

Utvalg:	Forskningsutvalget	Dato: 08.09.16
FU-sak:	FU 14/16	Arkivsaknr.:

Lagring og tilgjengeliggjøring av forskningsdata. Orientering

1 Innledning

Ved Universitetet i Bergen er det lagt til rette for åpen digital publisering av vitenskapelige artikler, såkalt Open Access. Open Access innebærer at vitenskapelige publikasjoner gjøres fritt, umiddelbart og permanent tilgjengelig for alle på internett. Hovedprinsippene er at publikasjonen blir publisert umiddelbart og med permanent åpen tilgang. Forfatter eller opphavsmann beholder opphavsretten til publikasjonen, men gir brukere, uten å forlange vederlag, tillatelse til å lese, laste ned, kopiere, distribuere, skrive ut, søke i eller lenke til fullteksten.

Universitetet i Bergen gjennomførte et pilotprosjekt med en budsjettpost for kostnader ved Open Access-publisering i årene 2013-2015. Prosjektet ble oppsummert som en suksess og budsjettposten er vedtatt videreført, men med nye retningslinjer gjeldene fra 01.10.16.

Fra eksterne kilder til finansiering av forskning, som Norges Forskningsråd og EU gjennom H2020, stilles det nå krav om at forskning finansiert av offentlig midler skal bygge på Open Access prinsipper om åpenhet og tilgjengeliggjøring av forskningsdata. Dette notatet gir en kort fremstilling av grunnlaget for og utfordringer knyttet til lagring og tilgjengeliggjøring av forskningsdata.

2 Prinsipper og definisjoner

2.1 OECD

OECD er blant de organisasjoner som har utarbeidet forslag til prinsipper og retningslinjer i sin Principles and Guidelines for Access to Research Data from Public Funding fra 2007. OECD legger vekt på tre sentrale forhold. I første rekke er det tale om forskningsdata som utgjør et primærmateriale for forskningen, at det er i all hovedsak tilgjengelig i et digitalt format, og at det er finansiert av offentlige midler.

In the context of these Principles and Guidelines, “research data” are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.

These Principles and Guidelines are principally aimed at research data in digital, computer-readable format. It is indeed in this format that the greatest potential lies for improvements in the efficient distribution of data and their application to research because the marginal costs of transmitting data through the Internet are close to zero. These Principles and Guidelines could also apply to analogue research data in situations where the marginal costs of giving access to such data can be kept reasonably low.

Research data from public funding is defined as the research data obtained from research conducted by government agencies or departments, or conducted using public funds provided by any level of government.

2.2 NFR

NFRs policy for tilgjengeliggjøring av forskningsdata ble lagt frem i 2014 og hadde som formål å legge til rette for økt grad av kvalitetssikring, bevaring, tilgjengeliggjøring og deling av forskningsdata blant forskere. I forhold til den mer åpne OECD-definisjonen gjør NFR viktige avgrensninger ved å påpeke at prinsipper og retningslinjer er avgrenset til å gjelde resultatdata som er generert i forskningsaktivitet som helt eller delvis er finansiert av Forskningsrådet. **Resultatdata** er definert som data som er generert gjennom forskning, og vil typisk være «fra eksperimenter, simuleringer, feltarbeid eller intervjuer. Resultatdata er alltid et direkte resultat av forskningen, uavhengig av om dataene baserer seg på kilde-data eller om de er samlet inn på nytt.

Data som allerede finnes, uavhengig av forskningen som skal gjennomføres, uavhengig av forskningen som skal gjennomføres er **kilde-data**, som ikke omfattes av definisjonen for data som skal åpent tilgjengeliggjøres (Open Access). Dette kan være informasjon som er innhentet til et annet formål (for eksempel forvaltningsdata, kliniske data eller værd-data) eller det kan være fysiske eller digitaliserte samlinger av objekter og tekster (for eksempel, tekstkorpus eller andre vitenskapelige samlinger).

