

Tilgjengeliggjøring av forskningsdata og data for forskning

Kunnskapsgrunnlag til KDs arbeid med en nasjonal strategi for
tilgjengeliggjøring av data

© Norges forskningsråd 2016

Norges forskningsråd
Postboks 564
1327 Lysaker
Telefon: 22 03 70 00
post@forskningsradet.no
www.forskningsradet.no/

Publikasjonen kan bestilles via internett:
www.forskningsradet.no/publikasjoner
eller grønt nummer telefaks: 800 83 001

Trykk: 07 Media AS

Oslo desember 2016.

ISBN 978-82-12-03566-9 (pdf)

Innhold

1	Sammendrag.....	4
2	Bakgrunn og mandat	5
2.1	Mandat og avgrensning.....	6
3	Tidligere og pågående prosesser internasjonalt og i Norge	7
4	Arbeidsform og metode	7
5	Kartlegging av norske datainfrastrukturer	9
5.1	Forskningsrådets finansiering av datainfrastrukturer	10
5.2	Norske tjenester for datalagring og -håndtering	20
5.3	Oppsummering.....	29
6	Finansieringsmodeller for datainfrastrukturer.....	29
6.1	Kilder til finansiering.....	29
6.2	Finansiering av norske datainfrastrukturer for forskningsdata	30
7	Avdekte behov	33
8	Drøftinger	34
8.1	Policyer/retningslinjer for datahåndtering	34
8.2	Samordning og rolleavklaring.....	35
8.3	Opplæring og kulturendring	35
8.4	Insentiver.....	35
8.5	Finansiering	36
9	Forskningsrådets innspill	38
10	Vedlegg	40
10.1	Definisjoner	40
10.2	Oversikt over institusjoner og informanter vi har konsultert	41
10.3	Intervjuguide	42
10.4	Invitasjonsbrev, utsendelseslister og program for to workshops i 2016	43

1 Sammendrag

Innen utgangen av 2017 skal Kunnskapsdepartementet (KD) utarbeide en strategi for økt tilgjengeliggjøring av forskningsdata. I den forbindelse har KD bedt Forskningsrådet om å etablere et kunnskapsgrunnlag.¹ I samråd med KD er utredningsoppgaven fortolket og avgrenset slik:

- Kartlegge Forskningsrådets finansiering av infrastrukturer som støtter tilgjengeliggjøring av forskningsdata
- Kartlegge hvilke tjenester for datalagring som finnes i dag
- Gi innspill til arbeidsdeling og rolleavklaring mellom tjenesteleverandører og til mulige finansieringsmodeller for nasjonalt viktige tjenester for lagring og deling av data

Vi har i samråd med KD valgt ikke å gå tungt inn i relaterte pågående prosesser der tilgjengeliggjøring av data for helserelatert forskning er sentralt.

Kunnskapsgrunnlaget er utarbeidet på basis av eksisterende dokumenter og dialog med representanter fra de fleste datainfrastruktur-prosjektene med finansiering fra Forskningsrådet og fra nasjonale leverandører av tjenester for lagring og tilgjengeliggjøring av data og UH-sektoren.²

Kunnskapsgrunnlaget supplerer Norsk veikart for forskningsinfrastruktur når det gjelder datainfrastrukturer og omfatter følgende:

- En kortfattet gjennomgang av relevante pågående prosesser nasjonalt og internasjonalt (primært i EU) knyttet til tilgjengeliggjøring av forskningsdata, herunder litt om hvorfor tilgjengeliggjøring av forskningsdata har kommet på den politiske agendaen.
- En samlet oversikt over datainfrastrukturer som hittil har fått finansiering fra Forskningsrådet, primært fra Nasjonal satsing på forskningsinfrastruktur (ca. 1 milliard kr i samlet kontraktssum hittil).
- En oversikt over viktige nasjonale tjenester for håndtering av forskningsdata og forvaltningsdata av interesse for forskning
- En gjennomgang av ulike finansieringsmodeller for datainfrastrukturer.
- En gjennomgang av behovene vi har identifisert gjennom samtaler og dialogmøter.
- En kortfattet drøfting av avdekte behov og mulige løsninger, samt innspill (se nedenfor).

Kartleggingen viser at det finnes en del barrierer for å få tilgang til data fra de datainfrastrukturer som i dag finnes/etableres. Dette gjelder både forsknings- og forvaltningsdata. **De viktigste barrierene** er:

- Manglende felles standarder for data og metadata kan gjøre gjenbruk av forsknings- og forvaltningsdata vanskelig.
- Flere forskningsinstitusjoner er restriktive med å tilgjengeliggjøre egne data for gjenbruk.
- Uthenting av data fra datainfrastrukturer eller registre der kostnadene for tilgang belastes forskningsprosjekter, har ofte manglende forutsigbarhet hva angår kostnader og tiden det tar.
- For helse og persondata er det mange ulike aktører som igjen er underlagt ulike forskrifter og formål, og hvor komplekse og strenge utleveringsbetingelser vanskeliggjør gjenbruk av data.

Kunnskapsgrunnlaget avdekker vesentlige strukturelle, tekniske og økonomiske behov som bør adresseres for å oppnå sikrere lagring og bedre tilgang til forskningsdata og forvaltningsdata for forskning. Noen av de viktigste **behovene aktørene har løftet fram** er:

- Strengere krav om deling av forskningsdata og om datahåndteringsplaner i forskningsprosjekter

¹ Bestillingen fra KD er formulert slik i tildelingsbrevet til Forskningsrådet for 2016: "Kunnskapsdepartementet ser behov for bistand fra Forskningsrådet i forbindelse med en handlingsplan/strategi for tilgjengeliggjøring av, og tilgang til, data. Dette gjelder både data som er resultater av forskning og data som er relevante for forskning. Den tematiske bredden i problemstillingen tilsier at ulike deler av Forskningsrådets organisasjon bør bidra med kunnskap, innsikt og øvrig rådgivning."

² Med "datainfrastruktur" mener vi forskningsinfrastrukturer med formål å innsamle, bearbeide, lagre, kuratere, tilgjengeliggjøre og/eller forbedre utnyttelse av data.

- Institusjonelle policyer for datahåndtering
- Samordning av nasjonale tjenester og aktører
- Utvikling og innføring av internasjonalt omforente standarder for datahåndtering
- Opplæring og kulturendring blant forskere og studenter
- Insentiver for lagring og tilgjengeliggjøring av datasett
- Langsiktighet og forutsigbarhet ved finansiering av datainfrastruktur
- Bevaring og tilgjengeliggjøring av allerede innsamlede datasett slik at disse kan gjenbrukes

Med utgangspunkt i kunnskapsgrunnlaget gir Forskningsrådet følgende **innspill til det videre arbeidet for økt tilgjengeliggjøring av forskningsdata:**

- Forskningsinstitusjonene bør i samarbeid med egne forskere lage retningslinjer for hvilke data som skal lagres, kurteres og tilgjengeliggjøres ved egen institusjon.
- Forskningsrådet og andre som bidrar med offentlige midler til FoU bør vurdere å kreve en datahåndteringsplan for prosjekter som genererer forskningsdata.
- Kunnskapsdepartementet bør vurdere å etablere et "forskningsdataforum" for samordning og videreutvikling av nasjonale løsninger for lagring, analyse og tilgjengeliggjøring av forskningsdata.
- Forskningsinstitusjonene bør vurdere å etablere felles kommunikasjonsarenaer hvor de kan diskutere fagspesifikke standarder som muliggjør deling og gjenbruk av data.
- Forskningsinstitusjonene bør vurdere å utvikle én felles, nasjonal tjeneste for sikker analyse og korttidslagring av sensitive forskningsdata.
- Institusjonene bør vurdere å legge bedre til rette for opplæring i datahåndtering og gjenbruk av data.
- Kunnskapsdepartementet bør vurdere å utrede insentiver for tilgjengeliggjøring av forskningsdata.
- Institusjoner med ansvar for datainfrastrukturer bør etablere langsiktige driftsfinansieringsmodeller.
- Forskningsrådet bør vurdere å utarbeide kriterier for hvordan Nasjonal satsing på forskningsinfrastruktur i spesielle tilfeller skal kunne bidra til grunnfinansiering av drift av datainfrastrukturer.
- Kunnskapsdepartementet, i samarbeid med relevante fagdepartementer, bør vurdere å utrede mulighetene for en strukturell finansiering som kan dekke kostnader til å gi forskerne innsyn i hvilke forvaltningsdata som finnes og avklare om forskningsprosjektet kan få tilgang til nødvendige data.
- KD bør vurdere å opprette en arbeidsgruppe med representanter fra sentrale forskningsinstitusjoner som kan utfylle eksisterende kunnskapsgrunnlag.

2 Bakgrunn og mandat

Mengden av digitale data som produseres tilsier at vi nå opplever en datarevolusjon på de fleste områder av samfunnet. Forskningen i dag er i økende grad datadrevet, og i en rapport fra EU-kommisjonen påpekes det at 90 % av alle data som er samlet inn har blitt generert de siste to årene (European Commission 2016³). Dette representerer utfordringer for forskningen som påvirker hele forskningskretsløpet, fra datainnsamling og analyse til publisering og tilgjengeliggjøring av dataene.

Tilgjengeliggjøring og gjenbruk av forskningsdata kan bidra til vitenskapelige fremskritt i form av økt etterprøvbarehet og transparens, og til å unngå at samme type datainnsamling må finansieres flere ganger. I tillegg kan økt tilgjengeliggjøring av forskningsdata bidra til innovasjon, økonomisk vekst og samfunnsutvikling (European Commission 2016⁴).

I Meld. St. 27 (2015-2016) Digital agenda for Norge – IKT for en enklere hverdag og økt produktivitet påpeker regjeringen viktigheten av å legge til rette for datadrevet innovasjon og teknologibruk. I arbeidet med å styrke tilgjengeliggjøringen av åpne offentlige data, vil regjeringen at det skal utarbeides strategier eller strategier for fem sektorer som er funnet spesielt samfunnsøkonomiske verdifulle. Disse er kultur, forskning og utdanning, statlige utgifter, transport og samferdsel, og kart og eiendom.

³ European Commission 2016a. "Realising the European Open Science Cloud". [hentet 18.11.16].

⁴ European Commission 2016b. "Open Innovation, Open Science, Open to the World. A vision for Europe".[hentet 18.11.16].

Statssekretærutvalget for Digital agenda har gitt Kunnskapsdepartementet (KD) ansvaret for å utforme en strategi for tilgjengeliggjøring av forskningsdata innen utgangen av 2017. I den forbindelse ba Kunnskapsdepartementet Forskningsrådet om å etablere kunnskapsgrunnlag.

Kunnskapsgrunnlaget/rapporten belyser Forskningsrådets finansiering av infrastrukturer som støtter tilgjengeliggjøring av forskningsdata, viktige tjenester for datalagring i Norge, arbeidsdeling og rolleavklaring mellom tjenesteleverandører, og mulige finansieringsmodeller for nasjonalt viktige tjenester knyttet til lagring og deling av data. Rapporten er utarbeidet på basis av interne dokumenter i Forskningsrådet, analyser og dialog med representanter for ulike institusjoner innen forvaltning og forskning.

Rapporten er strukturert på følgende vis: Kapittel 2 gjengir KDs bestilling til Forskningsrådet sammen med en nærmere definisjon og avgrensning av oppgaven. Kapittel 3 viser til relevante pågående prosesser nasjonalt og internasjonalt (primært EU) knyttet til tilgjengeliggjøring av forskningsdata. Videre redegjøres det for prosjektets arbeidsform og metoder i kapittel 4. Kapittel 5 beskriver Forskningsrådets finansiering av infrastrukturer som støtter tilgjengeliggjøring av forskningsdata hittil, og gir en oversikt over viktige norske tjenester for lagring og håndtering av forskningsdata og forvaltningsdata av interesse for forskning. Kapittel 6 beskriver ulike finansieringsmodeller for datainfrastrukturer. Kapittel 7 går gjennom de avdekte behovene som har blitt identifisert gjennom samtaler, intervjuer og dialogmøter med utvalgte fagpersoner og ledere. Med utgangspunkt i disse presenterer kapittel 8 drøftinger av avdekte behov og mulige løsninger, mens kapittel 9 presenterer Forskningsrådets innspill.

Spesifikke ord (som "datainfrastruktur") brukt i rapporten er definert samlet i vedlegg 10.1.

2.1 Mandat og avgrensning

Bestillingen fra KD er formulert som følger i tildelingsbrevet til Forskningsrådet for 2016:

"Kunnskapsdepartementet ser behov for bistand fra Forskningsrådet i forbindelse med en handlingsplan/strategi for tilgjengeliggjøring av, og tilgang til, data. Dette gjelder både data som er resultater av forskning og data som er relevante for forskning. Den tematiske bredden i problemstillingen tilsier at ulike deler av Forskningsrådets organisasjon bør bidra med kunnskap, innsikt og øvrig rådgivning."

I samråd med KD har Forskningsrådet fortolket og avgrenset bestillingen slik:

- Kartlegge Forskningsrådets finansiering av infrastrukturer som støtter tilgjengeliggjøring av forskningsdata
- Kartlegge hvilke tjenester for datalagring som finnes i dag
- Gi innspill til arbeidsdeling og rolleavklaring mellom tjenesteleverandører
- Gi innspill til mulige finansieringsmodeller for nasjonalt viktige tjenester knyttet til lagring og deling av data

Forskningsrådet tolker KDs mandat til å ha en nasjonal tilnærming, og har valgt å fokusere på norske tjenester for lagring og deling av data. Videre har vi i samråd med KD, valgt ikke å gå tungt inn i en del andre relaterte og pågående prosesser der bl.a. tilgjengeliggjøring av data for helserelatert forskning er sentralt:

- Helse og omsorgsdepartementet (HOD) ferdigstilte sin internrapport om Nasjonal helseregisterstrategi – status, utfordringer og veien videre 30. august.
- Agenda Kaupang leverte sin rapport til Forskningsrådet om "Persondata til helseforskning" 15. desember (hovedtrekk ble gjort kjent på et arrangement 8. november).
- E-helsedirektoratets utredning av Plattform for e-helse
- HOD mottar rapporten fra ekspertutvalget som skal utrede enklere tilgang og bedre utnyttelse av helsedata, den 30. juni 2017.

- Arbeidsgruppen for IKT-strategi og helhetlige løsninger i UH-sektoren leverte første delrapport til KD 29. juni og vil levere endelig rapport med forslag til IKT-strategi for UH-sektoren i desember. Spørsmålet om IKT-organisering utgjør en viktig komponent i oppfølgingen av rapporten fra Gjedrem og Fagernæs, "Kunnskapssektoren sett utenfra" (jan. 2016), om organisering av kunnskapssektoren.
- FIN har nedsatt et utvalg som skal vurdere statistikkloven og Statistisk sentralbyrås virksomhet. Utvalget skal levere sin rapport innen 15. desember 2017.

3 Tidligere og pågående prosesser internasjonalt og i Norge

Det er betydelig politisk fokus på den enorme mengden av digitale data som produseres i dagens samfunn. Forskningsdata blir i denne sammenhengen gjerne omtalt som "drivstoffet" som skal sikre samfunnsutvikling. Tabell 1 gir en oversikt over noen av de viktigste internasjonale og nasjonale "statements", retningslinjer, "soft-laws", policyer og rapporter som omhandler tilgjengeliggjøring av forskningsdata. Mange av disse dokumentene har vært førende med anbefalinger som har påvirket politikktutviklingen på området både i de nordiske landene og ellers i Europa.

4 Arbeidsform og metode

For å få et bilde av hvilke datahåndteringstjenester som finnes i dag, har vi laget en oversikt over hvilke datainfrastrukturer Forskningsrådet har finansiert. Videre har vi gjennomført samtaler med sentrale aktører som tilbyr tjenester for datalagring og/eller tilgjengeliggjøring av forskningsdata (se vedlegg 10.2-3). Aktørene omfatter både datainfrastrukturprosjekter med finansiering fra Forskningsrådet, nasjonale tjenesteleverandører, institusjoner med et nasjonalt ansvar for håndtering av forsknings- og/eller forvaltningsdata, og UH-sektor.

Hensikten med samtalene har vært å få informasjon om tjenestenes tilbud, deres organisering og rolleavklaring de ulike aktørene i mellom. I tillegg har vi spurt om hvordan aktørene selv mener man kan oppnå bedre tjenestetilbud innen datalagring og tilgjengeliggjøring av forskningsdata i fremtiden. Aktørene ble videre bedt om å nevne utfordringer de opplever relatert til drift, standardiseringer av tjenester, finansiering, og å komme med forslag til løsninger for bedre modeller. Samtalene ble gjennomført i perioden september – desember 2016. I tillegg ble det arrangert to dialogmøter der representanter for aktørene deltok (se vedlegg 10.4).

Samtaler med kollegaer i Forskningsrådets ulike fagavdelinger har blitt gjennomført for å innhente relevant informasjon om de ulike fagfeltene og for å kvalitetssikre kartleggingen av tjenester og aktører.

Vi har også gjennomgått nasjonale og internasjonale dokumenter og rapporter og deltatt på utvalgte nasjonale og internasjonale konferanser. En av rapportforfatterne har vært representant i en OECD-oppnevnt, internasjonal arbeidsgruppe som utreder finansieringsmodeller for datainfrastrukturer⁵.

⁵ Arbeidsgruppen er oppnevnt av OECD Global Science Forum og skal utrede "Sustainable Business Models for Data Repositories". De leverer sin rapport i første halvdel av 2017.

