> <https://www.regjeringen.no/no/dokumenter/kulturdepartementets-strategi-for-apne-data/id2576038>

## Boks 5.1 Nasjonale tjenester for åpne offentlig data

Det finnes flere nasjonale tjenester for åpne data fra det offentlige. *Data.norge.no* er den nasjonale katalogen over datasett som er gjort tilgjengelig for viderebruk. Formålet med portalen er å gjøre det enklere å oppdage, vurdere og eventuelt bruke åpne data fra det offentlige til nye formål. Portalen inneholder kun beskrivelser av datasettene, med pekere til hvor datasettene finnes. *Data.norge.no* støtter den europeiske standarden for utveksling av datasettbeskrivelser (DCAT-AP) og gjør alt innhold i selve datakatalogen tilgjengelig for den europeiske åpne data-portal <https://www.europeandataportal.eu/>. Både Digitaliseringsrundskrivet og Regjeringens retningslinjer for tilgjengeliggjøring av offentlige data legger opp til at offentlige virksomheter skal synliggjøre sine åpne data gjennom *data.norge.no*.

*Datahotellet* er en gratis datadistribusjonstjeneste fra Difi hvor dataeiere kan gjøre egne datasett tilgjengelig for bruk uten å måtte gjøre investeringer i egen infrastruktur eller programvare. For brukere betyr dette at data som tidligere kun har vært tilgjengelig for nedlasting som regneark (eller ikke tilgjengelig i det hele tatt), blir gjort tilgjengelig gjennom et standardisert programmeringsgrensesnitt (API) som leverer data i flere populære maskinlesbare formater. Alle datasett på *datahotellet* skal være registrert på *data.norge.no* og søk og navigering skjer også via *data.norge.no*.

### **Regjeringen vil**

- be offentlige aktører legge til rette for gjenbruk av data til forskning, innovasjon og verdiskaping ved oppgradering og utvikling av systemer for datalagring
- legge vekt på tilgang til data for forskning i arbeid med relevante lov- eller forskriftsendringer

## 5.2 Bedre samspill mellom offentlige data og forskningsdata

I OECDs gjennomgang av digitaliseringen av offentlig sektor i Norge pekes det på at systemet er for fragmentert. Det utvikles sektorinterne løsninger uten at det i tilstrekkelig grad tas hensyn til om bedre løsninger allerede er utviklet andre steder eller til behovet for at egen løsning bør fungere sammen med andres løsninger. Denne fragmenteringen gjør at Norge ikke lykkes med å løse systemiske forvaltningsutfordringer på tvers av sektorgrenser.<sup>34</sup>

I OECDs landgjennomgang av det norske forsknings-systemet beskrives Norge som gode med hensyn til

inkrementelle forbedringer og innovasjoner innenfor enkeltområder, men det pekes på at systemer og verktøy for å skalere opp og spre videre vellykkede tiltak og initiativer mangler, spesielt i offentlig sektor. Større og mer omgripende endringer vil kanskje kreve andre verktøy og måter å organisere seg på. Direktoratet for IKT og forvaltning får kritikk for å fokusere for ensidig på digitalisering og overse andre faktorer som påvirker innovasjon i offentlig sektor (som organisering, reguleringer, eksperimentering og læring osv.). Mange av samfunnsutfordringene overskrider sektorgrenser og trenger en tverrfaglig tilnærming og bedre horisontal koordinering på myndighetsnivå for å løses.<sup>35</sup>

For å motvirke fragmenteringen og bidra til at data som samles inn for offentlige midler skal være til best nytte for samfunnet, er det behov for en møteplass der man kan samle særlig relevante aktører fra forskning og forvaltning som har data som er relevante for svært mange. Det er behov for bedre koordinerte råd om (i) fornuftig samordning og harmonisering av faglige og IKT-tekniske løsninger på tvers av sektorer, (ii) om datakvalitet, tilgangsstyring, rutiner med mer, samt (iii) om hensiktsmessige finansieringsmodeller for data-

<sup>34</sup> <https://www.oecd.org/gov/digital-government/digital-government-review-norway-recommendations.pdf>

<sup>35</sup> OECD Reviews of Innovation Policy: Norway 2017

infrastrukturer av nasjonal betydning som skal betjene både forskere, forvaltere, næringsliv og innbyggere.

