

Nasjonal strategi for forskningsinfrastruktur

Utvalg for databaser, registre, tidsserier og samlinger

Forslag til strategi

Norges forskningsråd 2007

Norges forskningsråd
Postboks 2700 St. Hanshaugen
0131 OSLO
Telefon: 22 03 70 00
Telefaks: 22 03 70 01
bibliotek@forskningsradet.no
www.forskningsradet.no/

Publikasjonen kan bestilles via internett:

www.forskningsradet.no/publikasjoner

eller grønt nummer telefaks: 800 83 001

Grafisk design omslag:

Foto/ill. omslagsside:

Trykk: Norges forskningsråds trykkeri, Oslo

Opplag: 50

Oslo, september 2007

ISBN 978-82-12-02504-2(trykksak)

ISBN 978-82-12-02505-9 (pdf)

Forord

Forskningsrådet skal i løpet av 2007 ferdigstille en nasjonal strategi for forskningsinfrastruktur som skal omfatte avansert vitenskapelig utstyr, elektronisk infrastruktur og vitenskapelige databaser og samlinger. På denne bakgrunn har Forskningsrådet oppnevnt "Utvalget for databaser, registre, tidsserier og samlinger" som på et overordnet nivå skal komme med forslag til en strategi for Forskningsrådets arbeid med databaser (inkludert registre og tidsserier) og samlinger som anses å ha stor samfunnsmessig verdi.

Utvalget har hatt seks møter fra april til august 2007. Utvalgets medlemmer består av representanter fra sentrale brukere og forskningsmiljøer og representerer ekspertise på hvert sitt fagfelt. Utvalget har ikke hatt medlemmer fra alle fagfelt og arbeidet er således ikke uttømmende.

Foreliggende rapport presenterer utvalgets arbeid, forslag og anbefalinger. Beskrivelsene av de ulike fagfeltene er laget ut fra hvert enkelt utvalgsmedlems kunnskap og mening om fagfeltet. Utvalgsmedlemmene har søkt innspill og bistand fra kollegaer innenfor sitt fagfelt og vil rette en takk til dem som har hjulpet til for å få rapporten ferdig.

Tor Borgar Hansen og Anders Ekeland fra NIFU STEP har vært sekretariat for utvalget.

September 2007

Helge Sagen, utvalgets leder

Bjarne Hodne

Gjert Kristoffersen

Vigdis Namtvedt Kvalheim

Magne Lystad

Per Magnus

Johan-Kristian Tønder

Nils Valland

Innhold

1	Forskningsrådets nasjonale strategi for forskningsinfrastruktur	9
1.1	Svært kort om hele strategien	9
1.2	Formålet med utvalget for databaser, tidsserier, registre og samlinger	9
1.3	Utvalgets mandat og sammensetning	9
1.4	Definisjoner	11
2	Databaser som forskningsinfrastruktur	13
2.1	Bakgrunn	13
2.2	Databaser blir stadig viktigere som forskningsinfrastruktur.....	14
3	Felles utfordringer	17
3.1	Finansiering	17
3.2	Metadata	17
3.3	Standardisering og interoperabilitet.....	18
3.4	Merittering	18
3.5	Digitalisering av samlinger.....	18
3.6	Bedre tilgjengeligheten.....	19
3.7	Organisering	19
3.8	Avlevering, arkivering og gjenbruk.....	19
3.9	Prosessgenererte data.....	19
4	En nasjonal strategi	21
4.1	Overordnet mål og delmål	21
4.2	Forskningsrådets ansvar.....	23
4.3	Kriterier som kan anvendes for å prioritere ressursinnsatsen	23
5	Handlingsplan – prioriterte oppgaver.....	25
5.1	Tiltak.....	25
5.2	Momenter til et mandat for Forskningsrådets databaseutvalg	26
	Litteraturliste	27
	Vedlegg I: Status og perspektiver for fagområdene	29
	Medisin og helseområdet.....	29
	Samfunnsvitenskap.....	31
	Humaniora	36
	Naturvitenskapene	39
	Vedlegg II: Beskrivelse av sentrale aktører	45
	Nasjonalt folkehelseinstitutt	45
	Norsk Samfunnsvitenskapelig datatjeneste	46
	Statistisk sentralbyrå.....	48
	Språk- og tekstvitenskapelige datasentre	49
	MUSIT	49
	DigForsk.....	50
	Meteorologisk institutt	50
	Norsk Marint Datasenter	53
	Artsdatabanken.....	53
	Norge Digitalt.....	54

Sammendrag

Utvalget foreslår at Forskningsrådet - i samarbeid med andre aktører - må:

- Sammen med forskningsmiljøene bidra til utvikling og permanent finansiering av formålstjenlige datatjenester for forskningen innenfor alle fagområder. På fagområder der det foreligger konkrete og omforente behov og planer, bør disse søkes iverksatt så snart som mulig. Der dette ennå ikke er gjort, må det iverksettes fagspesifikke kartlegginger av situasjonen. Resultatet må bli et overordnet veikart for hele feltet og en konkret handlingsplan for hvert fagområde.
- Bidra i et krafttak for digitalisering av samlinger, av papirbaserte data og av materiale i eksisterende biobanker.
- Stille som vilkår for økonomisk støtte til prosjekter, programmer og større satsinger at resultater og data arkiveres, dokumenteres og forvaltes slik at det blir mulig å gjenbruke data og gi innsyn i forskningens resultater.
- Finansiere lagring og formidling av data for gjenbruk.
- Finansiere opprydding/videreføring av databaser gjennom egne programmer.
- Kreve ryddige kontraktsforhold når det gjelder eiendomsrett til dataene og utforming av regelverk og prinsipper for datadeling mellom aktørene.
- Ta initiativ til forenkling av vilkårene for utlevering av data til forskning, og til effektiv saksbehandling for slik utlevering. Det gjelder lover og regler så vel som institusjonell praksis og kultur.
- Ta initiativ for å sikre at datainnsamling som foretas av offentlige og private institusjoner og bedrifter tilpasses forskningens behov, bl.a. gjennom felles identifikasjonssystemer.
- Koordinere arbeidet for å sikre en fullstendig og enhetlig dokumentasjon av og god informasjon om datakilder, dvs. metadata, som kan brukes i forskning.
- Bidra til internasjonalisering av norsk forskning ved å sikre langsiktig finansiering av datainnsamling knyttet til internasjonale prosjekter.
- Representere Norge internasjonalt for å sikre at norske forskere kan dra nytte av det infrastrukturarbeidet som gjøres internasjonalt. I de nærmeste årene peker ESFRI-samarbeidet seg ut som et naturlig satsingsfelt, faglig så vel som økonomisk.

Utvalget foreslår at det oppnevnes et *Forskningsrådets databaseutvalg* som et *rådgivende* organ for Forskningsrådet. Dette utvalget må ha et helhetlig ansvar nasjonalt.

Forskningsrådets databaseutvalg skal være et faglig bindeledd mellom forskningsmiljøene og Forskningsrådet i alle spørsmål som angår utvikling og vedlikehold av en best mulig nasjonal forskningsinfrastruktur.

Forskningsrådet må stille med sekretariat for utvalget forankret i Divisjon for vitenskap.

Dersom Forskningsrådet oppretter et Forskningsrådets utvalg for infrastruktur, forutsetter vi at dette utvalget overtar ansvaret for de funksjonene som vi har forutsatt at databaseutvalget skal ha.

1 Forskningsrådets nasjonale strategi for forskningsinfrastruktur

1.1 Svært kort om hele strategien

Norges forskningsråds arbeid med en nasjonal strategi for forskningsinfrastruktur må sees på bakgrunn av nasjonale og internasjonale aktiviteter på området. For Norge er det viktig at man satser tilstrekkelig på infrastruktur til at forskningsmiljøene har mulighet til å arbeide i den internasjonale forskningsfronten, og at nasjonale prioriteringer er etablert slik at disse kan legges til grunn for deltagelse i den markante EU-satsingen innen feltet. Forskningsrådets ledelse og styrende organer har derfor engasjert seg i dette som en viktig arbeidsoppgave i 2007.

Målsetningen for dette arbeidet er å utarbeide en nasjonal strategi (veikart) for forskningsinfrastruktur samt å etablere en egnet beslutningsstruktur for behandling av de ulike kategorier forskningsinfrastruktur som gir effektiv utnyttelse av ressursene. Strategiarbeidet er organisert gjennom tre utvalg som skal levere innspill til strategien innenfor områdene:

- tungt utstyr/installasjoner
- e-infrastruktur
- databaser, tidsserier, registre og samlinger

1.2 Formålet med utvalget for databaser, tidsserier, registre og samlinger

Et utvalg nedsatt av Forskningsrådet skal på et overordnet nivå komme med et forslag til en strategi for Forskningsrådets arbeid med databaser (inkl. registre, tidsserier og samlinger) som anses å ha stor samfunnsmessig verdi. Utvalget skal foreslå tiltak for forvaltning av ulike kategorier databaser, herunder nødvendige beslutningssystem og investeringer, samt Forskningsrådets rolle i forhold til dette.

1.3 Utvalgets mandat og sammensetning

Utvalget har i løpet av sine første møter aktivt arbeidet med forslaget til mandat som ble utarbeidet av Forskningsrådet i forkant av arbeidets oppstart. Utvalget har hatt anledning til å diskutere mandatet og formulere den endelige utformingen.

Et utvalg nedsatt av Forskningsrådet skal på et overordnet nivå komme med et forslag til en strategi for Forskningsrådets arbeid med databaser (inkl. registre, tidsserier og samlinger) som anses å ha stor samfunnsmessig verdi. Utvalget skal foreslå tiltak for forvaltning av ulike kategorier databaser herunder nødvendige beslutningssystem og investeringer for databaser samt Forskningsrådets rolle i forhold til dette. Utvalgets arbeid skal ferdigstilles innen 7.9.2007.

1. Kategorisere ulike typer eksisterende og fremtidige databaser, registre, tidsserier og samlinger (inklusive pan-europeiske)

Beskrive nasjonale behov for databaser og bedre infrastruktur for tilgjengelighet og anvendelse av data og samlinger. Dette innebærer en kategorisering av databaser på flere nivåer, fra data til metadata, satt inn i en nasjonal og internasjonal sammenheng. Lage kriterier og forslag til rammeverk for de ulike kategorier, definere de ulike forsknings- og forvaltningsinstitusjoners roller og ansvar mht. hierarkiet av databaser og beskrive prinsipper for god tilgjengelighet og anvendelse av dem.

Når det gjelder samlinger skal arbeidet ses i sammenheng med den prosessen Kunnskapsdepartementet legger opp til med hensyn til oppfølgingen av NOU 2006:8 (Kunnskap for fellesskapet).

2. Utarbeide forslag til kriterier for strategisk prioritering av ressursinnsatsen til databaser som anses å ha stor samfunnsmessig verdi

Utvalget skal foreslå kriterier til bruk for utvelgelse av databaser som anses å ha stor samfunnsmessig verdi. Kriteriene skal brukes til å velge ut hvilke områder hvor en må sette inn resurser for å opprette/utvikle databaser, registre, tidsserier og samlinger til bruk for forskning. Utvalget skal foreslå et system for en evaluering av prioriteringene som er gjort.

3. Beskrive prinsipper for databasenes nasjonale og internasjonale bidrag

Belyse databasenes nasjonale og internasjonale dimensjon samt klarlegge kriterier med hensyn til å sikre at norske data blir effektivt utnyttet og bidrar i internasjonale nettverk og databaser.

4. System for utnyttelse, tilgjengelighet og koordinering av relevante databaser

a. Anbefale system for bedre utnyttelse og koordinering

Foreslå hvordan man kan etablere et nasjonalt system for oversikt og tilgang til relevante databaser. Herunder hvordan en skal møte utfordringene med å knytte relevante databaser sammen og etablere metadatakataloger. Vurderinger av ansvar for drift og administrasjon; videre innsamling av data, sikkerhet; standardisering etc. bør inkluderes.

b. Anbefale kriterier i forhold til IPR og bedre tilgjengelighet av data

Forholde seg til nasjonale regler og internasjonale avtaler for tilgjengelighet av data.

Vurder kriterier for eierskapsforhold; personvern; krav til avlevering av data; plikter og rettigheter for den enkelte forsker, forskningsinstitusjoner og bevilgende myndigheter i forhold til de ulike kategorier av databaser.

Utvalget skal vurdere behovet for å stille nye krav til forvaltning og utveksling av data fra offentlig finansiert forskning. Utvalget bør beskrive på et overordnet nivå eksisterende forpliktelser og avtaler om forvaltning og utveksling av data i nasjonal og internasjonal sammenheng.

5. Gi råd om finansiering og forvaltning

Foreslå finansieringssystem som sikrer langsiktighet i datainnsamling og drift av databaser som anses å ha stor samfunnsmessig verdi slik det er beskrevet i utvalgets øvrige arbeid.

Gi råd om forvaltningen av disse med hensyn til ansvarsforhold, virkemidler og investeringsprofil, inkludert både nasjonale og internasjonale databaser.

Utvalget har bestått av følgende personer:

Helge Sagen, *Norsk marint datasenter (leder)*

Vigdis Namtvedt Kvalheim, *Norsk samfunnsvitenskapelig datatjeneste*

Johan-Kristian Tønder, *Statistisk sentralbyrå*

Gjert Kristoffersen, *Institutt for lingvistiske, litterære og estetiske studier, Universitetet i Bergen*

Nils Valland, *Artsdatabanken*

Per Magnus, *Nasjonalt folkehelseinstitutt*

Magne Lystad, *Meteorologisk institutt*

Bjarne Hodne, *Institutt for kulturstudier og orientalske språk, Universitetet i Oslo og Forskerforbundet*

NIFU STEP ved Anders Ekeland og Tor Borgar Hansen har vært utvalgets sekretariat.

1.4 Definisjoner

Utvalget skal arbeide med databaser, registre, tidsserier og samlinger. Etter utvalgets mening er det hensiktsmessig å dele dette i to hovedgrupper, den ene kalt databaser – som omfatter registre, tidsserier, surveydata og andre spesialiserte databaser på den ene siden og samlinger på den andre. Mens databaser er digitale er samlingene fysiske objekter som kan få en digital gjengivelse av noen av objektets egenskaper.

Databaser er etter utvalgets mening en fellesbetegnelse for ulike former for *strukturerte, systematiserte, digitalt lagrede* data, som for eksempel private og offentlige registre, tidsserier, surveydata, digitale bilder, tekster, lydfiler. Det er essensielt at innholdet er systematisert. En samling av tekstfiler, eller digitale bilder er i seg selv ikke en database. Informasjonen må være i et datahåndteringssystem som gjør det mulig å gjenfinne informasjon på en systematisk måte ved bruk av ulike søkekriterier.

Et viktig begrep er (institusjonelt) *distribuerte databaser*. Med dette menes databaser som finnes hos ulike institusjoner men som på grunn av koblingsnøkler og innholdsmessig standardisering utgjør én database fra et innholdsmessig synspunkt. Alternativet er *sentraliserte* databaser – hvor data leveres inn/kopieres inn i én databasestruktur.

Samlinger er en systematisk samling av en viss type objekter. Det kan være historiske gjenstander, bevarte artseksemplarer, blodprøver, fossiler. Det som skiller samlingen fra databasen er at selv etter at et maleri er fotografert, dvs. digitalisert og lagt i en base med annen kontekstinformasjon, vil det å oppbevare maleriet fortsatt ha en egenverdi, fordi digitaliseringen ikke uttømmer gjenstandens informasjonsinnhold. En kan og vil ofte få behov for å gå tilbake til gjenstanden for å studere (digitalisere) nye kjennetegn ved gjenstanden. Derimot vil en for eksempel når det gjelder måleresultater som foreligger i papirform ikke lide samme type tap om papirversjonen skulle gå tapt etter at informasjonen er digitalisert, all informasjon er overført til det digitale mediet.

Utvalgets arbeid gjelder databaser som forskningsinfrastruktur og digitalisering av samlinger, d.v.s. å gi dem en digital representasjon. Forhold som berører oppbevaring, drift, vedlikehold og utvikling av samlinger i egenskap av fysiske objekter ligger dermed utenfor utvalgets mandat.

2 Databaser som forskningsinfrastruktur

2.1 Bakgrunn

Internasjonalt har det de siste årene vært stort fokus på behovet for forskningsinfrastruktur. I USA ble rapporten “Our Cultural Commonwealth: The report of the American Council of Learned Societies Commission on Cyber Infrastructure for the Humanities and Social Sciences”, publisert 13. desember 2006. I Europa la European Strategy Forum on Research Infrastructures (ESFRI) frem sitt første veikart i oktober 2006 med 35 prosjekter innen et bredt spekter av fagområder. Norge har, representert ved Norsk samfunnsvitenskapelig datatjenestes (NSD) leder, vært involvert i arbeidet med begge rapportene, som henholdsvis International Advisor to the Commission og som Chair for Working Group Social Sciences and Humanities (SSH) i ESFRI.

Norges sterke posisjon på infrastrukturområdet med sentral deltakelse i de infrastruktur- prosesser som skjer i Europa, skyldes ikke minst at Norge fra slutten av 60-tallet har hatt et relativt stort fokus på å bygge ut infrastruktur tjenester for de samfunnsvitenskapelige fag. Forskningsrådet brukte sitt eget organ, NSD, som instrument for satsingen. Dette har skjedd i nær kontakt med forskningsmiljøene.

Hovedmålsetningen har vært at data som ble samlet inn, gjennom forskningsprosjekter eller nasjonale satsinger, skulle stilles allment til disposisjon som en felles ressurs for forskningsmiljøene. Parallelt bygde man opp datatilbud som skulle fungere som samlede ressurser på tvers av fag og institusjoner. Etter hvert er de samfunnsvitenskapelige tilbudene tatt i bruk av stadig flere fagfelt, særlig medisin og deler av humaniora.

På mange måter er dette spesielt for Norge og har medført en fri og god tilgang til et informasjonsgrunnlag som er en fellesressurs for forskning. Over de siste 30-40 år har Norge opparbeidet en posisjon når det gjelder innholdsmessige, økonomiske, juridiske og teknologiske forhold rundt tilgang, organisering og forvaltning av forskningsdata som er blant de fremste i Europa og verden for øvrig.

Flere forhold har gitt grunnlag for dette. Den viktigste er at forskningen selv har tatt ansvar og brukt ressurser på å legge forholdene til rette for at forskningsrelevant informasjon blir lagt til rette i et format som gjør at den kan brukes av forskere på en enkel måte. I tillegg har man fra 70-tallet lagt vekt på å få til et forpliktende samarbeid med viktige ”eksterne” samarbeidspartnere som Statistisk sentralbyrå og Datatilsynet.