Tabell 1: Oversikt over noen av de viktigste internasjonale og nasjonale dokumenter for "Open Access to Research Data"

Internasjonalt (Primært Europa)	Nasjonalt
	2004
<ul style="list-style-type: none"> • Declaration on Access to Research Data From Public Funding 	
	2006
<ul style="list-style-type: none"> • ERC Scientific Council Statement on Open Access 	
	2007
<ul style="list-style-type: none"> • ERC Scientific Council guidelines for open access • "Principles and Guidelines for Access to Research Data from Public Funding" 	
	2008
	2009
<ul style="list-style-type: none"> • "The Fourth Paradigm: Data Intensive Scientific Discovery"⁶ 	<ul style="list-style-type: none"> • Meld. St. 30 (2008-2009) "Klima for forskning"
	2010
<ul style="list-style-type: none"> • "Riding the Wave: How Europe can gain from the rising tide of scientific data" 	
	2012
<ul style="list-style-type: none"> • "Recommendation on Access to and Preservation of Scientific Information" • "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020" 	<ul style="list-style-type: none"> • Meld. St. 18 (2012-2013) "Lange linjer – kunnskap gir muligheter"
	2013
<ul style="list-style-type: none"> • "Data Harvest - How Sharing Research Data Can Yield Knowledge, Jobs And Growth" 	
	2014
	2015
<ul style="list-style-type: none"> • "Making Open Science a Reality" • EU kommisjonens strategi for et digitalt enhetlig marked i EU • "European Open Science Cloud for Research" 	<ul style="list-style-type: none"> • "Tilgjengeliggjøring av forskningsdata. Policy for Norges forskningsråd" • Meld. St. 27 (2015-2016) "Digital agenda for Norge" • "Dataeksplosjonen – en stor utfordring, og en gedigen mulighet!"
	2016
<ul style="list-style-type: none"> • EU-kommisjonens beslutning om at data skal være åpne som standard i H2020 • "Guidelines on FAIR Data Management in Horizon 2020" • EU-kommisjonen forslag til nytt copyright direktiv innenfor Digital Single Market. • "Realising the European Open Science Cloud" • «Open innovation, open science, open to the world - a vision for Europe" • "The FAIR Guiding Principles for scientific data management and stewardship". • "Open Access to research Data – Status, Issues and Outlook" • ERAC Opinion on Open Research Data 	<ul style="list-style-type: none"> • "Nasjonal helseregisterstrategi – status, utfordringer og veien videre" • "Persondata til helseforskning"⁷

⁶ Hey, T. 2012. "The Fourth Paradigm - Data-Intensive Scientific Discovery, E-Science and Information Management ". [hentet 15.12.16]

⁷ Agenda Kaupang 2016. "Persondata til helseforskning ". [Agenda Kaupang leverte rapporten desember 2016 til Forskningsrådet]

5 Kartlegging av norske datainfrastrukturer

I dette kapittelet gis en oversikt over våre kartlegginger av tjenester for datalagring og deling som finnes nasjonalt. Delkapittel 5.1. gir en oversikt over de prosjekter for etablering av datainfrastruktur om hittil har mottatt støtte fra Forskningsrådet, primært gjennom Nasjonal satsing på forskningsinfrastruktur (heretter kalt INFRASTRUKTUR) etablert i 2009. Det er viktig å poengtere at innenfor kategorien datainfrastruktur inngår det prosjekter som har datahåndtering som ett av flere sentrale element. Videre gir delkapittelet en oversikt over hvilke type data disse infrastrukturene håndterer når de er ferdig etablerte og kommet inn i en driftsfase.

Delkapittel 5.2. fokuserer på de viktigste norske datainfrastrukturene og tjenestene for datalagring og håndtering utover de Forskningsrådet har finansiert. Vi presenterer ingen uttømmende liste, men en fremstilling basert på Forskningsrådets samtaler med aktørene (se vedlegg 10.2). De kartlagte tjenestene tilbys av institusjoner med nasjonale oppgaver, og de innbefatter generiske- og fagspesifikke datainfrastrukturer som håndterer forskningsdata og/eller forvaltningsdata.

Det at en datainfrastruktur tilgjengeliggjør data, betyr ikke at tilgangen til data er åpen i form av å være fritt nedlastbar eller gratis. Flere av tjenestene som omtales opererer for eksempel med betaling for tilgjengeliggjøring av data. Tjenestene opererer også med ulike løsninger for innlogging og autentisering, se tabell 2. Tjenestene kan deles inn i ulike kategorier:

- Databaser som lagrer og tilgjengeliggjør primærdata fra måleinstrumenter eller fra store undersøkelser
- Tjenester som lagrer og tilgjengeliggjør data fra forskningsprosjekter
- Søke- og metadatatjenester som forbedrer tilgang til eksisterende data
- Tjenester som tilbyr sikker mellomlagring og analyse av sensitive data
- Nettside med innsynsløsninger for data

Autoriseringsløsningene som de ulike datainfrastrukturene benytter vil bli spesifisert for hver enkelt tjeneste i delkapitlene 5.1. og 5.2.

Tabell 2: Oversikt over ulike autentiseringsløsninger⁸.

Autentisering/ innloggingsløsning	Beskrivelse
Tilgang uten bruker-ID	Nedlasting av data fra åpen portal eller nettside uten å registrere seg
Egen løsning for tilgang basert på bruker-ID/passord	Nedlasting av data fra åpen portal eller nettside etter å ha opprettet eget brukernavn og passord
Tilgang med FEIDE/eduGAIN	Feide (Felles Elektronisk IDEntitet) er Kunnskapsdepartementets valgte løsning for sikker identifisering i utdanningssektoren ⁹ . Avgjørelsen om en bruker skal få tilgang til tjenesten er basert på de opplysningene tjenesten får fra vertsorganisasjonen. eduGAIN er en internasjonal autentiseringsløsning for de forsknings- og utdanningsinstitusjoner som er godkjente medlemmer av denne.
Tilgang fra godkjente institusjoner	Denne kategorien omfatter ulike løsninger: <ul style="list-style-type: none">• Søknads- og godkjenningssløsninger som gjelder for tilgang til helsefaglige og personsensitive data.• Løsninger der tjenesteleverandører gir tilgang kun til ansatte ved spesifiserte institusjoner.
Åpen søke- og metadatatportal	Åpen portal som tilgjengeliggjør metadatabeskrivelser og informasjon om hvor datasett ligger lagret.
Nettside med innsynsløsninger for data	Gir ikke tilgang til data, men innsynsløsninger og grafisk representasjon av data

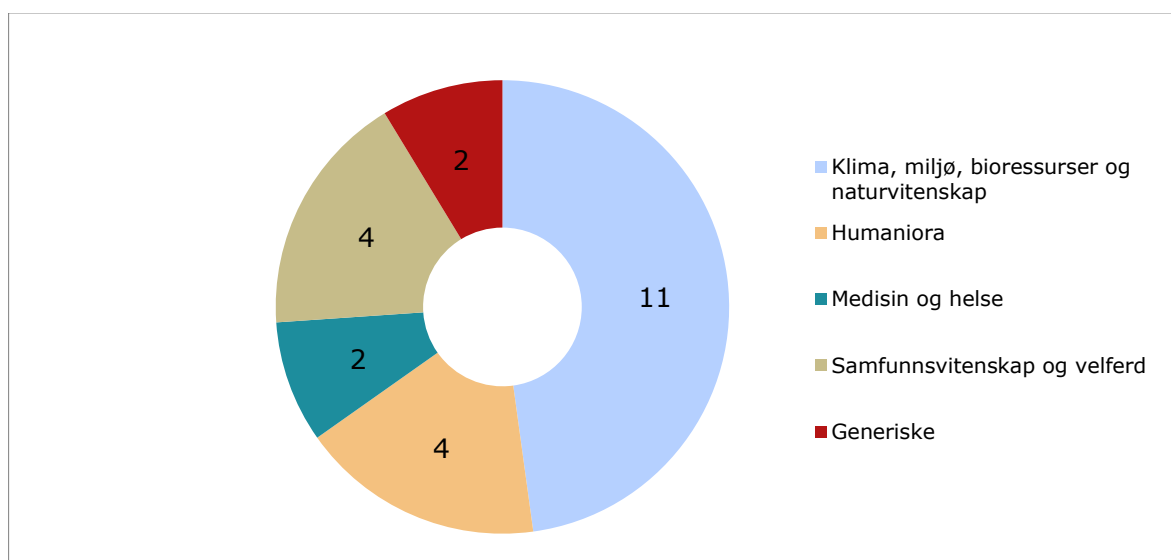
⁸ Flere av tjenestene utvikler nå løsninger for å legge til rette for pålogging fra 3. part IDPa (for eksempel Facebook, Google eller ORCID). Dette er ment som et supplement til vanlig Feide, spesielt med tanke på forskningsprosjekter med eksterne forskere.

⁹ Feide: Brukere registrerer seg hos egen vertsorganisasjon som er universiteter, høyskoler, kommuner og fylkeskommuner samt private skoleeiere. Vertsorganisasjon gir brukeren et brukernavn og passord, og er ansvarlig for autentisering.

5.1 Forskningsrådets finansiering av datainfrastrukturer

Forskningsrådet har i perioden 2010-2015 bevilget 4,1 milliarder kroner gjennom INFRASTRUKTUR. Av disse infrastrukturprosjektene har ca. én milliard kroner gått til et tjuetalls infrastrukturer der håndtering og tilgjengeliggjøring av data er det primære eller sekundære målet. INFRASTRUKTUR har et overordnet mål om å bidra til at norske forskningsmiljøer og næringsliv skal ha tilgang til relevant og oppdatert infrastruktur som understøtter forskning av høy kvalitet. Datainfrastrukturer som Forskningsrådet har finansiert, omfatter både generisk og fagspesifikk e-infrastruktur, vitenskapelige databaser og samlinger, og tjenester for metodeutvikling og dataanalyse.

At de fleste datainfrastrukturprosjektene er finansiert gjennom INFRASTRUKTUR innebærer at de er forankret i Forskningsrådets såkalte "områdestrategier"¹⁰ og til sammen dekker et bredt spekter av faglige behov innen følgende områder: Bioressurser, Bioteknologi, Humaniora, IKT, Klima og miljø, Medisin og helse, Naturvitenskap og teknologi, og Samfunnsvitenskap og velferd. Figur 1 gir en oversikt over antall datainfrastrukturer per område. Informasjons- og kommunikasjonsteknologi (IKT) er ikke inkludert som eget område i figuren ettersom dette representerer et viktig element i alle datainfrastrukturer.



Figur 1: Antall datainfrastrukturer finansiert av Nasjonal satsning på forskningsinfrastruktur (INFRASTRUKTUR) innenfor ulike fagområder.

Tabellene 3-7 nedenfor gir en oversikt over den **totale** kontraktsummen og -periode for prosjekter med støtte fra Forskningsrådet hvor etablering av datainfrastruktur er et hovedformål eller et av formålene med prosjektet. Vi presiserer m.a.o. at beløpene i tabellene 3-7 ikke henviser til beløp benyttet til etablering av datainfrastruktur alene, men den totale kontraktsummen for de respektive prosjektene.

I tabellene 3-6 er de ulike infrastrukturene delt inn på følgende måte (én tabell for hver): naturvitenskapelig forskning, helsefaglig og bioteknologisk forskning, samfunnsvitenskapelig forskning og humanistisk forskning. Etter henholdsvis tabell 3 og tabellene 4-6 følger figurene 2 og 3 som kategoriserer de ulike datainfrastrukturene i henholdsvis "database", "søke- og metadatatportal" eller "analyseverktøy". I tillegg gir disse figurene informasjon om hvorvidt infrastrukturene gir tilgang til data (og metadata) eller kun metadata. Mens figur 2 gjelder datainfrastrukturer for naturvitenskapelig forskning, har vi samlet helsefaglig, samfunnsvitenskapelig og humanistisk forskning i figur 3. Etter hver av disse to figurene har vi kort omtalt hver enkelt datainfrastruktur nevnt i de forutgående tabellene/figurene.

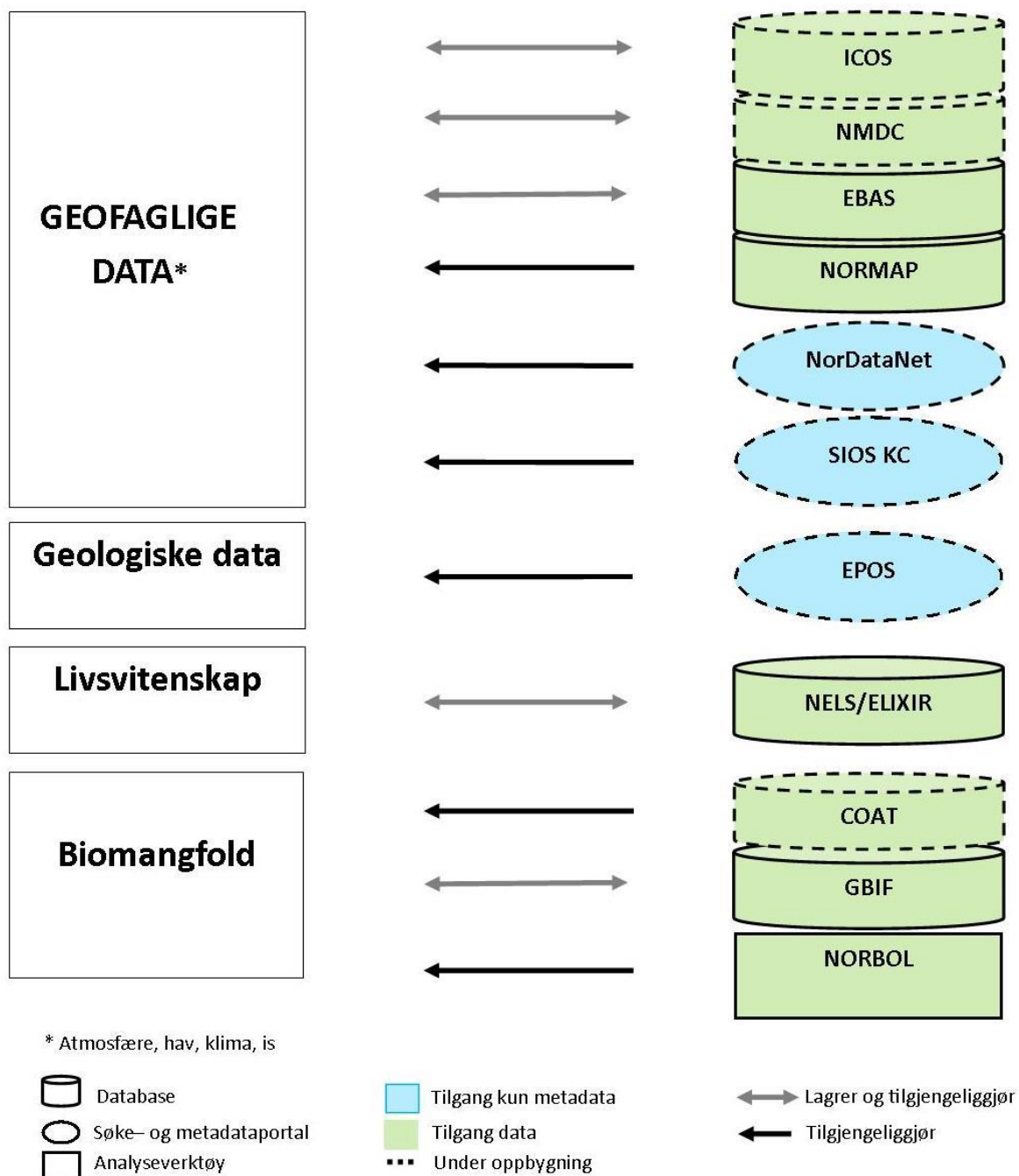
¹⁰ Områdestrategiene er en integrert del av Norsk veikart for forskningsinfrastruktur utarbeidet av de ulike fagavdelingene i Forskningsrådet i tråd med gjeldende strategier og strategier innenfor ulike fagfelt (se <http://www.forskningsradet.no/prognost-infrastruktur/Omradestrategier/1253972393843>).

Tabell 3: Kontraktsum og -periode for prosjekter med støtte fra Forskningsrådet for etablering av datainfrastrukturer for naturvitenskapelig forskning. Beløpene for Forskningsrådet gjelder primært etablering (ikke drift) av datainfrastrukturen.

Akronym	Navn	Institusjoner med i prosjektet	Sum totalt [mill. kr]	Kontraktsum, F.rådet [mill. kr]	Kontraksperiode, F.rådet	Tilgang
ICOS	Integrated Carbon Observation System Data Portal	UNI Research, CICERO, NILU, NIVA, NERSC, SKOG og LANDSKAP, HI, BIOFORSK, UiB, NPI	111,8	96,3	2016-2020	Tilgang uten bruker-ID Under oppbygging
NMDC	Norwegian Marine Data Centre	Havforskningsinstituttet, NTNU, CMR, FFI, UiN, Met.no, NERSC, NGU, NINA, Norsk polarinstitutt, Statens kartverk, UiB, UiO, UiT, NIVA, Akvaplan NIVA, UNI Research	64	48	2012-2015	Tilgang uten bruker-ID Under oppbygging
EBAS	A global harmonized database for atmospheric composition research	NILU	5,9	3,5	2011-2015	Tilgang uten bruker-ID
NORMAP	Norwegian Satellite Earth Observation Database for Marine and Polar Research	Nansensenteret i Bergen (NERSC), CERSAT, Kongsberg Satellite Services, Meteorologisk institutt	50,7	25	2010-2016	Tilgang uten bruker-ID for enkeltdatasett. Tilgang med bruker-ID/passord for transformasjoner o.l.
NorDataNet	Norwegian Scientific Data Network	Meteorologisk institutt, Norsk polarinstitutt, Havforskningsinstituttet, Nansensenteret, Norsk institutt for luftforskning	26	20	2015-2020	Åpen søke- og metadatatportal Under oppbygging
SIOS KC	The Svalbard Integrated Arctic Earth Observing System Knowledge Centre	UNIS, NERSC, Meteorologisk institutt, Norsk polarinstitutt	68,1	35	2015-2018	Åpen søke- og metadatatportal Under oppbygging
EPOS	European Plate Observing System – Norway (EPOS)	UiB, NORSAR, NGU, UiO, CMR, Statens kartverk	84,3	51	2016-2020	Åpen søke- og metadatatportal Under oppbygging
NELS/ELIXIR.NO	Norwegian e-Infrastructure for Life Science	UiB, UiO, NTNU, UiT, NMBU	82,7	50	2012-2017	Tilgang med FEIDE/eduGAIN
COAT	Climate-Ecological Observatory for Arctic Tundra	UiT, BIOFORSK, UNIS, NINA, NPI, NTNU, Met.no, HiH, UiO	42,3	35,5	2016-2020	Tilgang uten bruker-ID Under oppbygging
GBIF Norge	The Global Biodiversity Information Facility Norge	GBIF, UiO, Artsdatabanken	33,1	23,7	2005-2016	Åpen tilgang med og uten ID/passord

(Tabell 3 fortsetter)

NorBOL	Norwegian Barcode of Life	NTNU Vitenskapsmuseet, Naturhistorisk museum UiO, Bergen Museum UiB, Tromsø Museum UiT, Biodiversity Institute of Ontario	98,6	25,6	2014-2018	Tilgang uten bruker-ID gjennom iBOL
SUM			667,5	413,6		



Figur 2: Kategorisering av datainfrastrukturene oppsummert i tabell 3 (naturvitenskap) i "database", "søke- og metadataportal" eller "analyseverktøy", samt informasjon om hvorvidt infrastrukturen gir tilgang til data (og metadata) eller kun metadata.

