

# Støttetenester for forskingsdatahandtering på UiT Noregs arktiske universitet – erfaringar og forslag til beste praksis

*Best practice article*

**Philipp Conzett\*, Lene Østvand**

*UiT The Arctic University of Norway*

## Abstract

---

This article describes how support services for research data management have been developed and are run at UiT The Arctic University of Norway (UiT). We go through several areas of interest: technical solutions, researcher involvement, division of work between different units at our institution as well as co-operation with national and international stakeholders, data curation, skills development, training and curriculum development, dissemination, and policymaking. Each section concludes with a set of best practice recommendations.

**Keywords:** Information literacy, information seeking, evidence-based practice, cross-collaborative supervision, nursing work placements

**\*Contact:**

Philipp Conzett

e-mail: [philipp.conzett@uit.no](mailto:philipp.conzett@uit.no)

## Bakgrunn og tilnærming

Arbeidet med støttetenester for forskingsdatahandtering på UiT Noregs arktiske universitet (UiT) byrja for alvor då Universitetsbiblioteket på UiT (UB) i 2013 fekk ein førespurnad om å utvikla eit ope arkiv for forskingsdata og statistisk kode innanfor språkvitskap. Arkivet blei i 2014 lansert under namnet The Tromsø Repository of Language and Linguistics (TROLLing). Det blei i 2016 del av ei generell teneste for arkivering og deling av opne forskingsdata på UiT (UiT Open Research Data). Sommaren 2017 blei denne tenesta omorganisert til ei fleirinstitusjonell plattform for forvaltning av opne forskingsdata (DataverseNO). Parallelt med utviklinga av desse tekniske løysingane har vi bygd opp tilhøyrande tenester som står opp om ei effektiv, forsvarleg og framtidsretta forvaltning av forskingsdata.

Sjølv hadde vi lita eller inga erfaring med forskingsdatahandtering då vi gav oss i kast med det. Men internasjonalt har det skjedd ei rivande utvikling på dette området det siste tiåret. Det er allereie blitt utvikla støttetenester av ulike slag. Som eit overordna prinsipp for arbeidet vårt har vi derfor prøvd å halda oss til slagordet «adopt, adapt, develop», som kan samanfattast slik:

- Bruk eksisterande løysingar.
- Tilpass eksisterande løysingar der det er naudsynt.
- Som siste utveg, utvikle eigne løysingar frå botnen av.

Nedanfor skal vi gå nærare inn på ulike aspekt ved arbeidet vårt med å utvikla og drifta støttetenester for forskingsdatahandtering på UiT. Arbeidet så langt har i hovudsak fokusert på arkivering og deling av forskingsdata. Vi har delt framstillinga inn i seks område: forankring i fagmiljø, tekniske løysingar og standardar, arbeids- og rollefordeling, kuratering og oppfølging, kompetanseheving og opplæring, og institusjonelle retningslinjer og standardar. Innanfor kvart område skal vi avslutningsvis leggja fram forslag til beste praksis. I siste kapittelet summerer vi opp hovudkonklusjonane.

## Forankring i fagmiljø

Forankring i fagmiljø blir ofte framheva som ein av hovudnøkklane til å lykkast med utviklinga og etableringa av nye støttetenester. Kevin Ashley, direktør på Digital Curation Centre ved Universitetet i Edinburgh, framheva i eit føredrag på eit introduksjonsseminar om forskingsdata (Ashley, 2016) at tiltak som ikkje får støtte frå leiinga og som ikkje involverer fagmiljø, er daudfødde. UB har i fleire år formidla fordelane med open tilgang til forskingsresultat overfor forskarane. Det har ført til at stadig fleire forskarar er blitt medvitne om at UB er ein viktig aktør og støttespelar på dette området.

Det første initiativet til samarbeid om forskingsdatahandtering på UiT kom frå fagmiljøa sjølve. Sommaren 2013 fekk UB ein førespurnad frå ei forskingsgruppe i lingvistik på UiT om hjelp til å laga ei teneste der språkvitskaplege data og statistisk kode kunne arkiverast ope og delast med andre. Tenesta skulle vera open for språkvtarar internasjonalt. UB oppretta eit utviklingsprosjekt med interne ressursar og finansiering gjennom eigne driftsmidlar, og arbeidet skjedde i nært samarbeid med fagmiljøet. The Tromsø Repository of Language and Linguistics (TROLLing) blei lansert 18. juni 2014 som, etter det vi veit, er det første opne arkivet for språkvitskaplege forskingsdata. I samråd med fagmiljøet utvikla UB også rettleiingar på korleis arkivet skulle brukast. Vi la særleg vekt på korleis dei arkiverte dataa skulle strukturerast og dokumenterast i høve til internasjonale standardar for metadata, og val av filformat. Kanskje like viktig var – og er – arbeidet med å gjera tenesta kjend i fagmiljøet.

For UB var TROLLing ein test på kva vi kunne byggja av tenester for arkivering av forskingsdata. Det var òg ei reell prøve på kva slags krav og kompetanse som måtte til for å kunna tilby ei generell dataarkivteneste på UiT. Desse erfaringane tok UiT med seg vidare då vi hausten 2015 gjekk i gang med utviklinga av ei generell arkivteneste for forskingsdata på UiT. Arbeidet med denne nye tenesta, som seinare skulle få namnet UiT Open Research Data, var organisert som eit pilotprosjekt i samarbeid mellom UB og Avdeling for IT (ITA), med Avdeling for forskning og utviklingsarbeid (AFU) som partner. Prosjektrapporten beskriv målet med pilotprosjektet slik:

Hensikten med prosjektet var å få erfaring med tilrettelegging av teknisk infrastruktur for et institusjonelt arkiv for forskningsdata, kartlegge behov for kompetansebygging og støttetjenester for forskere som skal bruke en slik tjeneste, foreslå en forvaltningsmodell for forskningsdata på institusjonen og beskrive hva som må til for å iverksette en slik tjeneste i full skala (Longva, Amundsen, Dragseth, Østhus, & Høydalsvik, 2016, s. 4).