2.3 H2020

I EUs program for forskning og utvikling, Horisont 2020, er det et obligatorisk krav at publikasjoner skal publiseres Open Access. For forskningsdata er det imidlertid introdusert en prøveordning, Open Research Data Pilot. I arbeidsprogrammet fra 2014-16 var kun utvalgte fagområder med i ordningen, men fra 2017 er den utvidet til å omfatte alle tematiske områder i H2020. Å delta i ORD Pilot følger prinsippet om "as open as possible, as closed as necessary" og er i første rekke et tiltak som skal bidra til god datahåndtering som en essensiell del av «best practice» innen forskning.

Sentralt i datahåndteringen er en såkalt Data Management Plan (DMP) som bl.a. skal inkludere:

- •the handling of research data during & after the end of the project
- •what data will be collected, processed and/or generated
- •which methodology & standards will be applied
- •whether data will be shared/made open access and
- •how data will be curated & preserved (including after the end of the project).

Planer og ordninger for datahåndtering er på denne måten blitt integrert i det institusjonelle arbeidet med å søke om og å sikre inntekter fra EUs rammeprogram for forskning og utvikling.

3 NFRs anbefalinger

NFR policy for tilgjengeliggjøring av forskningsdata ble lagt frem i 2014 og hadde som formål å legge til rette for økt grad av kvalitetssikring, bevaring, tilgjengeliggjøring og deling av forskningsdata blant forskere.

NFR legger «åpen som standard» prinsippet til grunn for tilgjengeliggjøring av forskningsdata, men gjør viktige unntak. Hensyn til enkeltmenneskers eller nasjonal sikkerhet, hensyn til gjeldende regelverk for personvern og dersom tilgjengeliggjøring strider med andre juridiske bestemmelser skal

forskningsdata ikke gjøres åpent tilgjengelige. Data som har kommersiell verdi og er generert i prosjekter der en bedrift har kontrakt, kan unntas fra det generelle prinsippet om åpen tilgang.

Fremlegget har et sett av retningslinjer for arkivering, tilgjengeliggjøring og arkivering av forskningsdata og vil bli fulgt opp gjennom NFRs virkemidler for finanseringa av forskningsprosjekter gjennom nasjonal satsing på infrastruktur. I hovedtrekk er retningslinjene følgende syv hovedpunkt:

- Forskningsdata bør lagres/arkiveres på en sikker måte
- Forskningsdata bør gjøres tilgjengelig for videre bruk
- Forskningsdata bør gjøres tilgjengelig på et tidlig tidspunkt
- Forskningsdata bør utstyres med standardiserte metadata
- Forskningsdata bør utstyres med lisenser for tilgang, gjenbruk og videredistribusjon
- Forskningsdata bør gjøres tilgjengelig til lavest mulig kostnad
- Forskningsdata bør utstyres med en langtidsplan

4 Datalagring ved Universitetet i Bergen

4.1 BORA

BORA Bergen Open Research Archive (BORA) er Universitetet i Bergens (UiB) åpne forskningsarkiv. Alt materiale i BORA er åpent tilgjengelig i fulltekst på internett (Open Access). BORA sitt formål er å tilgjengeliggjøre og synliggjøre vitenskapelige publikasjoner, doktorgradsavhandlinger og masteroppgaver produsert av ansatte og studenter ved institusjonen. BORA skal også sørge for sikker langtidslagring av det materialet som tilgjengeliggjøres. I tilknytning til lagring av publikasjoner kan det i begrenset grad være anledning til å lagre forskningsdata i BORA.

4.2 Lie – utvalget

En arbeidsgruppe for sikker lagring av og tilgang til forskningsdata ved UiB ble oppnevnt av rektor i februar 2013. Bakgrunnen var bl.a. universitetsstyrets behandling av universitetets implementering av lov om medisinsk og helsefaglig forskning i forbindelse med revisjon av universitetets internkontrollsystem for behandling av personopplysninger i forskningsprosjekter. Gruppen var ledet av professor Rolv Terje Lie fra Det medisinsk-odontologiske fakultet. Et viktig grunnlagsdokument var rapporten fra «Rokne-gruppen» fra 2011 om Implementering av helseforskningsloven og revisjon av internkontrollsystemet for behandling av personopplysninger ved Universitet i Bergen».