Datainfrastrukturer for naturvitenskapelig forskning:

- **Integrated Carbon Observation System, Norway (ICOS)** er et ESFRI-prosjekt der den norske delen skal bidra inn med norske karbonobservasjoner fra målestasjoner og havgående fartøy. Prosjektet skal etablere en dataportal og tjenester for visualisering og utvikling av dataprodukter.
- **Norwegian Marine Data Centre (NMDC)** etablerer en portal og database for å gi tilgang til forskningsdata om havet som er spredt hos ulike forskergrupper og institusjoner i Norge. Dette skal både fungere som en søke og lagringstjeneste. Målet er at dataseriene enklere skal kunne utveksles mellom forskningsinstitusjoner.
- **A Global Harmonized Database for Atmospheric Composition Research (EBAS)** er en internasjonal database driftet av NILU, som tar imot og tilgjengeliggjør atmosfæriske data fra mer enn 40 internasjonale forskningsprosjekter og program.
- **Norwegian Satellite Earth Observation Database for Marine and Polar Research (NORMAP)** tilrettelegger satellittdata i en tverrfaglig database. Her skal observasjonsdata fra de nordlige havområdene og områdene rundt Arktis bearbeides, kvalitetssikres og tilrettelegges i et mer brukervennlig og tilgjengelig format for interesserte forskere. En dataportal er etablert hvor data kan lastes ned fra distribuerte datasentre.
- **Norwegian Scientific Data Network (NorDataNet)** skal gjøre eksisterende klima- og miljødata tilgjengelige gjennom en felles portal. Et viktig mål er å gi sømløs tilgang til data fra distribuerte datasentre gjennom en portal.
- Gjennom **The Svalbard Integrated Arctic Earth Observing System (SIOS)** skal forsknings og overvåkningssystemer på Svalbard samordnes og oppgraderes i et felles observasjonssystem. Dette omfatter feltdata, data fra bakkenære observasjonsplattformer og satellittdata. En komponent av SIOS er å utvikle et system for å integrere og gi tilgang til metadata fra andre databaser og arkiver.
- **European Plate Observing System - Norway project (EPOS-N)**, er det norske nasjonale bidraget i ESFRI-prosjektet EPOS. Et av målene for prosjektet er å gi forskere lettere tilgang til forskningsdata om jordens fysikk på tvers av landegrenser. En nasjonal webportal er etablert hvor det pekes til data som ligger lagret i ulike databaser.
- **Norwegian e-Infrastructure for Life Sciences (NeLS)** er en norsk node innen det pan-europeiske ESFRI prosjektet ELIXIR. NeLS utvikler en web-portal og database, samt analyserings- og visualiseringsverktøy for ikke sensitive biologiske data.
- **Climate- Ecological Observatory for Arctic Tundra (COAT)** har som formål å støtte forskning på klimaendringenes effekt på terrestriske økosystemer i Arktis gjennom fjernmåling og observasjoner. Infrastrukturen inngår i et internasjonalt nettverk med et datalagringsystem.
- **Global Biodiversity Information Facility (GBIF Norge)**. Den norske noden i GBIF mottar løpende finansiering fra Forskningsrådet, men ikke gjennom INFRASTRUKTUR. GBIF Norge tilgjengeliggjør data fra norske samlinger og institusjoner. Prosjektet har en portal som gir tilgang til primærdata om jordens biologiske mangfold fra hele GBIF nettverket.
- **Norwegian Barcode of Life Network (NorBol)** er en datainfrastruktur innenfor biologisk mangfold der artskartlegging utgjør en viktig komponent gjennom DNA-analyser. Dataene tilgjengeliggjøres gjennom et verdensbibliotek over jordens arter, International Barcode of Life (iBOL).

Selv om disse infrastrukturene spenner over svært ulike fagområder, har flere av dem en fellesnevner i at de utvikler eller tilbyr løsninger for å gi tilgang til metadata og/eller data på tvers av eksisterende databaser. NorDataNet, og SIOS KC og EPOS skal utvikle søke- og metadataportaler som forenkler tilgangen til eksisterende data. NMDC, ICOS Data portal og COAT skal alle etablere databaser og tjenester for tilgang til data. Siden flere er under oppbygging er det kun gjennom NORMAP og NELS forskere får tilgang til data, i tillegg til NorBOL som tilgjengeliggjør data gjennom International Barcode of Life (iBOL). Når det gjelder muligheter for datalagring for relevante forskningsprosjekter vil ICOS, NMDC og NELS tilby dette.

Tabell 4: Kontraktsum og -periode for prosjekter med støtte fra Forskningsrådet for etablering av datainfrastrukturer for helsefaglig forskning. Beløpene for Forskningsrådet gjelder primært etablering (ikke drift) av datainfrastrukturen.

Akronym	Navn	Institusjon	Sum totalt [mill. kr]	Kontraktsum, F.rådet [mill. kr]	Kontraksperiode, F.rådet	Tilgang
Biobank Norway	A national infrastructure for biobanks and biobank related activity in Norway	NTNU, UiO, UiB, UiT Norges arktiske universitet, FHI, Folkehelseinstituttet, Helse Sørøst, Helse Vest, Helse Midt-Norge og Helse Nord	163	165,3	2010-2015	Åpen søke- og meta-dataportal. Under oppbygging
Helseregistre for forskning	Helseregistre for forskning	UiB, NTNU, UiT, FHI, Nasjonalt Kunnskapssenteret for Helsetjenesten, Helse Nord RHF, Helsedirektoratet	60	36,5	2015-2019	Åpen søke- og metadataportal Under oppbygging
INCF	International Neuroinformatics Coordinating Facility	Nettverket består av medlemsland med tilhørende nasjonale noder og har medlemsland på tvers av Europa, USA, Asia og Australia.	1,5	7,7	2006-2016	Tilgang for godkjente brukere
SUM			224,5	209,5		

Tabell 5: Kontraktsum og -periode for prosjekter med støtte fra Forskningsrådet for etablering av datainfrastrukturer for samfunnsvitenskapelig forskning. Beløpene for Forskningsrådet gjelder primært etablering (ikke drift) av datainfrastrukturen.

Akronym	Navn	Institusjon	Sum totalt [mill. kr]	Kontraktsum, F.rådet [mill. kr]	Kontraksperiode, F.rådet	Tilgang
RAIRD	Remote Access Infrastructure for Register Data	NSD, SSB	42,2	35	2012-2018	Tilgang for godkjente brukere
e-VIR	eInfrastructure for Video Research	UiO, USIT, NSD	21,5	11,6	2016-2019	Tilgang for godkjente brukere Under oppbygging
ACCESS	ACCESS Life Course Database	NOVA, SSB	8,5	6,0	2010-2014	Tilgang med FEIDE/Dataporten
ACDC	Advanced Conflict Data Catalogue	PRIO, Uppsala universitet	8,6	8,0	2010-2013	Nettside med innsynsløsninger for data
CESSDA	Council of European Social Science Data Archives	NSD nasjonal tjenesteleverandør, CESSDA-samarbeid mellom 14 europeiske land	36,7	36,7	2013-2018	Åpen søke- og metadataportal
ESS	ESS-Survey Norwegian Membership and Participation in the European Social Survey (ESS)	NSD	28	10	2012-2016	Tilgang med FEIDE/Dataporten
SUM			145,5	107,3		

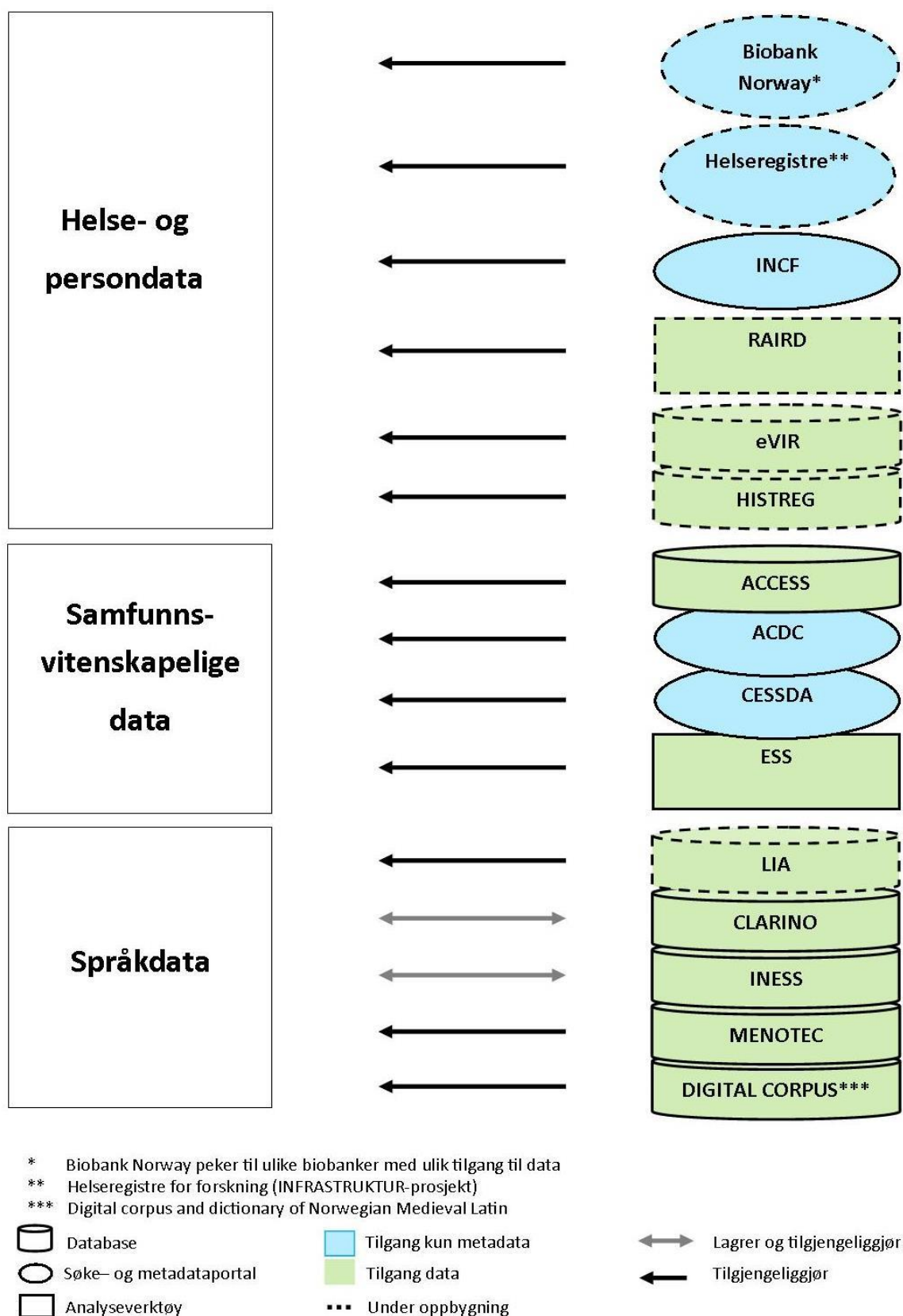
Tabell 6: Kontraktsum og -periode for prosjekter med støtte fra Forskningsrådet for etablering av datainfrastrukturer for humanistisk forskning. Beløpene for Forskningsrådet gjelder primært etablering (ikke drift) av datainfrastrukturen.

Akronym	Navn	Institusjon	Sum totalt [mill. kr]	Kontraktsum, F.rådet [mill. kr]	Kontraktperiode, F.rådet	Tilgang
LIA	Language Infrastructure made Accessible	UiO, UiB, UiT, NTNU, Nasjonalbiblioteket, Norsk Ordbok 2014, Humboldt-Universität zu Berlin, Pennsylvania State University, University of Wisconsin	40,9	27,3	2014-2019	Tilgang med FEIDE/eduGAIN
CLARINO ¹¹	Common Language Resources and Technology Infrastructure	UiB, UiO, NHH, UiT, NTNU, UNI Research	49,7	25	2012-2019	Tilgang med FEIDE/eduGAIN
INESS ¹²	Infrastructure for the Exploration of Syntax and Semantics	UiB, UNI Research, UiO, UiT, Nasjonalbiblioteket, UNINETT Sigma2, Kunnskapsforlaget, Microsoft	23,4	19	2010-2016	Tilgang med FEIDE/eduGAIN
Menotec ¹³	Medieval Norwegian Text Corpus	UiB og UiO	10	7	2010-2012	Tilgang med FEIDE/eduGAIN (gjennom INESS)
Digital corpus	Digital corpus and dictionary of Norwegian Medieval Latin	Nasjonalbiblioteket og Kunnskapsforlaget	10,7	7,8	2016-2022	Tilgang uten bruker-ID (gjennom Språkbanken og Bokhylla.no)
HISTREG	National Historical Population Register for Norway 1800-2020	UiT, Riksarkivet, SSB, NR, FHI, UiB, UiS, HiVolda, Lokalhist. Inst., Snøhetta forlag	63,4	25	2014-2023	Tilgang uten bruker-ID for data før 1920. Tilgang for godkjente brukere etter 1920 Under oppbygging
SUM			198,1	111,1		

¹¹ CLARINO omfatter flere noder, inkludert INESS og Menotec.

¹² Se fotnote 11.

¹³ Se fotnote 11.



Figur 3: Kategorisering av datainfrastrukturene oppsummert i tabell 4 (helsefag og bioteknologi), tabell 5 (samfunnsvitenskap) og tabell 6 (humaniora) i "database", "søke- og metadatatportal" eller "analyseverktøy", samt informasjon om hvorvidt infrastrukturen gir tilgang til data (og metadata) eller kun metadata.

Datainfrastrukturer for helsefaglig/bioteknologisk, samfunnsvitenskapelig og humanistisk forskning:

- **Biobank Norge** skal etablere en nasjonal biobankinfrastruktur for helseforskning, og har som ambisjon å bidra til økt bruk av biobanker for forskning, innovasjon og internasjonalt forskningssamarbeid. Biobank Norge skal også utvikle en nettportal med informasjon om tilgjengelige data gjennom et register over alle nasjonale biobanker, samt en beskrivelse av innholdet i biobankene. Utviklingen av denne web-baserte løsningen skal gjennomføres i samarbeid med datainfrastrukturprosjektet Helseregistre for forskning (se nedenfor). Alle har tilgang til å søke i metadata og se noen anonymiserte forskningsresultater fra biobankene. For å kunne benytte biologisk materiale i forskning må man ha godkjent forskningsprosjekt hos Regional Etisk Komité (REK) samt få godkjent prosjektet av biobankens styre.
- **Helseregistre for forskning** har som overordnet mål å gjøre helseregistrene mer tilgjengelig for forskning. Dette prosjektet arbeider med å styrke datasikkerheten og personvernet og skal fremme trygg og sikker lagring og analyse av forskningsdata. Prosjektet fokuserer på å bedre dokumentasjonen av helseregistrene og jobber blant annet for å utvikle bedre løsninger for sikker tilgang til dataene i registrene og økt datasikkerhet. I tillegg er ett av målene å forbedre kommunikasjon og øke muligheten til kobling mellom registrene ved å harmonisere variabler. Helseregistre for forskning skal videre etablere en online database for helseregister metadata. Prosjektet er koblet til 16 av de sentrale nasjonale registre, hvor 8 av 10 lovbestemte registre er inkludert i tillegg til en rekke kliniske registre.
- **International Neuroinformatics Coordinating Facility (INCF)** skal fremme, koordinere og implementere nevroinformatikk globalt ved å etablere, vedlikeholde, standardisere og oppdatere et distribuert databasesystem med nasjonale noder, koordinere og synkronisere aktivitetene ved de nasjonale nodene og implementere et eget arbeidsprogram. Alle deltakende land skal etablere en nasjonal node, som skal både koordinere aktiviteter nasjonalt og bidra inn i INCFs arbeidsprogram.
- **Remote Access Infrastructure for Register Data (RAIRD)** skal bli et web-basert forskningsverktøy som forenkler tilgang til og analyse av registerdata som er innsamlet med hjemmel i Statistikkloven. Prosjektet skal sikre konfidensialitet slik at forskerne slipper å sende søknader om bruk av dataene til Datatilsynet, REK og de ulike dataeierne. Prosjektet planlegges å være klart fra 2017.
- Prosjektet **e-infrastructure for Video Research (e-VIR)** skal utvikle en nasjonal videodatabase for lagring av videodata. Et mål er både å ivareta personvern og bidra til tilgjengeliggjøring av data. Databasen er underutvikling og løsninger for tilgjengeliggjøring og bruk er foreløpig ikke klart.
- **Advanced Conflict Data Catalogue (ACDC)** er en database for kvalitetssikrede og harmoniserte datasett over verdens væpnede konflikter. ACDC prosjektet har utviklet en konflikt bibliografi med referanser til litteratur om 253 konflikter i UCDP/PRIO Armed Conflict Dataset og dekker perioden 1946-2012. Bibliografien inneholder også korte beskrivelser av konflikter (UCDP database) og definisjoner (UCDP/PRIO Armed Conflict Dataset Codebook).
- Prosjektet **ACCESS Life Course** har som mål å forenkle tilgangen til data fra livsløpsstudiene "Den norske panelstudien om livsløp, aldring og generasjon" (NorLAG) og "Studien av livsløp, generasjon og kjønn" (LOGG). I tillegg til å forenkle tilgang til datasett, har formålet med prosjektet vært å øke kvalitet på dataene, harmonisere data innsamlet på ulikt tidspunkt, metodikkutvikling og fremme kunnskapsutveksling innenfor livsløpsforskning på tvers av forskningsinstitusjoner, disipliner og nasjoner. Data tilgjengeliggjøres gjennom Norsk senter for forskningsdata (NSD).
- **European Social Survey (ESS)** er en akademisk undersøkelse av demografisk og sosioøkonomiske forhold som gjennomføres annethvert år. I 2014/15 ble undersøkelsen gjennomført i 21 land. NSD har ansvar for å lagre, distribuere og behandle dataene fra undersøkelse. Ved utgangen av 2015 hadde 60 000 personer lastet ned ESS-data fra NSD¹⁴.