Norge har også satset på områder hvor internasjonalt forskersamarbeid er helt nødvendig, som for eksempel innen meteorologi og stratosfæriske data. En stadig større del av menneskelig aktivitet i samfunnet gjenspeiles i elektroniske registre. Dette gir enorm vekst i datamengden og dermed potensielt bedre muligheter for å analysere kulturelle og samfunnsmessige forhold. Behovet for data og samarbeid om data for å løse klima- og miljøproblemer er et annet eksempel på behov som krever en mer bevisst og målrettet strategi fra Forskningsrådet. Med unntak av for samfunnsvitenskapene har Forskningsrådet vært en lite aktiv premissleverandør og påvirkningsagent når det gjelder databaser som infrastruktur – til tross for de helt åpenbare gevinstene for bevilgende myndigheter, forskermiljøene og kunnskapssamfunnet. Det gjelder

både gjennom teknisk effektivisering, men ikke minst ved at en med nye typer data, lengre tidsserier osv. kan studere og gi bedre svar på både etablerte og nye problemstillinger. Forskningsrådet må i årene som kommer arbeide mye mer systematisk med alle sider av dette saksfeltet, som omfatter:

- finansiering av produksjon av nye data
- utvikling av standarder og metadata
- lover, regler og normer angående tilgjengelighet
- tilrettelegging for brukergrupper
- den institusjonelle strukturen som regulerer bruken av databaser som infrastruktur for forskning
- finansiering av digitalisering av eldre kilder innenfor samtlige fagfelt, og finansiere videreutvikling og drift når disse kildene er digitalisert

2.2 Databaser blir stadig viktigere som forskningsinfrastruktur

Databaser er av mange årsaker blitt en stadig viktigere del av det som omtales som infrastruktur for forskning. Det er flere årsaker til dette. Ikke bare øker bruken av databaser effektiviteten i forskningen ved at tilgang og bearbeiding av data går raskere og mer effektivt ved at arbeid med å finne og sammenstille data blir effektivisert. Noen typer data (folkeregistre, sosioøkonomiske mikrodata) finnes bare, og kan bare brukes som databasedata. For flere forskningsfelt fungerer dermed databasene på samme måte som nytt og bedre vitenskapelig teknisk utstyr tradisjonelt har gjort innen naturvitenskapene.

Et viktig mål for Databaseutvalget har derfor vært å fremskaffe oversikt over de ulike typer infrastruktur det er snakk om - innenfor hele bredden av forskningsområder.

En fornuftig organisert infrastruktur på dette området kan bidra til økt *effektivisering* av forskningen ved samarbeid om innsamling og utnyttelse av felles innsamlede data. Offentlig investering i slik infrastruktur vil bidra til å utvikle felles goder for forskersamfunnet og stimulere til større fellesprosjekter, effektiv forskeropplæring og bedre samordning av forskningen.

En slik infrastruktur bidrar også til å gjøre norsk forskning attraktiv i *internasjonal* forskningssammenheng, både i forhold til komparative undersøkelser, og generelt gjennom å fremskaffe større og unikt datatilfang, som er interessant for potensielle samarbeidsparter i andre land. Den vil bidra til at man kan se norske forhold i relasjon til forhold i andre land.

I Danmark og Sverige har en også i hhv. 2005 og 2006 utredet behovet for forskningsinfrastruktur herunder databaser. Fokus i begge rapporter¹ er humaniora, samfunnsvitenskap, medisin og helse og utnyttelsen av offentlige registerdata. Det er verdt å merke seg at i Danmark har KOR (Koordinerende Organ for Registerdata) et budsjett på 9 millioner DKK per år. Den danske rapporten bemerker at KOR har fått en veldig god mottakelse blant forskerne og er blitt møtt med stor interesse i andre land, blant annet at det svenske Vetenskapsrådet overveier å kopiere KOR modellen.

Så ser ut til å ha skjedd for Vetenskapsrådet opprettet DISC (Database Infrastructure Committee) i 1996. Fokus for DISC sin virksomhet er som for KOR offentlige registerdata (mikrodata) om personer og bedrifter. DISC samarbeider nært med Statistiska Centralbyrån og Svensk Nationell Datatjänst. DISC hadde et budsjett på 33 millioner SEK i 2006 og forventes å ha et budsjett på

¹”Fremtidens forskningsinfrastruktur – kortlegging av behov og forslag til strategi” fra det Strategiske forskningsråd. Desember 2005 og ”Vetenskapsrådet guide til infrastrukturen. En kartläggning och rekommendationer for svenska forskares tillgång till forskningsinfrastruktur på lång sikt”. Vetenskapsrådets rapportserie 14:2006

59,5 millioner i 2009. I 2006 gikk 1,5 millioner til drift av DISC, oppbygging og vedlikehold av databaser fikk 21 millioner, arkivering 5 millioner, mens teknikk og kvalitetsutvikling fikk 5 millioner. Økningen i DISCs budsjett er knyttet til at man ønsker at DISC skal ta et større ansvar for klima-, biosfære, litosfære- og miljøforskning.

Innenfor det europeiske forskningssamarbeidet vektlegges også samarbeid om videreutvikling av infrastruktur. Det sjette rammeprogrammet introduserte et eget delprogram "Support for Research Infrastructures" med nye aktiviteter og prosjektformer. Infrastrukturene kan være geografisk knyttet til én institusjon, distribuerte (f eks grid-nettverk) eller virtuelle, dvs. at tjenesten tilbys elektronisk.

Mye av denne type infrastruktur som benyttes i norsk forskning (og som omtales her) er ikke primært etablert for forskningsformål. NSD er opprettet for å ivareta forskningens behov, primært innenfor samfunnsvitenskapene. Men en rekke andre nasjonale samlinger/registre har andre eller todelte hovedformål (forvaltning, offentlig styring, overvåking, folkeopplysning med mer). Flere samlinger og registre kan med andre ord tjene flere formål og krever spesiell tilrettelegging for forskere. Forskningsrådets kartlegging i 2004 som et innspill til St. meld. Nr. 20 (2004-05) *Vilje til forskning* viser at omfanget av denne type infrastruktur er omfattende, og har stort potensial for videreutvikling av norsk forskning.

Den nasjonale satsingen gjøres for å gi et bedre tilbud nasjonalt, men også for å komme i posisjon internasjonalt i forbindelse med de nye ressurser som nå blir stilt til disposisjon via EUs 7. rammeprogram. Det arbeidet som skjer innen området i regi av ESFRI er viktig i denne sammenhengen.

For å sikre en fortsatt god tilgang til data av høy kvalitet er det med andre ord viktig og helt nødvendig at infrastrukturtenester videreutvikles og opprettholdes i tett samarbeid og i forståelse med forskersamfunnet og forskningsfinansierende myndigheter. Samtidig må det arbeides for at tilgjengeligheten stadig omfatter et bredere datagrunnlag. Dette krever en langsiktig og kontinuerlig strategisk innsats for å løse utfordringer knyttet til institusjonelle, økonomiske, juridiske og teknologiske barrierer for tilgang til data.

3 Felles utfordringer

I vedlegg I er det gitt en beskrivelse av status og perspektiver for de ulike fagområdene. Nettopp fordi det er ulike fagområder, som har ulike forskningsobjekter og ulike datakilder, vil mange utfordringer være særegne, både for fagområdene som helhet, men også for de ulike delområdene som hvert område omfatter. Men det er også utfordringer som er felles for flere – om ikke alle områder. Også disse utfordringene – og svarene på dem - vil være formet av den fagspesifikke konteksten. Med felles utfordringer menes her utfordringer på et forholdsvis overordnet nivå som er felles for flere fagområder – ikke nødvendigvis alle.

For forskeren er det nødvendig å få relevante data så raskt, komplett og billig som mulig, med tilhørende metadata og med kompetent service/veiledningstjeneste. I prinsippet skulle den som produserer data være den som best kunne tilfredsstille disse behovene. I praksis vil ikke situasjonen være optimal for ett eller flere av leddene ovenfor. Forskningsstøttende institusjoner som for eksempel NSD og NMD har bl.a. som oppgave å medvirke til å forbedre situasjonen på de punktene som ikke fungerer godt nok. Forskningsrådet har **og må fortsatt** delta i finansieringen slik at disse institusjonene kan tilby slike tjenester.

Behovet for tjenester vil variere i art og omfang mellom de ulike fagområdene. Også innenfor fagområdene kan det være ulike behov mellom ulike grupper av forskere. Dessuten vil behovene variere over tid fordi datakvalitet, kapasitet og kompetanse endrer seg som følge av endringene i oppgaver og ressurser hos de som produserer og leverer data til forskning. Det er derfor viktig at forskersamfunnet og Norges forskningsråd følger utviklingen i behovene for tjenester, og sørger for tjenesteytende institusjoner som til en hver tid har et samfunnsmessig effektivt tilbud som dekker forskernes behov.

3.1 Finansiering

Finansiering av arbeidet med databaser som infrastruktur for forskning er en viktig felles utfordring. Området har de siste ti-femten årene vært helt klart vært underfinansiert i forhold til både de kostnadsbesparelser som gode databaser gir, men ikke minst de gevinster en kan oppnå når det gjelder kvalitet i forskningen.

Gevinstene ved å investere i denne typen infrastruktur må derfor synliggjøres innad i Forskningsrådet gjennom Forskningsrådets egen satsing på feltet. Det bør utarbeides en oversikt over de investeringer som har vært gjort og de man vil gjøre når det gjelder databaser som infrastruktur. Dette er en viktig forutsetning for at Forskningsrådet på en effektiv måte skal kunne ta initiativer overfor andre myndigheter som direkte eller indirekte bevilger penger til forskning.

Det er viktig at finansieringen er langsiktig. Uferdige eller ikke-vedlikeholdte databaser har nesten ingen verdi, mens verdien av komplette og vedlikeholdte databaser er høy.

Et forsterket krav og mer konsekvent håndheving av et krav om arkivering for etterprøving og gjenbruk vil medføre visse kostnader som må finansieres.

3.2 Metadata

Med en voldsom økning i datamengden blir metadata, dvs. data om data svært viktig. En trenger metadata på mange nivåer. Ikke minst for å få oversikt over hva slags data som finnes på et fagområde, deretter dokumentasjon med ulike og økende detaljeringsgrad. På dette området åpner

utviklingen av semantiske teknologier, som regel basert på XML², bruk av ”semantisk Web”, web-services ol. Et eksempel på dette er Data Documentation Initiative³. Det er også en økende bruk av internasjonale standarder som Darwin Core 2. Det er svært viktig at det blir avsatt resurser til å følge – og påvirke – utviklingen på metadatafeltet og spre kunnskap og bruk av moderne metadata teknologier.

3.3 Standardisering og interoperabilitet

Nytten av det enkelte datasett øker sterkt hvis en benytter standarder og teknologiske plattformer som letter interoperabiliteten. Det er viktig at arbeidet med standardisering forsterkes. I den forbindelse er det helt avgjørende at det fra de som finansierer forskning blir stilt krav om bruk av standarder, om dokumentasjon, forsvarlig lagring og tilgjengelighet. Det er viktig at også informasjon om standardene på ulike fagområder er lett tilgjengelig.

En viktig forutsetning for interoperabilitet er at distribuerte databaser utformes slik at de kan gjøre dataene tilgjengelige og levere data gjennom søkbare tjenester i felles portaler innen fagområdene og på tvers av landegrensene. Ny teknologi muliggjør datadeling nasjonalt og internasjonalt uten å bygge opp institusjonell kapasitet som baserer seg på å forvalte kopier fra primærdatabaser. Sanntidstilgjengelighet i distribuerte baser krever imidlertid at primærdatabasene må oppgraderes til å kunne delta i denne type datakommunikasjon og det må lages tjenester med portalgrensesnitt som integrerer datakildene og gjør dataene interoperabile. Det bør satses sterkere på å utvikle slike tjenester og informasjonsflytarkitekturer for å effektivisere tilgjengeligheten av kvalitetsdata for forskning.

3.4 Merittering

Det å opprette og drifte databaser er å skape kollektive goder og det er viktig at incentivstrukturen rundt dette arbeidet er god. På den ene siden opplever en at høyt kvalifisert vitenskapelig personale bruker for mye av sin tid på dataarbeid som burde vært overlatt til personer med spesialisert kompetanse for dataarbeid. I slike tilfeller er arbeidet ”demeritterende”. På den andre siden er det eksempler på at dataarbeidet er så integrert med det vitenskapelige arbeidet at det *må* utføres av vitenskapelig personale. I disse tilfelle burde det regnes som meritterende. Praksis innen de ulike fagfelt er av historiske årsaker ulik. Innenfor enkelt fagområder gir utlevering av data til andre muligheter for å bli medforfatter.

3.5 Digitalisering av samlinger

En viktig felles utfordring for flere fagområder er digitalisering av samlinger og papirbaserte data. Det finnes felles teknologiske utfordringer hvor en kan vinne mye gjennom erfaringsutveksling om de meste effektive teknologiene for håndtering av digitale bilder, tekstgjenkjenning, koding osv. Når det gjelder databaser og tiltak for bedre og mer effektiv samlingsforvaltning ved universitetsmuseene, anbefalte Roganutvalget at museene gjennom tilrettelegging av museumsfaglige nettverk og samordning av programvareutvikling burde arbeide for en enhetlig og koordinert samlingsforvaltning og at det digitaliserte materialet må tilrettelegges for formidling, forvaltning og forskning. Også når det gjelder teknologier for god formidling til et bredere publikum kan det være mye å hente ved å utnytte faglige nettverk. Men den største utfordringen på dette feltet er finansiering av denne typen virksomhet som er arbeidskrevende og dermed kostbar, men samtidig klart samfunnsøkonomisk lønnsomt, jf. punktet om finansiering ovenfor.

² eXtensible Markup Language, en ”kraftigere” versjon av HTML

³ <http://www.ddialliance.org/codebook/index.html>

3.6 Bedre tilgjengeligheten

Felles for all infrastruktursatsing må være å sikre forskningsmiljøene bedre tilgang til de datakilder som i dag finnes og samtidig arbeide for at tilgjengeligheten stadig omfattes av et bredere datagrunnlag. Dette krever en kontinuerlig innsats for å løse utfordringer knyttet til økonomiske, institusjonelle og legale barrierer. Innsatsen bør, i den grad det er mulig, samordnes med institusjoner med lignende behov.

For at forskningsmiljøene skal kunne utnytte datakildene optimalt er det helt nødvendig at serviceapparat rundt data bygges videre ut. Dette apparatet må ha som sentrale oppgaver å etablere oversikter og metadatabaser med datadokumentasjon innrettet mot forskning på alle nivåer, informasjon om koblingsmuligheter, nærliggende datakilder, tidligere bruk og tilrettelegginger. Det ansees også som avgjørende at moderne kommunikasjonsteknologi kan anvendes for å gjøre databaser interoperabile, dette krever felles infrastrukturer og databasetjenester som inneholder nøkkelparametre som deles av databasene og som kan brukes som koplingsfelt.

Bedre tilgjengelighet innebærer også at det må avsettes ressurser til formidling og direkte betjening av forskerne og studentene. Erfaringsmessig utløser forskningsmessig bruk av data et stort behov for spesielle tilrettelegginger og utstrakt veiledning i bruk av data og mulige datakilder. Veiledningsbehovet varierer etter dataenes kompleksitet og brukernes erfaringsgrunnlag og ferdigheter.

3.7 Organisering

Når det gjelder databaser som infrastruktur er dagens situasjon preget av at ikke noen instans har et overordnet ansvar for dette. Dermed mangler det en instans som kan:

- arbeide med finansiering av databaser som forskningsinfrastruktur
- ha oversikt over hva som finnes av tilgjengelige ressurser og sørge for at disse blir utnyttet
- fange opp nye krav og behov – for data, for metadata, for utvekslingsstandarder og plattformer
- løse konflikter mellom fagmiljøer om tilgang til data
- koordinere norsk innsats i internasjonalt standardiseringsarbeid
- arbeide med lover, regler og den praktiske håndheving av dem

Derfor er det å skape en formålstjenlig institusjonell struktur en felles og viktig utfordring.

3.8 Avlevering, arkivering og gjenbruk

Norske forskningsprosjekter produserer mye data gjennom egne innsamlinger, koblinger og spesifiserte tilrettelegginger. I dag finnes regelverk og rutiner for å arkivere data som er samlet inn gjennom forskningsprosjekter finansiert av Norges forskningsråd. Forskningsrådet må være en pådriver overfor Universitets – og høyskolesystemet og instituttsektoren med sikte på at krav om arkivering av forskningsdata innføres i hele forskningssektoren. Det bør avsettes ressurser til dette arbeidet. Samtidig er det nødvendig at man bygger ut og raffinerer de rutiner som er etablert for dokumentasjon og informasjon om dataene. Gjennomføringen av et slikt arbeid forutsetter at man har rutiner som sikrer oversikt over hvilke prosjekter som produserer data, og som sikrer at dataene blir avlevert for arkivering. Ved siden av det som kan oppnås gjennom felles regler og avtaler, er det nødvendig å sette av langsiktige ressurser for å kontinuerlig arbeide for at data blir avlevert, tilrettelagt og vedlikeholdt på en måte som muliggjør senere gjenbruk.

3.9 Prosessgenererte data

Bruk av prosessproduserte data medfører ofte at forskerne møter dokumentasjonsproblemer, en stor grad av kompleksitet og metodiske utfordringer i tillegg til tilgang i seg selv.

Erfaringsmessig kan følgende forhold trekkes fram:

- juridiske utfordringer
- institusjonelle barrierer
- prispolitikk
- utfordringer knyttet til kobling av data fra ulike registre
- manglende eller vanskelig tilgjengelig dokumentasjon
- behovet for veiledning og oppfølging

Det er derfor viktig at man etablerer ordninger som bidrar til å løse disse problemene, ved å bygge videre på de ordninger, systemer og institusjoner om allerede finnes og som muliggjør forskning på denne datatypen.

Miljøene har behov for et godt organisert apparat som kan bistå og legge til rette for alle sider knyttet til bruk av data i forskningsprosessen. Denne helheten må bestå av bistand, hjelp og veiledning i alle faser av tilgang, tilrettelegging, dokumentasjon, utlevering og videre oppfølging. For at prosessen skal gå mest mulig effektivt er det viktig at forskeren eller studenten må forholde seg til færrest mulig instanser og at det er kompetanse og høy kvalitet knyttet til alle deltjenestene.

4 En nasjonal strategi

En nasjonal strategi på datasiden skal bidra til at norsk forskning har tilgang til relevante data av en slik kvalitet at de egner seg for forskning, ved

1. å sikre best mulig tilgang til eksisterende data, og
2. å ta initiativ til innsamling og/eller digitalisering av data på områder der det digitale datagrunnlaget er svakt eller mangler helt.