Utvalget kom med anbefalinger som var særlig knyttet til prosjekter som i hovedsak vil være UiB-interne og som behandler sensitive persondata.

4.3 Forelegg fra Datatilsynet

Universitetet i Bergen ble i februar 2016 ilagt et overtredelsesgebyr (bot) for urettmessig utlevering av helseopplysninger og mangelfull oppfyllelse av krav om internkontroll. Et forskningsprosjekt om «Innvandrerhelse fra begge sider» data fra Reseptregisteret, Fastlegedatabasen, Folkeregisteret, Helfo og Medisinsk fødselsregister. UiB hadde konsesjon for å bruke dataene internt ved universitetet, men ved en feil ble datafiler også delt med forskere ved andre institusjoner uten at UiB søkte om egen konsesjon. UiB meldte selv fra om dette til Datatilsynet, og kontaktet samarbeidspartnerne som bekreftet at de slettet dataene. Datatilsynet påpeker i et brev til UiB at universitetet har vist åpenhet i saken og gjort det de kan for å begrense skaden.

I den videre oppfølging ble Universitetet i Bergen pålagt å utarbeide dokumentasjon på internkontroll for forskning i tråd med helseforskningsloven. Datatilsynet fikk derfor en nærmere beskrivelse av den systematiske opplæring av prosjektledere, forskere og andre som har oppgaver relatert til behandling av personopplysninger i medisinsk og helsefaglig forskning, og de etablerte rutinene for informasjonsflyt og kunnskapsoppdatering

Revisjon av dagens rutiner vil skje som en del av et større arbeid med å samkjøre de to internkontrollsystemene for behandling av personopplysninger i forskning til ett kvalitetssystem. UiB skal utvikle et elektronisk system for risiko og etterlevelse av relevante lover og forskrifter (RETTE). Systemet er tenkt å fungere prosessstøttende til rutinene nedfelt i internkontrollsystemet. Systemet har som mål å gi UiB en samlet oversikt over pågående forskningsprosjekter som behandler personopplysninger

4.4 SAFE - bedre personvern i forskningsprosjekter

SAFE står for Sikker adgang til Forskningsdata og E-infrastruktur, og er UiBs elektroniske løsning for sikker behandling av sensitive personopplysninger i forskning. I løpet av 2016 blir SAFE innført på hele universitetet. SAFE bygger på Norm for informasjonssikkerhet i helse- og omsorgstjenestene (Normen) og sikrer at informasjonssikkerheten med hensyn til konfidensialitet, integritet og tilgjengelighet blir ivarettatt ved behandling av sensitive personopplysninger.

SAFE skal benyttes ved behandling av sensitive personopplysninger i forskning. En "personopplysning" er en opplysning eller vurdering som kan knyttes til en enkeltperson om:

- rasemessig eller etnisk bakgrunn, eller politisk, filosofisk eller religiøs oppfatning,
- at en person har vært mistenkt, siktet, tiltalt eller dømt for en straffbar handling,
- helseforhold,
- seksuelle forhold eller
- medlemskap i fagforeninger

Kravene til informasjonssikkerhet gjelder for direkte eller indirekte identifiserbare personopplysninger.

5 Utredning ved Universitetet i Oslo

Ved Universitetet i Oslo ble det nedsatt en arbeidsgruppe som leverte en rapport i mai 2015. Arbeidsgruppen fikk i oppgave å kartlegge eksisterende tjenester og praksis, forskernes behov for lagring og deling av forskningsdata ved UiO samt utvikle forslag til prinsipper og retningslinjer for lagring og deling/tilgjengeliggjøring av forskningsdata som ivaretar vitenskapelige ansattes rettigheter og å anbefale løsninger som samsvarer med finansieringskildenes økende krav til lagring og deling, og forskernes egne behov.

Rapporten påpeker at det til tross for god praksis og velfungerende lokale ordninger, er den generelle «situasjonen ved UiO (og nasjonalt) ikke tilfredsstillende ut fra målbildet. Et fåtall av systemene som brukes ved UiO til lagring av egne forskningsdata, kan samtidig brukes til åpen eller tilgangsstyrt deling. Flertallet av lagringssystemene som brukes på UiO er interne, hvor forskningsdata kun kan deles innenfor forskningsgruppen internt på UiO».