¹⁴ Norsk senter for forskningsdata (NSD) 2015. Årsmelding. [hentet 6.11.16]

- Det integrerte dataarkivsystemet **Council of European Social Science Data Archives (CESSDA)** omfatter dataarkiver i 14 europeiske land. Tjenestene som tilbys fra de ulike leverandørene integreres slik at man kan søke etter både data og dokumentasjon fra samme sted. For dette ESFRI-prosjektet er Norge vertsnaasjon.
- **National Historical Population Register for Norway 1800-2020 (HISTREG)** har som formål å etablere et digitalisert historisk befolkningsregister. Dette vil utvide eksisterende befolkningsregistre fra 1964 og bakover til år 1800. Tilgang til datamateriale fra perioden før 1920 er åpent for allmenheten, som har mulighet til å rette opp i eventuelle koblinger som er feil. Data fra perioden etter 1920 er kun tilgjengelige for forskning gjennom søknad. Dette materialet vil i stor grad være anonymisert.
- **Language Infrastructure Made Accessible (LIA)** skal skape et talekorpus basert på innsamlede opptak av norske dialekter. Således skal prosjektet redde mange av talemålsopptakene som er samlet inn av universitetene de siste 60 årene. Opptakene blir transkribert, digitalisert og utstyrt med metadata. Tilgangen til korpuset er fritt tilgjengelig for forskning, men man må søke om tilgang gjennom et nettskjema. Talesamlingen skal integreres i den norske Språkbanken som er utviklet og forvaltet av Nasjonalbiblioteket.
- **CLARINO** er den norske noden av ESFRI-prosjektet **CLARIN (Common Language Resources and Technology Infrastructure)**. CLARINO har fått midler fra INFRASTRUKTUR for å bygge opp historiske og nåtidige elektroniske språkressurser og koble sammen språk- og tekstdatabaser. Gjennom prosjektet kobles norske databaser sammen med utenlandske og forskere enkelt laste ned data som har åpne lisenser (Creative commons). For data med andre typer lisenser kreves innlogging med en ID godkjent av EduGAIN. For ikke-akademikere kan datainfrastrukturen opprette en lokal konto. Gjennom CLARINO kan brukere også laste inn eget material som deretter må godkjennes av en redaktør før det kan publiseres. En node i CLARINO er CLARINO Bergen Centre, som omfatter en lagrings- og kurateringsressurs og trebankinfrastrukturen **INESS**. INESS er sertifisert som et distribuert kunnskapscenter i CLARIN, og er del av den norske noden CLARINO. Videre omfatter CLARINO arkivet Medieval Nordic Text Archive (MENOTA) som tar imot middelaldertekst, det Forskningsrådsfinansierte datainfrastrukturprosjektet Medieval Norwegian Text Corpus (**MENOTEC**) og et online arkiv, CORPUSCLE, for språkkorpus som bare er tilgjengelig gjennom søk. Andre CLARINO-noder med lagringstjenester er Tekstlaboratoriet ved UiO, Termportalen ved Norges Handelshøyskole, Senter for samisk språkteknologi Giellatekno ved UiT og Språkbanken ved Nasjonalbiblioteket.
- **Infrastructure for the Exploration of Syntax and Semantics (INESS)** er en database av setningsstrukturer som tilbyr tilgang, søk og visualisering av språkdata for norsk og andre språk og utgjør en node i CLARINO. Det er mulig for brukere å bygge egne trebanker, men dette må godkjennes av prosjektledelsen og være i tråd med relevante standarder for metadata og lisensiering. Noen av trebankene er åpent tilgjengelige for søk, mens for de fleste av datasettene kreves innlogging med ID fra Clarin, eduGAIN eller med OpenIDP.
- **Medieval Norwegian Text Corpus (Menotec)** har tilgjengeliggjort norske middelaldertekster fra perioden 1150-1550 gjennom Arkiv for nordiske middelaldertekster (Menota). Menota er et nettverk av 18 nordiske arkiver, biblioteker og institutter som arbeider med middelaldertekster og håndskriftsamlinger. Menotec er en CLARINO-node. Tekstene er tilgjengeliggjort gjennom INESS portal.
- **Digital corpus and dictionary of Norwegian Medieval Latin (Digital Corpus)**. Infrastrukturen skal tilgjengeliggjøre norske latinske middelaldertekster gjennom et elektronisk korpus og en elektronisk ordbok. De digitaliserte latinske middelaldertekstene vil gjøres tilgjengelige gjennom Bokhylla.no som driftes av Nasjonalbiblioteket. Her har brukere med en norsk IP-adresse tilgang til alle bøker publisert før år 2000. Tekstkorpuset kan lastes ned gjennom Språkbanken.

Som det fremgår av tabell 4 og figur 3 er Biobank Norge og Helseregistre for forskning metadatatjenester som har som formål å forenkle forskningen gjennom og blant annet skape oversikt over infrastrukturene som finnes. Begge prosjektene er under oppbygging, men pilene i figur 3 indikerer at formål med prosjektene er å gi tilgang til metadata om henholdsvis nasjonale biobanker og nasjonale helseregistre.

Innen det samfunnsvitenskapelige feltet, benytter prosjektene RAIRD, eVIR, ESS og CESSDA NSDs tjenester for lagring av data. RAIRD er primært et analyseverktøy som skal forenkle tilgang til registerdata, mens eVIR skal bygge opp en database for audiovisuelle data der disse skal lagres hos NSD. ESS er en akademisk undersøkelse av demografiske og sosioøkonomiske forhold i ulike europeiske land, hvor NSD lagrer og tilgjengeliggjør data fra undersøkelsen. NSDs arkivtjeneste oppgraderes gjennom prosjektet NORD-i og her kan forskere fra ulike fagdisipliner få tilgang til og lagre data. CESSDA har en portal som gir oversikt over dataarkiv i 14 europeiske land, inkludert NSDs arkiv.

Innenfor humaniora tilbyr 5 av 6 prosjekter tilgang til språkdata. I de fleste tilfeller må brukerne av språkinfrastrukturene benytte ulike løsninger for autentisering. Begrensninger av tilgang skyldes lisenser knyttet til opphavsrett eller at materialet er av en personsensitiv karakter. I CLARINO har brukerne mulighet til å laste inn egne forskningsdata.

Tabell 7 gir en oversikt over kontraktsum og -periode for prosjekter med støtte fra Forskningsrådet for etablering og/eller drift av generiske datainfrastrukturer med relevans for flere fagområder.

Tabell 7: Kontraktsum og -periode for prosjekter med støtte fra Forskningsrådet for etablering og/eller drift av generiske datainfrastrukturer.

Akronym	Navn	Institusjon	Sum totalt [mill. kr]	Kontraktsum, F.rådet [mill. kr]	Kontraktsperiode, F.rådet	Tilgang
NORD-I	Norwegian Open Research Data Infrastructure	NSD, The University of Michigan	31	26	2016–2020	Tilgang med FEIDE/Dataporten
E-INFRA 2014	E-INFRA 2014 - a national e-Infrastructure for science	UNINETT Sigma2, UiT, UiO, NTNU, UiB	191	75,7	2016-2017	Ikke relevant
NorStore	NorStore Research Data Archive I og II	Uninett Sigma2	81,5	54	2014 2010-2013	Tilgang med FEIDE/Dataporten med eduGAIN Åpen søke- og metadata-portal
Sigma2 e-infra-structure	Sigma2 – e-infra-structure services	Uninett Sigma2	429,6	138,4	2015-2019	Ikke relevant
Sum			733,1	294,1		

Uninett Sigma2 og NSD tilbyr nasjonale, generiske løsninger for lagring og tilrettelegging av forskningsdata for forskningsprosjekter på tvers av fagdisipliner. Gjennom INFRASTRUKTUR har Uninett Sigma2 fått midler til å etablere **Norwegian Data Storage Infrastructure (NorStore)** som er en lagringstjeneste for forskningsdata. NorStore har kapasitet til å håndtere og lagre store datamengder, og tilbyr tjenester for gjenfinning, overføring, publisering og analyse av forskningsdata. NorStore er åpen for brukere fra universitets- og høgskolesektoren, forskningsinstitutter, helseforetakene og næringslivets forskningsprosjekter med finansiering fra Forskningsrådet eller gjennom andre offentlige finansieringskilder. Tilgang til lagringskapasitet fordeles etter søknad hvor faglig kvalitet og behovets begrunnelse vektlegges.

NSD har fått midler til **Norwegian Open Research Data Infrastructure (NORD-i)** som vil bidra til å fornye NSDs systemer for lagring og tilgjengeliggjøring av forskningsdata. NORD-i skal bidra til forenklet tilgang til allerede innsamlede data og mer automatiserte prosesser for opplasting og arkivering av data.

Forskningsrådets samlede finansiering indikert i tabellene 3-7 er oppsummert i tabell 8.

Tabell 8: Oppsummering av beløp i tabellene 3-7.

Datainfrastrukturprosjekter pr fagområde	Sum totalt [mill. kr]	Kontraktsum, F.rådet [mill. kr]
Naturvitenskapelige datainfrastrukturer	667,5	413,6
Datainfrastrukturprosjekter helse- og persondata, humaniora og samfunnsvitenskap	224,5 145,5 198,1	209,5 107,3 111,1
Generiske datainfrastrukturer	733,1	294,1
SUM	1968,7	1135,6

5.2 Norske tjenester for datalagring og -håndtering

Gjennom intervjuer med ulike aktører (vedlegg 10.2) har vi kartlagt sentrale tjenester for lagring og tilgjengeliggjøring av forskningsdata i Norge utover datainfrastrukturprosjektene med finansiering fra Forskningsrådet omtalt ovenfor. I tillegg er det et stort antall internasjonale tjenester for datalagring som norske forskere benytter seg av som denne rapporten ikke omfatter. Det finnes mange ulike datalagringstjenester, både kommersielle, institusjonelle, fagspesifikke og generiske. Vår tilnærming har vært å få de ulike aktørene vi har snakket med til å identifisere hvilke norske lagringstjenester de anser som de viktigste innen sitt fagområde. Videre har vi tatt forskerens perspektiv og undersøkt hvor de kan få tilgang til data og hvor de kan lagre data. Vi skiller mellom tjenester for forskningsdata og tjenester for forvaltningsdata med relevans for forskning.

For å ivareta forskerperspektivet har vi kategorisert tjenestene etter hvilken type datasett de håndterer, samt delt dem inn i tre grupper med relevans for henholdsvis: i) naturvitenskap, ii) helse- og persondata, samfunnsvitenskap og humaniora (språk) og iii) generiske tjenester. En fjerde gruppe er tjenester for datalagring og -håndtering relevante for forvaltningsdata.

Tabellene 9-12 oppsummerer hva de respektive fire gruppene av tjenestene gjør, hvilke institusjoner som forvalter dem. Tabellene 9-12 gir også informasjon om de respektive autentiseringsløsningene for tilgang. Etter hver tabell følger en figur (figurene 4-7).

Tabell 9: Tjenester for datalagring og -håndtering relevante for naturvitenskap.

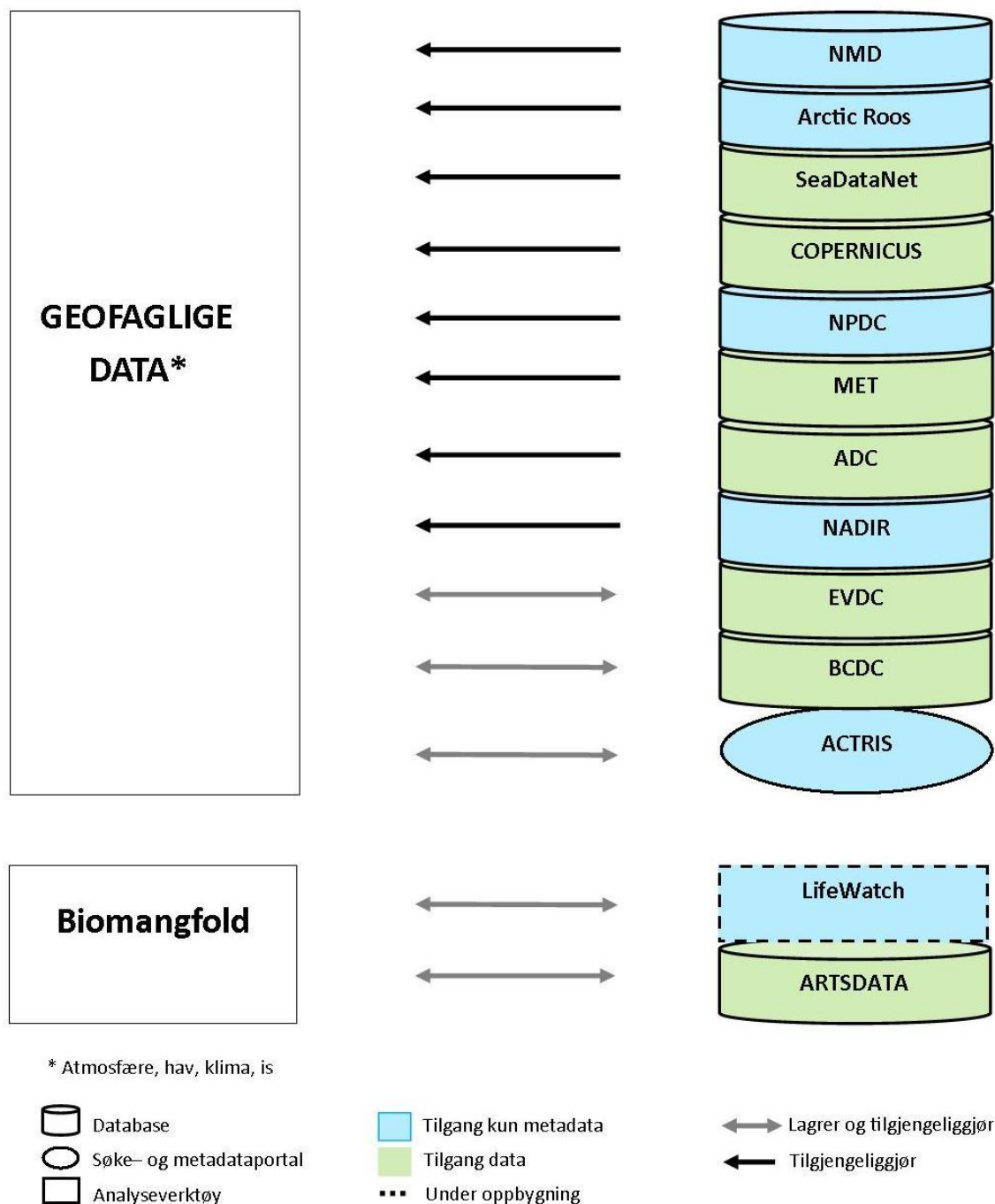
Akronym	Navn	Tjeneste	Institusjon og konsortier	Tilgang
NMD	Norsk marint datasenter	Nasjonalt datasenter for datahåndtering, fremstilling av dataprodukter og datalagring.	Havforskningsinstituttet (HI)	Nettside med innsynsløsninger for data
Arctic Roos	Arctic Regional Ocean Observing System	Observasjoner og grafisk representasjon av arktiske marine data.	Nansensenteret (NERSC), SMHI, Ifremer, HI, IOPAS, NIVA, DMI, MERCATOR, DAMTP, AWI, FMI, IUP, Met.no, NIERSC, NPI, GFI, FCOO	Nettside med innsynsløsninger for data
SeaDataNet	SeaDataNet Arctic	Norsk portal til EU SeaDataNet kvalitetssikra data	Havforskningsinstituttet	Tilgang med bruker-ID/passord
COPERNICUS	Copernicus in situ TAC Arctic	Portal for operasjonelle data i Arktis	Havforskningsinstituttet	Tilgang med bruker-ID/passord
NPDC	Norwegian Polar Data Centre	Database med oversikt over kart, datasett, fartøy, stedsnavn, tidsserier, prosjekt, publikasjoner, tracking over Svalbard/Arktis	Polarinstituttet	Tilgang med bruker-ID/passord

(Tabell 9 fortsetter)

Nedlastnings-tjenester MET ¹⁵	Nedlastningstjenester MET	Nedlastningstjenester med tilgang til vær- og klimadata og arktiske data	Meteorologisk institutt	Tilgang uten bruker-ID
ADC	Arctic Data Centre	Metadatakåndtering, distribusjon av data og visualisering	Meteorologisk institutt	Tilgang uten bruker-ID
NADIR	NILU's Atmospheric Database for Interactive Retrieval	Database atmosfæriske data fra forskningskampanjer rettet mot stratosfærisk ozon.	NILU	Tilgang for godkjente brukere
EVDC	ESA Validation Data Centre	Database for deling av CALVAL data for jordobservasjon.	NILU	Tilgang for godkjente brukere
BCDC	Bjerknes Climate Data Centre	Internasjonalt datasenter, fremstilling av dataprodukter (geologiske, oseanografiske og klimagasser), metadatakatalog og datalagring med hovedfokus på klimadata (modellering, hav, geologiske)	UiB, Uni Research, BCCR	Tilgang uten bruker-ID
ACTRIS Data Centre	The European Research Infrastructure for the observation of Aerosol, Clouds, and Trace gases	Forskningsinfrastruktur som harmoniserer og gir tilgang til atmosfæriske data. Benytter EBAS for «in-Situ målinger». Portalløsning gir tilgang til to tematiske databaser. Hovedstrukturen av ACTRIS Data Centre er lokalisert og driftet av NILU.	NILU ESFRI prosjekt	Åpen søke- og metadatatportal
LifeWatch ¹⁶	e-Infrastructure for Biodiversity and Ecosystem Research	ESFRI-prosjektet LifeWatch utvikler analyseverktøy for biodiversitetsdata	NINA, NIVA, HI, Artsdatabanken, GBIF og Naturhistorisk Museum/ UiO, (se fotnote 16) ESFRI-prosjekt	Analyseverktøy Under oppbygging
Artsdatabanken	Artsdatabanken	Dataportal, databaser med biodiversitetsdata.	KD, Artsdatabanken og GBIF	Tilgang uten bruker-ID

¹⁵ Meteorologisk institutts oversikt over nedlastningstjenester og frie data [hentet 15.12.16]

¹⁶ Prosjektet "Establishment of the LifeWatch Norwegian" er en del av Norges bidrag til prosjektet [ESFRI LifeWatch Europe](#), og har som hovedmål å utrede mulighetene for realisering av en norsk infrastruktur for deling av biodiversitetsdata.



Figur 4: Kategorisering av tjenestene oppsummert i tabell 9 (naturvitenskap), i "database", "søke- og metadatatportal" eller "analyseverktøy", samt informasjon om hvorvidt tjenestene gir tilgang til data (og metadata) eller kun metadata.

Tabell 9 og figur 4 viser at det innenfor naturvitenskap per i dag er ti av trettentjenester som gir tilgang til data. Av disse gir fem av ti tilgang til å laste ned data uten bruker-ID. De resterende gir tilgang gjennom registrering av brukernavn og passord, eller gjennom FEIDE/eudGAIN-autentisering. Flere av datainfrastrukturene tilbyr lagring, men tjenester som EBAS, ADC, NADIR og EVDC lagrer primærdatadata fra målestasjoner eller fra forskningsprogrammer og ikke fra individuelle forskningsprosjekter. Individuelle forskningsprosjekter kan lagre data hos følgende fire tjenester: EMEP, NELS, GBIF og Artsdatabanken. Enkelte av infrastrukturene som lagrer og samler inn datasett, gir tilgang kun til metadata for brukere utenfor egen institusjon.