4.1 Overordnet mål og delmål

Forskningsrådets datapolitikk må ha som overordnet mål å fremme, forenkle og forbedre empirisk forskning og samfunnets kunnskapsgrunnlag. Dette målet kan spesifiseres i følgende delmål, som vi kommenterer nærmere nedenfor ett for ett.

Delmål:

1. Dekke forskningens behov for data av høy kvalitet.
2. Sikre lik tilgang til data for alle forskningsmiljøer.
3. Bidra til at Norge og norske forskere gjennom sin tilgang til data knyttes opp mot europeiske og internasjonale forskningsaktiviteter/nettverk.

Dekke behov for data av høy kvalitet

En viktig forutsetning for en velfungerende infrastruktur er åpen og enkel tilgang til relevante data av høy kvalitet. For at en skal oppnå dette må dataene for det første finnes. For det andre må ansvaret for forvaltning og sikring av tilgang plasseres hos kompetente institusjoner og gis en tilfredsstillende finansiering.

Forskningsrådet må bidra til sikker arkivering av forskningsresultater og datagrunnlag, både for gjenbruks og kontrollformål. Vi tar som utgangspunkt at det innen de fleste fagmiljøer finnes institusjoner som har kompetanse på bygging og forvaltning av fagspesifikke, digitale forskningsressurser. Det kreves likevel en fagspesifikk kartlegging av situasjonen innenfor de ulike fagområdene som en grunnleggende del av strategien. Det må særlig legges vekt på hvordan disse institusjonene i dag finansieres, og Forskningsrådet må sammen med institusjonene og ev. i samarbeid med relevante departementer utarbeide en tilfredsstillende organiserings- og finansieringsplan både for nybygging av nødvendige ressurser og for forvaltning og tilgjengeliggjøring av slike ressurser.

Utvalget legger til grunn at koordinert dataforvaltning gjennom egne forskningsinfrastruktur-tjenester er en hensiktsmessig organisering av datatjenester for forskning. En koordinert innsats vil gi klare stordriftsfordeler for norsk forskning. Man oppnår bl.a. ofte en direkte økonomisk gevinst ved at datainnsamling, tilrettelegging og dokumentasjon kun må gjøres én gang. En nasjonal koordinering, med fokus på data som en felles ressurs, krever også og bidrar til at man får utviklet og etablert metoder, dokumentasjon og standarder som muliggjør sammenlignende analyser over tid, nivåer osv. Forskningsrådet må også bidra til at datainnsamling som foretas av offentlige og private institusjoner og bedrifter tilpasses forskningens behov, bl.a. gjennom felles standard kodeverk og identifikasjonssystemer.

Forskningsinfrastrukturen som forvalter og tilgjengeliggjør data, må ha grundig kjennskap til disse, hvordan de er generert og hvilke definisjoner og feilkilder som gjelder. Infrastruktur-tjenestene må i tillegg ha bred kompetanse på utvikling av nødvendig programvare, kjennskap

til data og datakilder kombinert med innsikt i faglige problemstillinger. De må samtidig ha nærhet til og legitimitet i de miljøene de skal betjene, slik at de fanger opp behov for data og andre tjenester som forskningsmiljøene etterspør, eller som de antar miljøene vil komme til å etterspørre. Det er også viktig at slike infrastrukturtenester utvikles og opprettholdes i samarbeid og forståelse med dem som finansierer forskning, dvs. institusjonene og Forskningsrådet.

Forskningsrådet må bidra til at det finnes fullstendig og enhetlig dokumentasjon (metadata) av datakilder som er relevante for forskning. Gjennom krav knyttet til bevilgninger og lignende virkemidler må Forskningsrådet bidra til at internasjonale standarder tas i bruk overalt hvor slike finnes, og bidra til at nasjonale og internasjonale standarder utvikles der de måtte mangle. Forskningsrådet bør også søke å få andre forskningsfinansierende institusjoner til å stille tilsvarende krav.

Lik tilgang til data for alle forskningsmiljøer

Dette punktet innebærer for det første at Forskningsrådets strategi må omfatte alle fagområder, og at det følgelig må sørges for tilstrekkelig finansiering av forskningsinfrastruktur innenfor alle. Viktig i denne sammenheng er spørsmålet om hvordan forskere og studenter skal kunne bruke tjenestene uavhengig av økonomi.

En vellykket datapolitikk forutsetter også at man i størst mulig grad oppnår å redusere økonomiske, tekniske, kunnskapsmessige, juridiske og administrative barrierer mellom brukere og tilgjengelige dataressurser. Forskningsrådet bør være den drivende kraften i dette arbeidet.

Hjelp til å forholde seg til lovverk og institusjonelle rutiner for deling og utlevering av data er en viktig del av en fungerende infrastrukturteneste. Forskere og studenter som ønsker å bruke et datasett 30 år fram i tid må være sikret tilgjengelige data og dokumentasjon i oppdaterte standardformat. På samme måte må et forpliktende internasjonalt samarbeid ha ordninger og systemer som sikrer hurtig og adekvat tilgang til viktige flernasjonale datakilder.

Forskningsrådet må bidra til ryddige kontraktsforhold når det gjelder eiendomsrett til dataene. Et helt grunnleggende ledd i fremtidens datapolitikk må i denne sammenhengen være at Forskningsrådet knytter krav til alle forskningsbevilgninger om at data etter prosjektslutt avleveres til institusjoner som har arkivering av data som oppgave, jf. forrige punkt, og at dataene, ev. etter noen års sperrefrist, gjennom disse institusjonene gjøres allment tilgjengelige for forskersamfunnet. Hensikten er å bidra til at forskningens tilgang til data ikke hindres av ineffektiv saksbehandling, restriktiv tolkning av lover, institusjonell praksis og kultur.

Internasjonal koordinering

I Forskningsmeldingen er det pekt på at det skal legges til rette for bedret utnyttelse av data både i nasjonalt og internasjonalt forskningssamarbeid. Det pekes spesielt på at norske databaser skal videreutvikles og at det skal arbeides med å etablere datatilbud som åpner for sammenlikninger med andre land. Skal Norge ligge i første rekke internasjonalt er det nødvendig at den nasjonale infrastrukturen inngår i et forpliktende samarbeid med ledende internasjonale miljøer. Dette skjer for eksempel gjennom aktiv deltakelse i ulike internasjonale fora, gjennom felles datainnsamling og gjennom utveksling og tilrettelegging. I dette arbeidet må man trekke inn forskere og infrastrukturtenester som har erfaring fra internasjonalt samarbeid.

4.2 Forskningsrådets ansvar

For å nå de delmålene ved hjelp av de virkemidlene som er nevnt foran, ser utvalget for seg tre roller som Forskningsrådet må ha når det gjelder forskernes tilgang til databaser og samlinger. Disse rollene er nevnt i forrige avsnitt. Her tydeliggjør vi dem.

Forskningsrådet som premissleverandør

Forskningsrådet må være premissleverandør overfor forskningsmiljøer, dataproduserende institusjoner og forskningens infrastruktur tjenester/datatjenester. Rådet må stille som vilkår for økonomisk støtte til infrastruktur tjenester, prosjekter og programmer at resultater og data arkiveres, dokumenteres og forvaltes i tråd med kravene nevnt ovenfor. Rådet må derfor støtte satsinger og prosjekter som bidrar til et bedre datagrunnlag for forskning.

Forskningsrådet som pådriver

Forskningsrådet må påta seg ansvaret som rådgiver overfor dataproduserende institusjoner og deres eiere. Rådet må søke å påvirke og samarbeide med forskningsstøttende institusjoner, dataeiere, datavernmyndigheter og andre myndigheter for å oppnå forenkling i vilkår og praksis for utlevering til forskning. Rådet bør også søke å påvirke innhold i og samordning av private og offentlige institusjoners datainnsamling og dokumentasjonen av data, slik at de best mulig tjener forskersamfunnets interesser, så vel som samfunnets interesser.

På noen områder er mangelen på data til forskning så stor at Forskningsrådet i samarbeid med offentlige og private institusjoner må ta initiativet til oppretting og drift av de nødvendige ressursene.

Forskningsrådet som finansieringsinstitusjon

All datainnsamling er kostnadskreven. For forskningsmiljøene blir egeninnsamling av data ofte begrenset av økonomi. Det er derfor viktig og nødvendig at man i størst mulig grad kan basere seg på at den datainnsamling som foregår i private og offentlige institusjoner kan stilles til disposisjon for forskning. Forskningsrådet vil bare ha begrenset økonomisk mulighet til å finansiere løpende datainnsamling og organisering av data i databaser med tilgang for forskere. Men Forskningsrådet bør ha det overordnede ansvaret for at bygging av (digital) forskningsinfrastruktur og forvaltning av data får tilstrekkelig finansiering. Utvalget ser derfor for seg at finansieringen av fremtidens forskningsinfrastruktur må bli et spleiselag mellom Forskningsrådet, institusjonene og departementene, der Forskningsrådet må påta seg ansvaret for å organisere dette spleiselaget, og selv være beredt til å inngå som én av finansieringskildene.

4.3 Kriterier som kan anvendes for å prioritere ressursinnsatsen

Når fremtidens forskningsinfrastruktur knyttet til databaser etc. skal utvikles, vil det nødvendigvis innebære en rekke prioriteringer. Uansett hvor romslig finansieringen blir, vil det ikke være tilstrekkelig til at alle gode ideer kan realiseres. Nedenfor foreslår vi en liste med 7 kriterier som kan anvendes ved slik prioritering. Vi understreker at listen ikke er uttømmende og at alle kriteriene ikke vil være like anvendbare innenfor alle fagområder. Kriteriene er ikke ment å skulle brukes til tallmessig skåring og rangering mellom databasene og infrastruktur tjenestene, men mer som en huskeliste for skjønnsmessig vurdering av samfunnsverdi og behov for ressursinnsats. Noen av kriteriene gjelder databasenes kvalitet og innhold. Andre kriterier omfatter deres bruksverdi, tilgjengelighet og funksjonalitet.

Det er viktig at Forskningsrådet tar de ulike fagområdene med på råd når prioriteringen skal foretas, f.eks. gjennom de nasjonale fakultetsmøtene. Samtidig minner vi om at delmålet formulert ovenfor om at alle fagområder skal sikres den samme tilgangen til sentrale data må gjelde uavkortet. Det er derfor ikke formålstjenlig å anvende kriterier for prioritering mellom

samfunnsområder så fremt det ikke er uforholdmessig store forskjeller på tilstand og status for databaser m.v. mellom disse.

Kriterium 1: Ressursen inneholder data som dekker et langt tidsrom

Lange tidsserier som er framkommet med samme og anerkjente metodikk over lengre tidsrom, har gjerne stor utsagnskraft og er derfor en viktig datakilde for forskning som studerer endring over lengre tidsrom. Det bør imidlertid tas i betraktning at tidsdimensjonen ikke er like sentral i all forskning.

Kriterium 2: Ressursen inneholder data som er representative for landet eller for regioner

Representativitet øker datakvalitet og relevans, enten det er en fullstendig landsdekkende registrering eller det er brukt en samplingsmetodikk som gir representative tall.

Kriterium 3: Ressursen inneholder data som er spesielt viktige som grunnlag for analyse, kunnskapsproduksjon, politikkutforming og forvaltning

Nøkkeldata på sentrale politikkområder er viktige kilder og grunnlag for forskningens rolle som uavhengig kunnskapskilde for samfunnet og dermed viktige for politikkutforming og forvaltning.

Kriterium 4: Ressursen inneholder data som har en nøkkelfunksjon innen fagområdet

Data som innen fagområdet er grunnleggende, nyskapende og derfor spesielt viktige, gjerne med opprinnelse i grunnforskning er viktig å ta vare på i databaser som har en varighet og dataforvaltning ut over forskningsprosjektets tidsramme.

Kriterium 5: Ressursen inneholder data og funksjonalitet som utgjør en felles og sentral infrastruktur innen fagområdet eller mellom fagområder

Felles data med koblingsnøkler som derfor brukes av mange forskningsinstitusjoner innen fagområdet eller mellom fagområder er viktige, særlig hvis databasene også har funksjoner som gjør delingen mulig i sanntid.

Kriterium 6: Ressursen inneholder data som utgjør grunnlaget for å oppfylle Norges internasjonale forpliktelser

Databaser som inneholder parametere som Norge måles på i forhold til oppfølgingen av internasjonale konvensjoner og avtaler er i utgangspunktet av stor samfunnsmessig verdi.

Kriterium 7: Ressursen inneholder data som er fritt tilgjengelige gjennom brukervennlige grensesnitt

Dataenes tilgjengelighet for andre forskere, forskningsmiljøer og forskningsinstitusjoner nasjonalt og internasjonalt er avgjørende for å oppnå hensikten med en strategi for forskningsinfrastruktur.

5 Handlingsplan – prioriterte oppgaver

Med utgangspunkt i strategien skissert i kapittel 4 foreslår vi i dette kapitlet en kort handlingsplan. Planen må i neste omgang tilpasses og konkretiseres i forhold til de ulike fagområdene, der behovene for tiltak kan være forskjellige, jf. kapittel 3 og vedlegg 1, og ikke minst må den tilpasses den totale strategien for utvikling av en formålstjenlig forskningsinfrastruktur i Norge, der innspill fra andre utvalg også skal tas hensyn til.

5.1 Tiltak

Som det eneste organet som kan ta et overordnet nasjonalt ansvar på feltet, som premissgiver, som pådriver og organisator, og som finansieringskilde i samarbeid med andre aktører, må Forskningsrådet:

- Sammen med forskningsmiljøene bidra til utvikling og permanent finansiering av formålstjenlige datatjenester for forskningen innenfor alle fagområder i tråd med målene satt opp i strategien i kapittel 4. På fagområder der det foreligger konkrete og omforente behov og planer, bør disse søkes iverksatt så snart som mulig. Der dette ennå ikke er gjort, må det iverksettes fagspesifikke kartlegginger av situasjonen. Resultatet må bli et overordnet veikart for hele feltet og en konkret handlingsplan for hvert fagområde.
- Bidra i et krafttak for digitalisering av samlinger, av papirbaserte data og av materiale i eksisterende biobanker.
- Stille som vilkår for økonomisk støtte til prosjekter, programmer og større satsinger at resultater og data arkiveres, dokumenteres og forvaltes slik at det blir mulig å gjenbruke data og gi innsyn i forskningens resultater.
- Finansierte lagring og formidling av data for gjenbruk.
- Finansierte opprydding/videreføring av databaser gjennom egne programmer.
- Kreve ryddige kontraktsforhold når det gjelder eiendomsrett til dataene og utforming av regelverk og prinsipper for datadeling mellom aktørene.
- Ta initiativ til forenkling av vilkårene for utlevering av data til forskning, og til effektiv saksbehandling for slik utlevering. Det gjelder lover og regler så vel som institusjonell praksis og kultur.
- Ta initiativ for å sikre at datainnsamling som foretas av offentlige og private institusjoner og bedrifter tilpasses forskningens behov, bl.a. gjennom felles identifikasjonssystemer.
- Koordinere arbeidet for å sikre en fullstendig og enhetlig dokumentasjon av og god informasjon om datakilder, dvs. metadata, som kan brukes i forskning.
- Bidra til internasjonalisering av norsk forskning ved å sikre langsiktig finansiering av datainnsamling knyttet til internasjonale prosjekter.
- Representere Norge internasjonalt for å sikre at norske forskere kan dra nytte av det infrastrukturarbeidet som gjøres internasjonalt. I de nærmeste årene peker ESFRI-samarbeidet seg ut som et naturlig satsingsfelt, faglig så vel som økonomisk.

Utvalget foreslår at det oppnevnes et *Forskningsrådets databaseutvalg* som et rådgivende organ for Forskningsrådet, sammensatt av representanter fra alle fagområder, fra universiteter og høyskoler og de institusjonene som arbeider med databaser og forskningsinfrastruktur. Dette utvalget må ha et helhetlig ansvar nasjonalt.

Forskningsrådets databaseutvalg skal være et faglig bindeledd mellom forskningsmiljøene og Forskningsrådet i alle spørsmål som angår utvikling og vedlikehold av en best mulig nasjonal forskningsinfrastruktur.

Skal et slikt rådgivende organ fungere, må det ha en hensiktsmessig organisasjonsstruktur i Forskningsrådet å forholde seg til. Forskningsrådet må stille med sekretariat for utvalget forankret i Divisjon for vitenskap.

5.2 Momenter til et mandat for Forskningsrådets databaseutvalg

Gi råd til Forskningsrådet om

- Utforming, implementering og rullering av strategi
- Prioritering av økonomiske ressurser
- Evaluering av forslag og initiativ fra forskningsmiljøene om tiltak
- Evaluering av analyser og utredninger bestilt av NFR

Dersom Forskningsrådet oppretter et Forskningsrådets utvalg for infrastruktur, forutsetter vi at dette utvalget overtar ansvaret for de funksjonene som vi har forutsatt at databaseutvalget skal ha.