Arbeidsgruppen lanserte retningslinjer i tråd med Forskningsrådet anbefalinger, nevnt ovenfor.

6 Norsk senter for forskningsdata

Norsk senter for forskningsdata (NSD) er en nasjonal institusjon med et av verdens største arkiv for forskningsdata og formidler av data og datatjenester til forskningssektoren.

Hovedmålsettingen er å forbedre muligheter og arbeidsvilkår for empirisk forskning som først og fremst er avhengig av tilgang til data.

Forsker/produsent avgjør selv hvor tilgjengelig dataene skal være for eventuell gjenbruk, men målet er at mest mulig data skal være tilgjengelig for videre forskning. Hovedmålet er at offentlig finansiert forskning skal være fritt tilgjengelig.

Datasekk som publiseres med data på NSDs nettsider er fritt tilgjengelig for nedlasting av bruker. Datasekk som publiseres uten data er tilgjengelig etter søknad til NSD og, i noen tilfeller, etter godkjenning fra forsker/produsent. Det gjelder egne regler for utlån av personidentifiserbare data.

NSD er tildelt Data Seal of Approval (DSA). DSA er del av et europeisk rammeverk for evaluering og sertifisering av digitale lagringstjenester og består av en rekke krav som må være oppfylt. Kravene er knyttet til prosedyrer for å hente inn data, hvordan data forvaltes og hvordan data distribueres. Gjennom sertifiseringen og det arbeidet som ligger bak, forsikres dataprodusenter og forskere om at data bevares og forvaltes på en sikker og pålitelig måte.

7 Avdelingsdirektørens kommentarer

Sikker bruk av data i forskning, ikke minst av personsensitive data, har fått mye oppmerksomhet ved UiB og er fulgt med konkrete tiltak. Lie-utvalget har gitt viktige anbefalinger og i det interne revisjonsarbeidet er det satt i gang et utvikling av et elektronisk system for risiko og etterlevelse av relevante lover og forskrifter (RETTE). Dette vil være et sentralt virkemiddel for fortsatt sikker bruk av personsensitive data.

Effektive ordninger for lagring og tilgjengeliggjøring av forskningsdata har imidlertid også et videre perspektiv som byr på nye utfordringer og muligheter. Datamengden i alle fag øker nærmest eksplodivt og krever utvikling og implementering av internasjonale, nasjonale og institusjonelle ordninger. Slike ordninger er av strategisk betydning fordi faglig utvikling, høy internasjonal kvalitet og innovasjon i forskningen forutsetter omfattende, oppdaterte, og relevante data som er tilgjengelige, siterbare og analyserbare.

Skal UiB nå sine strategiske mål er det avgjørende at det utvikles og etableres institusjonelle ordninger som sikrer trygg lagring, sikker bruk og åpen tilgjengeliggjøring av forskningsdata, i tillegg til det som i dag lagres i og er tilgjengelige i BORA og SAFE.

Det er naturlig at de anbefalte retningslinjene fra Norges Forskningsråd legges til grunn. Før slike retningslinjer kan vedtas er det imidlertid nødvendig å gjennomføre en kartlegging av behovene, som ventelig vil variere i samsvar med ulike forskningstradisjoner, metodologi og studieobjekter. Ved andre institusjoner ser man at bibliotek, IT-avdelinger og forskningsadministrasjon i fellesskap tilrettelegger for gode løsninger.

Universitetet i Bergen har truffet tiltak som vil sikre bruk av personsensitive data. Dette må opprettholdes og videreutvikles med ordninger og retningslinjer også for datalagring og tilgjengeliggjøring av andre forskningsdata.

Etter avdelingsdirektørens oppfatning er det i tillegg avgjørende at UiBs retningslinjer for legger vekt på enkeltforskernes rettigheter og samtidig gir gode muligheter for utvikling av institusjonelle ordninger for datahåndtering, slik det bl.a. legges opp til i H2020 og EUS fremtidige rammeprogrammer.

Saken legges med dette frem for utvalget til orientering og drøfting.