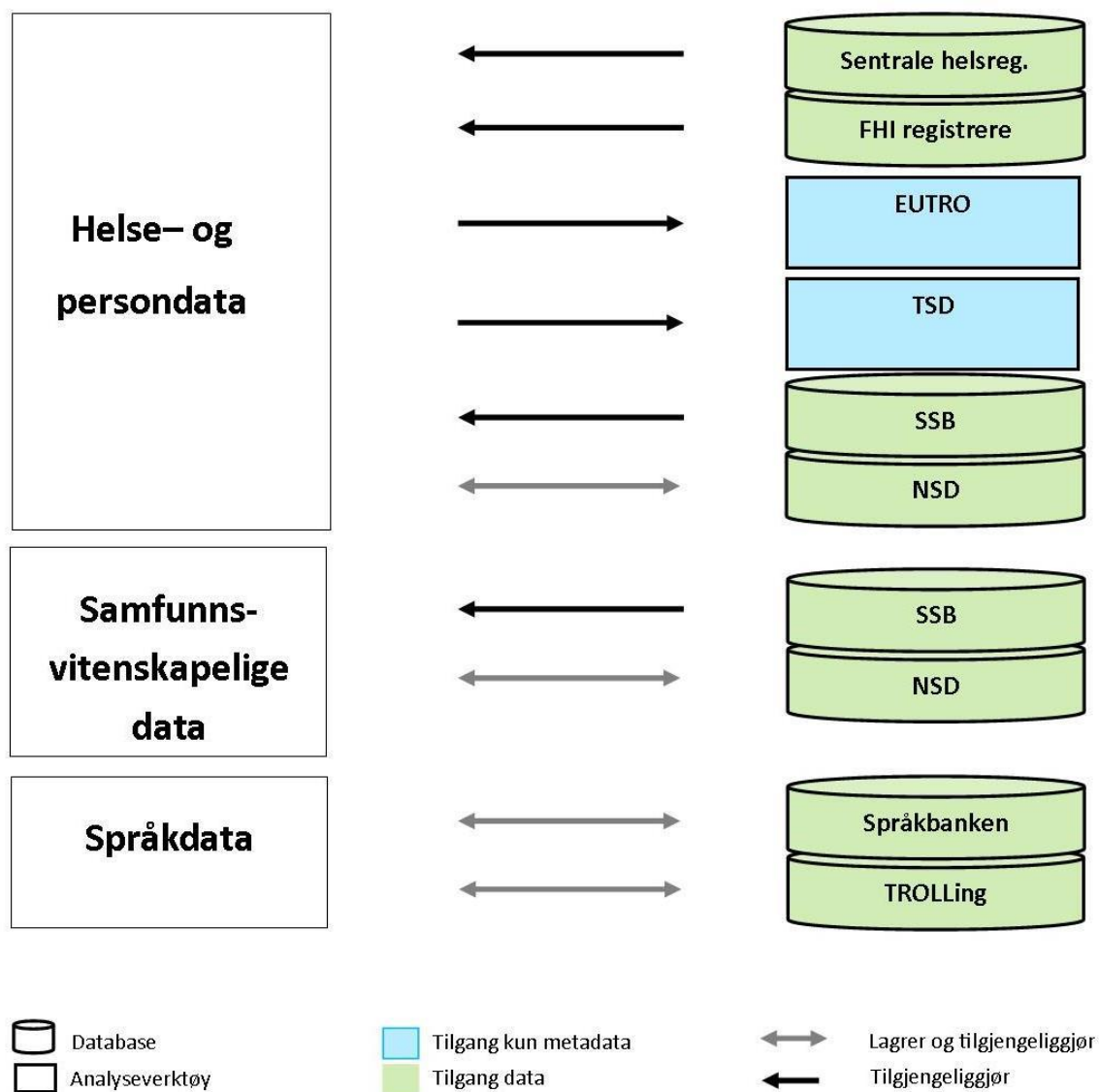
Tabell 10: Tjenester for datalagring og -håndtering relevante for helse- og persondata, samt humaniora og samfunnsvitenskap.

Akronym	Navn	Tjeneste	Institusjon og konsortier	Tilgang
Sentrale helseregistre ¹⁷	Se fotnote 17	Landsdekkende helseregistre. Helseovervåking, helsestatistikk og beredskap, kvalitetsforbedring av helsetjenester, forskning, administrasjon og styring	Se fotnote 17	Tilgang for godkjente brukere
FHI registre, helseundersøkelser og biobanker ¹⁸	Folkehelseinstituttets registre, helseundersøkelser og biobanker	Data fra helseregistre, helseundersøkelser og biobanker	Folkehelseinstituttet	Tilgang for godkjente brukere
EUTRO	EUTRO	Oversikt over prosjektinformasjon, prosjektdeltakere, prosjektdokumenter, budsjettoversikt, publikasjonsoversikt. Loggfører bruk av og endringer i datasett. Mellomlagring for aktive forskningsdata	UiT	Tilgang for godkjente brukere
TSD	Tjeneste for sensitive data	Brukere kan samle inn, lagre, dele og gjøre beregninger på sensitive data innenfor et lukket miljø	UiO/ USIT	Tilgang for godkjente brukere
SSB	Statistisk sentralbyrå	Utlånbare mikrodata	SSB	Tilgang for godkjente brukere (tidsbegrenset utlån)
NSD ¹⁹	Norsk senter for forskningsdata AS	Godkjent som arkiv for persondata, herunder også sensitive persondata	NSD	Tilgang med FEIDE/Dataporten
Språkbanken/ Nasjonalbiblioteket	Språkbanken	Arkiv for tilgjengeliggjøring (nedlasting) av digitale språkressurser med gjenbruksverdi for norsk språkteknologi.	Nasjonalbiblioteket	Tilgang uten bruker-ID (noen krever at du fysisk signerer en avtale før du får tilgang).
TROLLing	The Tromsø Repository of Language and Linguistics	Arkiv for tilgjengeliggjøring (nedlasting) av digitale lingvistiske data og statistisk kode	UiT	Tilgang med FEIDE/eduGAIN

¹⁷ Sentrale registre (inkludert lovbestemte registre) er opprettet etter vedtak i Stortinget: Dødsårsaksregisteret (DÅR), Forsvarets helseregister, Genetisk masseundersøkelse av nyfødte, Helsearkivregisteret, Hjerter- og karregisteret (HKR), Kreftregisteret, Medisinsk fødselsregister (MFR), Meldingssystem for smittsomme sykdommer (MSIS), Nasjonalt vaksinasjonsregister (SYSVAK), Norsk overvåkningssystem for antibiotikaresistens hos mikrober (NORM), Norsk overvåkningssystem for infeksjoner i sykehustjenesten (NOIS), Norsk pasientregister (NPR), Pseudonymt register for pleie- og omsorgssektoren (IPLoS), Register over svangerskapsavbrudd (Abortregisteret), Reseptbasert legemiddelregister (Reseptregisteret), Resistensovervåking av virus i Norge (RAVN). Databehandlingsansvarlige: Helsedir., FHI, Kreftregisteret, Forsvarsdep., Riksantikvaren og OUS.

¹⁸ For en oversikt, se FHIs side om [data fra helseregistre – store helseundersøkelser og biobanker](#)

¹⁹ NSD har vært godkjent som arkiv for persondata, herunder også sensitive persondata, siden Datatilsynet ble opprettet i 1980, og har avtaler med universitet- og høyskoler i Norge og alle de nasjonale forskningsinstituttene om å arkivere persondata på vegne av institusjonene. Tilsvarende har NSD avtale med SSB om å lagre og formidle data fra SSB til forskning. I 2014 inngikk Riksarkivet en formell avtale med NSD om langtidslagring og formidling av forskningsdata som faller inn under arkivloven.



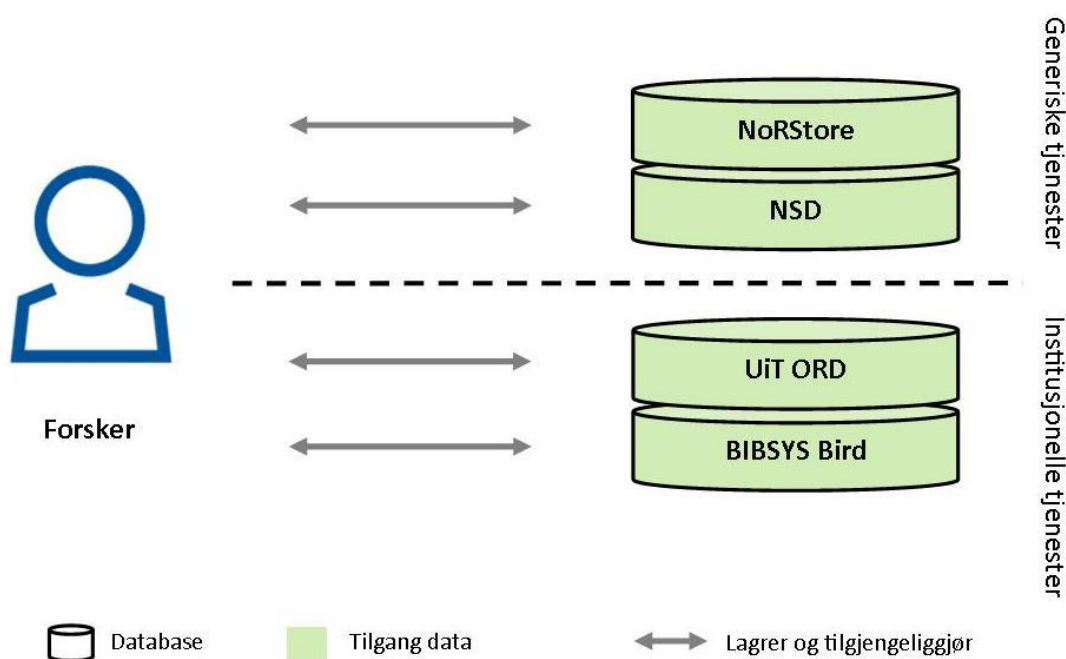
Figur 5: Kategorisering av tjenestene oppsummert i tabell 10 (helse- og persondata, samt humaniora og samfunnsvitenskap), i "database" eller "analyseverktøy", samt informasjon om hvorvidt tjenestene gir tilgang til data (og metadata) eller kun metadata.

Tabell 10 og figur 5 viser at de sentrale helseregistrene og FHI registrene er markert som grønne og indikerer tilgang til data, men datatilgang reguleres av ulike lovverk og tilgangsprosedyrer. EUTRO og TSD er løsninger som tilbyr sikker korttidslagring og analyse av sensitive data. SSB og NSD forvalter datainfrastrukturer som gir tilgang til registerdata for forskning. NSD lagrer persondata og sensitive persondata for universitet- og høyskoler i Norge og de nasjonale forskningsinstituttene.

Språkbanken og TROLLing gir tilgang til digitale språkressurser, og tilbyr lagring for relevante forskningsprosjekter.

Tabell 11: Generiske tjenester for datalagring og -håndtering.

Akronym	Navn	Tjeneste	Institusjon og konsortier	Tilgang
NorStore	NorStore Research Data Archive	Lagringsressurs for vitenskapelige data. Håndtere, arkivere og bevare vitenskapelige data. Programvare for å organisere og bevare data.	Uninett Sigma2	Tilgang med FEIDE/Dataporten med eduGAIN Åpen søke- og metadata-portal
NSD	Norsk senter for forskningsdata AS	NSD rådgir, håndterer, arkiverer og bevarer alle typer forskningsdata fra ulike fagfelt, både åpne og lukkede data (e.g. sensitive persondata).	NSD	FEIDE/Dataporten
UiT Open Research Data	UiT Open Research Data	Dataarkiv for studenter og ansatte ved UiT	UiT	Tilgang med FEIDE
BIRD	BIRD Research Data Repository	Dataarkiv for ansatte ved registrerte institusjoner	BIBSYS	Tilgang med FEIDE/eduGAIN



Figur 6: Informasjon om hvorvidt de generiske tjenestene oppsummert i tabell 11 gir tilgang til data (og metadata) eller kun metadata.

Tabell 11 og figur 6 viser at NorStore og NSD er de eneste tjenestene som kan defineres som generiske nasjonale dataarkiv. UiT Open Research Data er et institusjonelt arkiv, mens BIRD kan fungere som arkiv for institusjoner de har inngått avtale med.

Tabell 12: Tjenester for datalagring og -håndtering relevante for forvaltningsdata.

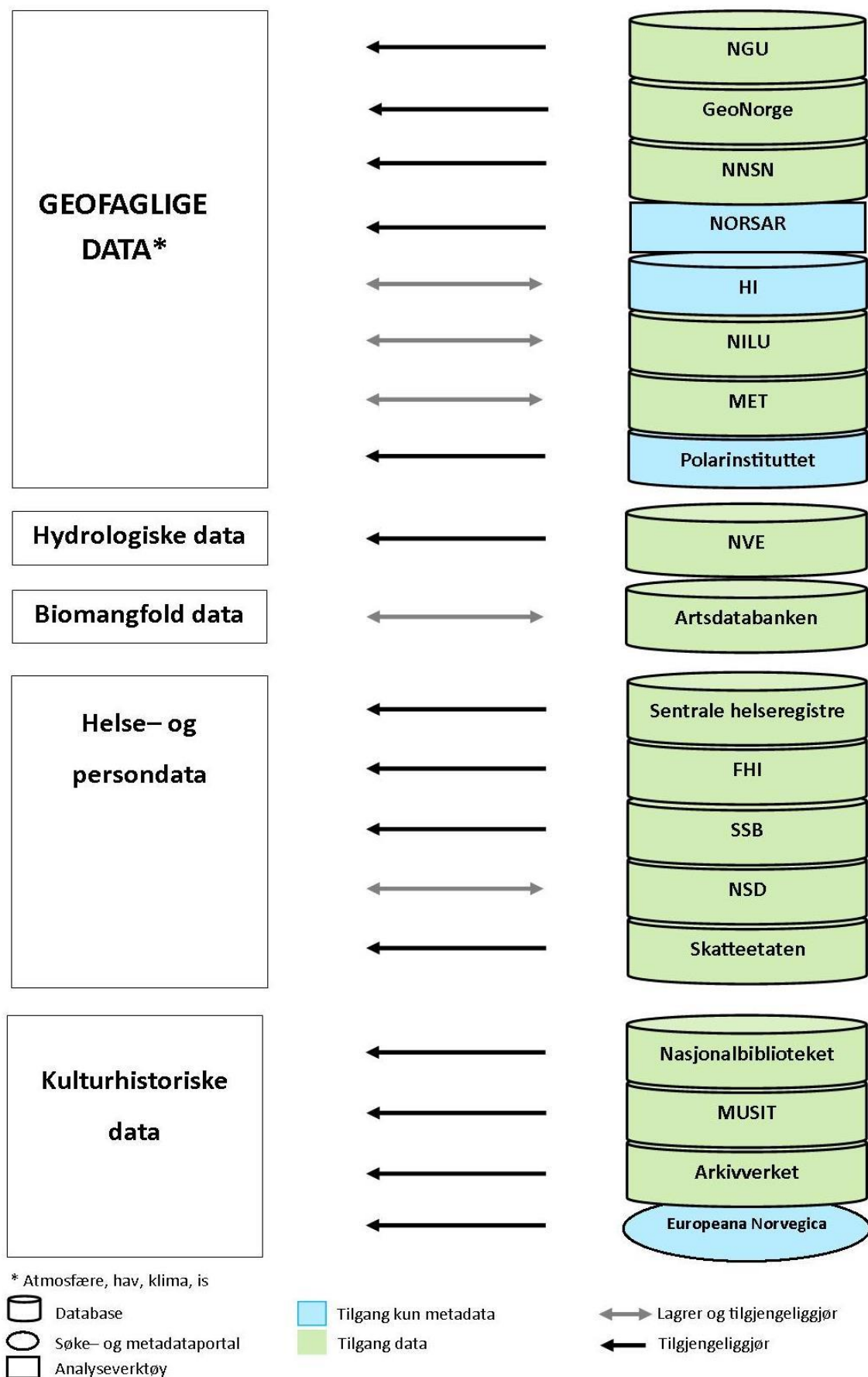
Akronym/forkortelse	Institusjon	Tjeneste	Tilgang
NGU	Norges geologiske undersøkelse	Tilbyr kart- og geodata	Tilgang uten bruker-ID
GeoNorge	Kartverket	Nasjonale nettstedet for kartdata og annen geografisk stedfestet informasjon	Tilgang uten bruker-ID for en del data/ BAAT innlogging på noen data
NNSN	Norsk nasjonalt seismisk nettverk/ Universitetet i Bergen og Oljeindustriens Landsforening	Seismiske antenner som registrerer jordskjelv. Data tilgjengeliggjøres gjennom NORSAR og internasjonale sentre	Tilgang uten bruker-ID
NORSAR	Stiftelsen NORSAR	Software-løsninger, visualisering og data	Tilgang uten bruker-ID
HI	Havforskningsinstituttet	Samler inn og håndterer data fra alle norske havområder. Metadata er tilgjengelig gjennom en datakatalog.	Nettside med innsynsløsninger for data
NILU	Norsk institutt for luftforskning	Samler inn og håndterer data knyttet til atmosfærens sammensetning, klimaendringer, luftkvalitet og miljøgifter. Tilbyr databaseaktiviteter i en rekke EU-finansierte prosjekter.	NILU forvalter ulike databaser og portaler med ulik tilgang til data (for eksempel NADIR, EVDC, ACTRIS Data Centre og EBAS)
MET	Meteorologisk institutt	Samler inn og håndterer meteorologiske data. Nedlastningstjenester med tilgang til vær- og klimadata og arktiske data.	Tilgang uten bruker-ID
Polarinstituttet	Polarinstituttet	Forskning, miljøovervåking og kartlegging av Arktis og Antarktis. Driver Norwegian Polar Data Centre som er en database med oversikt over kart, datasett, fartøy, stedsnavn, tidsserier, prosjekt, publikasjoner, tracking over Svalbard/Arktis.	Tilgang med bruker-ID/passord
NVE	Norges vassdrags- og energidirektorat	Kartdata, skreddadata, hydrologiske data	Tilgang uten bruker-ID
Artsdatabanken	Artsdatabanken	Dataportal, databaser med biodiversitetsdata.	Tilgang uten bruker-ID
Sentrale Helseregistre	Sentrale helseregistre ²⁰	Landsdekkende helseregistre. Helseovervåking, helsestatistikk og beredskap, kvalitetsforbedring av helsetjenester, forskning, administrasjon og styring	Tilgang for godkjente brukere
FHI	Folkehelseinstituttets registre, helseundersøkelser og biobanker	Data fra helseregistre, helseundersøkelser og biobanker	Tilgang for godkjente brukere
SSB	Statistisk sentralbyrå	SSB låner ut mikrodata til forskningsprosjekter, og har data knyttet til personer, boliger, kommuner, virksomheter og foretak med mer ²¹	Tilgang for godkjente brukere (tidsbegrenset utlån)

²⁰ Sentrale registre (inkludert lovbestemte registre) er opprettet etter vedtak i Stortinget: Dødsårsaksregisteret (DÅR), Forsvarets helseregister, Genetisk masseundersøkelse av nyfødte, Helsearkivregisteret, Hjerter- og karregisteret (HKR), Kreftregisteret, Medisinsk fødselsregister (MFR), Meldingssystem for smittsomme sykdommer (MSIS), Nasjonalt vaksinasjonsregister (SYSVAK), Norsk overvåkningssystem for antibiotikaresistens hos mikrober (NORM), Norsk overvåkningssystem for infeksjoner i sykehustjenesten (NOIS), Norsk pasientregister (NPR), Pseudonymt register for pleie- og omsorgssektoren (IPLOS), Register over svangerskapsavbrudd (Abortregisteret), Reseptbasert legemiddelregister (Reseptregisteret), Resistensovervåking av virus i Norge (RAVN). Databehandlingsansvarlige: Helsedir., FHI, Kreftregisteret, Forsvarsdep., Riksantikvaren og OUS.

²¹ For mer informasjon om mikrodata som SSB låner ut, se <http://www.ssb.no/#statistikk-hovedmeny>.