Litteraturliste

- Aukrust, Knut og Bjarne Hodne (red.) (1998): Fra skuff til skjerm. Om universitetenes databaser for språk og kultur, Universitetsforlaget
- Det Strategiske Forskningsråd (2005): Fremtidens forskningsinfrastruktur - kortlægning af behov og forslag til strategi, Baggrundsrapport
- European Strategy Forum on Research Infrastructures (ESFRI) (2006): European Roadmap for research Infrastructures. Report 2006
- Godøy, Øystein, Helge Sagen, Stein Tronstad og Aasmund Fahre Vik (2007): National and international dataflow during the International Polar Year, 24 April 2007
- Hægeland, T. et. al. (2003): Infrastruktur i samfunnsvitenskap - Om tilgang til registerdata for forskningsformål. Rapport til Norges forskningsråd fra en utredningsgruppe, 20. august 2003
- Kristiansen, Nina og Christian-Emil Ore (1998): Dokumentasjonsprosjektet – Sluttrapport 1992-1997, Universitetet i Oslo
- Norges forskningsråd (2003): Lange tidsserier for miljøovervåkning og forskning – Rapport nr.1 - Viktige klimadaserier
- Norges forskningsråd (2003): Lange tidsserier for miljøovervåkning og forskning – Rapport nr.2 - Viktige terrestriske og limniske daserier
- Norges forskningsråd (2004): Lange tidsserier for miljøovervåkning og forskning – Rapport nr.3 - Viktige marine daserier
- Norges forskningsråd (2004): Behov for vitenskapelig utstyr, databaser, samlinger av vitenskapelig materiale og annen infrastruktur, Utredning gjennomført av Norges forskningsråd som innspill til forskningsmeldingen, Oslo 18.5.2004
- Norges forskningsråd (2007): Forskningens fellesformue forvitrer. Lange tidsserier for miljøovervåkning og forskning
- Norges offentlige utredninger (1997): Tilgang til helseregistre, Ot.prp.nr.5 (1999-2000). Om lov om helseregistre og behandling av helseopplysninger (helseregisterloven), NOU 1997:26
- Norges offentlige utredninger (2005): God helse, bedre forskning. Lov om medisinsk og helsefaglig forskning, som involverer mennesker, humant biologisk materiale og helseopplysninger (helseforskningsloven), NOU 2005:1
- Norges offentlige utredninger (2006): Kunnskap for fellesskapet. Universitetsmuseenes utfordringer, NOU 2006:8

OECD (2007): OECD Principles and Guidelines for Access to Research Data From Public Funding, Paris

Oftedal, L. et. al. (2006): Innstilling fra Arbeidsgruppen for museumsdatabasene, Blindern, 29.4.2006

Ore, Christian-Emil og John Herstad (2004): Infrastrukturtiltak for humanistisk forskning, notat 27.4 2004

Ore, Christian-Emil og Oddrun Ragnsæter (2007): Sluttrapport for Museumsprosjektet, foreløpig versjon, 7. mai 2007

Stortingsmelding nr. 20 (2004-2005): Vilje til forskning

Vetenskapsrådet (2006): Vetenskapsrådet guide til infrastrukturen. En kartlegging och rekommendationer for svenska forskares tilgang til forskningsinfrastruktur på lång sikt, Vetenskapsrådets rapportserie 14:2006

Vedlegg I: Status og perspektiver for fagområdene

Hensikten med vedlegget er å gi en oversikt over status og viktige utfordringer på ulike fagområder. På grunn av den korte tiden og utvalgets sammensetning har det ikke vært mulig å dekke alle fagområder og delområder innen hvert fagområde. Som påpekt i handlingsplanen er det å skaffe oversikt over tilgjengelige ressurser, over behov og utfordringer en viktig oppgave i det videre arbeidet med forskningsinfrastruktur.

Medisin og helseområdet

Det er to sentrale forskningsområder som særlig utnytter databaser og samlinger. Det er på den ene siden klinisk forskning som følger pasientgrupper for å oppnå bedre diagnostikk og behandling, og det er den befolkningsbaserte årsaksforskningen (analytisk epidemiologi) som har som formål å forebygge utvikling av sykdom.

I et internasjonalt perspektiv har Norge, sammen med de øvrige nordiske land, særlige fortrinn. Dette skyldes det offentlige helsevesen, fødselsnummersystemet, og den oversiktlige og godt utdannede befolkningen som tradisjonelt har en positiv innstilling til medisinsk forskning. Dette komparative fortrinnet har Norge bare i liten grad utnyttet forskningsmessig.

For å forstå årsaker til sykdom og forløp av sykdom i moderne medisin er biobanker en nødvendighet. Biobanker er samlinger av humant biologisk materiale som tillater analyser av eksponeringer (infeksjoner, kostholds faktorer, miljøgifter osv.), analyser av gener og analyser av tidlige og sene tegn til sykdomsutvikling gjennom undersøkelse av genekspresjon og mønstre av proteiner og metabolitter. Biobanker er svært dyre å opprette og drifte. Full automatisering av prosesseringen, lagringen og uttakene krever millioninvesteringer for selv små forskningsbiobanker.

Biobanker kan aldri fullt ut digitaliseres, selv om man kan komme langt med å gjøre genotyping, konsentrasjonsmålinger og beskrivelser av makromolekylers struktur, og deretter registrere resultatene. Det er en rask teknisk utvikling innen molekylær medisin som innebærer at det er nødvendig å ta vare på det opprinnelige materialet for å kunne nyttiggjøre seg de nye teknikker. Kliniske biobanker, for eksempel patologisk-anatomiske samlinger, vil ha lagret resultatet av den undersøkelsen som var aktuell i den diagnostiske utredningen, men det er vesentlig at man kan gå tilbake til det opprinnelige preparat i for å besvare forskningsspørsmål.

Eksposeringene (de potensielle årsaksfaktorene) kan også hentes fra databaser (for eksempel luftforurensning, utdanning, yrke) og helseutfall kan hentes fra helseregistre eller fra databaser i primær- eller sekundærhelsetjenesten. For å gjøre analytiske studier er tidsdimensjonen vesentlig. Man må kunne følge den enkelte person gjennom ulike livsfaser. Eksposeringene som igangsetter kreftutvikling kan ligge 20-30 år forut for sykdomsdebut. Hvis dette skal fanges opp må man ha omfattende, prospektive, longitudinelle observasjonsundersøkelser (kohortstudier). For å forstå pasientkarrierer må man kunne følge en person gjennom ulike behandlingstilstander og flyttemønstre. Spesielt for sjeldne sykdommer, må hele den norske befolkning være bakgrunns-populasjon. Det tilsier at en nasjonal infrastruktur må benyttes for å bringe sammen data fra ulike kilder og koble dem sammen ved hjelp av fødselsnumre. Forskere trenger imidlertid ikke tilgang til personidentifiserbare data.

Aktuelle, sentrale helseregistre er Kreftregisteret, Fødselsregisteret, Pasientregisteret, Meldesystemet for infeksjonssykdommer, Vaksinasjonsregisteret (SYSVAK), Reseptregisteret og Dødsårsaksregisteret. Flere av disse registrene er i dårlig forfatning når det gjelder kvalitet og dokumentasjon, og det er nødvendig med et krafttak for å heve kvaliteten.

Videre er det nødvendig med en sterk, sentral serviceinstitusjon der forskerne kan henvende seg for å få tilgang til data. Sett fra en forskers side er den ideelle situasjon at det fantes en web-portal der man kunne få full informasjon om alle tilgjengelige databaser og biobanker med beskrivelser av variablene i databasene og deres validitet og typen og kvaliteten av det biologiske materialet. Videre kunne det være informasjon om hva dataene tidligere er brukt til med publikasjons-henvisninger, og kunnskap om hvilke søknadsprosedyrer, formelle regler og kostnader som er forbundet med tilgang. Den sentrale institusjon skal kunne sammenstille de data og utlevere de biologiske prøver som er nødvendig for å kunne besvare problemstillingen.

I dag er Nasjonalt folkehelseinstitutt pålagt å sørge for å gjøre sine data (befolkningsundersøkelser og nasjonale helseregistre) tilgjengelig for forskning, og har bygget opp juridisk og teknisk kompetanse i den anledning. Men langt de fleste kliniske data finnes relativt utilgjengelig i primærhelsetjenesten og de ulike helseforetak. For å virkeliggjøre forskningen må disse data gjøres tilgjengelig gjennom moderne IT-nettverk. Det vil si at rutinedata om diagnoser, behandlinger og mer detaljerte data fra kvalitetsregistre må inngå. Nasjonalt folkehelseinstitutt kan utvikle sin kompetanse videre slik at det også kan ha aksess til kliniske data gjennom distribuerte databaseløsninger. Det innebærer ikke at alle data skal samles et sted eller at alle databaser må ha samme programspråk eller struktur, men at de etterspurte data på en fleksibel måte den kan settes sammen i en sentral institusjon når det er behov for det. En egen profesjonell Divisjon for databaser og biobanker ved Nasjonalt folkehelseinstitutt som på denne måten gir service til alle forskere vil være for medisinsk forskning hva NSD er for samfunnsvitenskap.

Spørsmål om eierskap og rettighetene til å kunne forske på helsedata er avklart når det gjelder de sentrale helseregistrene gjennom helseregisterloven. Videre er det gitt klare føringer for å lette helsefaglig forskning gjennom odelstingsproposisjonen som ble fremlagt i juni i år.

Samfunnsvitenskap⁴

“Det har lenge vært høy bevissthet om infrastrukturspørsmål for samfunnsforskning i Norge, sammenlignet med andre land. Gjennom de siste 30–40 år har Norge opparbeidet en sterk posisjon når det gjelder innholdsmessige, økonomiske, juridiske og teknologiske forhold rundt tilgang til, organisering og forvaltning av samfunnsvitenskapelige forskningsdata. Dette har gitt gode muligheter for å studere det norske samfunnet langsiktig og komparativt. Statistisk sentralbyrå og Norsk samfunnsvitenskapelig datatjeneste har spilt en avgjørende rolle for å sikre et bredt datagrunnlag for norsk samfunnsvitenskapelig forskning. Forskningsmiljøene på sin side har lagt ressurser i infrastrukturiltak for å utnytte tilbudene etter sine egne behov. Samlet har dette medført at man i Norge har klart å se statistikkbehov, offentlige utredningsbehov og forskningens behov for empirisk grunnlag i sammenheng”. (Stortingsmelding 20, Vilje til forskning, s. 62)

Dagens infrastruktur

Utviklingen av den samfunnsvitenskapelige infrastrukturen i Norge har vært basert på en stor og bevisst satsing fra Norges forskningsråd. Den bygger i stor grad på et forpliktende samarbeid mellom forskningsmiljøene representert ved NSD og sentrale dataprodusenter, særlig Statistisk sentralbyrå (SSB). NSD har for forskningsmiljøene spilt en viktig rolle som koordinator, datatilrettelegger, dataformidler og dataarkiv.

Det forpliktende samarbeidet mellom forskningsmiljøene og Datatilsynet er også et viktig element i den infrastrukturen som er etablert for norsk forskning. Samarbeidet ble etablert da personregisterloven trådte i kraft i 1980. Den langsiktige oppbyggingen av NSD som kompetansesenter for personvern og personvernombud for norsk forskning har vært avgjørende for norsk empirisk samfunnsforskningens gode rammevilkår.

For å sikre en fortsatt god tilgang til data er det nødvendig at infrastrukturtenester og serviceapparatet rundt data bygges videre ut i tett samarbeid og i forståelse med forskersamfunnet og forskningsfinansierende myndigheter. Dette apparatet må ha som sentrale oppgaver å etablere oversikter og metadatabaser med datadokumentasjon innrettet mot forskning på alle nivåer – og informasjon om koblingsmuligheter og tidligere bruk av data. Bedre tilgjengelighet innebærer også at det må avsettes ressurser til formidling og direkte betjening av forskerne og studentene. For å få til alt dette kreves det kontinuerlig innsats for å løse utfordringer knyttet til institusjonelle, økonomiske, juridiske og teknologiske barrierer.

Forskningsgenererte data

Data fra spørreundersøkelser, ofte i kombinasjon med registerdata, er en viktig kilde for samfunnsforskningen. Slike undersøkelser, særlig der de skal kombineres med registerdata, stiller store krav til utformingen av det materialet som sendes ut og til de tillatelse som må innhentes. I den forbindelse er det en utfordring å lære opp/bevisstgjøre både forskere og de som samler inn data på oppdrag fra forskningsprosjekter, om at det er innholdet i informasjonsskriv og ev. samtykkeformular som legger de viktigste føringene på videre bruk av data. Veiledningstjenester som personvernombudet hos NSD, er viktig i denne sammenhengen.

⁴ I denne teksten har utvalget ikke lyktes med å komme frem til et felles syn på status og perspektiver for prosessgenererte data som en del av forskningsinfrastrukturen på det samfunnsvitenskapelige området. I teksten presenteres derfor to versjoner, den første forfattet av Vigdis Namtvedt Kvalheim og den andre av Johan-Kristian Tønder.

Forskningsrådet bør vurdere om det er behov for å satse mer systematisk og strategisk i forhold til å bygge opp datagrunnlaget på ulike samfunnsområder gjennom særskilt tilrettelegging av data (tematiske satsninger) som omfatter både forskningsdata og registerdata.

Arkivering

Norske forskningsprosjekter produserer mye data gjennom egne innsamlinger, koblinger og spesifiserte tilrettelegginger. Det er klart et behov for å øke innsatsen knyttet til datafangst og arkivering av slike prosjektdata. Arkivering for gjenbruk er viktig for å ta vare på og på best mulig måte utnytte de ressurser samfunnet har investert i forskning og dataproduksjon. Arkivering er videre en forutsetning for å muliggjøre etterprøving og innsyn i forskningens resultater.

I dag finnes krav til avlevering og arkivering av data hos NSD for forskningsprosjekter innenfor samfunnsfag, humaniora, medisin og helse og miljø - og utviklingsfag som er finansiert av Norges forskningsråd. Det er imidlertid nødvendig å rette fokus på hva som kan gjøre for å sikre at arkiveringsplikten blir oppfylt i tråd med kontraktsvilkårene, og styrker arbeidet med datafangst, dokumentasjon og informasjon om data. Datafangstarbeidet forutsetter at man har prosedyrer som gir informasjon om hvilke prosjekter som produserer data og som sikrer at dataene blir avlevert for arkivering. Ved siden av det som kan oppnås gjennom felles regler og avtaler, er det nødvendig å sette av langsiktige ressurser for å kontinuerlig arbeide for at data blir avlevert, tilrettelagt og vedlikeholdt på en måte som muliggjør senere gjenbruk. De viktigste stikkordene for åpen og fri (demokratisk) tilgang til data er (prosjektekstern) arkivering, dokumentasjon og informasjonstilgang. Dette kan bare sikres ved at data overføres til en sentral arkivinstusjon som har den nødvendige permanens, kompetanse og serviceapparat til å sikre forsvarlig oppbevaring og effektiv tilgang til data.

Dokumentasjon

For at infrastruktur og datatjenester skal fungere best mulig må man se på tjenestene og data-tilbudet som et kollektivt gode for forskningsmiljøene. Gjennom koordinert innsats får en klare stordriftsfordeler både faglig og økonomisk for norsk forskning.

En nasjonal koordinering, med fokus på data som en felles ressurs, krever at man får utviklet og etablert metoder, dokumentasjon og standarder som muliggjør sammenlignende analyser over tid, nivåer osv. Det må derfor avsettes midler til dokumentasjon av data og metadata og til oversettelse til engelsk. Med de nye teknologiske mulighetene vil norske forskere få en mye bedre tilgang til data for komparative analyser, analyser av tidsserier og forløp. Dette krever imidlertid at dokumentasjonen legges opp på en slik måte at disse mulighetene ivaretas, gjennom bl.a. felles internasjonale dokumentasjonsstandarder. På det samfunnsvitenskapelige området er Norge allerede aktivt involvert i slikt arbeid, gjennom NSDs deltakelse i Data Documentation Initiative (DDI Alliance). DDI er et samarbeid mellom institusjoner, særlig i USA og Europa med formål å utvikle en ny standard for dokumentasjon av samfunnsvitenskapelige numeriske data.

Internasjonalisering av norsk forskning

I Forskningsmeldingen er det pekt på at man har unike dataserier innen samfunnsfagene og at det skal legges til rette for bedret utnyttelse av slike data både i nasjonalt og internasjonalt forsknings-samarbeid. Det pekes spesielt på at norske databaser skal videreutvikles og at det skal arbeides med å etablere datatilbud som åpner for sammenlikninger med andre land.

Dette er viktig ut fra særlig tre hensyn:

- For å sikre at data om Norge kan inngå i tverrnasjonale analyser fordi gode, veldokumenterte data om Norge foreligger klar til bruk.
- For å sikre norske forskere tilgang til utenlandske og internasjonale data.

- For å sikre at norske forskere kan delta i internasjonale prosjekter hvor de har norske data med som “medgift”.

Innenfor samfunnsvitenskapene er det konkret pekt på behovet for å delta regelmessig i omfattende internasjonale intervjuundersøkelser som European Social Survey (ESS), International Social Survey Programme (ISSP), World Values Survey (WVS) og for at Norge supplerer de runder av Eurobarometer som EU selv finansierer i Norge. Andre undersøkelser kunne vært nevnt, f. eks. European Union - Survey on Income and Living Conditions (EU-SILC), Community Innovation Survey (CIS).

Det er viktig å sørge for langsiktig finansiering av slike aktiviteter. Norges forskningsråd må følge opp dette i sin infrastruktursatsing.

Videre er det viktig å bygge opp nye datatilbud basert på spesialkompetanse i Norge. En viktig forskningsressurs kunne være en ny europeisk regionaldatabase basert på erfaringer med NSDs Kommunedatabase. Denne ville kunne gi europeiske forskere et meget viktig supplement til andre datakilder på et område hvor Norge allerede er i første rekke.

Det eksisterer i dag mange samarbeidsprosjekter mellom norske og utenlandske forskere på norske registerdata. Den utenlandske forskeren kommer til Norge og arbeider her en tid med samme tilgang til data som sine norske kolleger. Dersom den utenlandske forskeren reiser til hjemlandet, kan vedkommende ikke fortsette forskningen med direkte tilgang til mikrodata levert av SSB, men må ha med seg et sett detaljerte tabeller som vedkommende kan arbeide videre med. Dette er ofte tilstrekkelig til å gi god forskning, og tilgangen på gode mikrodata i Norge gjør at utenlandske forskere ønsker samarbeid med norske kolleger selv med denne restriksjonen på tilgangen til mikrodata. Men det er klart at slike ordninger legger en viss begrensning på forskernes fleksibilitet og samarbeid. Forskningsrådet bør derfor ta opp med SSB om en kan finne ordninger innenfor regelverket som gjør det mulig for utenlandske forskere å få tilgang til norske mikrodata selv om de arbeider i utlandet.

En annen begrensning i dagens regelverk er muligheten for å opprette anonymiserte ”public use files” (PUF) på SSBs registerdata. Dette er kanskje en større begrensning i det internasjonale forskningssamarbeidet for enkelte fagområder enn tilgangen til mikrodata nevnt ovenfor. Forskningsrådet bør derfor gi høy prioritet til arbeidet med å finne en løsning på dette i samarbeid med SSB.

Prosessgenererte data – versjon 1

Prosessgenererte data er data som produseres av forvaltningen om forvaltningen selv, om saksområder som forvaltningen dekker og eventuelt om de personer og instanser som berøres av forvaltningens saksbehandling og vedtak. Bruken av slike prosessgenererte data til forskningsformål er ennå i sin begynnelse, mye fordi utnyttelse av denne type data er krevende for produsent, formidler og forsker. utfordringene kan være tekniske, metodiske, analytiske og personvernmessige, i tillegg til lang saksbehandlingstid og pris knyttet til datatilgang. Det er derfor viktig at Forskningsrådet arbeider for å løse utfordringer knyttet til faglige, økonomiske, institusjonelle og legale barrierer.