Regjeringen ser behov for et nasjonalt forum for gjenbruk av forskningsdata, med representasjon fra relevante offentlige organer som eier eller er ansvarlige for forskningsdatainfrastrukturer eller datainfrastrukturer og samlinger opprettet for forvaltningsformål som er særlig relevante for forskere. Aktuelle deltakere kan for eksempel være Norsk senter for forskningsdata, Uninett, Helsenett, Direktoratet for e-helse, Norges forskningsråd, Statistisk sentralbyrå, Datatilsynet, Direktoratet for forvaltning og IKT, Riksarkivet og Nasjonalbiblioteket. Før det nye organet for høyere utdannings- og forskningstjenester er kommet på plass og sammensetningen av organets overordnede strategiske råd er kjent, er det vanskelig å konkludere endelig om hvordan behovet for et nasjonalt forum for gjenbruk av forskningsdata mest hensiktsmessig kan ivaretas. Kunnskapsdepartementet vil følge opp behovet for et slikt forum i samarbeid med det nye organet for høyere utdannings- og forskningstjenester og sentrale interessenter.

#### **Regjeringen vil**

- vurdere å etablere et nasjonalt forum for gjenbruk av forskningsdata med representasjon fra relevante offentlige organer for å få råd i spørsmål om samordning, datakvalitet, tilgangsstyring, rutiner og finansiering av nasjonale datainfrastrukturer som skal betjene både forskere, forvaltere, næringsliv og innbyggere

### **5.3 Enklere tilgang til data fra Statistisk sentralbyrå for forskningsformål**

Data som er verdifulle for produksjon av statistikk, er også gjerne verdifulle for forskning. Statistikkloven fastsetter at en av Statistisk sentralbyrås (SSBs) oppgaver er å gi opplysninger til statistisk bruk for forskningsformål, innenfor rammene av blant annet personvernet og statistisk konfidensialitet. SSBs data er viktige kilder i en rekke typer forskning i Norge (samfunnsvitenskapelig og humanistisk forskning, helseforskning osv.). Selv om bruken av slike data allerede utgjør en stor verdi for norsk forskning i dag, er det muligheter for flere forbedringer for å kunne

utnytte potensialet på en best mulig måte. Mulige forbedringer er særlig knyttet til å gjøre prosessene for datatilgang mer forutsigbare og effektive.

En rekke forskningsinstitusjoner har pekt på utfordringer med å få utlevert mikrodata fra registrene til SSB. Vilklårene for tilgang virker uoversiktlige, rollefordelingen mellom ulike instanser i godkjenningprosessen oppleves som uklar, det er høyere tidsbruk og kostnader for utlevering enn forskerne mener burde forventes, og utleveringsprosedyrene oppleves som lite tilpasset behovet for å gjenbruke data.<sup>36</sup>

SSBs data inkluderer både opplysninger SSB har innhentet for statistisk bruk direkte fra privatpersoner og foretak til å lage statistikk, og opplysninger fra forvaltningens administrative datasystemer (registerdata). Begge typer er data som i grunnregelen ikke kan deles åpent når de er innhentet for behandling til statistiske formål. Data som SSB innhenter i første hånd, forutsetter i mange tilfeller konfidensialitet. Dersom det oppstår usikkerhet om hvordan opplysningene blir håndtert i etterkant, kan det føre til at personer eller bedrifter unnlater å svare på spørsmål eller gir uriktige opplysninger, og dataene vil kunne bli av dårligere kvalitet. Det kan imidlertid være mulig å levere ut opplysninger dersom det på forhånd er innhentet samtykke. Registerdata inneholder ofte sensitive opplysninger, og tilgang til disse er som regel strengt regulert. I henhold til statistikkloven skal datagrunnlaget for offisiell statistikk gjøres tilgjengelig for forskning. Mottaker er ansvarlig for å behandle opplysningene i henhold til statistikklovens taushetsplikt og annet regelverk for behandling av personopplysninger. God informasjonssikkerhet og godt personvern er avgjørende for befolkningens tillit til forvaltningen og aksept for at det utvikles omfattende registre.

For registerdata vil det ofte være mulig å få adgang til opplysningene direkte fra etaten som forvalter det aktuelle registeret i den grad reguleringene åpner for at opplysningene kan brukes til forskning. Forskere finner det likevel ofte mest hensiktsmessig å hente opplysningene fra SSB, for eksempel fordi man skal ha opplysninger fra flere registre på en gang, samt SSBs kvalitetssikring av registerdataene. Kvalitetssikringen som gjøres av SSB, er av stor verdi, og det er lite effektivt at denne gjøres flere steder.