(Tabell 12 fortsetter)

NSD	Norsk senter for forskningsdata	Nasjonalt arkiv for forskningsdata, låner ut data fra SSB (registerdata, utvalgsundersøkelser, kommunedata), ulike databaser på individdata-, regionaldata- og institusjonsdatanivå, data fra forskningsprosjekter	Tilgang uten bruker-ID
Skatteetaten	Skatteetaten	Skatteetaten er ansvarlig for et fullstendig og oppdatert folkeregister som omfatter nøkkelopplysninger om alle personer som er eller har vært bosatt i Norge. Registeret danner grunnlaget for blant annet skattemanntallet, valgmanntallet og befolkningsstatistikken	Tilgang for godkjente brukere
Nasjonalbiblioteket	Nasjonalbiblioteket	Digitalt arkiv og sikkert oppbevaringssted for digitalt materiale både for Nasjonalbiblioteket og for andre kulturinstitusjoner	Fysisk tilstedeværelse hos Nasjonalbiblioteket for tilgang til data for forskning. Tilgang med ID/passord
MUSIT	Universitetsmuseenes IT-organisasjon	Drifter og tilgjengeliggjør data for naturhistorie via GBIF og Artsdatabanken. Kulturhistorisk fotobase.	Tilgang uten bruker-ID
Arkivverket	Arkivverket	Arkiverer og tilgjengeliggjør arkivmateriale fra statlige virksomheter. Arkivverkets ambisjon er å digitalisere ti prosent av samlingene innen 2030.	Fysisk tilgang til materialet i Arkivverkets lokaler. Tilgang uten bruker-ID til Digitalarkivet
Europeana Norvegica	Europeana Norvegica	Multimedie- fremstillinger av kulturhistoriske objekter	Åpen søke- og metadatatportal



Figur 7: Kategorisering av tjenestene for forvaltningsdata oppsummert i tabell 12 i "database", "søke- og metadatatportal" eller "analyseverktøy", samt informasjon om hvorvidt tjenestene gir tilgang til data (og metadata) eller kun metadata.

Det fremgår av tabell 12 og figur 7 at mange av datainfrastrukturene tilgjengeliggjør forvaltningsdata av interesse for forskning. Videre ser vi at 11 av 19 tilgjengeliggjør data uten at brukerne trenger bruker-ID/autorisering.

5.3 Oppsummering

Gjennomgangen i dette kapittelet viser at det finnes en stor variasjon av datainfrastrukturer som håndterer, lagrer og tilgjengeliggjør data. Den faktiske tilgangen til både forsknings- og forvaltningsdata varierer imidlertid. Flere av datainfrastrukturene forvalter personsensitive data hvor tilgangsbegrensning er i tråd med gjelde regelverk.

For de øvrige datainfrastrukturene indikerer kartleggingen at det finnes en del barrierer for å få tilgang til data. Dette gjelder både forsknings- og forvaltningsdata. Flere forskningsinstitusjoner er restriktive med å tilgjengeliggjøre egne data for gjenbruk. Andre har prosedyrer som innebærer at kostnadene for tilgjengeliggjøring belastes forskningsprosjekter og det er ofte uforutsigbart hva den endelige kostnaden blir. Videre kan komplekse og strenge utleveringsbetingelser for helse og persondata gjøre at saksbehandlingstiden for tilgangssøknader kan bli lang. Uforutsigbarhet knyttet til kostnader og søknadsbehandling kan dermed vanskeliggjøre gjenbruk av data.

Gjennomgangen viser også at det finnes relativt få tjenester hvor individuelle forskningsprosjekter kan lagre data.

6 Finansieringsmodeller for datainfrastrukturer

6.1 Kilder til finansiering

Research Data Alliance (RDA) og ICSU World Data System (WDS) har gjennomført en kartlegging av eksisterende og mulige fremtidige finansieringsmodeller ved 22 internasjonale arkiver for forskningsdata²². Arbeidet blir videreført av en ekspertgruppe oppnevnt av OECD Global Science Forum, RDA og ICSU Committee on Data for Science and Technology (CODATA)²³. Denne gruppen skal gjennomføre en større kartlegging og drøfting av finansieringsmodeller. Rapporten fra gruppen vil imidlertid ikke være ferdig før våren 2017. Vi har derfor valgt å basere kartleggingen av finansieringskilder på metodikken til RDA/WDS, med tilpassing til norske forhold og det som til nå er kjent om arbeidet til gruppen. Videre er observasjoner og vurderinger i rapporter fra German Council for Scientific Information Infrastructures²⁴, Royal Irish Academy²⁵ og Knowledge Exchange²⁶ lagt til grunn for diskusjonen nedenfor.

På basis av dette kan vi inndele finansieringsmodellene for datainfrastrukturer i syv ulike typer:

1. **Strukturell finansiering:** Langsiktig finansiering uten krav om direkte gjenytelser. Den vil normalt komme som støtte fra et departement eller fra Forskningsrådet. Innretning og omfang på støtten kan være gjenstand for regelmessige vurderinger, men den vil ikke bli avsluttet uten at dette er kjent god tid i forveien.
2. **Forsknings- og utviklingsprosjekt:** Konkrete og tidsavgrensede prosjekter og aktiviteter, der formålet normalt vil være å etablere eller videreutvikle datainfrastrukturen. Støtte fra

²² Dillo, I., Simon, H., and Waard, A. 2016. "Income Streams for Data Repositories, RDA-WDS Interest Group on Cost Recovery for Data Centres". [hentet 15.12.16]

²³ OECD GSF/RDA/CODATA Expert Group on Business Models for Data Repositories. [hentet 15.12.16]

²⁴ German Council for Scientific Information Infrastructures. 2016. "Enhancing Research Data Management: Performance through Diversity. Recommendations regarding structures, processes, and financing for research data management in Germany" Göttingen.

²⁵ Kitchin, R., Collins, S. and Frost, D. 2015. "Funding models for Open Access Repositories". [hentet 15.12.16]

²⁶ Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Data. 2016. "Funding research data management and related infrastructures". [hentet 7.11.2016]

INFRASTRUKTUR inngår her. Støtten kan komme direkte fra forskningsfinansiører, eller via en forskningsutførende organisasjon som er ansvarlig for FoU-prosjektet. Regnskapsmessig vil dette være bidragsprosjekter.

3. **Engangskostnad for deponering** (Data Processing Charge – **DPC**, også kalt Deposit fee). Vi velger forkortelsen DPC på grunn av analogien til avgift for publisering av artikler, APC – Article Processing Charge.
4. **Kostnad for tilgang** (Data Access Charge – **DAC**). Dette er en kostnad for tilgang til data eller for verdiøkende tjenester. Det er normalt ikke kostnader knyttet til selve dataene; det ville være i strid med prinsippet om åpen tilgang til offentlig finansierte data. Derimot kan det være betaling for kostnader ved å gi tilgang til data, for eksempel uttrekk av sensitive data. Det kan også være kostnader ved verdiøkende tjenester, som for eksempel generering av metadata, kobling av data, visualisering og analyse av data. Det er også arkiver i utlandet som har en vesentlig del av inntektene knyttet til lisensiering av egenutviklet programvare.
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. I medlemskapsmodellen går flere organisasjoner sammen om å dekke kostnader knyttet til deponering og/eller tilgang til data. 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.
7. **Oppdrag**. Dette er oppdrag knyttet til deponering, tilgang og/eller verdiøkende tjenester.

En finansieringsmodell vil normalt være sammensatt av flere av disse kildene.

6.2 Finansiering av norske datainfrastrukturer for forskningsdata

Fagspesifikke datainfrastrukturer med støtte fra INFRASTRUKTUR

Kartleggingen viser at datainfrastrukturer som har mottatt finansiering fra Forskningsrådet gjennom INFRASTRUKTUR til nå i hovedsak baserer seg på to finansieringskilder:

- a) Forsknings- og utviklingsprosjekt (INFRASTRUKTUR og i noen tilfeller støtte til ESFRI-prosjekt gjennom Horisont 2020) som dekker investeringer og i noen grad drift i etableringsfasen
- b) Støtte fra vertsinstitusjon

Bakgrunnen for dette er innretningen på INFRASTRUKTUR. Prosjekter som innvilges mottar finansiering fra Forskningsrådet i inntil fem år (etableringsfasen). Dette er ofte i kombinasjon med en andel egenfinansiering fra vertsinstitusjonene (selv om INFRASTRUKTUR ikke har prinsipielle krav om dette). De fleste datainfrastrukturene er fortsatt i denne fasen. Dette fordi et stort flertall av data-infrastrukturprosjektene fikk tildeling etter utlysningene i 2010, 2012 og 2014. De første prosjektene fikk tildeling etter 2009-utlysningen, med tidligste oppstart i 2011. Disse er nå i ferd med å gå over i driftsfasen.

Prosjekter med støtte fra INFRASTRUKTUR skal fremlegge planer for hvordan bærekraftig drift kan sikres når finansieringen fra Forskningsrådet opphører og prosjektet går over i driftsfasen. Brukerbetaling forventes å utgjøre en del av driftsfinansieringen. I alle søknader om forskningsprosjekter til Forskningsrådets øvrige finansieringsordninger kan utgifter til bruk av forskningsinfrastruktur dekkes. Finansieringsplaner Forskningsrådet har mottatt fra prosjekter tyder på at direkte finansiering fra vertsinstitusjoner og partnere vil utgjøre en større andel av driftsfinansieringen enn brukerbetaling. Dette kan dels skyldes at dataene som skal arkiveres i datainfrastrukturer med støtte fra INFRASTRUKTUR i stor grad har opphav hos vertsorganisasjon/partnerne. Videre kan det skyldes at brukerbetaling, spesielt for tilgang til data, vurderes å være i konflikt med etablert praksis innenfor fagfeltet. Noen av datainfrastrukturene har imidlertid inntekter i form av oppdrag (økonomisk aktivitet).

For enkelte forskningsinfrastrukturer er ikke støtte fra vertsinstitusjon og brukerbetaling fra prosjekter tilstrekkelig til å dekke driftsutgifter. Derfor har Forskningsrådet åpnet for at INFRASTRUKTUR i særskilte tilfeller kan gi langsiktig støtte til grunnfinansiering av drift. Ingen datainfrastrukturer har foreløpig søkt om slik støtte. Det er mulig at noen datainfrastrukturer ser dette som en mulighet ved neste utlysning av midler, og at det for noen oppleves som utfordrende å dekke kostnadene til drift uten slik støtte.

Generiske datainfrastrukturer

Fem av de kartlagte tjenestene kan betegnes som generiske:

- NSDs tjenester
- Uninett Sigma2s tjenester for datahåndtering
- Tjenester for Sensitive Data (TSD)
- UiT Open Research Data (UiT ORD)
- BIBSYS Bird

Disse tjenestene har annen sammensetning av finansieringskilder enn de fagspesifikke tjenestene omtalt ovenfor, se figur 8.

	Strukturell	FoU	Deponering	Tilgang	Vertsinstitusjon	Medlemskap	Oppdrag
NSD	D	I	d				d
Sigma2	D/i	I	[d]			D/i	[d]
TSD		I	d		D/i		d
UiT ORD					D/I		
BIBSYS Bird					i	D/I	

- D: Stor grad av denne typen støtte/inntekt for drift av datainfrastruktur
d: Liten grad av denne typen støtte/inntekt for drift av datainfrastruktur
I: Stor grad av denne typen støtte/inntekt for investering i datainfrastruktur
i: Liten grad av denne typen støtte/inntekt for investering i datainfrastruktur
[d]: Planlagt å dekke noe av driftskostnaden gjennom denne typen inntekt

Figur 8: Finansieringskilder for de generiske datainfrastrukturene

NSD og Uninett Sigma2 har strukturell finansiering fra Forskningsrådet. De har også betydelige bevilgninger fra INFRASTRUKTUR og noen inntekter fra internasjonale FoU-prosjekter, hovedsakelig Horisont 2020. For Uninett Sigma2 kommer en vesentlig andel av inntektene fra medlemsorganisasjonene (de fire universitetene UiO, UiB, NTNU og UiT Norges arktiske universitet). Også NSD har inntekter gjennom en medlemskaps-/abonnementsmodell, men disse dekker NSDs rolle som personvernombud for rundt 150 forsknings- og utdanningsinstitusjoner. NSD har til gjengjeld inntekter knyttet til håndtering av offentlige databaser og registre. For disse to organisasjonene har ikke støtte fra vertsinstitusjon noen mening. Dette fordi de er opprettet for å levere tjenester til UH-sektor og kun har inntekter knyttet til disse tjenesteleveransene (ingen direkte bevilgning fra staten).

Oppbygging av TSD er finansiert gjennom prosjekter (INFRASTRUKTUR og diverse FoU-prosjekter med støtte fra Forskningsrådet) og finansiering fra UiO. Driften vil finansieres gjennom brukerbetaling (markert som deponering selv om det ikke er dekkende for TSDs virksomhet), oppdrag og støtte fra UiO.

UiT ORD er en institusjonell datainfrastruktur hvor kostnadene dekkes av UiT Norges arktiske universitet.

BIBSYS Bird er tjeneste for forvaltning av forskningsdata utviklet gjennom et samarbeid mellom BIBSYS og BI. Infrastrukturen er et tilbud til utdannings- og forskningsinstitusjoner som ønsker å tilby sine forskere et felles verktøy for administrasjon av forskningsdata.

Datainfrastrukturer med forvaltning som hovedformål

Flere offentlige etater har ansvar for datainfrastrukturer/-samlinger som er opprettet for forvaltningsformål, men som i stor grad også benyttes til forskning. Disse institusjonene har strukturell finansiering (bevilgning fra staten) for å dekke sine formålsbestemte oppgaver, men dette inkluderer ikke nødvendigvis støtte til forskning som utføres utenfor etaten selv. Vi illustrerer i figur 9 hvordan disse etatene finansierer tilgang til data fra forskere utenfor etaten gjennom å se på Nasjonalbiblioteket (NB), Arkivverket, Meteorologisk institutt (MET), Folkehelseinstituttet (FHI) og Statistisk Sentralbyrå (SSB).

	Strukturell	FoU	Deponering	Tilgang	Vertsinstitusjon	Medlemskap	Oppdrag
NB	(D/I)	i			D/I		
Arkivverket	(D/I)				D/I		
MET	(D/I)	i			D/I		d
FHI	(D/I)	i		D			d
SSB	(D/I)	i		D			d

- D: Stor grad av denne typen støtte/inntekt for drift av datainfrastruktur
d: Liten grad av denne typen støtte/inntekt for drift av datainfrastruktur
I: Stor grad av denne typen støtte/inntekt for investering i datainfrastruktur
i: Liten grad av denne typen støtte/inntekt for investering i datainfrastruktur
(D/I): Stor grad av denne typen støtte/inntekt for investeringer i og drift av datainfrastrukturen, men forskning er ikke primærformålet med infrastrukturen

Figur 9: Finansieringskilder for datainfrastrukturer med forvaltning som hovedformål

Tre av disse institusjonene (NB, Arkivverket og MET) finansierer i stor grad forskernes tilgang til data gjennom egne midler. I noen grad er utvikling av selvbetjeningsløsninger finansiert gjennom FoU-prosjekter og eventuelt oppdrag. Det er forskjeller på hvor langt de tre institusjonene har kommet i å utvikle selvbetjeningsløsninger. MET har gjort alle sine offentlige data tilgjengelig via elektroniske løsninger (API og manuell visning/nedlasting). NB har digitalisert sitt analoge materiale gjennom å omdisponere eksisterende bevilgninger fra statsbudsjettet, og jobber nå for å øke den digitale pliktavleveringen. De deltar i flere infrastrukturprosjekter for tilgjengeliggjøring av data og metadata og drifter Språkbanken som er en nasjonal datainfrastruktur for språkteknologi og forskning. Arkivverket skal ta vare på arkivmateriale fra statelig virksomheter for ettertiden og gjøre dette tilgjengelig for bruk. En stor og viktig oppgave er nå å ivareta og tilgjengeliggjøre digitalt skapt materiale. Det er satt i gang et arbeid med å planlegge hvilke materiale som skal digitaliseres framover. I 2014 var kun 2 % av arkivdokumentene digitalisert.

FHI tar betalt for verdiøkende tjenester knyttet til administrasjon og utlevering av data fra registrene og fra Den norske mor og barn undersøkelsen. Prosjekter betaler en avgift for tilgang til data. Ved uthenting av biologisk materiale kommer et tillegg for biobankens arbeid. FHI har utarbeidet en kostnadsmodell basert på at den oppdragsbaserte virksomheten har krav om å være selvfinansiert. Alle oppdrag (også bidragsfinansiert aktivitet hos den forskningsutførende institusjonen) må dermed være med å dekke inn de generelle felleskostnadene.

SSB er også et eksempel på en institusjon som får grunnfinansiering over statsbudsjettet, og som tar betaling for verdiøkende tjenester, som uthenting og tilrettelegging av mikrodata for forskning. I dag er det kostnadsfritt å få tilgang til data, men SSB tar betalt for saksbehandling av søknader, samt tilrettelegging og overføring av data.

7 Avdekte behov

Vi har gjennomført individuelle samtaler (se vedlegg 10.2-3) med representanter fra ulike datainfrastrukturer for å få ett innblikk i datainfrastrukturenes praksis, tjenester og finansiering. I samtalene ba vi informantene bl.a. om å drøfte behov for endringer og mulige løsninger som kan forbedre dataflyten både nasjonalt og internasjonalt. Vi snakket også med informanter ved de fem største universitetene. Fokus da var institusjonenes policyer, rutiner for datahåndtering og gjenbruk av forskningsdata.

De viktigste behovene aktørene trakk frem er oppsummert nedenfor. Vi nevner for ordens skyld at disse utsagnene gjenspeiler informantenes synspunkter, og representerer derfor ikke nødvendigvis Forskningsrådets meninger eller holdninger.

Behov for datahåndteringsplan

- Det er behov for å innføre krav om at forskningsprosjekter som genererer data skal ha en datahåndteringsplan (Data Management Plan – DMP). IT-tekniske løsninger for å generere DMPer bør være generiske med noen like og faste informasjonsbiter, være av internasjonal standard og være tilpasset fagfeltene på detaljnivå.

Behov for krav om – og bedre tekniske muligheter for – lagring og tilgjengeliggjøring av forskningsdata, samt insentiver for å gjøre dette

- Finansiørene bør stille strengere krav til deling av forskningsdata ved kontraktinngåelse.
 - Det er behov for at forskningsinstitusjonene tilbyr løsninger for datalagring, enten som kort-, mellom- eller langtidslagring.
 - Det må utarbeides ulike insentiv-ordninger for tilgjengeliggjøring av datasett.

Behov for samordning og nasjonal løsning

- Det bør bli bedre samordning og samarbeid mellom relevante aktører nasjonalt og internasjonalt.
- Det er behov for felles nasjonale retningslinjer, føringer og løsninger for lagring og/eller tilgjengeliggjøring av data.
- Et forum hvor aktører fra relevante datainfrastrukturer, ulike institusjoner og fagmiljø kan samles regelmessig for å samkjøre ulike faglige og IT-tekniske løsninger kan være nyttig for å bedre mulighetene for gjenbruk av data.

Behov for IT-løsninger som snakker sammen

- Det er behov for gode selvbetjeningsløsninger for lagring, gjenfinning og gjenbruk av datasett.
- Det mangler gode tjenester hvor man kan få informasjon om hvilke datasett som finnes og hvor man kan få tak i disse.
- Forskningsinstitusjonene etterlyser pakkedløsninger for analyse og lagring av data.
- Det er behov for å utvikle nye, eller videreutvikling av eksisterende, IT-tekniske løsninger for å muliggjøre internasjonal kobling av datasett. Dette forutsetter at det finnes internasjonale metadata- og datastrukturer, og at datasett utstyres med entydige identifikatorer (DOI).

Behov for metadatastandarder

- Metadatastandardene må faglig forankres nasjonalt og internasjonalt i tillegg til at man må finne standarder som fungerer på tvers av fag.
- Det er behov for tilrettelegging og standardisering av data og metadata for at forvaltningsdata skal kunne gjenbrukes i forskning.