Regelverk og institusjonell praksis

Det er et problem for internasjonalt forskningssamarbeid at registerdata ikke kan sendes ut av landet og at SSB ikke kan opprette "public files" på grunnlag av registerdata. Det er også et problem at norske data ikke kan distribueres som en del av internasjonale datasett fordi de inneholder enkelte demografiske variabler fra offentlige registre. Når man sammenligner med praksis på helseregister – og biobankområdet, der forskningssamarbeid og utveksling av data er

mulig - også på tvers av landegrense, er det vanskelig å forstå begrunnelsen for en slik restriktiv praksis.

Forskningsrådet må ta initiativ overfor SSB og andre offentlige dataeiere for å finne fram til mer effektive ordninger for formidling av data innenfor gjeldende regelverk. På områder der lovverk eller praktiseringen av dette er til hinder for tilgang og en hensiktsmessig organisering av data-tjenester for forskning, må Forskningsrådet være en pådriver for å endre dette.

Organisering av effektive datatjenester

For at prosessproduserte data skal kunne utnyttes på en samfunns – og forskningsmessig effektiv måte er det helt nødvendig å bygge videre på de ordninger, systemer og institusjoner som allerede finnes og som muliggjør forskning innenfor gjeldende regelverk. Miljøene har behov for et bredt anlagt apparat som kan bistå å legge til rette for alle sider knyttet til bruk av data i forskningsprosessen, fra bistand, hjelp og veiledning i arbeidet med søknad om tilgang til tilrettelegging, dokumentasjon, utlevering og videre oppfølging. For at prosessen skal gå mest mulig effektivt er det viktig at forskeren eller studenten må forholde seg til færrest mulig instanser og at det er kompetanse og høy kvalitet knyttet til alle deltjenestene.

Erfaringer har også vist at oppbygging av fellestjenester på registerdatasiden etter KIRUT-modellen ⁵ (utvalgsdatabaser), med NSD som tilrettelegger og formidler av avidentifiserte og anonymiserte data, er en effektiv måte å løse databehovene til store deler av forskningsmiljøene. Slike løsninger vil gi verdifulle alternativ til SSBs heldekkende registre. Tilsvarende organisering og utvikling av nye datatjenester er foreslått på det medisinske området ⁶. NSD og SSB er i dialog om ulike muligheter og løsninger for tilrettelegging og formidling av registerdata til forskningsformål.

Forskningsrådets strategi må fortsatt være at data tilrettelagt med offentlige midler skal gjøres tilgjengelig for forskning, bredt, raskt, gratis og sikkert gjennom NSD eller institusjoner som kan yte tilsvarende service. Dersom forskeren eller doktorgrads- og masterstudenten skal kjøpe data og service vil mange studenter og forskere ikke kunne benytte seg av de datatilbud som etableres. Den frie forskningen er avhengig av at det bygges opp datatilbud som er åpen for alle. Jo større barrierer det er når det gjelder tilgang til data (pris, regelverk og byråkrati) jo større er sannsynligheten for at forskningen monopoliseres ressurssterke forskergrupper, noe som vil bidra til å redusere både mangfold og kvalitet.

Prosessgenererte data – versjon 2

Forvaltningen av de "goder og plikter" (ordninger) som det offentlige tildeler og pålegger så vel fysiske som juridiske personer (enheter) i Norge, forutsetter at det offentlige må samle inn en rekke informasjoner om enhetene og deres aktivitet, og må fatte beslutninger som berører enhetene. I løpet av de siste vel 40 år har det meste av denne informasjonen blitt digitalisert. Ettersom flere offentlige etater har behov for opplysninger for de samme enhetene, har det blitt utviklet felles identifikasjonssystemer for enhetene i form av personnummer, bedriftsnummer og adresser, og sentrale registre som forvalter dette identifikasjonssystemet på en enhetlig måte er etablert. Saksbehandlingen i forvaltningen krever dessuten tidfesting av aktiviteter som rapporteres og de beslutninger som gjøres. Data som produseres som en følge av slike prosesser, kalles ofte for "prosessgenererte data".

⁵ KIRUT – Klientstrømmer inn i, rundt i og ut av trygdesystemet, er en database som inneholder opplysninger om hvordan enkeltindivider beveger seg mellom arbeid, de ulike trygdeordningene og sosialhjelp. Data ble levert fra Arbeidsdirektoratet, Rikstrygdeverket og Statistisk sentralbyrå. Basen inneholder opplysninger for et representativt utvalg på 10 % av befolkningen i Norge mellom 16 og 67 år, for årene 1989 til 1996. Fødselsnummer er erstattet av pseudonym.

⁶ NOU 1997:26 Tilgang til helseregistre, Ot.prp.nr.5 (1999-2000) Om lov om helseregistre og behandling av helseopplysninger (helseregisterloven).

Lovene som regulerer de offentlige organers forvaltning av ordningene, gir i de fleste tilfeller adgang til bruk av opplysningene som genereres i statistikk og forskning. Statistikkloven gir Statistisk sentralbyrå en særskilt rett til å utnytte den informasjonen som offentlige institusjoner samler inn, til offisiell statistikk. Statistikkloven åpner også for at mikrodata samlet inn til offisiell statistikk på bestemte vilkår kan brukes i forskning. Gjennom tilrettelegging av dataene for statistiske formål, blir dataene i de fleste tilfeller lettere tilgjengelige for forskning enn om den enkelte forsker skulle kontakte den enkelte forvaltningsorgan og få levert data fra disse som hun/han så skulle sette sammen. SSB har også utviklet både tjenesteyting og dokumentasjon, og er den viktigste leverandøren av mikrodata fra offentlig forvaltning til norsk samfunnsforskning. Også samfunnsmedisinsk og historisk forskning benytter seg av SSBs mikrodatatjenester.

Bruken av prosessgenererte data var lenge begrenset til visse miljøer og fagområder. Dette skyldes bl.a. at kunnskap om kildene og muligheten for å få tilgang til slike data har vært lite kjent, at kompetansen til å utnytte data i forskning har vært mangelfull, og at kostnadene ved å få tilgang til data og bearbeidingen av dem har vært for høye. Det har også vært en viss usikkerhet omkring regelverket for tilgang og bruk av prosessgenererte data. Forskningsrådet og SSB finansierte i 2005 – 2006 et prosjekt om "Informasjon om mikrodata" som har bedret situasjonen når det gjelder dokumentasjon av kildene og informasjon om mulighetene for tilgang og bruk, men det står nok fortsatt en del hindre igjen før mulighetene er fullt utnyttet. Både SSB og tjenesteytende institusjoner som NSD må arbeide videre for å bygge ned slike hindre. Nedenfor tar vi opp noen utfordringer og skisserer noen veier til løsninger.

Dataenes relevans og kvalitet i forskning

Som regel vil de administrative behov for informasjon og kvaliteten på denne informasjonen falle sammen med statistikernes og forskernes, men i noen tilfeller vil det være ønske om endringer i definisjoner mv. eller nyttig med andre opplysninger. Statistikkloven gir SSB retten til å påvirke innhold og omfang av forvaltningens registre. I samarbeid med SSB kan forskerne påvirke innholdet i forvaltningens datainnsamling og dermed innholdet i de mikrodata som er tilgjengelig for forskningen.

Mikrodata tilrettelagt for forskning

Som nevnt er dataene i SSB til en viss grad tilrettelagt for forskning gjennom bearbeidingen for statistiske formål. I tillegg til statistiske grunndata foreligger det i SSB en del "mellomprodukter" for analytiske formål. Eksempler på det er forløpsdatabaser for eksempel sysselsetting og utdanning. Det pågår en diskusjon i SSB om en skal fortsette med utviklingen av slike "mellomprodukter", om en skal satse på et sett av ferdig utviklede og oppdaterte mikrofiler, eller om en skal tilrettelegge grunndataene på en måte som gir muligheter for alle typer uttak av mikrofiler til statistikk og forskning direkte fra grunndata ved hjelp av effektive uttakssystemer. Spørsmål om å tilby forskerne data basert på et utvalg av enheter (personer, bedrifter) blir også diskutert. Denne typen utvalgsdatabaser kan løse databehovene til noen av forskningsmiljøene, og vil være et supplement til Statistisk sentralbyrås heldekkende registerdatatjenester. Dette er diskusjoner som forskerne bør engasjere seg i.

Vilkår for tilgang til forskning

SSBs adgang til utlevering av mikrodata til forskning skjer som nevnt på visse vilkår. Et av vilkårene er at det forvaltningsorganet som har levert data til SSB for offisiell statistikk skal godkjenne en eventuell utlevering av slike mikrodata til forskning. Forskeren må altså søke forvaltningsorganet om å få dataene fra SSB. Dette krever tid både til å skrive søknad og til å vente på svar fra forvaltningsorganet. Det siste vil ofte være avhengig av forvaltningsorganets saksbehandlingskapasitet, og den kan i mange tilfeller være begrenset. Forskningsrådet bør ta opp med SSB spørsmålet om det er mulig å finne fram til mer effektive ordninger innenfor gjeldende regelverk, eventuelt om regelverket kan endres.

Arkivering av prosessgenererte mikrodata brukt til forskning

Et annet vilkår for at forskeren skal få adgang til forvaltningens data gjennom SSB, er at mikrodataene bare skal brukes til det prosjektet forskeren søker forvaltningsorganet om, og at dataene slettes når prosjektet er ferdig. Dette gjelder også når registerdata er koblet mot data fra spørreundersøkelser. Det betyr at forskningsprosjekter basert på mikrodata fra SSB ikke vil kunne etterleve Forskningsrådets vilkår for støtte, nemlig at de data forskningen er basert på skal arkiveres med sikte på at forskningsresultatet skal kunne kontrolleres og at dataene skal kunne brukes til annen forskning i ettertid. Forskningsrådet må ta opp med SSB mulig å komme fram til endringer i regelverket som kan gi grunnlag for arkivering og hensiktsmessig gjenbruk av mikrodata levert til forskning.

Finansiering av datauttak

Et annet problem når det gjelder tilgang til registerdata er prisen på data. SSB har som prinsipp at alle uttak av mikrodata må dekke de marginalkostnadene som uttaket fører med seg. Avhengig av hvordan databehovet passer med SSBs tilrettelegging av de aktuelle data, vil prisen kunne variere. Forskernes evne til å betale vil også kunne variere, og en kan komme til å oppleve at nyttig forskning ikke blir gjennomført fordi forskeren ikke har midler til uttak av data. Dette vil særlig gjelde doktorgrads- og mastergradsstudenter, men kan også ramme den frie forskningens tilgang til data. Forskningsrådet bør utrede i hvor stor grad det skal gå inn i finansieringen av datauttak, etter hvilke prinsipper finansieringen skal skje, og hvordan den skal administreres.

Kompetanse på mikrodataforskning

De fleste forskerne som bruker SSBs mikrodata tar direkte kontakt med SSB når de skal bestille data. En mindre del av brukerne blir henvist til SSB via NSD. I sitt samarbeid med SSB om å finne løsninger på de utfordringene som er nevnt ovenfor, er det viktig at den gruppen av forskere som står for den alt overveiende del av forskningen på prosessgenererte mikrodata i Norge, blir sterkt representert i tillegg til NSD og Forskningsrådets administrasjon.

Humaniora

Humaniora er som alle andre vitenskapelige områder sammensatt av flere forskjellige disipliner. Her har vi gjort en inndeling i språkforskning og historisk orientert forskning. Det er fundamentalt sett ikke noe vanntett skott mellom historie og samfunnsfagene, jf avsnittet om historiske data som f.eks. folkeregisterdata ovenfor.

Språkrelatert forskning - Tekst og taledata

Digitalisering åpner for enorme muligheter innenfor all forskning på tekst, det være seg fra et innholdsperspektiv eller et formperspektiv.

Det finnes en god del digitale språkressurser i dag, og det finnes sterke fagmiljøer ved universitetene i Oslo, Bergen og Trondheim som både utvikler og bruker slike ressurser. Men behovet er langt fra dekket, jf. punktene om Norsk språkbank og et digitalt kilderegister nedenfor. Eksempler på hva som finnes på området i dag følger nedenfor. (Skillet mellom fokus på form og innhold må ikke forstås slik at et korpus ikke kan være nyttig *både* fra et form- og innholdsperspektiv.)

- Tekstkorpus med fokus på innhold, som f.eks. Wittgensteinarkivet i Bergen og den nye, nasjonale Ibsen-utgaven, der den elektroniske kritiske utgaven vil utgjøre en sentral forskningsressurs for fremtidens Ibsen-forskning.
- Tekstkorpus med fokus på form, inkludert grammatisk annotering, som muliggjør kvalitativ og kvantitativ språkvitenskapelig forskning og språkteknologisk forskning og utvikling, Eksempler på slike korpus er Norsk aviskorpus (UiB), og UiO-korpusene Oslo-korpuset av taggedede, norske tekster, Norsk Ordboks nynorsk korpus og bokmålskorpus ved Leksikografisk avdeling. Ved Universitetet i Tromsø finnes det et samisk tekstkorpus som omfatter 6,65 mill. ord, mesteparten nordsamisk. Ca. 1/5 er parallellført med norsk originaltekst i et parallellkorpus.
- Talekorpus, som f.eks. Norsk talespråkskorpus, Oslodelen (UiO). Et større dialektkorpus for norske og skandinaviske dialekter er under oppbygging i et samarbeid mellom UiO, UiT og NTNU samt fem skandinaviske universiteter (ScanDia-korpuset). Slike korpus er viktige for grunnforskning så vel som for mer anvendt forskning, som f.eks. talegjenkjenning. Slike korpus er meget dyre å utvikle (1 t tale = minst 15 t transkribering). Talekorpus med fokus på innhold kan også tenkes å spille en rolle innenfor kulturhistorisk forskning.
- Andre typer tekstkorpus, som parallellkorpuset Oslo Multilingual Corpus (UiO), Norsk Andrespråkskorpus (UiB), The Sofie Treebank (UiO).
- Leksikalske ressurser, som f.eks. de ved Norsk Ordbank ved Universitetet i Oslo.

Området mangler:

- Planer for en balansert og systematisk oppbygd norsk språkbank ble utviklet rundt årtusen-skiftet, men er ennå ikke realisert. De leksikalske ressursene og taleressursene som ble kjøpt av konkursboet etter Nordisk språkteknologi på Voss, utgjør bare en sped begynnelse. Uten en slik bank vil språkteknologisk forskning og utvikling på norsk være underlagt store begrensninger i forhold til internasjonale samarbeidspartnere.
- Det er behov for flerspråklige, parallelle korpus, innlærerkorpus, typer av annoterte korpus (syntaktisk annoterte, innholdsannotering av ulike slag). Mange av de eksisterende korpusene trenger å forstørres og berikes på ulike måter.
- Det er også behov for samiske språkressurser. Tekstkorpuset nevnt ovenfor bør utvides til ti ganger dagens størrelse. Det eksisterer i tillegg lydbånd med flere hundre timer dialektmateriale, storparten ved Samisk institutt i Kautokeino, en del ved de samiske samlinger i Karasjok, og en liten del ved Universitetet i Tromsø, som må digitaliseres eller transkriberes. Utvalget antar at lignende behov finnes for kvensk, men mangler oversikt over dette.
- Det er behov for ny- og videreutvikling av språkteknologiske leksikon, særlig mht semantisk klassifisering.
- Det er et stort behov for digitalisering av kildene for norsk språkhistorie. Noe er digitalisert, men mye finnes bare i papirformat og som ikke transkriberte innspillinger fra siste del av det tjuende århundret, arkivert i mer eller mindre sovende målføre- og dialektarkiver.
- Det samme gjelder digitalisering av nasjonallitteraturen i vid forstand, dvs. både fag- og skjønnlitteratur. Nasjonalbiblioteket har store ambisjoner på dette området, men det er uklart i hvilken grad resultatene vil være optimale for forskningsformål. Av de store forfatterne er

per i dag bare Ibsen gjennomdigitalisert etter grunnleggende filologiske prinsipper, mens det for andre forfattere finnes digitaliserte versjoner uten variant- og noteapparat.

Deling av data er lite problematisk, men det ligger større utfordringer knyttet til forvaltning og sikring.

Standardiseringsarbeid

Selv om det finnes (tilløp til) internasjonale standarder for metadata, er de i ulik grad tatt i bruk i utviklingen av norske ressurser. CLARIN-prosjektet i ESFRI-programmet er meget viktig i denne sammenheng. Forhåpentligvis vil en del grunnleggende standardiserings- og metadata-problemer kunne løses her.

Personvern og opphavsrettigheter

Her er det fortsatt viktige spørsmål å avklare, ikke minst praktiseringen av lover og regler for personvern. I dag er det slik at ved opptak av mest mulig naturlig tale mellom f.eks. to personer må det være en tredje person til stede som passer på at Datatilsynets retningslinjer blir fulgt. Disse retningslinjene skal hindre at ikke opplysninger som kan true personvernet kommer fram i de innspilte samtaleene/intervjuene. Tilstedeværelsen av en slik "sensor" legger sterke begrensninger på engasjerende samtaleemner og gjør målet om å få fram mest mulig naturlig tale svært vanskelig å oppnå. Dette gjør materialet tilsvarende mindre brukbart, ikke bare for språkforskning, men også i noen tilfeller for historisk forskning, når f.eks. steds- og personnavn enten må unngås eller "sladdes".

Språkteknologiske anvendelser

De språkressursene en bygger opp vil ikke bare være nyttig for grunnforskningen. De er også helt uunnværlige i språkteknologisk utviklingsarbeid, som for eksempel å bedre universell tilgjengelighet ved å øke mulighetene for talestyring av programmer, for diktering på norsk, for tekst til tale (maskinell opplesning av skrevet tekst). Dette er basisteknologier som har bred anvendelse, ikke minst i arbeidet med universell tilgjengelighet. Fraværet av en framtidig og oppdatert norsk språkbank vil få negative følger for norsk konkurransekraft i forhold til applikasjoner på store fremmedspråk som engelsk. Vi kan med andre ord oppleve at talestyring av programvare forutsetter at brukeren behersker engelsk godt nok til å bli forstått av den engelskbaserte applikasjonen.

Kulturhistoriske data

Digitaliseringsarbeid

Både på det naturhistoriske og det kulturhistoriske området har det allerede vært gjennomført større tiltak for å digitalisere universitetsmuseenes samlinger. Arbeidet ble påbegynt i Dokumentasjonsprosjektet (<http://www.dokpro.uio.no/>) og til en viss grad fortsatt i Museumsprosjektet (<http://www.muspro.uio.no/>). Noen samlinger er ferdig digitalisert, men mye gjenstår og Roganutvalget understreket i NOU 2006:8 "Kunnskap for fellesskapet" behovet for å fortsette digitaliseringsinnsatsen. Universitetsmuseenes videre arbeid er sikret gjennom Revita-planer som er initiert av universitetsmuseene selv. Slike planer finnes ved UiO, NTNU og er under utarbeidelse ved UiT. Imidlertid er planene ikke sikret finansielt.

