



Tilgjengeliggjøring av forskningsdata

Policy for
Norges forskningsråd

Tilgjengeliggjøring av forskningsdata

Policy for
Norges forskningsråd

© Norges forskningsråd 2014

Norges forskningsråd
Postboks 564
1327 Lysaker
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: Design etc
Foto/ill. omslagsside: Digital Vision

Oslo, september 2014

ISBN 978-82-12-03361-0 (pdf)

Innhold

Tilgjengeliggjøring av forskningsdata.....	1
Innhold.....	3
Forord	4
1 Innledning.....	5
Mål	5
Forskningsrådets rolle	5
Avgrensing	6
Bruk av data i forskningsprosessen.....	6
2 Prinsipper og retningslinjer i Forskningsrådets policy	8
Hovedprinsipp.....	8
Unntak.....	8
Forskningsrådets retningslinjer	9
Begrepsavklaringer	10

Forord

Åpenhet og kunnskapsdeling står i sentrum for forskningen. Den raske teknologiske utviklingen fører til endringer i hvordan forskning utføres og resultatene deles. Forskningen blir i større grad drevet framover ut fra tilgangen til nye og store datamengder. Resultatene av offentlig finansiert forskning er et fellesgode som har verdi både for vitenskapen og samfunnet. Bedre tilgang til forskningsdata vil styrke kvaliteten i forskning, både ved at resultater kan valideres og etterprøves på en bedre måte, men også ved at datasett kan brukes på nye måter og i kombinasjon med andre datasett. Ny og utvidet bruk av forskningsdata vil legge til rette for utvidet tverrfaglig forskning, men også nyskaping i samfunns- og næringsutvikling. I tillegg vil deling av data føre til mer fornuftig ressursbruk innenfor forskningssektoren fordi man unngår duplisering av arbeid og fordi datasett kan benyttes i andre sammenhenger enn opprinnelig tenkt. Åpen tilgang til resultater og data vil også være verdifullt for utdanning og føre til at forskningen blir mer synlig i samfunnet.

Policyen indikerer hvordan Forskningsrådet vil benytte sine virkemidler til å fremme åpen tilgang til forskningsdata. Prinsippene og retningslinjene som presenteres er også ment som en veiledning for forskere som skal planlegge sine forskningsprosjekter. Policy og tilhørende retningslinjer vil være gjenstand for revisjon basert på erfaringer med implementeringen, internasjonal utvikling, endringer i juridisk rammeverk og nye teknologiske muligheter.

1 Innledning

Mål

Med denne policyen ønsker Forskningsrådet å legge til rette for økt grad av kvalitets-sikring, bevaring, tilgjengeliggjøring og deling av forskningsdata blant forskere. Dette skal bidra til:

- Forbedret kvalitet i forskningen gjennom bedre mulighet til å bygge på tidligere arbeider og sammenstille data på nye måter
- Gjennomsiktighet i forskningsprosessen og bedre mulighet for etterprøvnbarhet av vitenskapelige resultater
- Økt samarbeid og mindre duplisering av forskningsarbeid
- Økt innovasjon i næringsliv og offentlig sektor
- Effektivisering og bedre utnyttelse av offentlige midler

Forskningsrådets rolle

Å bevare og gjøre forskningsdata tilgjengelige for videre bruk må skje i samspill mellom forskere, forskingsinstitusjoner, vitenskapelige tidsskrifter, infrastrukturtilbydere og myndighetene. Dette krever nye måter å tenke på hos aktørene, kompetanseutvikling og opplæring og nye infrastrukturer og verktøy. Forskningsrådet ønsker å være en pådriver for bevaring og deling av forskningsdata. Dette betyr at vi vil tilrettelegge for samarbeid mellom aktørene, finansiere relevante aktiviteter og veilede forskningsmiljøene gjennom å:

