



Tilgjengeliggjøring
av forskningsdata
Revidert 2017

Policy for
Norges forskningsråd

Tilgjengeliggjøring av forskningsdata

Revidert 2017

Policy for
Norges forskningsråd

© Norges forskningsråd 2017

Norges forskningsråd
Besøksadresse: Drammensveien 288
Postboks 564
1327 Lysaker
Telefon: 22 03 70 00
post@forskningsradet.no
www.forskningsradet.no/

Publikasjonen kan lastes ned her:
www.forskningsradet.no/publikasjoner

Oslo, desember 2017

ISBN 978-82-12-03653-6 (pdf)

Innhold

Forord	2
1. Innledning.....	3
Mål.....	3
Forskningsrådets rolle	3
Definisjoner og avgrensninger	3
Bruk av data i forskningsprosessen	4
2. Prinsipper og retningslinjer i Forskningsrådets policy.....	5
Hovedprinsipp	5
Unntak for åpen tilgang.....	5
Forskningsrådets retningslinjer	6
Begrepsavklaringer	7

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 av 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 forskningen, 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 næringsutviklingen. I tillegg kan deling av data føre til at man unngår duplisering av arbeid og at datasett kan brukes i andre sammenhenger enn opprinnelig tenkt. Åpen tilgang til resultater og data vil også være verdifullt for utdanning og kunne føre til at forskningen blir mer synlig i samfunnet.

Policyen indikerer hvordan Forskningsrådet vil bruke virkemidlene sine for å fremme åpen tilgang til forskningsdata. Prinsippene og retningslinjene som presenteres, er også ment som en veiledning for forskere som skal planlegge forskningsprosjektene sine.

Policyen, som kom i 2014, vil oppdateres med passende mellomrom basert på erfaringer med implementeringen, internasjonal utvikling, endringer i juridisk rammeverk og nye teknologiske muligheter. Utgangspunktet for denne oppdateringen (2017), er at stadig flere forskningsfinansierer, inkludert EU, krever at FoU-prosjekter skal levere en datahåndteringsplan som beskriver hvordan datamaterialet i et prosjekt skal håndteres, fra innsamlingen av dataene til etter prosjektslutt. I tråd med dette har Forskningsrådet i 2017 innført nye retningslinjer for datahåndtering. Prinsippene og retningslinjene som presenteres i policyen er i tråd med regjeringens «Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata» (desember 2017). Hovedprinsippene fra 2014 er i all hovedsak de samme, men følgende nye momenter er tatt inn:

- Forskningsrådet har innført retningslinjer om at FoU-utførende institusjoner eller foretak skal vurdere om prosjekter som mottar finansiering fra Forskningsrådet skal utarbeide en datahåndteringsplan. Videre oppfordrer Forskningsrådet til at datahåndteringsplaner skal være offentlige og dermed bør publiseres åpent av FoU-utførende institusjon. Dette kan stimulere til økt åpenhet og mulighet for at fagmiljøer bedre kan følge og lære av sine fagfellers praksis.
- De internasjonale FAIR-prinsippene (se side 4) for viderebruk av forskningsdata, er tatt inn som et hovedprinsipp i policyen.
- Flere av retningslinjene på side 5 er nå omformulert fra «bør-» til «skal-utsagn». Dette er for å stimulere til at FAIR-prinsippene for viderebruk av forskningsdata følges.

1. Innledning

Mål

Med denne policyen ønsker Forskningsrådet å legge til rette for økt grad av kvalitetssikring, lagring, tilgjengeliggjøring og deling av data. Dette skal bidra til:

- forbedret kvalitet i forskningen gjennom å kunne bygge på tidligere arbeider og sammenstille data på nye måter
- gjennomsiktighet i forskningsprosessen og forbedret mulighet for etterprøvbare av vitenskapelige resultater
- økt samarbeid og mindre duplisering av forskningsarbeid
- økt innovasjon i næringslivet og i offentlig sektor
- effektivisering og bedre utnyttelse av offentlige midler

Forskningsrådets rolle

Lagring og tilgjengeliggjøring av forskningsdata for videre bruk må skje i samspill mellom forskere, forskningsinstitusjoner, vitenskapelige tidsskrifter, infrastrukturtilbydere og myndighetene.