Behovet for digitalisering finnes også i Arkivverket, der det meste av materialet ennå ikke er digitalisert. Sett fra forskningens ståsted er hovedargumentene for å videreføre digitaliseringsarbeidet at dataene bidrar både til historydybde og har en positiv effekt på internasjonal forskning. I tillegg kan det hevdes at digitalisering bidrar til en demokratisering og tilbakeføring til det lokale miljø, noe som er i tråd med den Kulturelle Skolesekken.

Standardiseringsarbeid

Databasene som er opprettet følger internasjonale standarder. Derimot er det et problem at Riksarkivet følger sin egen standard.

Det finnes en god del digitalisert materiale som ikke er standardisert (se Ore og Ragnsæter (2007) for detaljer). I og med at mye av digitaliseringsarbeidet som pågår antagelig ikke er særlig koordinert, som for eksempel digitalisering av malerier og objekter (eks. vikingskip), er dette et viktig prosjekt.

Naturvitenskapene

Fagområdene som omhandles i punktet bærer preg av utvalgets medlemmer og noen fagområder er ikke representert, som for eksempel geologi, sjøkartlegging, veterinærmedisin og landbruksvitenskapene.

Fagområdene meteorologi og oseanografi har tradisjonelt hatt høy fokus på dataforvaltning og datautveksling. To organisasjoner som spiller en viktig rolle internasjonalt er World Meteorological Organization, WMO og Intergovernmental Oceanographic Commission, IOC. På 1960-tallet ble det etablert et system for internasjonalt samarbeid og datautveksling, verdensdatasentre, regionale datasentre og nasjonale datasentre. Dette systemet er fortsatt aktivt og møtes annet hvert år i en internasjonal konferanse med 60-70 medlemsland.

En annen internasjonal viktig organisasjon er det internasjonale havforskningsråd, International Council for the Exploration of the Sea – ICES. ICES koordinerer og promoterer forskning i nord Atlanteren og har 20 medlemsland. Det arbeides i ca 100 arbeidsgrupper som rapporterer til overordnede organer deriblant en gruppe for data og informasjons forvaltning, WGDIM Working Group for Data and Information Management. ICES har også et eget datasenter tilknyttet sekretariatet.

Meteorologi og luftforurensning

Innenfor meteorologiområdet er man avhengig av internasjonalt samarbeid og datautveksling. Datasamarbeid skjer primært gjennom WMO der en jobber etter veletablerte standarder. Bruk av andre lands data er helt nødvendig innen meteorologi og dette har etablert en åpen datapolitikk for bruk av meteorologiske data. Meteorologisk institutt forvalter egne innsamlede data og data fra andre aktører i en moderne database som er tilgjengelig 24 timer i døgnet alle dager i året. Av særlig viktighet er observasjoner som inngår i internasjonale nettverk og som har pågått etter samme metode i lang tid. Slike måleserier er spesielt viktige for klima- og miljøovervåking. Målinger fra Arktis har stor betydning for å sikre en global datadekning.

Et prinsipp er at alle data produsert med offentlige midler skal være offentlig tilgjengelig. Imidlertid har meteorologiske institutter som driver kommersiell virksomhet en utfordring i tilgjengeliggjøring av data, og dette fører nok til at noe data holdes tilbake. Selv om meteorologiske institusjoner har utvekslet data gjennom mange år, finnes det et forbedringspotensiale innen metadata. En forbedring på dette området vil også gjøre dataene mer tilgjengelig for forskning.

Nasjonalt råd for å fremme nasjonalt samarbeid om innsamling av data har vært forsøkt opprettet, men en har ikke lyktes i arbeidet.

Klimaforskning og stratosfærisk/osonforskning bedrives på internasjonalt nivå og preges også av høy grad av datautveksling. På dette området råder en åpen datapolitikk.

Eksempler på store databaser er NADIR og EMEP, disse driftes av NILU som datasenter. Et klassisk problem er at data ofte ligger på flere steder i ulike versjoner og det er varierende hvor mye metadata som legges inn. Det mangler standarder for metadata.

Området trenger sannsynligvis ingen NSD-liknende institusjon, men en felles webportal – one stop shopping – ville være nyttig.

Marine forskningsdata

Havforskningsinstituttet samler inn og forvalter store mengder marine forskningsdata innen fagfeltene marin biologi, marin kjemi, plankton, oseanografi, akvakultur, genetikk, fiskeriakustikk og andre. Havforskningsinstituttet er vertsinstitusjon for Norsk Marint Datasenter som er organisert som en av 19 faggrupper ved instituttet. NMD har som overordnet mål å være det nasjonale kontaktpunktet for marine data, også for data som ikke eies av Havforskningsinstituttet. NMD har som hovedoppgave å lagre og gjøre tilgjengelig forskningsdata fra instituttet og andre datakilder. Norsk Marint Datasenter (NMD) ble etablert som Norsk Oseanografisk Datasenter (NOD) i 1971. Datasenteret har hatt ansvaret som National Oceanographic Data Centre NODC i Norge siden etableringen.

Havforskningsinstituttet har gjennomført en intern kartlegging av hva som eksisterer av forskningsdatasett og lange tidsserier. Dette har resultert i en datakatalog på 500 datasett og lange tidsserier som forvaltes av NMD. En søkbar datakatalog er laget og blir publisert på internett høsten 2007. En utfordring som blir en følge av å lage datakatalogen blir å tilgjengeliggjøre forskningsdataene på en hensiktsmessig måte. Mange av datasettene eksisterer kun på papirform i journaler, mens andre foreligger på elektronisk form, men ikke nødvendigvis i en søkbar database.

Metadata som er registrert om datasettene og tidsseriene er gjort etter en europeisk standard, utviklet i et EU-finansiert forskningsprosjekt, EDMED European Directory for Marine Environmental Datasets. Registreringene er i utgangspunktet gjort på norsk, men informasjonen er oversatt til engelsk for å kunne rapporteres i et nytt EU-finansiert forskningsprosjekt, SeaDataNet.

Viktige marine dataserier

Forskningsrådet fikk i perioden 2003 - 2004 gjennomført et omfattende arbeid for å kartlegge og vurdere "verneverdi" av lange tidsserier for miljøovervåkning og forskning. Det ble nedsatt 3 arbeidsgrupper som hver utarbeidet en rapport, Lange tidsserier for miljøovervåkning og forskning:

- Viktige klimadataserier. Norges forskningsråd 2003. Meteorologisk institutt hadde koordineringsansvaret
- Viktige terrestriske og limniske dataserier. Norges forskningsråd 2003. Direktoratet for naturforvaltning hadde koordineringsansvaret
- Viktige marine dataserier. Norges forskningsråd 2004. Havforskningsinstituttet hadde koordineringsansvaret

En utfordring som gjenstår etter denne kartleggingen er en samkjøring av informasjonen i de tre rapportene. Ulike definisjoner av kringinformasjon er benyttet og informasjon som er registrert i en rapport er ikke nødvendigvis tilgjengelig i de andre rapportene. En koordinering av disse oversiktene er nødvendig samt å gjøre informasjonen tilgjengelig på internett.

MAREANO og Marin geografisk infrastruktur

MAREANO skal kartlegge alle norske havområder, dvs. et samlet areal på 2 057 826 km². Ifølge miljøvernminister Helen Bjørnøy er målet med kartleggingen å hindre alle aktiviteter som kan true miljøet. Fram til 2010 konsentrerer MAREANO all innsats i et 142 000 km² stort område av Barentshavet og Lofoten, for å kunne bidra med informasjon til en revidert forvaltningsplan for

Barentshavet. Hovedmålet for MAREANO er å gi ”kunnskapsbasert beslutningsstøtte for økosystembasert forvaltning av norske kyst og havområder” gjennom kartlegging av våre havområder. Dette kan i prinsippet gjøres uten å lage effektive forvaltnings- og formidlings-systemer for geodata. Dette løses i prosjektet gjennom å lage et nettbasert formidlings-systemsystem basert på en distribuert databaseløsning. De tre hovedpartnerne i prosjektet er Norges geologiske undersøkelse, Statens Kartverk Sjø og Havforskningsinstituttet.

Kunnskapen skal systematiseres i en offentlig tilgjengelig marin arealdatabase for norske havbunnsområder, som åpner for muligheten for all annen bruk av innsamlede og forvaltede data fra Mareano. Et viktig fremtidig delmål i MAREANO vil være å stille data til disposisjon for et stort antall brukere utenfor MAREANO.

Partnerne skal innen rammen av statens kartpolitikk gjøre sine datasett tilgjengelig for eksterne brukere, og samarbeide om drift av www.mareano.no som den sentrale innfallsporten til kunnskap. Gjennom en slik tilnærming til en Marin geodata-infrastruktur (MSDI: Marine spatial data infrastructure) blir det behov for avklaringer av datarettigheter, kostnader, tilgang, interoperabil teknologi, behov for standardbaserte løsninger, tilrettelegging for dataintegrasjon, fokus på datakvalitet, egnethetsvurderinger osv.

Partnerne er i ferd med å danne en felles forståelse rundt utfordringene og deres innbyrdes prioriteringer. Neste skritt er et mer strukturert opplegg med eventuelt arbeidsgrupper som har ansvar for ulike tema som utvikles videre mot konkrete anbefalinger.

Det internasjonale polarår 2007 - 2008 - IPY

Under det internasjonale polaråret skal det gjennomføres forskning som ingen enkelt nasjon vil kunne finansiere eller utføre alene. Innsatsen fra minst 50.000 forskere og teknikere fra vel 60 land gjør polaråret 2007-2008 til det største internasjonale forskningsprogrammet noensinne. Norge og norske forskere har en framtrødende rolle under polaråret 2007-2008, ikke minst i kraft av sterke forskningsmiljøer med et omfattende internasjonalt nettverk og gode fasiliteter og logistikkstøtte både i Arktis og i Antarktis. Med en bevilgning på 330 millioner over 4 år, er Norge også blant de største bidragsyterne i internasjonal sammenheng.

Det er i underkant av 30 forskningsprosjekter finansiert i Norge og disse prosjektene vil alle generere store mengder forskningsdata. I de tre foregående polarårene var det ikke organisert en koordinert dataforvaltning som har ført til at det er vanskelig å bruke dataene som ble samlet inn. I det fjerde polaråret forsøker Norges Forskningsråd å få dette koordinert gjennom å finansiere dataforvaltning. Et hovedmål vil være å sikre langsiktig forvaltning av forskningsdata generert av norske forskningsprosjekter under IPY og sikre at data og informasjon er gratis og åpent tilgjengelig i tråd med IPYs internasjonale føringer. Dette kan gjøres gjennom å:

- Etablere en nasjonal, koordinert oversikt over nasjonalt finansierte IPY data som kan videreføres i en langsiktig oversikt over polare forskningsdata.
- Etablere kontakt og underbygge samarbeid mellom eksisterende datahåndteringsinstitusjoner, og beskrive og implementere internasjonalt aksepterte og standardiserte grensesnitt mellom dem for enkel utveksling og høsting av data og metadata.
- Etablere en langsiktig forvaltning av IPY-finansierte data og metadata i egnede datasentre gjennom tilrettelegging av dataformater og lagringsfasiliteter, og sikre at data dokumenteres og publiseres i samsvar med aksepterte standarder for metadata.

Terrestriske og limniske dataserier og naturhistoriske samlinger

Noen av de lengste tidsseriene, f.eks. Landskogstakseringen fra 1919, ble igangsatt ut fra behovet for data i primærressursforvaltningen. Behovet har vært til stede siden etableringen og aktiviteten

som har generert dataene har vedvart og også utvidet etter som nye problemstillinger har tilkommet. De fleste av landbrukets dataserier baserer seg på samplingsmetodikk som gir arealrepresentative data.

Produksjonen av en rekke av andre lange dataserier ble startet opp fordi politikerne og forvaltningen reiste miljø- eller ressurspolitiske problemstillinger som forskningsinstitusjonene fikk i oppdrag å finne svar på. Dataforvaltningen ble godt ivaretatt så lenge forskningsaktiviteten var høy. Etter noen år, eller tiår, mistet etter hvert flere slike miljøpolitisk begrunnede forskningsprogrammer politisk interesse og det ble en utfordring for forskningsinstitusjonene å fortsette aktivitetene, gjerne med mer begrensede ressurser eller under annet navn. Det er i avslutningen, i overgangen til andre programmer eller andre aktører at dataforvaltningen lett kan svikte. Derved kan det være en stor utfordring å finne eller gjenskape dataseriene, selv om rådataene er dokumentert i analoge tabeller og rapporter. For flere av de lange dataseriene som har miljøpolitisk interesse, er det en fornyet interesse, spesielt siden klimaforandringene genererer behov for å dokumentere trender og forklaringsmodeller på effekter i naturen.

Det er identifisert nærmere 300 dataserier innen terrestrisk og limnisk fagområder i en utredning initiert av Forskningsrådet (LangeTidsserier_terr-limn.pub-270974.pdf) Arbeidsgruppen som identifiserte disse har valgt ut 30 tidsserier som har særlig stor verdi for disse fagområdene ut fra kriterier som kvalitet, tilgjengelighet og bruk. Metainformasjon om dataseriene er tilgjengelig på internett (<http://tidsserier.dirnat.no/>)

Innsamling av objekter til de naturhistoriske samlinger ved universitetsmuseene har foregått i nesten 300 år. Denne innsamlingen og i moderne tid, deres tilhørende databaser, er kommet i stand som følge av museenes mandat og oppgaver. Disse er rimelig stabile over tid, selv om tilveksten varierer som følge av ujevn ressurstilgang og variasjon av fokus i universitetsmuseene, herunder konservatorenes faglige arbeidsområder. Når det gjelder digitalisering av historisk materiale i samlingene har denne blitt forsert de 10 siste årene som følge av større krav tilgjengelighet av samlingene, dette har delvis foregått gjennom Universitetsmuseenes nasjonale databaseprosjekt/Museumprosjektet. I de naturhistoriske samlingene på de 4 universitetsmuseene er det i pr. 2006 til sammen digitalisert om lag 5 mill objekter/observasjoner av totalt om lag 15 mill. objekter eller observasjoner.

For naturhistoriske samlinger ligger verdien i at de bl.a. er en dokumentasjon av tidligere utført forskning på arter, et grunnlag for framtidig artsforskning og en database over arter funnet i norsk natur samlet gjennom en lang periode.

Metoder og standardisering

Gjennomgående er dataserier godt dokumentert i forhold til metoder, parametrene i disse følger som regel standarder som er akseptert innen fagfeltet. Mange av seriene er også "konservative" i metodevalget for at seriene skal være konsistente. Når det gjelder nomenklatur er det større variasjon, i alle fall mellom dataseriene.

Standardiseringen av naturhistoriske samlinger følger internasjonale standarder når det gjelder taksonomisk nomenklatur, svært mange andre parametre i samlingene varierer fra fagområde til fagområde. For distribusjon er data fra primærdatabasene kodet med den internasjonale standarden Darwin Core 2.

Lange tidsserier er metadatastandardisert i forhold til søk <http://tidsserier.dirnat.no/>, dette er imidlertid ikke tilfelle med data fra naturhistoriske samlinger.

Et godt eksempel på viktigheten av standardisering for datautveksling er Global biodiversity information facility (GBIF), som er et internasjonalt OECD-initiert "mega-science"-prosjekt, www.gbif.org. Det er etablert et nettverk mellom forsknings-institusjoner for deling av artsdata. GBIF har som formål å gjøre primærdata om jordas biologisk mangfold fritt tilgjengelig for alle over internett. Målgrupper er primært grunnforskning innen biosystematikk og biogeografi, sekundært forvaltningen for bevaring av jordas biologiske mangfold og ressurser. Til nå er det 27 land, hvorav 16 europeiske, som er med som voting participants i GBIF, i tillegg kommer 21 land og 35 organisasjoner som er assosierte medlemmer. Den norske GBIF participant node ble etablert som et treårig prosjekt av Forskningsrådet i 2005, ved Naturhistorisk Museum - Universitetet i Oslo.

Estimatene over antall arter på jorden varierer mellom 5-30 mill. arter. Nettverket har pr. dato ca. 130 mill. sted- og tidfestede artsobservasjoner fra over 200 dataleverandører, disse deles gjennom en felles portal <http://newportal.gbif.org/> med taksonomisk og geografisk søk. Det er i dag lagt ut data om over 2 mill. objekter fra norske naturhistoriske samlinger i GBIF. I Norge samarbeider GBIF-noden med Artsdatabanken og i løpet av 2007 vil alle universitetsmuseene og de fleste forskningsinstitusjonene dele stedfestede artsdata gjennom GBIF. Darwin Core 2 er den internasjonale kodenstandarden for datasettene og den taksonomiske nomenklaturen i Norge forvaltes av GBIF og Artsdatabanken i felleskap gjennom Artsnavnebasen.

Datautveksling

Dersom en med utveksling mener å dele data med forskere i andre institusjoner og land, foregår dette i stor grad på noen områder. Utlån av objekter i naturhistoriske samlinger er helt vanlig og er en forutsetning for kvaliteten i samlingene. Datautveksling av natur-historiske samlingsdata internasjonalt foregår i dag gjennom GBIFs nettverkløsninger basert på standardiserte parametre (Darwin Core 2).

Når det gjelder delingsmekanismene for dataserier vil være mer på individuelt forskernivå eller institusjonsnivå. Der slike avtaler inngås vil datautvekslingen være betydelig, men parameterutvalget vil som regel være avgrenset til det enkelte samarbeidsprosjekt. Her er det ingen standardiserte delingsmodeller.

Av datatjenester innen naturhistoriske museer har nylig universitetsmuseene opprettet en felles dataforvaltningstjeneste MUSIT. Denne tjenesten vil på sikt kunne bidra sterkt til standardisering av dataforvaltningen. For deling av artsdata fra både universitetsmuseene og fra forskningsinstituttene har Artsdatabanken oppgaver i å formidle data fra disse i et felles grensesnitt. Det er nylig laget felles infrastruktur for de naturvitenskapelige institusjonene som gjelder felles taksonomisk nomenklatur i Artsnavnebasen, felles stedfestet artsdatadeling gjennom en webkarttjeneste og en felles Naturtypedatabase.

For dataserier er det i dag ingen felles datatjeneste utenom metadataten, og foreløpig ikke standard utvekslingssystemer. Delingen av data skjer på basis av individuelle peer-to-peer nettverk eller bilaterale institusjonelle avtaler med fysisk deling av rådata eller bearbejdede data. I EU-prosjektet Lifewatch har noen av de største naturfaglige forskningsinstitusjonene i Norge under planlegging en ny delingsmodell for lange tidsserier.