- investere i infrastrukturer som støtter kvalitetssikring, langsiktig bevaring, deling og gjenbruk av forskningsdata
- bidra til god arbeidsdeling mellom offentlige infrastruktur- og tjenestetilbydere gjennom møteplasser og rådgivningsvirksomhet
- implementere prosedyrer i søknadsbehandlingen som sikrer at relevante søknader inneholder planer for datahåndtering
- implementere prosedyrer i prosjektoppfølgningen som fører til at planene for datahåndtering blir fulgt av prosjektene som har fått midler
- videreføre praksisen med krav i våre kontrakter om at forskningsdata skal arkiveres på en forsvarlig måte i minimum 10 år
- tilby oppdatert oversikt over godkjente, nasjonale (eventuelt internasjonale) infrastrukturer for lagring, arkivering og tilgjengeliggjøring av data
- oppfordre institusjonene til å utforme egne policyer og retningslinjer for økt tilgjengeliggjøring av forskningsdata

Avgrensning

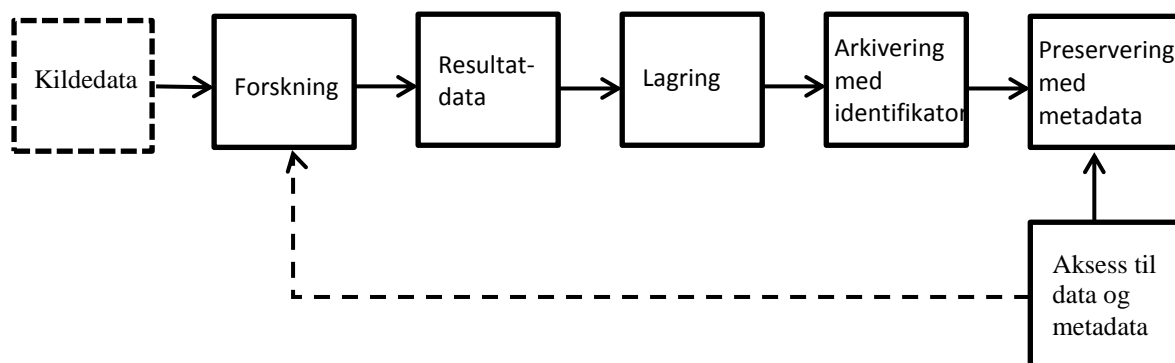
Kildedata er data som allerede finnes, uavhengig av forskningen som skal gjennomføres. Dette kan være informasjon som er innhentet til et annet formål (for eksempel forvaltningsdata, kliniske data eller værdata) eller det kan være fysiske eller digitaliserte samlinger av objekter og tekster (for eksempel biblioteker, tekstkorpus eller andre vitenskapelige samlinger). Informasjon på internett vil også kunne kalles kildedata i denne sammenheng, og her er informasjonen svært heterogen. Kildedata som brukes som inngangsdata i forskningen, men som er innsamlet, generert eller bearbeidet av andre enn forskerne eller forskningsinstitusjonene som gjennomfører forskningen, vil normalt ikke omfattes av retningslinjene i denne policyen.

Resultatdata er data som er generert gjennom forskning. Dette kan være data som er generert gjennom ny analyse eller sammenstilling av allerede eksisterende kildedata, men det kan også være helt nye data som er generert gjennom ny datainnsamling. Typisk vil slike data være fra eksperimenter, simuleringer, feltarbeid eller intervjuer. Resultatdata er alltid et direkte resultat av forskningen, uavhengig av om dataene baserer seg på kildedata eller om de er samlet inn på nytt. Resultatdata fra tidligere gjennomført forskning som brukes som input til ny forskning, betegnes derfor som kildedata ut fra definisjonen i avsnittet ovenfor.

Prinsippene og retningslinjene i denne policyen er avgrenset til å gjelde for resultatdata som er genererte i forskningsaktivitet helt eller delvis finansiert av Forskningsrådet.

Bruk av data i forskningsprosessen

Det er vanlig å bruke mange datakilder i forskning, og resultatdata fra forskning blir gjenbrukt, kombinert med andre data og distribuert både direkte og gjennom andres bruk av dataene. For at resultatdata skal være mulig å gjenbruke, må de bearbeides etter at de er generert. Vi kan illustrere dette med følgende fasedelte prosess¹:



Den første fasen etter generering er lagring på en fysisk infrastruktur. Når dataene er forsvarlig lagret og sikret i form av reservekopier, må de arkiveres. Denne fasen innebærer blant annet å sikre at dataene ikke kan endres og gjøre dataene gjenfinnbare ved hjelp identifikatorer. Neste fase består i å øke kvaliteten på dataene i form av å katalogisere dem og gjøre dem mer anvendelige. Denne fasen kan kalles preserveringsfasen, og det er her dataene berikes med metadata, som er beskrivende informasjon om

