

Hvordan håndterer du forskningsdata?

Hvordan unngå at forskningsdata går tapt, ikke minst når forskere går av med pensjon? Kronikkforfatterne slår et slag for god datahåndtering.

Hva er de praktiske konsekvensene av at både EU (Horizon 2020) og norske myndigheter (Norges Forskningsråd og Kunnskapsdepartementet) i det siste har markert tydelig at ikke bare publikasjoner, men også forskningsdata som genereres gjennom offentlig finansierte prosjekter, skal være tilgjengelige for alle? Undertegnede ønsker å dele noen erfaringer som kanskje kan være nyttige for andre, både på institusjons- og på enkeltforskernivå.

Moralsk plikt

God forvaltning, og deling av forskningsdata, er helt sentralt for å fremme transparens og kvalitet i forskningen – diskusjonene rundt dette er over. Vi er alle enige om at ordlydene fra forskningsfinansiererne som sier «As open as possible, as closed as necessary» og «Åpen som standard» er både viktige og riktige. For eksempel professor Robert T. Barrett ved UiT og Tromsø Museum – han nærmer seg pensjonsalderen og har samlet inn data om sjøfugler og trekkfugler i over 40 år. Nå blir alle hans data fra et langt forskerliv tilgjengelig for alle. Barrett mener det er flott at alle hans innsamlede forskningsdata kan benyttes av andre i framtida. «Det er bare positivt og artig om mine data og tidsserier kan komme andre til gode, og ekstra fint blir det jo om de refererer til meg. Jeg får en sitering, og samtidig vet andre at det faktisk var jeg som samlet inn dataene.» Per i dag har han arkivert og delt nærmere 30 datasett i UiT Open Research Data-arkivet. «Vi forskere kan alltid skjule oss bak at vi ikke har tid, men vi har faktisk en moralsk plikt. En ting er hva vi har publisert, men man

«Arkivering og deling av data har en veldig sterk smitteeffekt mellom forskere.»



AV LARS FIGENSCHOU, STEIN HØYDALSVIK OG LEIF LONGVA, Universitetsbiblioteket i Tromsø, UiT Norges arktiske universitet

burde være såpass stolt av det man har gjort i løpet av forskerkarrieren, at man lar alt materialet være offentlig tilgjengelig. Til glede for yngre forskere, og alle våre etterfølgere.»

Åpne data ved UiT

Parallelt, og i forlengelsen av arbeidet med åpen tilgang for artikler, har det vært naturlig for oss ved UiT å gjøre det samme med forskningsdata. Vi er enige om at «Open Science» er målet, og derfor måtte vi ha på plass en egen institusjonell policy. Ikke for å påføre forskerne nok en administrativ byrde, men som et styrende dokument i institusjonens streben etter en mer «up to date» håndtering av forskningsdata. Det er jo faktisk arbeidsgiverne våre som i hovedregel har eierskap til de forskningsdataene som forskerne våre genererer. Så ved UiT skal nå forskningsdata gjøres åpent tilgjengelige, såfremt det ikke er juridiske, etiske, sikkerhetsrelaterte eller kommersielle grunner til ikke å gjøre det. For å hjelpe forskerne våre har vi utviklet ulike støttetjenester, og da har det vært helt naturlig å ta utgangspunkt i de såkalte FAIR-prinsippene («Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data»), som nettopp er et sett med retningslinjer for hvordan man kan gjøre forskningsdata mest mulig tilgjengelige (se Conzett & Østvand, 2018 for en mer detaljert beskrivelse av utviklingen av støttetjenester ved UiT).

Det ble utviklet og etablert et eget arkiv for åpne data (UiT Open Research Data) for våre

KRONIKK & DEBATT



Send debattinnlegg og kronikkforslag til aksel.vidnes@forskerforum.no
Maks lengde på kronikk: 9000 tegn med mellomrom. Honorar for trykte kronikker: kr 2000.
Maks lengde på debattinnlegg: 7000 tegn med mellomrom, men kortere innlegg har større sjanse for å bli trykket.
Debattinnlegg honoreres ikke.

egne UiT-forskere, og deretter etablerte vi et arkiv for eksterne institusjoner og forskere (DataverseNO). Parallelt fant vi det nødvendig å lokalisere et passende arkiv og/eller tjeneste for lagring eller publisering av datahåndteringsplaner, noe vi fortsatt jobber med. Videre etablerte vi et undervisningsopplegg hvor forskerne og ph.d.-studentene kunne lære hvordan man utarbeider en datahåndteringsplan, og hvordan man på best mulig måte, etter FAIR-prinsippene, arkiverer og deler data. Dette undervisningsopplegget måtte matche de krav og forventninger som EU og NFR nå la på våre forskere og ph.d.-studenter. I tillegg ble det laget en egen forskningsdataportal ved UiT, hvor forskerne skal finne all nødvendig informasjon og hjelp.