Behov for opplæring og kulturendring

- Det er behov for opplæring for forskere knyttet til de ulike tjenestene, og veiledning rundt rettigheter, forskrifter og lovverk.

- Forskerne bør bli tilbudt opplæring i hvordan man kan klargjøre dataene for lagring og tilgjengeliggjøring og hvordan man gjenbraker og siterer datasett.
- Det er behov for å utdanne såkalte "data stewards" eller datakuratorer med kompetanse innen et fagfelt med kompetanse til å videreutvikle datahåndteringstjenester og veilede forskere.

Behov for langsiktig og forutsigbar finansiering

- Datainfrastrukturer har behov for langsiktig og forutsigbar finansiering. Data som deponeres i arkiv skal ha en lang og forutsigbar levetid.

Behov for forutsigbar og kostnadseffektiv tilgang til data innsamlet for forvaltningsformål

- Betalingsmodeller for uthenting av data viser at det er lite forutsigbarhet knyttet til tidsbruk og kostnader ved søknad om tilgang til data.
- Det er behov for en finansieringsmodell som sikrer brukerne tilgang på like vilkår og til en lavest mulig kostnad.

Behov for forenkling av tilgang til helsedata i forskning

- Det er behov for gjennomgang av de ulike registres forskrifter og omforent tolkning av forskriftene for å lette tilgangen og gjenbruk av data. En effektivisering av tekniske løsninger og mer opplæring av forskerne bør også prioriteres. I tillegg etterspørres det en større forutsigbarhet i forhold til tidsbruk, kostnader og tilgang ved søknad om tilgang til register- eller forvaltningsdata.

Behov for å redde/kuraterer og tilgjengeliggjøre data som allerede er innsamlet

- Allerede innsamlede data som er lagringsverdige må kurateres for lagring og tilgjengeliggjøring. Dette er et svært tidkrevende og kostbart arbeid, og det er behov for en nasjonal dugnad for å få gjort disse dataene gjenfinn- og gjenbrukbare.

8 Drøftinger

8.1 Policyer/retningslinjer for datahåndtering

Kostnader knyttet til datainnsamlingen har de siste årene blitt kraftig redusert, mens kostnaden for datahåndtering har økt. Det vil trolig ikke være kostnadseffektivt å ivareta alle forskningsdataene som produseres. Dataeierne må beslutte hvilke data som skal kurateres for lagring og tilgjengeliggjøring. Policyer/retningslinjer på institusjonsnivå er viktige verktøy for å prioritere og å sette rammer for datahåndtering og for å etablere tydelig praksis for egen institusjon.

Internasjonalt opplever forskerne at forskningsfinansiører i økende grad krever datahåndteringsplaner hvor det reflekteres i hvilken grad dataene kan tilgjengeliggjøres og hvordan. Forskningsrådets policy for tilgjengeliggjøring av forskningsdata er formulert som en rekke anbefalinger fremfor absolutte krav. Det er i stor grad opp til forskningsprosjektene selv om de tilgjengeliggjør sine datasett. Gjennom samtaler har vi fått informasjon om at flere sentrale fagmiljøer er restriktive med tilgjengeliggjøring av forskningsdata utenfor egen institusjon.

Forskningsrådet har gjennom sine finansieringsmekanismer virkemidler for å sikre at prosjekter som får støtte i økende grad tilgjengeliggjør forskningsdata. Gjennom å stille krav om at finansierte prosjekter skal fremlegge en DMP ved kontraktinngåelse, kan Forskningsrådet bidra til en tidlig bevisstgjøring rundt tematikken og til økt tilgjengeliggjøring og gjenbruk.

For å få en omforent linje for dette kan man tenke seg at også andre FoU-finansiører kan innføre lignende krav. For å sikre nødvendig institusjonell forankring er det viktig at dataeierne selv godkjenner en slik DMP og påser at institusjonelle policyer for datalagring og tilgjengeliggjøring følges. For at det skal oppleves enklest mulig for forskeren bør det tilbys DMP-verktøy som er i tråd med hva som kreves av forskningsfinansiører internasjonalt. Et slik DMP-verktøy bør være generisk med mulighet for å tilpasses ulike fag.

8.2 Samordning og rolleavklaring

Flere store nasjonale datainfrastrukturer leverer i dag fellestjenester innen IKT til forskningsinstitusjonene. I tillegg illustrerer kartleggingen ovenfor at det finnes mange fagspesifikke datainfrastrukturer og at det er en del overlapp i tjenestene. Gjennom våre samtaler med aktørene fremkom det at dagens tjenestetilbud oppleves som uoversiktlig og at det er behov for tydeligere rolleavklaring mellom de nasjonale tjenestetilbydere. Det er viktig med god samordning mellom aktørene for å utvikle det nasjonale tjenestetilbudet og diskutere rolleavklaring. På dette grunnlaget mener vi det er behov for etablering av et forskningsdataforum på nasjonalt nivå hvor aktører fra relevante datainfrastrukturer, institusjoner og fagmiljø kan samkjøre både de ulike faglige og IT-tekniske løsningene. Et slikt forum kan fungere som et rådgivende organ for myndigheter og FoU-institusjoner og delta i internasjonale prosesser.

Både forskningsinstitusjoner og universiteter etterlyste gode og enkle lagrings- og tilgjengeliggjøringsløsninger som kan tilbys deres forskere. Løsningene for god dataflyt (kobling av datasett mot forskningsartikler, prosjekt, forsker, institusjon og arkiv) finnes allerede, men må tilrettelegges og samordnes for brukerne. I den grad det er mulig bør internasjonale data- og IT-tekniske standarder benyttes. Videre bør det unngås at det brukes offentlige midler på å utvikle konkurrerende og overlappende tjenester. Eksempelvis kan det være hensiktsmessig å utvikle én felles nasjonal tjeneste for sikker analyse og korttidslagring av sensitive forskningsdata. Løsningene som utvikles må fungere for forskerne og basere seg på deres behov. Det er viktig at forskere innenfor beslektede fagområder samordner seg og blir enige om felles datastandarder.

8.3 Opplæring og kulturendring

Økte krav til lagring, tilgjengeliggjøring og gjenbruk av forskningsdata vil føre til større behov for kunnskap om datahåndtering på flere nivå. Avhengig av hvilke type forskningsdata som skal håndteres/benyttes må forskerne forholde seg til ulike utfordringer av juridisk, teknisk og etisk art. EU-kommisjonen har påpekt at det er et stort behov for å utdanne såkalte datarøkttere (data scientists) som har faglig og it-teknisk forståelse²⁷. Riktig og målrettet opplæring er nødvendig for at forskere og studenter skal kunne tilgjengeliggjøre og gjenbruke forskningsdata i større grad. For å bidra til en kulturendring vil det være viktig å inkludere kurs i datahåndtering på master- og PhD-nivå.

Økt kompetanse blant forskere og studenter vil på sikt redusere tidsbruk og kostnader knyttet til lagring/kuratering og tilgang til datasett. Vertsorganisasjoner for datainfrastrukturer bør være aktive i å søke INFRASTRUKTUR og andre finansieringsordninger om utvikling av selvbetjeningsløsninger. Slike løsninger vil både forenkle forskernes tilgang til- og deponering av data/metadatas og redusere kostnadene knyttet til drift av datainfrastrukturer.

8.4 Insentiver

Sett fra forskerens synspunkt er det i dag få insentiver knyttet til å dele data, mens arbeidet med å gjøre data tilgjengelig for andre forskere kan være omfattende og gå utover egen vitenskapelig produksjon. Nordisk institutt for studier av innovasjon, forskning og utdanning (NIFU) leverte i desember 2016 rapporten "Vekt på forskningskvalitet"²⁸, som nå er på høring med høringsfrist 1.3.2017. Rapporten anbefaler å innføre en siteringskomponent av artikler i publiseringsindikatoren. Rapporten trekker frem at åpen publisering ofte fører til hyppigere sitering. Dersom man inkluderer sitering av datasett i publiseringsindikatoren kan dette bidra økt fokus på tilgjengeliggjøring av data i forskningsmiljøene.

²⁷ European Commission 2016a. "Realising the European Open Science Cloud". [hentet 18.11.16].

²⁸ Sivertsen, G. 2016. "Vekt på forskning. En mulig utvidelse av publiseringsindikatoren med en siteringsindikator". [hentet 15.12.16]

8.5 Finansiering

Datainfrastrukturer må 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 (mer enn 10 år). Dette krever at det finnes et element av langsiktighet i finansieringsmodellen, et element som i mange tilfeller ikke er der i dag.

Et prosjekt er per definisjon rettet mot å løse konkrete oppgaver innenfor en gitt tidsramme.

Prosjektfinansiering, for eksempel gjennom INFRASTRUKTUR, er derfor velegnet til etablering og senere investeringer/oppgraderinger, men ikke til langsiktig finansiering av drift av datainfrastrukturer.

Finansieringsmodeller for drift kan inneholde en kombinasjon av de øvrige seks finansieringskildene.

Noen egenskaper ved disse kildene er:

- **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 og det er få insentiver for forbedringer og innovasjon. Stor andel strukturell finansiering bør derfor følges av jevnlig evalueringer (som også kan resultere i at finanseringen 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 Open Access tankegang hvor det er normalt at man må betale for publisering i åpne journaler. En utfordring vil være å avklare hva som er en riktig kostnad (den faktiske kostnaden for arkivering og kuratering over en lang tidsperiode vil være ukjent) og at datainfrastrukturer som tar betalt for arkivering kan tape i konkurransen med datainfrastrukturer som ikke benytter denne mekanismen. I samtaler med aktørene for datainfrastrukturene kom det frem at betaling for deponering kan være egnet for prosjekter med store behov for lagringskapasitet, mens arkivering av data som skal tilgjengeliggjøres bør i størst mulig grad være gratis for prosjektene.
- **Betaling for tilgang** til data er i tråd med prinsippet om at brukeren betaler for de tjenestene han ønsker. Dersom det tas betalt for selve dataene vil slik betaling være i strid med Open Access prinsippet og prinsippet om åpen tilgang til offentlig finansierte data. En kostnad knyttet til uttrekket av data er mulig uten å bryte disse prinsippene. Det er også mulig å ta betalt for verdiøkende tjenester, som generering av metadata, kobling av data, visualisering og analyse av data. Noen aktører er bekymret for at brukerbetaling for tilgang til data vil redusere den faktiske bruken, og det blir påpekt at det er institusjonene, ikke enkeltprosjektene, som bør betale. Det blir også fremhevet at betaling for tilgang til forskningsdata kan være en hemsko for innovasjon og kan gjøre internasjonalt samarbeid vanskelig. Det kan føre til at data generert i Norge blir lagret ved internasjonale datasentre som ikke har betaling for tilgang. Det er derimot en viss aksept for at verdiøkende tjenester som tilrettelegging av data kan dekkes av prosjektene.
- **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 de 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 kostander 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 internt 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. Statsstøtteregelverket setter dessuten grenser for hvor stor økonomisk aktivitet en datainfrastruktur kan ha.

For både *fagspesifikke og generiske datainfrastrukturer for forskningsdata*, bør medlemskapsmodellen vurderes som et element i finansieringsmodellen for drift. Modellen legger til rette for god forankring i ledelse og brukergrupper hos de sentrale interessentene, langsiktig finansiering og gode muligheter for reduserte kostnader gjennom samarbeid. Andre elementer som bør vurderes er bidrag fra vertsinstitusjon (mest aktuelt der infrastrukturen er geografisk lokalisert til et sted) og inntekter fra verdikjende tjenester. Det siste kan være tjenester både til akademiske og kommersielle brukere. Brukerbetaling for deponering/lagringskapasitet kan være et element for forskningsprosjekter med særskilt store behov, men bør brukes med forsiktighet når hovedformålet med deponeringen er å gjøre data tilgjengelig for andre. Betaling for tilgang bør unngås med mindre dette er etablert praksis innen det aktuelle fagfeltet, jamfør diskusjonen over.

I helt spesielle tilfeller kan det søkes om støtte til langsiktig grunnfinansiering av drift gjennom INFRASTRUKTUR. For å være kvalifisert til å motta slik støtte, må en rekke kriterier knyttet til infrastrukturen og finanseringen av denne være oppfylt²⁹. Kriteriene er generelle og selv om flere av kriteriene er relevante også for datainfrastrukturer, fanger de neppe i tilstrekkelig grad opp en del utfordringer som spesifikt er knyttet til drift av nasjonale datainfrastrukturer. Forskningsrådet bør derfor vurdere å lage et kriteriesett som også fanger opp slike spesifikke utfordringer. Elementer i et slikt kriteriesett (i tillegg til de som allerede er der) kan være at datainfrastrukturene skal følge FAIR-prinsippene (Findable, Accessible, Interoperable, Reusable)³⁰ og at datainfrastrukturene skal (eller bør) være basert på samarbeid/medlemskapsmodell innenfor sektoren/fagfeltet.

Støtte til drift fra INFRASTRUKTUR kan nå gis for en periode på inntil fem år, og videre støtte kan gis etter ny utlysning. Det bør vurderes å gi driftsstøtte til datainfrastrukturer på samme vilkår som angitt i medlemskapsmodellen. Dette vil åpne for å gi støtte gjennom en såkalt "5+5-ordning". Her tildeles støtten for en periode på 10 år. Kontrakten forlenges så for nye 10 år allerede etter fem år forutsatt at tjenesten da får en god evaluering. Dersom evalueringen er mindre god, foreslås korrektive tiltak, eventuelt avvikling, i løpet av de kommende fem årene. Denne ordningen er brukt i finansieringen av Uninett Sigma2, og både universitetene (medlemmene) og Forskningsrådet har forpliktet seg til denne.

For *datainfrastrukturer med forvaltning som hovedformål* bør det etableres finansieringsmodeller som gir vertsorganisasjonen dekning for den delen av driftskostnaden som er knyttet til tilrettelegging av data for forskningen – i den grad slike kostnader ikke er dekket gjennom statens generelle bevilgning til vertsorganisasjonen. Et element i disse modellene kan være brukerbetaling for utlevering av data. Fordi slik betaling kan redusere bruken og bidra til at tilgang ikke gis på like vilkår, bør mulighetene for bidrag gjennom strukturell finansiering vurderes for viktige nasjonale datatjenester. Slik finansiering kan for eksempel dekke kostnader til å gi forskerne innsyn i hvilke data som finnes og avklare om forskningsprosjektet kan få tilgang til nødvendige data. Strukturell finansiering kan også benyttes til delvis dekning av kostnader til utlevering av data. Medlemskapsmodell er ikke så godt egnet for disse datainfrastrukturene, siden det vil forsterke ulikhetene i tilgang til forvaltningsdata. Forskere i medlemsorganisasjoner vil ha fri/rimeligere tilgang enn ansatte i andre organisasjoner.

²⁹ Forskningsrådet. 2016. "Hva det kan søkes om. Vedlegg til utlysning i INFRASTRUKTUR med søknadsfrist 12. oktober 2016". [hentet 7. 12.2016].

³⁰ Wilkinson, M. D. et al. 2016. [The FAIR Guiding Principles for scientific data management and stewardship](#). [hentet 15.12.16]

9 Forskningsrådets innspill

I dette kapittelet foreslår vi ulike innspill til hvordan det kan legges bedre til rette for økt tilgjengeliggjøring og gjenbruk av data for forskning med utspring i de avdekte behovene beskrevet i kapittel 6. Innspillene er knyttet til nasjonal samordning, klarere føringer for tilgjengeliggjøring, opplæring og kulturendring, insentiver, og finansieringsmodeller.

1. Forskningsinstitusjonene bør i samarbeid med egne forskere lage retningslinjer for hvilke data som skal lagres, kurteres og tilgjengeliggjøres ved egen institusjon.
 - Forskningsinstitusjonene bør tydeliggjøre hvilke løsninger de anbefaler at forskerne benytter for håndtering av ulike datasett.
2. Innføre krav om datahåndteringsplaner for forskningsprosjekter
 - Forskningsrådet og andre som bidrar med offentlige bidrag til FoU bør vurdere å kreve en datahåndteringsplan for prosjekter som genererer forskningsdata.³¹
 - Forskningsinstitusjonene bør ha mekanismer for å godkjenne DMP for sine forskningsprosjekter.
3. Kunnskapsdepartementet bør vurdere å etablere et "forskningsdataforum" for samordning og videreutvikling av nasjonale løsninger for lagring, analyse og tilgjengeliggjøring av forskningsdata.
 - Forskningsdataforumet kan bestå av representanter fra sentrale forskningsinstitusjoner og NSD, Uninett Sigma2, BIBSYS og CRISTin.
 - Forskningsdataforumet bør tilstrebe samarbeid mellom datainfrastrukturer og brukere på tvers av fag.
 - Forumet bør gi råd om videreutvikling av ulike IT-tekniske løsninger, og adressere behov for harmonisering av tjenester.
 - Forumet kan fungere som et rådgivende organ for nasjonale myndigheter i spørsmål knyttet til datahåndtering, og bidra i relevante internasjonale fora.
4. Etablere fagspesifikke fora innenfor beslektede fag for samordning og videreutvikling av datastandarder
 - Forskningsinstitusjonene bør vurdere å etablere felles kommunikasjonsarenaer hvor de kan diskutere fagspesifikke standarder som muliggjør deling og gjenbruk av data.
 - Disse faglige foraene bør ha tett samarbeid med forskningsdataforumet som jobber med de nasjonale løsningene.
5. En nasjonal tjeneste for sikker håndtering av sensitive forskningsdata
 - Forskningsinstitusjonene bør vurdere å utvikle én felles, nasjonal tjeneste for sikker analyse og korttidslagring av sensitive forskningsdata. Denne kan finansieres gjennom en medlemskapsmodell.
6. Tilrettelegge for opplæring
 - Institusjonene bør vurdere å legge bedre til rette for opplæring i datahåndtering og gjenbruk av data.
 - Nasjonale arkiv og databehandlere bør vurdere behovet for å tilby kurs og opplæring av forskere og studenter i bruk av egne tjenester.
 - Utdanningsinstitusjoner bør utrede behovet for å innføre nye kurs- og utdanningstilbud for å utdanne "datarøktere" med teknisk og faglig veiledningskompetanse.
7. Kunnskapsdepartementet bør utrede insentiver for tilgjengeliggjøring av forskningsdata gjennom for eksempel siteringsindeks.
8. Etablering av langsiktige finansieringsmodeller for drift av datainfrastrukturer for forskningsdata

³¹ For Forskningsrådet betyr dette å ta signalene i policyen "[Tilgjengeliggjøring av forskningsdata. Policy for Norges forskningsråd](#)" et skritt videre. Hvordan et slikt krav bør implementeres kommer vi tilbake til etter nærmere vurdering i Forskningsrådet.