Kompetanseutvikling, opplæring og investering i nye infrastrukturer og verktøy er nødvendig for å oppnå dette. Forskningsrådet ønsker å være en pådriver for økt lagring og tilgjengeliggjøring 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 lagring, deling og bruk av forskningsdata
- bidra til god arbeidsdeling mellom offentlige infrastruktur- og tjenestetilbydere gjennom møteplasser og rådgivningsvirksomhet
- videreføre praksisen med krav i våre kontrakter om at forskningsdata skal arkiveres på en forsvarlig måte i minimum 10 år
- oppfordre institusjonene til å utforme egne policyer og retningslinjer for sikker lagring, håndtering og økt tilgjengeliggjøring av forskningsdata
- legge til rette for at de ansvarlige FoU-utførende institusjonene eller foretakene tar stilling til om prosjekter som mottar finansiering fra Forskningsrådet skal utarbeide en datahåndteringsplan

Definisjoner og avgrensninger

Prinsippene og retningslinjene i denne policyen er avgrenset til å gjelde for forskningsdata som er genererte i forskningsaktivitet helt eller delvis finansiert av Forskningsrådet. Policyen fokuserer på forskningsdata som er tilgjengelig i digitalt format.

Med *forskningsdata* mener vi her registreringer/nedtegnelser/rapporteringer i form av tall, tekster, bilder og lyder som genereres eller oppstår underveis i forskningsprosjekter. Dette kan for eksempel være data som er generert gjennom ny analyse, sammenstilling av eksisterende data, eller helt nye data generert gjennom ny datainnsamling. Forskningsdata er alltid et direkte resultat av forskningen, uavhengig av om dataene baserer seg på eksisterende data eller om de er samlet inn på nytt.

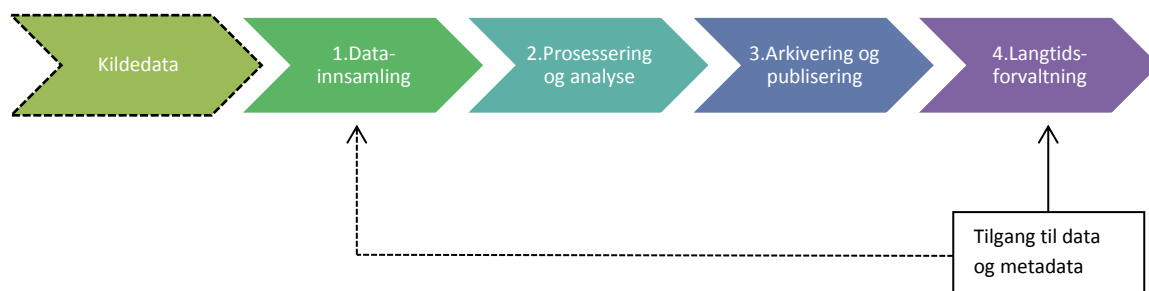
Data som allerede finnes uavhengig av forskningen som skal gjennomføres, betegner vi som *kildedata*. Dette kan være informasjon som er innhentet til et annet formål (for eksempel

forvaltningsdata, kliniske data eller værdata) eller det kan være fysiske eller digitaliserte samlinger av objekter og tekster (for eksempel biblioteker, eller data som gjenbrukes fra tidligere prosjekter, f.eks. tekstkorpus eller andre vitenskapelige samlinger). Informasjon på internett vil også kunne kalles kilddata i denne sammenhengen, og her er informasjonen svært heterogen. Kilddata 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.

Bruk av data i forskningsprosessen

For at forskningsdata skal være mulig å gjenbruke, må de bearbeides etter at de er generert.

Verdikjeden for digitale forskningsdata er basert på fire hovedtrinn:



1. Det første steget er innsamling eller produksjon av data, der data lagres på et arbeidsområde. Det er god praksis å bruke aksepterte dokumentasjonsstandarder allerede i dette steget, da det gjør senere arbeid med langtidslagring og publisering enklere.
2. I det andre steget kvalitetssikres dataene. I denne prosessen avgjøres det om rådata skal tas vare på. Videre arbeid i denne fasen er analyse av de kvalitetssikrede dataene.
3. Det tredje steget er tilrettelegging av data for langtidslagring (arkivering) og publisering. Dette innebærer disiplinsspesifikk dokumentasjon og koding av data (metadata) som beskriver hvem som er ansvarlig for dataene, hva som er i dataene, hva som er gjort med dataene, hvem som kan bruke dem til hva, osv. Disse metadataene er grunnlaget for andre forskeres mulighet til å finne og bruke dataene. Relevant informasjon om dataene vil indekseres i søkemotorer/kataloger når dataene overleveres til et dataarkiv. Når dataene lagres i et dataarkiv, tildeles datasettet en identifikator som følger datasettet hele levetiden. Dataene er nå tilgjengelig for andre forskere.
4. Når dataene er lagret i et arkiv bevares de for fremtiden. Dette innebærer vanligvis en lagringsperiode på minimum 10 år og oppover. Teknologier endrer seg – det gjør at data i et langtidslager må tilpasses nye teknologier og formater for å sikre at de er søkbare og anvendelige også i fremtiden.