Fordelene ved god datahåndtering

Vi har gjennomgående samarbeidet med de mest motiverte forskerne, og så brukt disse som ambassadører når tjenestene våre etter hvert skulle gjøres kjent. I tillegg har det selvfølgelig vært nødvendig å formidle hvilke fordeler forskerne har av god forskningsdatahåndtering. Vårt arbeid med å bevisstgjøre fagmiljøene og forskerne, samt alle implementerte opplæringstiltak, har så vært fulgt opp med to ulike prosjektbaserte tiltak. I 2015 kjørte vi i gang et pilotprosjekt hvor vi tok tak i alle positive forskere, uavhengig av disiplin, som ville arkivere og gjøre sine forskningsdata tilgjengelige for andre. Hensikten med dette prosjektet var å (i): få erfaring med tilrettelegging av teknisk infrastruktur for et institusjonelt arkiv for forskningsdata, (ii): kartlegge behov for kompetansebygging og støttetjenester for forskere som skal bruke en slik tjeneste, (iii): foreslå en forvaltningsmodell for forskningsdata ved institusjonen, og sluttelig (iv): prøve å beskrive hva som må til for å iverksette en slik tjeneste i full skala. I 2017 startet vi så «Seniorprosjektet» (et treårig prosjekt) som har som formål «å sikre at verdifulle forskningsdata ikke går tapt når en seniorforsker ved institusjonen slutter, og å gjøre dataene anvendbare i fremtidig

«Vil du prøve å være en aktiv og åpen forsker, som ikke går baklengs inn i fremtiden?»

forskning». Det finnes mye empiri på at dataene som ligger hos forskerne som er på vei ut av forskningen (pensjon), er de som er mest utsatte for å gå tapt, og dette er et kjempeproblem for institusjonene. I dette prosjektet hjelper vi enkeltforskere med tilrettelegging av deres egne forskningsdata, nettopp med henblikk på trygg lagring og tilgjengeliggjøring. Vi vurderer tilstand, ev. mangel på struktur og metadata, sårbarhet, hvilke formater dataene er lagret på, graden av unikhet og så videre. Seniorprosjektet vårt har nådd ut til mange, men vi merker at det er en slags treghet i systemet. Med det mener vi at veien er veldig lang fra det øyeblikk en forsker bestemmer seg for at «det å ta vare på og dele data» er en god ide – til det øyeblikket det faktisk kan publiseres ett nytt datasett i UiT Open Research Data-arkivet. Dette kan sikkert ha mange grunner. De som imidlertid synes å være de mest presserende årsakene, sett fra vårt ståsted, er at mange (eldre) gjennom en lang forskerkarriere ikke har tenkt på struktur, det å dokumentere dataene sine godt, og andre beskrivelser og metadata. Derfor er det faktisk en ganske stor faglig jobb som må gjøres, før vi i prosjektet kan ta over og gjøre vår del av jobben (det å være en hjelper og data-kurator). Det skal dog sies at det også finnes eksempler på det motsatte. På den korte tiden prosjektet har vært i gang, ser vi også at det har skjedd mye positivt ved vår institusjon, UiT, og særlig med hensyn til de generelle holdningene ovenfor «Open Science». Prosjektet har fått veldig god drahjelp fra den nye forskningsdatapolicyen, og samtidig opplever vi at for mange forskere (og kanskje særlig de eldre) er dette med arkivering og deling av forskningsdata en modnings sak. Når de så

får bruke den tiden det tar å tenke igjennom hva dette innebærer, så er de aller fleste enige med oss. Økt publisering av forskningsdata gir økt synlighet og er til fordel, både for egen forskning, for institusjonen og for samfunnet. Samtidig ser vi også at arkivering og deling av data har en veldig sterk smitteeffekt mellom forskere.

Hva kan du selv gjøre?

Når det gjelder håndtering av forskningsdata, så er det ikke så mye som må til. Det er nok at den gjennomsnittlige forskeren endrer bitte litt på arbeidsrutinene sine. Tar man hensyn til, og har FAIR-prinsippene i bakhodet når man starter et nytt forskningsprosjekt, går egentlig det meste av seg selv. Hva kan jeg gjøre for at dataene mine skal være lett å finne for andre? Hva kan jeg gjøre for at dataene skal være lett tilgjengelige? Er arkivet jeg bruker, i stand til å «snakke» med andre arkiver og søkemotorer, slik at mine data faktisk dukker opp når det sitter en ung student i Etiopia og søker etter nettopp det mine data sier noe om? Og sluttelig, hva kan jeg gjøre for å sikre at mine data lettest mulig kan gjenbrukes av andre? Det finnes meget enkle oppskrifter for å oppfylle disse standardene og/eller kravene, og de er bare ett tastetrykk eller to unna. Undertegnede anslår at dersom man som forsker investerer en time eller to (nei, ikke noe særlig mer!) så bør man være i stand til å behandle sine egne data på en FAIR måte. Så vil du, når det gjelder forskningsdataene dine (ja, ups ... glemte i farta at det er jo arbeidsgiveren din som i hovedsak eier «dine data»), fortsette å jobbe slik du alltid har gjort, eller vil du prøve å være en aktiv og åpen forsker, som ikke går baklengs inn i fremtiden? Valget er ditt (nei, ups ... valget er jo ikke ditt lenger). De som finansierer forskningen, sier jo både at du «bør», og at du «skal». Så brett opp ermene – og hiv deg rundt (som vi sier i Tromsø)!