Dersom en forsker i en dataeier institusjon avleverer primærdata fra dataserier til en annen forsker i en annen institusjon praktiseres medforfatterskap dersom medforfatteren har hatt et reelt bidrag i det vitenskapelige arbeidet.

Delvis oppdragsfinansierte forskningsinstitusjoner vegrer seg mot fritt å dele rådata og bearbejdede data med høy oppløsning. Grunnen til dette er risikoen de løper ved å dele "gratis" data som senere kan være deres kapital i framtidige oppdrag. Det er nærliggende for disse

institusjonene også å innlemme data for statlige forsknings- og overvåkningsprogrammer som i prinsippet skal være allment tilgjengelige. De forklarer dette med merverdien dataene tilføres ved den vitenskapelige bearbeiding, kontroll og kontekst med henvisning til åndsverksloven. Noen forskningsinstitusjoner er reelle konkurrenter på noen overlappende områder. Dette kan også føre til vegring mot deling av data.

Offentlighetslovens bestemmelser griper ganske langt inn med krav om å dele data, da lovens dokumentbegrep omfatter databaser. Uklarhet om den operative håndhevelsen av offentlighetsloven er noen av hindrene for å dele data:

Det er en allmenn oppfatning at det også er gode grunner for ikke å dele data:

- Sensitiv artsinformasjon (jfr. Offentlighetslovens § 6)
- Dårlig eller udokumentert datakvalitet
- Data som inngår i aktive forskningsarbeider som ennå ikke er publisert

Tiltak innenfor terrestriske og limniske dataserier samt naturhistoriske samlinger

De viktigste tiltakene Forskningsrådet bør gjennomføre er å:

1. Allokere mer penger til prioriterte databaser m.v. gjerne organiseres gjennom et langsiktig program med lengre varighet på minst 10 år. Det bør prioriteres å:
 - a. Etablere og videreføre databaser på dataserier som er temporært avsluttet, men som kan bli aktuelle igjen
 - b. Rydde opp i gamle databaser- og tidsserier og lage moderne databaser
 - c. Lage funksjonelle databaser i etterkant av forskningsprosjekter og -programmer
2. Finansiere bygging av overbyggende felles infrastrukturer og datatjenester som kan gjøre dataene i institusjonenes primærdatabaser tilgjengelige gjennom felles grensesnitt.
3. Knytte datapolitikk til bevilgning for at:
 - a. Dataene blir forvaltet i databaser med ivaretar krav til etterprøvbarhet, kvalitet og datasikkerhet
 - b. Ilagt etiske retningslinjer for innlegging av data, bruk av data i forskning og sekundær informasjonsproduksjon.
 - c. Dataene blir tilgjengelige i databaser der innholdet er åpent for andre forskningsinstitusjoner i rimelig tid etter at resultatene er publisert.
4. Utrede behovet for en nasjonal ordning for den private instituttsektoren som omfatter mer likeverdige finansielle rammevilkår for ivaretagelse av de samfunnsmessige oppgavene, herunder deling av data.

Vedlegg II: Beskrivelse av sentrale aktører

Hensikten med vedlegget er å beskrive noen sentrale aktører når det gjelder databaser som infrastruktur for forskning. På grunn av den korte tiden og utvalgets sammensetning har det ikke vært mulig å lage en fullstendig oversikt. I mange tilfelle vil institusjoner har flere roller samtidig, de er både dataprodusenter, forskningsinstitusjoner og leverandører av viktige infrastruktur-tjenester. Det er en viktig oppgave i forbindelse med kartlegging og utarbeidelse av handlingsplaner for fagområdene å beskrive de sentrale aktørene på hvert område, ikke minst de aktører som forholder seg til flere fagområder.

Nasjonalt folkehelseinstitutt

Det finnes ingen institusjon i dag som kan samle data fra helsevesenet på en enkel måte og gjøre dem tilgjengelig for forskning. Det nærmeste man kommer er Nasjonalt folkehelseinstitutt.

Hvilke data finnes? Hvis man skal forstå befolkningens helse finnes kanskje de viktigste data i primærhelsetjenesten. Det er også mye data i spesialisthelsetjenesten, men det er i dag ikke vanlig å sette sammen data fra primær- og spesialisthelsetjenesten. Stortinget vedtok 1. februar 2007 å opprette et personidentifiserbart Nasjonalt pasientregister basert på polikliniske undersøkelser og innleggelse ved sykehus. Dette sentrale helseregisteret kommer i tillegg til 7 andre, slik som Kreftregisteret, Medisinsk fødselsregister og Dødsårsaksregisteret. I tillegg finnes noen større helseundersøkelser som har vært drevet frem av ulike forskergrupper, dels ved universitetene og dels ved institusjoner i den sentrale helseforvaltning. Det er også enkelte kvalitetsregistre for noen spesielle sykdommer og behandlinger.

Den sentrale helseforvaltningen ble fornyet i 2002 ved at det ble opprettet et Nasjonalt folkehelseinstitutt, et Sosial- og helsedirektorat og et Statens helsetilsyn. Sosial- og helsedirektoratet har som oppgave å gjennomføre og iverksette nasjonal politikk i helse- og sosialsektoren, mens Statens helsetilsyn skal følge med på hvordan tjenestene og personellet utøver sin virksomhet og gripe inn overfor virksomheter og helsepersonell som utøver virksomheten i strid med lovgivningen.

Nasjonalt folkehelseinstitutt (heretter forkortet FHI) ble satt sammen av store deler av Statens institutt for folkehelse, Statens helseundersøkelser, Medisinsk fødselsregister og Enhet for legemiddelstatistikk- og metodologi fra Norsk medisinaldepot. FHI skal bidra til å bedre befolkningens helsetilstand, først og fremst ved å fremskaffe kunnskap og gi råd og tjenester som gjør det mulig å forebygge sykdom og helseskade. En viktig oppgave for FHI er å gjøre data fra helseregistre og andre kilder lettere tilgjengelig for forskning. FHI består, i tillegg til ledelse, stab og en administrasjonsdivisjon, av fem fagdivisjoner for epidemiologi, miljømedisin, psykisk helse, smittevern og retts toksikologi og rusmiddelforskning. Ved alle fagdivisjonene drives det egen forskning. FHI har databehandlingsansvar for 6 av de nasjonale helseregistrene og driver noen store helseundersøkelser i befolkningen. Den mest omfattende undersøkelsen er Den norske mor og barn undersøkelsen som inkluderer snart 100 000 familier med spørreskjemadata og biologisk materiale. Hensikten er å finne årsaker til alvorlige sykdommer for å kunne bidra til nye og bedre forebyggende tiltak.

Ved Divisjon for epidemiologi er det opprettet en egen avdeling for forskningsdata som skal bidra til at data og biologisk materiale blir gjort tilgjengelig for forskere internt i instituttet og ved andre forskningsinstitusjoner nasjonalt og internasjonalt. Avdelingen har ansvar for godkjenning av

søknader og utlevering av data fra helseregistre og helseundersøkelser, og gir juridisk hjelp og sikrer at gjeldende lover og regelverk følges, samtidig med at personvern og andre rettigheter ivaretas for deltakere i helseundersøkelser. Det er opprettet en omfattende elektronisk prosjekt-database for å holde orden på de mange forespørslene.

Norsk Samfunnsvitenskapelig datatjeneste

”Norsk kompetanse når det gjelder infrastruktur for forskning har vist seg å være blant de fremste i Europa, noe som ikke minst skyldes arbeidet til Norsk samfunnsvitenskapelig datatjeneste (NSD)”⁷

Formål

NSD er et av verdens største arkiv for forskningsdata og formidler av data til forskere og studenter i Norge og utland. I tillegg er NSD et kompetansesenter som veileder forskere i forhold til datainnsamling, dataanalyse, metode, personvern og forskningsetikk. Hovedmålsettingen er å forbedre muligheter og arbeidsvilkår for empirisk forskning som først og fremst er avhengig av tilgang til data. Dette gjøres ved å bygge ned økonomiske, tekniske, kunnskapsmessige, juridiske og administrative barrierer mellom brukere og dataressurser.

Organisering

NSD ble etablert i 1971 og var frem til 2002 institusjonelt tilknyttet Norges forskningsråd. Etter vedtak i Stortinget ble NSD med virkning fra 1.1.2003 fristilt fra NFR og organisert som aksjeselskap heleid av Kunnskapsdepartementet. NSD har 65 heltidsansatte.

Norges forskningsråd har siden etableringen av NSD, gjennom sine grunn- og prosjektfinansieringer, spilt en sentral rolle for både omfanget og kvaliteten på de tjenester NSD tilbyr sine brukere. Hovedtyngden av brukerne kommer fra de samfunnsfaglige fagområdene ved universitetene, høyskolene og instituttsektoren. NSD betjener i økende grad også andre fagområder, særlig medisin og helsefag.

Nasjonalt samarbeid

NSD skal tilby forsknings- og utdanningsmiljøene et bredt, helhetlig og kvalitetsmessig godt tilbud innen de områder NSD arbeider. For å nå denne målsettingen legger NSD vekt på og drar nytte av nært samarbeid med flere norske instanser.

NSDs muligheter som nasjonal infrastruktur for norske forskningsmiljøer er uløselig knyttet til miljøenes tillit til og bruk av NSDs tilbud. At forskningsmiljøene i Norge har støttet valget om å konsentrere innsatsen på feltet ved én institusjon, har vært helt avgjørende for at NFR har gitt NSD gode og trygge rammevilkår til å løse sine primæroppgaver. Gjennom satsingen på NSD har Norges forskningsråd etablert en rekke felles ressurser som har forbedret muligheter og arbeidsvilkår for empirisk forskning. Det er bygd opp omfattende arkiver med data om individer, regioner og politiske og administrative systemer, og NSD har satset på internasjonalisering av norsk forskning gjennom deltakelse i internasjonale intervjuundersøkelser og i tilrettelegging av komparative regionale data. Datatilbudet er blitt stadig mer omfattende.

Forskningsrådet har også etablert og vært hovedbidragsyter til driften av Datafaglig sekretariat. Sekretariatets oppgaver som nå er videreført av Personvernombudet for forskning, har gitt NSD spisskompetanse innen området forskning og personvern og lagt grunnlaget for de omfattende oppgaver NSD har på dette området i dag.

⁷ Norges forskningsråd, Nytt om EU-forskningen, Nr. 2, juni 2007

I 2003 ble NSD personvernombud for universitetene, de vitenskapelige og statlige høgskolene og en rekke helseforetak og forskningsinstitutter. NSD er i dag personvernombud for i overkant av 135 institusjoner, deriblant alle universitet og vitenskapelige høgskoler, statlige og private høgskoler, instituttsektoren og flere helseforetak. Personvernombudsordningen innebærer at plikten til å innhente forhåndssamtykke (konsesjon) fra Datatilsynet langt på vei erstattet av en meldeplikt som medfører at NSD blir siste instans med ansvar for å veilede, vurdere og godkjenne forskningsprosjekter i forhold til personopplysningsloven og helseregisterloven. Den brede oppslutningen om ordningen og siste års vekst i antall institusjoner som slutter seg til, tyder på at norsk forskning har sett seg tjent med å ha et rådgivnings- og serviceorgan med spesialkompetanse på spørsmål om forskning og personvern.

I forslaget til ny forskrift til personopplysningsloven var NSDs viktige rolle og kvalitetssikrende funksjon fremhevet som et argument for å bygge ned konsesjonsplikten og begrense Datatilsynets forhåndskontroll. Det tyder på at Datatilsynet har tillit til at personverninteressene har vært tjent med den kontroll norsk forskning har pålagt seg selv med obligatorisk veiledning og forhåndsvurdering hos NSD.

Et omfattende og nært samarbeid med Norges forskningsråd vil også i fremtiden være helt avgjørende for både omfanget og kvaliteten på de tjenester NSD kan tilby sine brukere. Avtalen med Norges forskningsråd om langsiktig finansiering av infrastruktur for forskning, er derfor viktig for institusjonen. I avtalen slår NFR fast at fra Norges forskningsråds side representerer NSD en viktig satsing på infrastruktur og datatjenester, i første rekke for samfunnsvitenskapelig forskning, men også for andre forskningsområder.

NSD tilrettelegger og formidler data fra en rekke ulike kilder. NSDs avtale med Statistisk sentralbyrå om formidling av data til forskningsformål står sentralt i denne sammenheng. Ved å bruke NSD til å forvalte data til forskningsmiljøene har man fått etablert prosedyrer som ivaretar personvernet samtidig som det er åpnet for relativt utstrakt bruk av individdata til forskningsformål. Samarbeidet er unikt i internasjonal sammenheng.

Internasjonalt samarbeid

NSDs bidrar til internasjonalisering av norsk forskning gjennom deltakelse i internasjonale organisasjoner og prosjekter. Gjennom forpliktende samarbeid i en rekke prosjekter, deltar NSD i arbeidet med å bygge opp et europeisk og internasjonalt datagrunnlag for komparative analyser. Dette samarbeidet legger til rette for norsk deltakelse i internasjonale forskningsprosjekt ved at norske forskere får tilgang til internasjonale data, samtidig som data om norske forhold blir tilgjengelig for omverden og Norge kan inkluderes i utenlandske analyser.

NSD er en sentral internasjonal aktør innen sitt fagfelt. NSD mottok i 2005 en prestisjetung europeisk forskningspris som en av partnerne i European Social Survey som ble tildelt EUs forskningspris (Descartes-prisen) for 2005. Ved kåring av vinnere ble det lagt vekt på at man har oppnådd eksepsjonelle resultater gjennom europeisk samarbeid og at prosjektet er et resultat av fremragende forskning. NSD er ansvarlig for dataarbeid, dataservice og for de internettorienterte undervisnings- og infrastrukturoppleggene i European Social Survey (ESS). De første tre årene har om lag 12 000 akademiske brukere benyttet seg av disse tjenestene. ESS er nå bygget opp rundt NSDs brede kompetanse innen teknologi, dokumentasjon av data og tilrettelegging og bruk av survey data.

Statistisk sentralbyrå

Statistisk sentralbyrå (SSB) har hovedansvar for innsamling, bearbeiding og formidling av offisiell, norsk statistikk. Offisiell statistikk og analyser basert på statistikk skal gi allmennheten, næringslivet og myndighetene kunnskap om samfunnets struktur, utvikling og virkemåte.

SSB ble etablert i 1876. Det er en faglig uavhengig institusjon som administrativt ligger under Finansdepartementet. SSB har en egen lov, "Lov om offisiell statistikk og Statistisk sentralbyrå" fra 1989. Loven gir SSB rett til å bestemme hva som skal være offisiell statistikk og ansvar for å samordne all offisiell statistikk i Norge.

ssb.no er hovedkanalen for publisering av offisiell statistikk. Her legges ny statistikk ut hver dag, og brukerne kan selv hente tall fra Statistikkbanken. Statistikkloven pålegger offentlige organer som skal opprette eller endre større administrative datasystem, å melde dette til SSB. SSB kan fremme forslag om hvordan systemene skal utformes for å ivareta statistikkhensyn. Videre har SSB rett til å utnytte offentlige administrative datasystemer som grunnlag for offisiell statistikk.

Data som er samlet inn eller registrert for bruk i forvaltningen er underlagt taushetsplikt etter forvaltningsloven. I tillegg finnes det taushetsbestemmelser i ulike særlover som også vil tre inn dersom de aktuelle data er underlagt dette regelverket, for eksempel helsepersonelloven, arbeids- og velferdsforvaltningsloven, sosialtjenesteloven, barnevernloven og folkeregisterloven. Utgangspunktet er at dataene bare skal benyttes til det formål de ble innhentet/registrert for, men at det åpnes for bruk til forskning og statistikk.

SSB har med hjemmel i statistikkloven tilgang til de fleste administrative registre, d.v.s. registre opprettet for bruk i offentlig forvaltning, for utarbeidelse av offisiell statistikk og forskning.

Forskere som ønsker tilgang til data vil kunne henvende seg enten direkte til det forvaltningsorgan som eier det administrative register, eller de kan henvende seg til SSB. SSBs kopi av registrene er som regel bedre tilrettelagt for bruk i forskning. Før forskeren kan få tilgang til dataene må det gis dispensasjon fra taushetsplikten. Som hovedregel er det forvaltningsorganet som eier registeret, eller registereiers overordnede organ, som kan gi dispensasjon fra taushetsplikten slik at data kan utleveres til forskning. Dette gjelder uavhengig av om forskeren henvender seg direkte til registereier eller SSB for tilgang til data. Dersom forskeren skal ha tilgang til data fra flere administrative registre, må forskeren følgelig henvende seg til flere instanser. Erfaring hittil viser at saksbehandlingskapasiteten i noen organ er presset, og at det derfor kan ta lang tid før nødvendig dispensasjon blir gitt.

Når det gjelder de sentrale helseregistre opprettet i medhold av helseregisterloven er det gitt nærmere regler for tilgang til dataene i forskrift til det enkelte register, og et sentralt vilkår er at forskningen må ligge innenfor registerets formål.

Dersom forskeren ønsker tilgang til personopplysninger må forskeren oppfylle personopplysningslovens krav. Dette medfører bl.a. at forskningsprosjektet skal meldes til Datatilsynet eller et personvernombud, evt. at det må innhentes konsesjon fra Datatilsynet eller tilrådning fra personvernombud dersom det er tale om sensitive personopplysninger.

Dersom forskeren henvender seg til SSB for tilgang til data, vil i tillegg til overnevnte vilkår også statistikklovens bestemmelser regulere forholdet. Forskeren må være tilknyttet en godkjent forskningsinstitusjon, og avidentifiserte/anonyme data som lånes ut skal bare benyttes til det formål/prosjekt, og av angitt personell, som er omtalt i søknaden. Etter prosjektslutt skal data slettes eller returneres SSB.

For forskere som ser at anonyme data er tilstrekkelige for formålet, kan SSB overlate til NSD arbeidet med å tilrettelegge aidentifiserte data til anonym form etter forskerens ønske.

Nærmere opplysninger: www.ssb.no under overskriften Forskning/Utlevering av data til forskning.

Språk- og tekstvitenskapelige datasentre

Tekstlaboratoriet ved UiO

<http://www.hf.uio.no/tekstlab/>

Tekstlaboratoriet er en enhet ved Institutt for lingvistiske og nordiske studier som er engasjert i språkteknologi med vekt både på språk og teknologi. Fokus er på forskning og utvikling av nye ressurser som programvare og verktøy for brukergrensesnitt, søking, databaser, tekstforskning, språkforskning og tekstbaser/korpus. Tekstlaboratoriet tilbyr praktisk og teoretisk støtte i språkteknologiske spørsmål til språkforskere, stipendiater og masterstudenter ved Det humanistiske fakultet. Det har også felles forsknings- og utviklingsprosjekter med andre institusjoner i inn- og utland.