¹ Fasene etter generering av resultatdata er basert på <http://www.clir.org/initiatives-partnerships/data-curation>

dataene. Dette kan være beskrivelser av eksperimentet eller simuleringen hvor dataene framkom, referanser til standarder som er brukt i arbeidet og kategorisering av dataene. Det kan også være beskrivelser av kvaliteten på data og beskrivelser som gjør det mulig å søke etter dataene, som tidsrom, geografisk område, fagfelt og tema for forskningen. Uten gode metadata er det vanskelig å søke etter data, og de kan ikke brukes av andre fordi det vil være for stor usikkerhet knyttet til hva dataene faktisk representerer.

Etter preservering kan dataene brukes av andre, men det kan være nødvendig/ønskelig med en siste fase som består av kuratering av dataene (ikke vist på figuren). Kuratering innebærer å vedlikeholde og oppdatere dataene og gjøre dem relevante for videre bruk over tid.

Vi ser at lagring og arkivering er en forutsetning for å kunne gjenfinne resultatdata. Videre er generering av metadata en forutsetning for at resultatdata skal ha langsiktig verdi og kunne gjøres tilgjengelig for andre. Konsekvensen av dette er at en policy for tilgjengeliggjøring av forskningsdata også må omfatte lagring, arkivering og generering av metadata.

2 Prinsipper og retningslinjer i Forskningsrådets policy

Hovedprinsipp

Åpen tilgang

Forskningsrådets policy følger "åpen som standard"-prinsippet når det gjelder tilgang til forskningsdata. Forskningsrådet vil derfor bidra til at forskningsdata i utgangspunktet skal gjøres åpent tilgjengelig, men at det gjøres unntak for data som ikke kan eller bør gjøres tilgjengelig. I Horisont 2020 sin definisjon² av åpen tilgang er det i tillegg et krav at tilgangen skal være gratis. Forskningsrådet har her valgt å basere seg på at brukeren bør dekke de faktiske kostnadene knyttet til uthenting av data, og vi legger oss dermed nærmere OECDs definisjon³ av åpen tilgang, som sier at tilgang skal gis til lavest mulig kostnad, fortrinnsvis ikke mer enn marginalkostnadene for tilgjengeliggjøringen.

Unntak

Det finnes flere utfordringer knyttet til det å gjøre enkelte datasett åpent tilgjengelig. Årsaker til å begrense tilgjengeligheten kan være:

Sikkerhetshensyn

I tilfeller hvor tilgjengeliggjøring av dataene kan true enkeltmenneskers eller nasjonal sikkerhet, **skal** datasettene ikke gjøres åpent tilgjengelig.

Personsensitiv data

I tilfeller hvor tilgjengeliggjøring av dataene er i strid med gjeldende regelverk for personvern, **skal** datasettene ikke gjøres åpent tilgjengelig.

Andre juridiske forhold

I tilfeller hvor tilgjengeliggjøring av dataene strider med andre juridiske bestemmelser, **skal** datasettene ikke gjøres åpent tilgjengelige.

Kommersielle forhold

Data som har kommersiell verdi og er generert i prosjekter der en bedrift har kontrakt med Forskningsrådet, **kan** unntas fra det generelle prinsippet om åpen tilgang. I disse tilfellene anbefales det at dataene gjøres tilgjengelig etter en periode, forslagsvis etter 3 eller 5 år.

Andre forhold

I tilfeller hvor tilgjengeliggjøring av data får store økonomiske eller praktiske konsekvenser for dem som har generert/samlet inn dataene, **kan** datasettene unntas fra det generelle prinsippet om åpen tilgang dersom det argumenteres tilfredsstillende for dette.

² European Commission: "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020", Version 1.0, 11. Desember 2013. Tilgjengelig her som [PDF](#). [Lastet 16.04.2014].

³ OECD: "Principles and Guidelines for Access to Research Data from Public Funding", April 2007. Tilgjengelig her som [PDF](#). [Lastet 16.04.2014].