- Institusjoner med ansvar for datainfrastrukturer bør etablere langsiktige driftsfinansieringsmodeller. Finansiering gjennom medlemskap bør være en av finansieringskildene i en slik modell.
 - Forskningsrådet bør vurdere å utarbeide kriterier for hvordan Nasjonal satsing på forskningsinfrastruktur i spesielle tilfeller bør kunne bidra til grunnfinansiering av drift av datainfrastrukturer.
9. Utrede strukturell finansiering for å sikre forskerne innsyn i og tilgang til forvaltningsdata på like vilkår
- Kunnskapsdepartementet, i samarbeid med relevante fagdepartementer, bør vurdere å utrede mulighetene for en strukturell finansiering som kan dekke kostnader til å gi forskerne innsyn i hvilke forvaltningsdata som finnes og avklare om forskningsprosjektet kan få tilgang til nødvendige data.
 - Utredningen bør også vurdere delvis dekning av kostnader til utlevering av data, for i større grad enn i dag å gi tilgang til dataene på like vilkår.
10. Etablere en arbeidsgruppe med oppgave å foreslå retningslinjer for åpen tilgang til forskningsdata
- KD bør vurdere å opprette en arbeidsgruppe med representanter fra sentrale forskningsinstitusjoner som kan utfylle eksisterende kunnskapsgrunnlag. Dette kan blant annet omfatte kartlegging av forskernes bruk av og behov for datahåndteringstjenester, insentiver for å tilgjengeliggjøre datasett og relevante internasjonale prosesser.

10 Vedlegg

10.1 Definisjoner

Begrep	Forklaring
API	Application Programming Interface (API) refererer til et grensesnitt i en programvare slik at spesifikke deler av denne kan aktiveres fra en annen programvare.
Datainfrastruktur	Forskningsinfrastrukturer som har som viktig formål å innsamle, bearbeide, lagre, kuratere, tilgjengeliggjøre og/eller forbedre utnyttelse av data.
DMP	Data Management Plan (datahåndteringsplan) skal vise hvordan dataene generert fra prosjektet skal kurateres.
ESFRI	European Strategy Forum for Research Infrastructures
Forskningsinfrastruktur	Anlegg, ressurser og tilknyttede tjenester som forskningsmiljøer bruker for å drive forskning på de respektive områdene, herunder vitenskapelig utstyr eller instrumentsett, kunnskapsbaserte ressurser, for eksempel samlinger, arkiver eller strukturerte vitenskapelige opplysninger, IKT-baserte infrastrukturer, for eksempel nett, databehandling, programvare og kommunikasjon, eller andre enheter av en særlig art som er nødvendig for å gjennomføre forskning. Slike infrastrukturer kan være samlet på ett sted eller spredt (et organisert nettverk av ressurser) i samsvar med artikkel 2 bokstav a) i rådsforordning (EF) nr. 723/2009 av 25. juni 2009 om Fellesskapets rettslige ramme for et konsortium for en europeisk forskningsinfrastruktur (ERIC-konsortium).
Forskningsdata	Registeringer/nedtegnelser/rapporteringer i form av tall, tekster, bilder og lyder som genereres eller oppstår underveis i forskningsprosjekter.
Forvaltningsdata	Data som er samlet inn eller generert med primærformål å benyttes i offentlig forvaltning.
Kildedata/ data for forskning	Data som allerede finnes, uavhengig av forskningen som skal gjennomføres. Dette omfatter forskningsdata, forvaltningsdata og andre typer data av relevans for forskning.
Kuratering	Prosesser og aktiviteter knyttet til organisering og integrering av data samlet fra ulike kilder og bevaring av datasett slik at dataenes verdi er vedlikeholdt over tid for å sikre at dataene forblir tilgjengelige for gjenbruk og lagring.
Metadata	Data om data, informasjon som beskriver annen informasjon, gjerne en elektronisk fil som et tekstdokument, bilde eller film.
Søke- og metadatatjeneste	Tjeneste som tilgjengeliggjør metadatabeskrivelser og informasjon om hvor datasett ligger lagret.
Tilgjengeliggjøring	Å tilgjengeliggjøre data for viderebruk handler i mange tilfeller om mer enn å publisere informasjon slik at det er mulig å bla i data på en nettside. Viderebruk handler også om at rådata gjøres tilgjengelig i det som kalles «maskinlesbare formater», slik at datamaskiner kan brukes til å tolke og analysere datamaterialet. Rådata er data som kan prosesseres maskinelt, tas fra hverandre, blandes med andre data og brukes i nye sammenhenger.

10.2 Oversikt over institusjoner og informanter vi har konsultert

Datainfrastruktur	Informanter
Artsdatabanken	Nils Valland
European Plate Observing System – Norway (EPOS)	Kuvvet Atakan
Integrated Carbon Observation System Data Portal (ICOS)	Truls Johannessen, Benjamin Pfeil og Are Olsen
Norsk institutt for luftforskning (NILU)	Cathrine Lund Myhre, Kjetil Tørseth og Markus Fiebig
Norwegian Scientific Data Network (NorDataNet)	Øystein Godøy
Uninett Sigma2	Andreas Jaunsen
The Tromsø Repository of Language and Linguistics (TROLLing)	Helene Andreassen og Phillip Conzett
Tjeneste for sensitive data (TSD)	Gard Thomassen
Norwegian e-Infrastructure for Life Science (Elixir.no)	Kjell Petersen
Registreringsentral for historiske data (HISTREG)	Gunnar Thorvaldsen
Language Infrastructure made Accessible (LIA)	Kristin Hagen
Nasjonalbiblioteket	Thomas Langvann
Språkbanken/Nasjonalbiblioteket	Hege Stensrud Høsøien
Norsk senter for forskningsdata (NSD)	Vigdis Kvalheim
Statistisk sentralbyrå (SSB)	Elisabetta Vassenden, Svein Johannessen og Thorleiv Valen
Common Language Resources and Technology Infrastructure (CLARINO)	Koenraad De Smedt
Folkehelseinstituttet (FHI)	Gun Peggy Knudsen
The Global Biodiversity Information Facility Norge (GBIF Norge)	Dag Endresen
Infrastructure for the Exploration of Syntax and Semantics (INESS)	Victoria Rosén
Norsk faglitterær forfatterforening	Tore Slåtta og Olav Torvund
BIRD Research data Repository – BIBSYS (BIRD)	Jan Erik Garshol
Universitetet i Oslo (UiO)	Lex Nederbragt, Gard Thomassen, Margaret Fotland, Live Kvale og Elin Stangeland
Universitetet i Tromsø (UiT)	Helene Andreassen
Norges miljø og biovitenskapelige universitet (NMBU)	Solveig Fossum-Raunehaug
Universitetet i Bergen (UiB)	Kuvvet Atakan
Norges teknisk-naturvitenskapelige universitet (NTNU)	Rune Brandshaug, Henrik Karlstrøm, Thor Bjørn Arlov og Knut Noralf Husby

10.3 Intervjuguide

Hovedtema	Undertema
Hvilke lagringstjenester er de viktigste for ditt fagområde?	Hvordan fungerer de lagringsløsningene som finnes i dag? Er det behov for nye/andre lagre?
Hvilke standarder på data og metadata benyttes i dag?	Hvordan fungerer dagens tjenester med tanke på gjenfinning, interoperabilitet mellom datasett og kvalitet på datasett og metadata? Hvordan er kvaliteten på dataene din tjeneste håndterer?
Hvordan finansieres tjenesten i dag?	Er finansieringsmodellen bærekraftig? Hvilke svakheter har modellen? Hvordan burde en ideell modell ha vært? Skal de som leverer data til lagring betale for å gjøre dette? Skal de som får tilgang til data betale for dette?
Hvem skal finansiere opplæring?	Hvem bør ha ansvar for kurs og opplæring?
Dersom vi skal få til en slik god nasjonal løsning for lagring, tilgjengeliggjøring, bruk og sitering av data;	Hvordan skal en slik struktur se ut? (med tanke på rolleavklaring og finansiering) Det trengs kanskje noen endringer for å få til dette, hva tenker dere at deres rolle er i dette? Hvordan kan dere videreutvikle deres tjeneste for å komme nærmere et slikt ideal? Hvordan får man til felles standarder for data og metadata?
Ansvar og rollefordeling	Innenfor ditt fagfelt, hvilke lagre trenger vi og hvem har ansvaret for dette? Hvem burde ha ansvaret for å sikre interoperabilitet? Hvem setter standarden for web-grensesnittet/API? Hvem skal utvikle "discovery-tjenester" som muliggjør datakobling på tvers av fag? Hvem har ansvaret for at forskere/studenter/instituttet får opplæring? Hvem har ansvaret for at forskningsdataene publiseres/lagres/gjøres tilgjengelig? (Er det forskerne? Prosjektleder? Instituttleder? Fakultet?) Hvem har ansvaret for at en DMP følges? Hvem har ansvaret for at metadataene følger "korrekt" standard?

10.4 Invitasjonsbrev, utsendelseslister og program for to workshops i 2016

Invitasjon til workshop om lagring og tilgjengeliggjøring av forskningsdata, 7. juni 2016

Et viktig mål i norsk forskningspolitikk er større grad av åpen tilgang til offentlig finansierte forskningsdata. Dette vil kreve utvikling av standarder og nye løsninger. I dag finnes det en rekke tjenester for arkivering, kuratering og tilgjengeliggjøring av forskningsdata i Norge. Mens noen av disse tjenestene er generiske, er andre spisset mot spesifikke fagområder.

Forskningsrådet ønsker å samle tjenesteleverandører for lagring og tilgjengeliggjøring av forskningsdata i Norge til dialog om det videre arbeidet med tilgjengeliggjøring av forskningsdata. Et første skritt i dette er å invitere til en workshop den 7. juni, se vedlagt program. Formålet med workshopen er både å bli kjent med hverandres tjenester og å diskutere hvordan tjenestetilbudet kan videreutvikles på tvers av ulike fagfelt.

I tillegg til inviterte innledere vil workshopen inneholde gruppearbeid med diskusjoner rundt følgende to hovedtema:

1. Hvordan videreutvikle dagens tjenester for datahåndtering?
 - Hvordan videreutvikle nåværende tjenester?
 - Hva bør tilbys som basistjenester og hva bør tilbys som fagspesifikke tjenester?
 - Hvordan bør driften av tjenestene finansieres?
2. Hvordan stimulere til økt bruk av tjenestene?
 - Hvordan nå fram til potensielle brukere
 - Hvordan sørge for tverrfaglig bruk av forskningsdata?
 - Hvordan bør ansvarsdeling mellom tjenestetilbyderne og fagmiljøene være med tanke på tilbud om kurs, opplæring og utdanning knyttet til tjenestene?

Forskningsrådet vil benytte kunnskap fra workshopen i sine innspill til Kunnskapsdepartementets arbeid med en strategi for åpen tilgang til forskningsdata. Denne skal ferdigstilles i 2017.

Etablering av en dialogarena?

Et ytterligere formål med workshopen i juni er å sondere den videre interessen for å etablere en arena for litt mer jevnlig dialog der aktuelle problemstillinger og tema kan drøftes. Følgende problemstillinger, deltagelse og møtefrekvens kan her være aktuelle:

Formål: Deling av erfaringer og diskusjoner rundt hvordan det kan skapes et helhetlig tjenestetilbud for deling og gjenbruk av forskningsdata. Problemstillinger for dialog kan være:

- Koble data på tvers av fagdisipliner og skape helhetlige løsninger for forskerne
- Norsk bidrag til European Open Science Cloud (EOSC)
- Basis-infrastruktur horisontalt, fagspesifikk infrastruktur vertikalt; her må basis-infrastrukturen være godt kjent og fylle etterspurte behov
- Basis-infrastrukturer som henger sammen
- Søke å dekke behov i alle felt, samtidig som man ikke får for mye overlapp

Deltagelse: Forskningsrådet ønsker at dette skal være en dialog der deltagerne har høy kompetanse på lagring og tilgjengeliggjøring av forskningsdata. Vi har derfor sendt denne invitasjonen til kun 25 personer som enten er leder for en infrastruktur-tjeneste eller for en institusjonen som tilbyr en slik tjeneste (se vedlagt utsendelsesliste). Forskningsrådet foreslår at én representant fra hver av disse infrastrukturene eller institusjonene med høy kompetanse på lagring og tilgjengeliggjøring av forskningsdata deltar.

Møtefrekvens: Vi foreslår at medlemmene i dialogarenaen møtes én gang hvert halvår, men vi er åpne for å justere møtefrekvensen (opp eller ned) ved behov.

Program, 7. juni 2016

Klokkeslett	Presentasjon	Innleder
09:30-10:00	Registrering og kaffe	
10:00-10:10	Velkommen	Forskningsrådet
10:10-10:30	Presentasjonsrunde av alle deltagerne	
10:30-10:50	Lagring og tilgjengeliggjøring av forskningsdata - med spesiell fokus på NORD-i	NSD
10:50-11:10	Lagring og tilgjengeliggjøring av forskningsdata gjennom NorStore	UNINETT Sigma2
11:10-11:30	Pause med frukt	
11:30-11:50	Åpen tilgang til data gjennom åpen publisering	CRISStin
11:50-12:10	Å gjøre forskningen synlig. Tildeling av DOI-nr. i Norge	BIBSYS
12:10-13:00	Lunsj	
13:00-13:20	Informasjon om European Open Science Cloud	Forskningsrådet
13:20-13:40	KDs strategi for tilgjengeliggjøring av data	Kunnskapsdepartementet
13:40-15:20	Gruppearbeid* i 4 grupper inkl. 20 min pause	
15:20-15:40	Oppsummering gruppearbeid i plenum	Forskningsrådet
15:40-16:00	Videre arbeid. Etablering av dialogarena?	
16:00	Slutt	

*) Tema for gruppearbeid:

- Hvordan videreutvikle dagens tjenester for datahåndtering?
- Hvordan stimulere til økt bruk av tjenestene?

Invitasjonsliste, 7. juni 2016

SSB	Terje Risberg
NSD	Vigdis Kvalheim
Arkivverket	Olav Hagen Sataslåtten
Norsk Helsearkiv	Tom Kovig
Artsdatabanken	Ivar Myklebust
NINA	Norunn Myklebust
ELIXIR	Inge Jonassen
CLARINO	Koenraad De Smedt
UNINETT Sigma2	Gunnar Bøe
Helseregisterforskning	Stein Emil Vollset
LIA	Janne Bondi
Norwegian Marine Data Centre	Helge Sagen
NorBol	Torbjørn Ekrem
Biobank Norway	Kristian Hveem
SIOS	Ole Jørgen Lønne
CRISStin	Katrine Weisteen Bjerde
Folkehelseinstituttet	Camilla Stoltenberg
BIBSYS	Frode Arntsen
MET	Øystein Godøy
Tjeneste for sensitive data, USIT	Gard Thomassen
EPOS	Kuvvet Atakan
ICOS	Truls Johannessen
Elinfrastruktur for Video Research	Kirsti Klette
Registreringsentral for historiske data	Gunnar Thorvaldsen
INESS	Victoria Rosen

Dialogmøte/Workshop, lagring og tilgjengeliggjøring av forskningsdata, 1. november 2016

Et av tiltakene i Meld. St. 27 (2015-2016) *Digital agenda for Norge – IKT for en enklere hverdag og økt produktivitet* er å styrke arbeidet med tilgjengeliggjøring av offentlige data. Meldingen anser bl.a. forskningsdata som samfunnsøkonomisk særlig verdifulle, og et av tiltakene er at det skal utarbeides en strategi for deling og tilgjengeliggjøring av slike data. I den forbindelse har Kunnskapsdepartementet (KD) bedt Forskningsrådet om å etablere et kunnskapsgrunnlag. Dette skal benyttes i KDs arbeid med å utforme en nasjonal strategi for deling og tilgjengeliggjøring av forskningsdata, som skal lanseres i 2017.

Den 7. juni i år var de fleste av mottagerne av denne invitasjonen samlet til en første workshop for å dele synspunkter og gi innspill til et slikt kunnskapsgrunnlag. Forskningsrådet har siden da arbeidet videre med innspillene, og vi har gjennomført intervjuer med en rekke sentrale aktører.

I workshopen den 1. november ønsker Forskningsrådet å motta synspunkter på hovedelementene i det kunnskapsgrunnlag vi hittil har etablert, og vi utfordrer deltagerne til å komme med innspill til mulige anbefalinger i den kommende nasjonale strategien.

For å innlede til diskusjonen vil Forskningsrådet presentere status for eget kunnskapsgrunnlag. I tillegg vil fire av møtedeltagerne – alle fra ulike fagfelt – presentere sine tanker om hvilke tiltak som skal til for å forbedre forskernes tilgang til forskningsdata.

Program, 1. november

Klokkeslett	Presentasjon								
10:00-10:30	Velkommen og Presentasjonsrunde								
10:30-12:00 (inkludert kort pause)	"Hvordan bedre forskernes tilgang til forskningsdata?" 4 x 15 minutters presentasjoner, hver etterfulgt av kort diskusjon								
(pause) →	<table> <tr> <td>1) Kjell Petersen og Eivind Hovig</td> <td>ELIXIR-Norway</td> </tr> <tr> <td>2) Koenraad De Smedt</td> <td>CLARINO</td> </tr> <tr> <td>3) Kuvvet Atakan</td> <td>EPOS</td> </tr> <tr> <td>4) Dag Endresen</td> <td>GBIF Norway</td> </tr> </table>	1) Kjell Petersen og Eivind Hovig	ELIXIR-Norway	2) Koenraad De Smedt	CLARINO	3) Kuvvet Atakan	EPOS	4) Dag Endresen	GBIF Norway
1) Kjell Petersen og Eivind Hovig	ELIXIR-Norway								
2) Koenraad De Smedt	CLARINO								
3) Kuvvet Atakan	EPOS								
4) Dag Endresen	GBIF Norway								
12:00-12:45	Lunsj								
12:45-13:45	Presentasjon av status for Forskningsrådets kunnskapsgrunnlag til KDs strategi for deling og tilgjengeliggjøring av forskningsdata								
13:45-15:20	Gruppearbeid med formål om å kommentere Forskningsrådets kunnskapsgrunnlag, samt å gi innspill til mulige anbefalinger								
15:20-16:00	Presentasjon av gruppearbeidene								
16:00	Slutt								

Invitasjonsliste, 1. november

Artsdatabanken	Nils Valland	INESS	Victoria Rosén
BIBSYS	Frode Arntsen, Sigrid Smedsrud	LIA	Kristin Hagen
	Jan Erik Garshol	Nasjonalbiblioteket	Thomas Langvann
Biobank Norway	Kristian Hveem	NILU	Kjetil Tørseth, Markus Fiebig
CLARINO	Koenraad De Smedt	NINA	Roald Vang
CRISStin	Katrine Weisteen Bjerde	Norsk Helsearkiv	Jens Georg Gleditsch
Difi	Pia Josendal	NMDC	Helge Sagen
eVIR	Kirsti Klette	NSD	Bjørn Henrichsen
ELIXIR	Kjell Petersen		Vigdis Kvalheim
	Eivind Hovig	HISTREG	Gunnar Thorvaldsen
EPOS	Kuvvet Atakan	SSB	Terje Risberg
Folkehelseinstituttet	Gun Peggy Knudsen		Svein Johansen
GBIF Norge	Dag Endresen	TSD/USIT	Gard Thomassen
Helseregistre	Stein Emil Vollset	UNINETT Sigma2	Andreas Jaunsen
ICOS	Truls Johannessen	TROLLing	Helene Andreassen
	Benjamin Pfeil		

ISBN 978-82-12-03566-9