Enhet for digital dokumentasjon ved UiO

<http://www.edd.uio.no/>

Enhet for digital dokumentasjon (EDD) er opprettet for å vedlikeholde og videreutvikle databasene og de elektroniske samlingene fra Dokumentasjonsprosjektet. EDD tilbyr en rad ulike søkbare databaser på web. Mest kjent er de store håndordbøkene. Fagmiljøene har i tillegg tilgang på flere andre databaser via en spesialutviklet programvare. EDD tilbyr også enhetene ved det Humanistiske fakultetet faglig/metodisk-støtte og teknisk støtte til å lage store digitale dokumentasjonssamlinger og databaser. EDD har lang erfaring med å håndtere lyd, bilde og tekst. EDD har kompetanse innenfor leksikografi, korpuslingvistikk, navnegransking, historiske dokumentarkiv, foto, arkeologiske- og kulturhistoriske gjenstandssamlinger, samt naturhistorie.

Avdeling for kultur, språk og informasjonsteknologi (Aksis) ved UiB

<http://www.aksis.uib.no/>

Aksis initierer og administrerer bl.a. eksternt finansierte forskningsprosjekter knyttet til språk- og tekstteknologi. Avdelingen består av tre forskergrupper, hvorav to er relevante i denne sammenhengen:

Forskergruppe for språkteknologi driver forskning innen datalingvistikk, korpuslingvistikk, språkressurser og språkteknologi. Porteføljen omfatter prosjekter i maskinoversettelse, terminologi, leksikografi, nettordbøker, grammatikker og trebanker.

Forskergruppe for tekstteknologi arbeider med utvikling av informasjonsteknologiske løsninger for tekstkoding, digital utgivelsesfilologi og klassifisering og tilrettelegging av massive tekstsamlinger.

MUSIT

MUSIT er et samarbeid mellom universitetsmuseene i Bergen, Oslo, Trondheim og Tromsø med det formål å tilrettelegge for forvaltning av og gi tilgang til museenes digitaliserte samlinger. MUSIT skal også utvikle og drifte felles databaseløsninger. Den daglige driften av MUSIT ivaretas av Drifts- og utviklingsgruppen (DUG) som er administrativt lagt til Universitetet i Oslos senter for informasjonsteknologi, USIT.

MUSIT består av to koordineringsgrupper, én for naturfag og én for kulturfag, og flere faggrupper. Med unntak av DUG, som er permanent, og fungerer som sekretariat for MUSITs styre, opprettes faggruppene etter behov og rapporterer til koordineringsgruppene eller styret. MUSIT bygger videre på det omfattende arbeidet som ble påbegynt i Museumsprosjektet, et nasjonalt samarbeidsprosjekt for å imøtekomme krav til gjenstandsforvaltning, feltarbeid, forskning og formidling i de norske universitetsmuseene. Museumsprosjektet ble opprettet våren 1998 og formelt avsluttet 31.12.2006.

DUG er lokalisert ved Universitetet i Oslos Senter for informasjonsteknologi (USIT), i Seksjon for utdannings og forskningsstøtte (SUF). MUSIT ledes av et styre nedsatt av de fire museenes eiere, dvs. universitetene i de ovennevnte byene. Styret rapporterer til Årsmøtet som har funksjon som MUSITs generalforsamling. På årsmøtet møter representanter for de fire universitetene. MUSIT finansieres av universitetene i Bergen, Oslo og Tromsø og Norges teknisk-naturvitenskapelige universitet.

DigForsk

DigForsk AS er et nasjonalt senter for digitalisering av vitenskapelige, kulturhistoriske samlinger og arkiver med relevans for forskning, utdanning, forvaltning og allmennheten. Selskapet er heleid av Universitetet i Oslo og organisert som et datterselskap av UNIRAND AS.

DigForsk AS ble opprettet 1.januar 2006 for å koordinere og samordne digitalisering av forskningsmateriale på tvers av ulike fag- og forskningsmiljø. Selskapet har spesialisert seg på digitalisering av forsknings- og kulturmateriale, bibliotek –, museum – og arkivsamlinger. DigForsk AS er dermed et verktøy som forskningsmiljøene og andre kan bruke for å gjøre sine arkiv og samlinger digitale, det være seg forskningsresultater eller urørte kulturhistoriske samlinger.

DigForsk AS bygger på erfaringer fra et tidligere prosjekt ved universitetene i Norge. Prosjektet UniReg (Universitetenes registreringsentraler) ble etablert som en del av Det nasjonale dokumentasjonsprosjektet i 1992 og har hatt ansvar for digitalisering av samlinger og arkiver for forskningsmiljøene. Selskapet som nå etableres er en videreføring og omorganisering av UniReg. DigForsk AS består av 2 avdelinger med til sammen 6 underavdelinger i Finnmark, foruten hovedkontoret i Oslo.

UniReg og senere DigForsk har i 13 år samarbeidet med museer og universiteter, noe som har resultert i at en rekke fysiske arkiv og gjenstadssamlinger fra museer, forskningsarkiver og bibliotek nå foreligger i elektronisk form.

For resultater av digitaliseringsarbeidet, henvises det til Museumsprosjektets databaser. Siden 1998 har DigForsk AS tatt på seg digitaliseringsoppdrag koordinert av det nasjonale Museumsprosjektet på vegne av ulike universitetsmuseer. Museumsprosjektet har siden laget databaser over det digitaliserte materialet. Materiale som ligger i disse databasene er i hovedsak digitalisert av DigForsk AS.

Meteorologisk institutt

Meteorologisk institutt er et statlig forvaltningsorgan under Kunnskapsdepartementet. Meteorologisk institutt utfører målinger og beregninger for allmennheten til bruk i:

- værvarsling og klimastudier
- forskning
- kommersielle tjenester

Formålet er uttrykt i vedtektene for Meteorologisk institutt, som innledes slik: Meteorologisk institutt står for den offentlige meteorologiske tjeneste for sivile og militære formål. Instituttet skal arbeide for at myndigheter, næringslivet, institusjoner og allmennheten best mulig kan vareta sine interesser for sikring av liv og verdier, for planlegging og for vern av miljøet.

Det innebærer følgende oppgaver:

- Utarbeide værvarsler.
- Studere Norges klima og gi klimatologiske utredninger.
- Innhente meteorologiske data i Norge, nærliggende havområder og på Svalbard.
- Drive forsknings- og utviklingsarbeid.
- Leverer flyværtjenester.
- Formidle resultatene av arbeidet.
- Utføre oppdrag og yte spesialtjenester.
- Delta i det internasjonale meteorologiske samarbeid.

Meteorologisk institutt utfører landsdekkende observasjoner av atmosfæriske variable fra ca. 700 stasjoner i Norge og Arktis. De aller fleste måleseriene finnes digitalt fra 1957; noen helt tilbake til 1800-tallet. I tillegg mottas observasjoner fra satellitt og radar og fra et globalt observasjonsnett. Alle dataene er tilgjengelig 24 timer i døgnet i en moderne database.

Instituttet legger vekt på å bearbeide og stille til rådighet klimadata slik at de kommer samfunnet til nytte i planlegging, konsekvensanalyser og klimastudier. Instituttet bidrar til bærekraftig samfunnsutvikling ved å utarbeide scenarier om mulige klimaendringer i framtiden. Høy kvalitet på data er avgjørende, og instituttet har et effektivt system for å kontrollere nye data før de legges i databasen.

Via SeNorge-portalene (<http://senorge.no/>) får publikum tilgang til digitale kart tilbake til 1960 for tema innen vær, vann, snø og klima. SeNorge er et samarbeidsprosjekt mellom Statens kartverk (SK), Norges vassdrags- og energidirektorat (NVE) og Meteorologisk institutt. Arbeidet med portalen har vært støttet av Norges forskningsråd.

Digitale data er tilgjengelig via internett. Men det er fortsatt store mengder data på analog form (bøker, kart, diagrammer) som ikke er tilgjengelig elektronisk. Her pågår det et arbeid med digitalisering av de viktigste lange måleseriene.

I samarbeid med NRK utvikles det en værportal som skal øke tilgjengeligheten av data og grafikk for værvarsling, klima og forskning. Det vil bli utviklet grensesnitt og produkter tilpasset brukernes (samfunnets) behov. Se: <http://yr.no/>

Forskning og utvikling utføres innenfor fagområdene meteorologi, klimatologi, oseanografi og transport av forurensning i atmosfære og hav. Numeriske værvarsler og varsler av bølger, strøm, vannstand og sjøis er sentrale oppgaver. Observasjoner, modeller og modellresultater anvendes også for å øke kunnskapsnivået om viktige prosesser i atmosfære og hav.

Instituttet har i dag en modell for beregning av langtransporterte luftforurensninger (EMEP-modellen). Modellen brukes for å kvantifisere den grenseoverskridende luftforurensningen mellom landene i Europa. Resultatene brukes til å støtte internasjonale forhandlinger om reduksjon av utslipp til luft.

Internasjonalt samarbeid

Virksomheten er helt avhengig av internasjonalt samarbeid. De viktigste samarbeidende organisasjoner er:

- WMO (grunnleggende værvarsling og klima)
- ECMWF (prognoser og forskning)
- EUMETSAT (satellittdata)
- EUMETNET (samarbeid mellom meteorologiske institutter i Europa)
- ECOMET (kommersielle tjenester)
- NORDMET (nordisk samarbeid værvarsling)
- HIRLAM (prognosemodeller værvarsling)

Instituttet deltar aktivt i WMOs arbeid med å utvikle nytt system for global utveksling av data, basert på moderne informasjonsteknologi. Instituttet har også en spesiell oppgave innenfor WMOs regional sammenslutning for Europa når det gjelder koordinering av observasjonssystemet innenfor regionen.

Internasjonal datapolitikk er formulert i resolusjoner i hver organisasjon. WMO er viktigst for met.no. Observasjoner distribueres globalt på dedikerte kanaler. Nesten alle data kan brukes fritt for værvarsling, klima og forskning.

Nasjonalt samarbeid

1. Norsk klimasenter

Meteorologisk institutt (**met.no**) og Bjerknessenteret for klimaforskning (BCCR) har formalisert et omfattende samarbeid innenfor klimaforskning under navnet *Norsk klimasenter*.

Samarbeidet vil blant annet bestå i forskning på grunnleggende forståelse for mekanismene bak klimaendringer. Senteret vil også gjennomføre vær- og klimavarsler for tidsperioder på sesong, tiår og 100 år. Målet er å komme fram til forskningsresultater som er relevante og brukbare for samfunnet.

Målet med Norsk klimasenter er

- å forstå årsakene til klimaendringer
- å kunne varsle disse klimaendringene
- å på best mulig måte bidra med kunnskap om hvordan klimaendringer vil påvirke fremtidig samfunnsutvikling nasjonalt og globalt

2. CIENS – Forskningssenter for miljø og samfunn

CIENS (<http://www.ciens.no/>) er et strategisk forskningssamarbeid mellom selvstendige forskningsinstitutter og Universitetet i Oslo. Senteret er basert på felles faglige strategier og forskningsprogram.

Fagprogrammet har detaljerte planer for forskningssamarbeid på følgende 6 fagtema:

- Klimaendringer
- En bærekraftig atmosfære
- Integreert vassdrags- og kystsoneforvaltning
- Bærekraftig byutvikling
- Bærekraftig transport
- Naturens produksjon av varer og tjenester

Deltakende institutter er:
CICERO
Senter for klimaforskning,

met.no
Meteorologisk institutt

NIBR
Norsk institutt for by- og regionforskning

NILU
Norsk institutt for luftforskning

NINA
Norsk institutt for naturforskning

NIVA
Norsk institutt for vannforskning

TØI
Transportøkonomisk institutt
Stiftelsen Norsk senter for
samferdselsforskning

UiO
Universitetet i Oslo Institutt for geofag

Assosiert medlem

NVE
Norges vassdrags- og energidirektorat

Norsk Marint Datasenter

Norsk Marint Datasenter (NMD) ble etablert som Norsk Oseanografisk Datasenter (NOD) i 1971. Datasenteret har hatt ansvaret som National Oceanographic Data Centre NODC i Norge siden etableringen. NMD er organisert som en underavdeling innen Havforskningsinstituttet med 12-18 årsverk som en gjennomsnittlig årlig innsats. Datasenteret langtidslagrer forskningsdataene ved Havforskningsinstituttet og tilbyr lignende tjenester med dataforvaltning av marine forskningsdata for andre norske institusjoner.

NMD sine aktiviteter organiseres internt i tre hovedområder som samsvarer med dataflytmodellen som foreslås.

Data inn

langtidslagring av forskningsdata,
operasjonell datahåndtering

Data bearbeiding

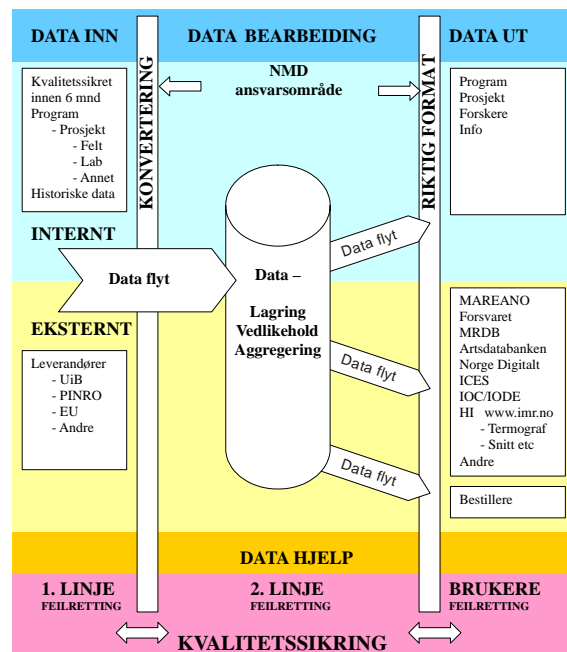
utvikle databasesystemer, 2. linje
kvalitetssikring

Data ut

søk i marine data, karttjenester, internett
presentasjon av data

NMD skal være det nasjonale
kontaktpunktet for marine data, også for
data som ikke eies av

Havforskningsinstituttet. NMD har som
hovedoppgave å lagre og gjøre tilgjengelig forskningsdata fra instituttet og andre datakilder. Lagringsmediet og databasestrukturer vil velges slik at de til enhver tid er hensiktsmessig.



Artsdatabanken

Artsdatabanken ble i 2004 opprettet som et resultat av at Regjering og Storting har ønsket et sterkere fokus på den kunnskapen som skal ligge til grunn for arbeidet med biologisk mangfold. Artsdatabanken er underlagt et styre oppnevnt av Kunnskapsdepartementet, og startet arbeidet i

januar 2005. Artsdatabanken holder til på Kalvskinnet i Trondheim, i tilknytning til NTNU. Artsdatabanken har 9 ansatte.

Artsdatabanken er en nasjonal kunnskapsbank for biologisk mangfold. Den viktigste oppgave er å forsyne samfunnet med oppdatert og lett tilgjengelig informasjon om norske arter og naturtyper.

I magasiner og databaser hos museer, forskningsinstitutter, forvaltningsorganer og frivillige organisasjoner ligger informasjon og data samlet inn gjennom mer enn hundre år. Dette er kunnskap som samfunnet har stor bruk for, men som det så langt ikke har vært mulig å ta i bruk på grunn av manglende kapasitet og infrastruktur.

Artsdatabanken samarbeider tett med de biologiske fagmiljøene for å bygge en faglig og teknisk plattform og infrastruktur som skal gjøre denne informasjonen synlig og tilgjengelig for allmennheten.

Målet er at alle som trenger det skal kunne skaffe seg eksisterende kunnskap om biologisk mangfold fra informasjon og tjenester som ligger på www.artsdatabanken.no.

I Norge er det kjent rundt 40 000 arter av dyr, planter og andre organismer. Det totale antallet er trolig nærmere 60 000 - hver og en avhengig av sitt spesielle livsmiljø og naturtype. Å skulle formidle kunnskap om biologisk mangfold i Norge er derfor en formidabel oppgave.

Artsdatabankens hovedsatsingsområder er:

1. Dokumentere status for truede arter (Norsk rødliste 2006)
2. Dokumentere økologisk risiko av fremmede arter (Norsk svarteliste 2007)
3. Utvikle en norsk naturtypeinndeling og drifte en tilhørende innholdsdatabase
4. Etablere og drifte en artsnavnebase for alle norske arter
5. Utvikle og drifte en webkarttjeneste for kvalitetssikrede geografiske artsdata
6. Utvikle og drifte en internettbasert registreringstjeneste for artsobservasjoner for publikum

Flere av tjenestene fungerer som felles infrastruktur for alle universitetsmuseene og andre forskningsinstitusjoner som til sammen har 40-50 databaser som skal levere innhold til formidlingstjenestene. Artsdatabanken har tett samordning med samarbeidspartene under utvikling av informasjonstjenestene.

Norge Digitalt

De fleste naturvitenskapelige forskningsinstitusjonene er med i Norge Digitalt, den felles infrastrukturen med geografiske data som gjør tilgjengelig både topografiske grunnkartdata og temadata (http://www.statkart.no/?_to=914). Norge Digitalt er et bredt samarbeid mellom virksomheter som har ansvar for å fremskaffe stedfestet informasjon og/eller er store brukere av slik informasjon. Samarbeidet som er andelsfinansiert, ble iverksatt fra 1. januar 2005.

Samarbeidet har sin forankring i Stortingsmelding nr. 30 (2002-2003) "Norge digitalt" – et felles fundament for verdiskaping. Statens kartverk er sekretariat for samarbeidet.

Norge digitalt er også knyttet til Inspire - Infrastruktur for geografisk informasjon i Europa. Inspire er et nytt EU-direktiv for geografisk infrastruktur i Europa. Direktivet trådte i kraft 15. mai 2007. Inspire skal bygge på de enkelte lands nasjonale geografiske infrastrukturer.

Norge digitalt og Inspire bygger på prinsippet om distribuerte databaser tilgjengelige i sanntid, bruk av enhetlige utvekslingsstandarder (metadata, WMS og WFS) og portaltjenester der brukeren kan søke og laste ned datasettene på tvers av tjenesteleverandører. Forskningsrådet har gjennom Høykom-satsingen bidratt vesentlig til at webservice-teknologien har blitt innført i de institusjonene som har fått støtte til utviklingsprosjekter.