<sup>36</sup> Se f.eks. felles innspill til Statistikklovutvalget fra Universitets- og høyskolerådet og Forskningsinstituttene fellesarena: <https://www.abelia.no/contentassets/d02b8111abb04dff89fcd863ee25b6a/innspill-til-statistikklovutvalget-17-08-17.pdf>

For å effektivisere håndtering av mikrodataoppdrag reorganiserte SSB dette arbeidet i 2013 i en egen seksjon for utlån av mikrodata. Samtidig endret de prismodellen for oppdragene, og fra

2014 av har merkostnader ved utlevering blitt fakturert oppdragsgiver fullt ut. Tabell 5.1 viser noen nøkkeltall for mikrodataoppdragene i SSB de seneste årene.

**Tabell 5.1 Mikrodataoppdrag i SSB**

|                                  | 2012  | 2013  | 2014  | 2015  | 2016  |
|----------------------------------|-------|-------|-------|-------|-------|
| Oppdrag levert                   | 176   | 158   | 267   | 220   | 238   |
| Oppdrag under behandling 31.12.  | ..    | > 250 | 75    | 79    | 100   |
| Besvarte henvendelser/veiledning | ..    | 191   | 270   | 386   | 436   |
| Totalt utfakturert, mill. kr     | 2,4   | 8     | 12,5  | 10,4  | 12,5  |
| Laveste/høyeste pris, 1000 kr    | 2-300 | 2-600 | 2-336 | 2-661 | 2-630 |
| Gjennomsnittspris, 1000 kr       | 14,2  | 61    | 57    | 42,1  | 52    |

Tallene gjelder utlån fra seksjon for mikrodata. Noen få oppdrag håndteres fremdeles i andre deler av SSB. Besvarte henvendelser/veiledning er antallet som er gitt rundt ufullstendige søknader/bestillinger. Oppdrag under behandling for 2013 er antallet på det meste i løpet av året, ikke per 31.12.

Kilde: Statistisk sentralbyrås årsrapporter for 2016 og 2014

Effektive prosesser for utleveringsoppdrag stiller krav både til oppdragsgiverne og til SSB. I tillegg til prosessene for å forbedre sin interne oppgaveløsning som har gjort det mulig å behandle flere oppdrag enn tidligere, har SSB også lagt vekt på å forbedre veiledning og informasjon for at forskere skal kunne utarbeide fullstendige søknader av god kvalitet. SSB har også blitt enige med registereierne og Datatilsynet om en forenkling i prosedyrene for dispensasjoner fra taushetsplikten etter forvaltningsloven, og fra oktober 2015 har forskere kunnet søke dispensasjon direkte fra SSB for data som brukes til statistikk, der de tidligere måtte søke enkeltvis til de ulike eierne av registrene dataene er hentet fra.

I perioden 2012–2017 har Norsk senter for forskningsdata (NSD) og SSB samarbeidet om prosjektet RAIRD (Remote Access Infrastructure for Register Data), som er støttet av Forskningsrådets infrastruktursatsing. Prosjektet utvikler en teknisk løsning for å forbedre tilgangen til registerdata for forskningsformål. Løsningen vil muliggjøre statistiske analyser av mikrodata uten å gi tilgang til selve dataene, og inneha ulike sperrer for at analysene kan resultere i identifiserbare opplysninger. Løsningen vil dermed gjøre det mulig å forske på administrative mikrodata uten behandlingsgrunnlag, og med betydelig enklere prosedyrer for godkjenning. Denne muligheten vil ikke fullt ut erstatte behovet for også å utlevere selve mikrodata til forskning. Det vil likevel trolig være et

aktuelt alternativ for noen av dem som søker om slike data i dag, noe som kan gi mer ressurser til å håndtere de oppdragene der denne løsningen ikke strekker til.