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 2020s definisjon¹ av åpen tilgang, er det krav om at tilgangen – i den grad det er mulig – skal være gratis. Forskningsrådet har her valgt å basere seg på at brukeren bør dekke de faktiske kostnadene knyttet til uthenting av data. Vi legger oss dermed nærmere OECDs definisjon² av åpen tilgang, som sier at tilgang skal gis til lavest mulig kostnad, fortrinnsvis ikke mer enn marginkostnadene for tilgjengeliggjøringen.

FAIR-prinsippene for god datahåndtering

De internasjonale FAIR-prinsippene er utarbeidet som et sett av retningslinjer for å tilrettelegge for videre bruk av forskningsdata. FAIR er et akronym for ordene *findable*, *accessible*, *interoperable* og *reusable*³. Forskningsdata skal med andre ord være av en kvalitet som gjør dem tilgjengelige, gjenfinnbare og gjenbrukbare. Videre ligger det i begrepet *interoperable* at både data og metadata skal kunne håndteres maskinelt og at man skal bruke konsistente vokabular.

Unntak for åpen tilgang

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.

¹ European Commission: "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020", Version 3.2, 21 March 2017 [Lastet 20.11.2017].

² OECD: "Principles and Guidelines for Access to Research Data from Public Funding", April 2007 [Lastet 20.11.2017] og OECD: "Making Open Science a Reality" 2015. OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris.

³ Wilkinson, Mark D. mfl. (2016) "The FAIR Guiding Principles for scientific data management and stewardship". Scientific Data. 3(160018). <https://doi.org/10.1038/sdata.2016.18>.

Se også: <https://www.force11.org/group/fairgroup/fairprinciples>

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.

Forskningsrådets retningslinjer

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

1.0 Forskningsdata skal lagres/arkiveres på en sikker måte

- 1.1 Dataene skal lagres i sikre arkiver, enten sentralt ved egen institusjon eller i nasjonale arkiver
- 1.2 Forskningsutførende institusjon har ansvar for å velge relevante arkivløsninger for ulike forskningsdata som er i samsvar med FAIR prinsippene
- 1.3 Prosjektene bør ha et bevisst forhold til hvordan forskningsdata av langsiktig verdi skal forvaltes
- 1.4 Prosjektene bør ta stilling til hvordan data som er vurdert til ikke å ha langsiktig verdi, skal forvaltes, eventuelt destrueres etter en viss tid

2.0 Forskningsdata skal gjøres tilgjengelig for videre bruk

- 2.1 Forskningsdata skal 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 skal utstyres med standardiserte metadata

- 4.1 Metadataene skal 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 skal 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 Håndtering av forskningsdata skal beskrives i en datahåndteringsplan


- 7.1 Forskningsutførende institusjon skal godkjenne at datahåndteringsplaner er i tråd med institusjonens krav og forventninger. Planen bør oppfylle FAIR-prinsippene
- 7.2 Datahåndteringsplaner bør så langt det er mulig være offentlige og publiseres åpent av forskningsutførende institusjon slik at fagmiljøer bedre kan følge sine fagfellers praksis

FAKTABOKS

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.

Forskningsdata	<i>Med forskningsdata mener vi i denne policyen registreringer/hedtegnelser/rapporteringer i form av tall, tekster, bilder og lyder som genereres eller oppstår underveis i forskningsprosjekter.</i>
Kildedata	<i>Kildedata er data som allerede finnes, uavhengig av forskningen som skal gjennomføres.</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>
Datahåndteringsplan	<i>En datahåndteringsplan (på engelsk: data management plan, forkortet DMP) er et dokument som beskriver hvordan forskningsdataene i et prosjekt skal håndteres, fra prosjektets start til slutt.</i>
Metadata	<i>Metadata er data som definerer eller beskriver andre data.</i>
Digital Object Identifier (DOI)	<i>DOI er en varig og unik identifikator for forskningsdata og publikasjoner</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, desember 2017

ISBN 978-82-12-03653-6 (pdf)