Forskningsrådets retningslinjer

Forskningsrådet har utarbeidet et sett av retningslinjer for arkivering, tilgjengeliggjøring og deling av forskningsdata. Retningslinjene vil bli fulgt opp gjennom Forskningsrådets virkemidler for finansiering av forskningsprosjekter og gjennom Nasjonal satsning på forskningsinfrastrukturer.

1.0 Forskningsdata bør lagres/arkiveres på en sikker måte

- 1.1 Dataene bør lagres i sikre arkiver, enten sentralt ved egen institusjon eller i nasjonale arkiver

2.0 Forskningsdata bør gjøres tilgjengelig for videre bruk

- 2.1 Forskningsdata bør gjøres tilgjengelige for alle relevante brukere, under like vilkår, så fremt det ikke er juridiske, etiske eller sikkerhetsmessige grunner til ikke å gjøre det

3.0 Forskningsdata bør gjøres tilgjengelig på et tidlig tidspunkt

- 3.1 Dataene som ligger til grunn for vitenskapelige artikler gjøres tilgjengelig så tidlig som mulig, og aldri senere enn ved publiseringstidspunkt
- 3.2 Andre data som kan være av interesse for annen forskning, bør gjøres tilgjengelig innen rimelig tid, og aldri senere enn tre år etter endt prosjekt

4.0 Forskningsdata bør utstyres med standardiserte metadata

- 4.1 Metadataene bør gjøre andre i stand til å søke etter og ta i bruk dataene
- 4.2 Metadataene bør følge internasjonale standarder
- 4.3 Metadata bør gi en beskrivelse av datakvaliteten

5.0 Forskningsdata bør utstyres med lisenser for tilgang, gjenbruk og videredistribusjon

- 5.1 Lisensene bør være internasjonalt anerkjente
- 5.2 Lisensene bør legge så få begrensninger som mulig på tilgang, gjenbruk og videredistribusjon av dataene

6.0 Forskningsdata bør gjøres tilgjengelig til lavest mulig kostnad

- 6.1 Metadata bør gjøres tilgjengelig uten kostnad og publiseres slik at de kan høstes maskinelt og brukes i søk etter forskningsdata
- 6.2 Forskningsdata bør fortrinnsvis gjøres tilgjengelig uten kostnad
- 6.3 Prisen for tilgang til forskningsdata bør aldri være høyere enn de faktiske kostnadene knyttet til tilgjengeliggjøring


7.0 Forskningsdata bør utstyres med en langtidsplan

- 7.1 Det bør utarbeides en plan for hvordan data som er vurdert til å ha verdi på lang sikt, skal forvaltes
- 7.2 Prosjektene bør ha et bevisst forhold til hvordan forskningsdata som er vurdert til ikke å ha langsiktig verdi, skal forvaltes, eventuelt destrueres etter en viss tid

Begrepsavklaringer

Noen av de sentrale begrepene innenfor tilgjengeliggjøring av forskningsdata er ikke alltid entydige. Det er derfor nødvendig å redegjøre for hva Forskningsrådet legger i disse begrepene for å tydeliggjøre hvilke betingelser som ligger til grunn for disse prinsippene og retningslinjene.

Forskningsdata	<i>Med forskningsdata mener vi i denne policyen registreringer/nedtegnelser/rapporteringer i form av tall, tekster, bilder og lyder som genereres eller oppstår underveis i forskningsprosjekter.</i>
Offentlig finansiert	<i>Alle prosjekter og aktiviteter som er helt eller delvis finansiert med offentlige midler regnes som offentlig finansierte.</i>
Åpen tilgang	<i>Med åpen tilgang menes det at forskningsdata skal være tilgjengelige for relevante brukere, på like betingelser, og til lavest mulig kostnad. Tilgangen skal være enkel, brukervennlig og, hvis mulig, internettbasert.</i>



Publikasjonen kan lastes ned fra
www.forskningsradet.no/publikasjoner

Norges forskningsråd

Drammensveien 288
Postboks 564
1327 Lysaker

Telefon: 22 03 70 00
Telefaks: 22 03 70 01
post@forskningsradet.no
www.forskningsradet.no

Omslagsdesign: Design et cetera AS
Foto/ill. omslagsside: Digital Vision

Oslo, september 2014

ISBN 978-82-12-03361-0 (pdf)