Overgangen til en prismodell der forskningsprosjektene må fullfinansiere SSBs merkostnader ved datautleveringen, førte til en prisøkning fra tidligere praksis der SSB delfinansierte dette fra sin egen driftsbevilgning. At kostnadene synliggjøres og belastes forskerne, kan bidra til økt effektivitet ved at man bedre avveier kostnadene mot nytten før man søker, og ved at søknader utformes med høyere kvalitet. Samtidig skaper høyere priser utfordringer, særlig når prisene er vanskelige for forskere å beregne på forhånd, noe som gjør det vanskelig å planlegge budsjetter for prosjektsøknader eller tilbud i oppdragskonkurranser. Prisene kan også skape utfordringer for typer av forskning der oppdragsmarkedet og muligheter for tilskudd er små. Statistikklovutvalget, som ble nedsatt i september 2016 og skal avgi rapport innen 15. mars 2018, skal blant annet vurdere prinsipper for oppdragsvirksomheten til SSB og deres prising av eksterne oppdrag, blant annet til forskning.

#### **Regjeringen forventer**

- at forskningsinstitusjonene bidrar til bedre bestillerkompetanse hos egne forskere og støttepersonell

### **Regjeringen vil**

- utrede en finansieringsmodell for tilgang til data fra Statistisk sentralbyrå til forskning som ivaretar både hensynet til en effektiv bruk av ressurser på datatilgangen og samfunnsmessige gevinster av forskningen
- støtte opplæringstiltak for forskere og støttepersonell i regi av Statistisk sentralbyrå
- vurdere forslag fra Statistikklovutvalget og innspill fra høringen med sikte på å bedre forskeres tilgang til data fra Statistisk sentralbyrå

## **5.4 Enklere og sikrere tilgang til helsedata**

Norge har et internasjonalt fortrinn i og med det kunnskapsgrunnlaget som er samlet gjennom helseregistre, biobanker og befolkningsundersøkelser og andre data som regelmessig frembringes gjennom helsetjenestene. Det er bakgrunnen for at lettere tilgang til og økt utnyttelse av helsedata er ett av fem prioriterte områder i HelseOmsorg21-strategien. Regjeringen har fulgt opp med tiltak for å gjøre helsedata lettere tilgjengelig for forskning og andre formål i handlingsplanen for oppfølging av HelseOmsorg21-strategien. Målet er å sikre at de store investeringene i helseregistre, biobanker og befolkningsundersøkelser blir brukt til å utvikle ny kunnskap og bedre tjenester. Dette kan og skal skje innenfor personvernmessig trygge rammer.

I dagens system kan det være ressurs- og tidkrevende å få tilgang til helsedata. Særlig kan det være utfor-

drende når det er behov for kobling av data fra flere kilder og når data skal brukes til forskningsprosjekter med en mer eksplorativ tilnærming. Utfordringene er blant annet beskrevet i Forskningsrådets rapport *Enklere tilgang – mer forskning* fra desember 2016 og i Helsedatautvalgets rapport *Et nytt system for enklere og sikrere tilgang til helsedata* fra juni 2017.

Forskningsrådets rapport viser til at det er bred enighet om at det er for tungvint og tar for lang tid å få tilgang til koblede registerdata for forskningsformål. Videre heter det at kartleggingen som er gjennomført i sammenheng med rapporten sannsynliggjør at dette medfører mindre norsk helseforskning. Det vises til at andre nordiske land over tid har etablert nasjonale infrastrukturer og fellesløsninger for helseanalyse, tilgjengeliggjøring av data og forskning som setter dem bedre i stand til å utnytte helsedata enn i Norge.

Helsedatautvalget peker på en rekke utfordringer knyttet til organisering, rettslig regulering og dagens godkjenningssystem som bidrar til lang saksbehandlingstid, kompleksitet og inkonsistente avgjørelser. Videre mener utvalget at statistikkloven og SSBs praksis gjør at muligheten for å få sammenstilt demografiske og sosioøkonomiske data med helsedata til forskningsformål er tungvint. Utvalget konkluderer med at det er for komplisert og tidkrevende å få utlevert helsedata til sekundærbruk og viser til at disse utfordringene har vært grundig diskutert og dokumentert i en årrekke.

## Boks 5.2 Helsedatautvalgets konklusjoner

Helsedatautvalget ble nedsatt av Helse- og omsorgsdepartementet i 2016 for å utrede et bedre og mer effektivt system for behandling av helsedata for statistikk, planlegging, helseanalyser, kvalitetsforbedring og forskning, innovasjon og næringsutvikling. Utvalget overleverte sin rapport til Helse- og omsorgsdepartementet i juni 2017.

Et samlet utvalg mener at pågående tiltak for å gjøre tilgangen til helsedata enklere og sikrere må gjennomføres så raskt som mulig. Utvalget mener imidlertid at tiltakene ikke er tilstrekkelige til å løse de grunnleggende utfordringene. Utvalgets flertall anbefaler derfor at det innføres et nytt system for tilgang til helsedata som innebærer at:

- dagens krav om forhåndsgodkjenning fra Datatilsynet og/eller de regionale komiteene for medisinsk og helsefaglig forskningsetikk avvikles

- det etableres en nasjonal tjeneste som ivaretar alle forvaltningsoppgaver for å tilgjengeliggjøre helsedata
- en første versjon av en helseanalyseplattform realiseres raskt og bygges ut over tid under et strengt kvalitetsgodkjennings- og informasjonssikkerhetsregime
- demografiske og sosioøkonomiske data på individnivå gjøres tilgjengelig på en helseanalyseplattform, enten fra Statistisk sentralbyrå eller fra andre kilder.
- Utvalgets mindretall mener at mulige handlingsalternativer og konsekvensene av tiltakene ikke er tilstrekkelig utredet. Etter mindretallets vurdering er det derfor ikke grunnlag for å konkludere med at tiltakene bør gjennomføres. Dette gjelder særlig pkt. 2 og 3.

Rapporten ble sendt på høring 1. september (frist 1. desember) for å sikre forankring og innspill fra en større bredde av aktører før departementet tar stilling til videre oppfølging.

Det pågår et viktig arbeid gjennom Helsedataprogrammet som Direktoratet for e-helse i 2017 fikk ansvar for å etablere og følge opp. Helsedataprogrammet skal både bidra til å gjøre helsedata lettere tilgjengelig for forskning og annen sekundærbruk og til bedre personvern og informasjonssikkerhet. Direktoratet skal, i regi av Helsedataprogrammet, utvikle en rekke fellestjenester for helseregistrene som vil gjøre bruken av helsedata enklere, for eksempel en tjeneste for å etterspørre samtykke og en oppslagstjeneste for kodeverk, metadata, og registervariabler. Det vil bli etablert en nasjonal løsning for publisering av kodeverk, terminologi og metadata, med hovedvekt på sekundærbruk av data til forskning og kvalitetsarbeid.

En sentral oppgave i Helsedataprogrammet er å etablere en plattform for tilgjengeliggjøring og analyse av helsedata (helseanalyseplattformen). Formålet er å forenkle og effektivisere tilgangen til helsedata. Forskningsrådet og Direktoratet for

e-helse inngikk i mars 2017 en samarbeidsavtale om å utvikle helseanalyseplattformen til beste for forskning. I tillegg samarbeider en lang rekke aktører om å finne gode løsninger for en slik plattform gjennom Helsedataprogrammet.

### **Regjeringen forventer**

- at databehandlingsansvarlige for helseregistre og Statistisk sentralbyrå gjennom Helsedataprogrammet bidrar til at data i større grad skal kunne analyseres sikkert og effektivt på tvers av datakilder

### **Regjeringen vil**

- vurdere forslagene fra Helsedatautvalget og innspill fra høringen med sikte på å etablere et enklere og sikrere system for tilgang til helsedata
- vurdere etablering av en plattform for helseanalyser, jf. oppdrag til Direktoratet for e-helse om å utrede etablering av en slik plattform

## 6 Økonomiske og administrative konsekvenser

---

Oppfølging av forventningene og tiltakene i strategien vil i stor grad skje gjennom forskningsinstitusjonenes og offentlige dataforvalteres ordinære utviklingsarbeid og finansieres innenfor gjeldende budsjetttrammer. Enkelte tiltak kan ha økonomiske

og administrative konsekvenser. Forslag som innebærer tiltak med budsjetteffekter vil bli vurdert i den ordinære budsjettprosessen, blant annet i lys av handlingsrommet i budsjettpolitikken.



Utgitt av:  
Kunnskapsdepartementet  
Offentlige institusjoner kan bestille flere eksemplarer fra:  
Departementenes sikkerhets- og serviceorganisasjon  
Internett: [www.publikasjoner.dep.no](http://www.publikasjoner.dep.no)  
E-post: [publikasjonsbestilling@dss.dep.no](mailto:publikasjonsbestilling@dss.dep.no)  
Telefon: 222 40 000

Publikasjonskode: F-4442 B  
Omslagsfoto: Colourbox  
Design og ombrekking: Departementenes sikkerhets- og serviceorganisasjon  
Trykk: Departementenes sikkerhets- og serviceorganisasjon  
12/2017 – opplag 100