Eit viktig element i pilotprosjektet var altså å få kartlagt kva behov forskarane på UiT hadde når det gjeld handtering og arkivering av forskingsdata. Inspirert av eit liknande prosjekt på University of Minnesota (Johnston, 2014), sende dei fagansvarlege på UB ut ein open invitasjon til forskarar i sine respektive fagmiljø til å vera med på prosjektet. Målet var å skaffa eit avgrensa, men så representativt som mogleg, utval av forskingsdata frå forskarar som var interesserte i å delta i utviklinga av ei arkivløyning for opne forskingsdata. Fjorten personar melde si interesse. Alle bortsett frå éin var fagleg tilsette. I ein første avklaringsrunde viste det seg at sju av interessentane var ueigna for pilotprosjektet ettersom dei hadde sensitive data eller andre typar data som ikkje utan vidare kunne gjerast ope tilgjengelege. Dette gav oss ein nyttig peikepinn på kor viktig kompetanse på opphavsrett og personvern kom til å bli i ei fullskala støttetjeneste. Etter denne silinga stod vi att med åtte datasett frå åtte ulike fagområde innafor humaniora, samfunnsfag og naturfag. Blant deltakarane var det både enkeltforskarar og forskingsgrupper. Ut ifrå ei ressursvurdering enda vi til slutt opp med å konsentrera oss om seks deltakarar. Sjølv om det var få deltakarar, så var utvalet etter vår vurdering stort nok for å kunna tena som datagrunnlag for pilotprosjektet. Med utgangspunkt i desse pilotdataa fekk vi undersøkt ei rekkje aspekt knytte til handtering av forskingsdata:

- Kartleggja relevante metadatabeskrivingar for dei ulike fagområda
- Vurdera behovet for hjelp til standardisert beskriving og dokumentasjon av data og til formathandtering/-konvertering
- Kartleggja alternative fagspesifikke eller generelle dataarkiv, nasjonalt og internasjonalt
- Tilrådingar, råd og rettleiing om lisensar for forskingsdata

Erfaringane frå pilotprosjektet tente som nyttige innspel i det vidare arbeidet med utviklinga av eit generelt arkiv for opne forskingsdata på UiT med tilhøyrande støttetjenester.

Kontakten mellom UB og fagmiljø skjer vanlegvis igjennom UB sine fagansvarlege, anten i form av formelle eller uformelle møte eller ulike typar opplæring. Dei fagansvarlege kjenner til fagmiljøa og veit kva typar data dei brukar i forskinga si. Nokre fagansvarlege deltek (sporadisk) også på forskingsgruppemøte og på møte med instituttleiarar. Der får dei høve til å presentera UB sine støttetjenester for forskingsdatahandtering.

Primærmålgruppa for støttetjenester for forskingsdatahandtering er forskarar og doktorgradsstudentar. Det seier seg derfor sjølv at det er viktig å forankra slike tenester i dei ulike fagmiljøa på institusjonen. Vi har følgjande råd for å lykkast med å få forskarane med på laget:

- Involver fagmiljø i ein tidleg fase.

- Bruk nettverket og kompetansen til dei fagansvarlege på UB for å nå ut til fagmiljøa.
- Start samarbeidet med (små grupper av) motiverte forskarar, og bruk dei seinare som ambassadørar når tenesta skal gjerast kjend.
- Formidle fordelane som forskarane har av god forskingsdatahandtering.

## Tekniske løysingar og standardar

Då vi fekk utfordringa om å laga eit ope arkiv for språkdata, måtte vi sjå oss om etter ei teknisk løysing. Frå før brukte vi DSpace for å drifta kunnskapsarkivet vårt, Munin. Det finst andre liknande digitale arkivapplikasjonar, til dømes EPrints og Fedora, men dei er, eller i alle fall var i 2013, arkivapplikasjonar for tilgjengeleggjering av digitale forskingspublikasjonar, ikkje forskingsdata.

Vi ville helst bruka ein applikasjon som var utvikla spesielt med tanke på arkivering og tilgjengeleggjering av forskingsdata, ettersom vi lokalt ikkje har kapasitet til omfattande tilpassing og vedlikehald. I tillegg hadde vi fleire funksjonelle og tekniske krav til ein slik applikasjon.

Som nemnt bestemde vi oss for å konsentrera oss om opne forskingsdata i starten. Det overordna målet fram til no har altså vore at UiT skal leggja til rette for at forskingsdata som er genererte på UiT, kan gjerast tilgjengelege på best mogleg måte.

Applikasjonen vi var ute etter, ville bli eit viktig ledd i arbeidet med tilgjengeleggjering. For å beskriva krava som generelt bør stillast til ein slik applikasjon er det naturleg å ta utgangspunkt i dei såkalla «FAIR»-prinsippa (FORCE11, 2014). Det er eit sett med retningsliner for korleis ein kan gjera forskingsdata mest mogleg tilgjengelege. «FAIR» er ei forkorting som står for «Findable», «Accessible», «Interoperable», og «Re-usable», og kan i hovuddrag samanfattast slik:

«**Findable**» vil seia at forskingsdata er lett å finna. For å oppnå det bør applikasjonen ha støtte for

- tildeling av såkalla persistente identifikatorar (PID),
- registrering av gode metadata,
- indeksering eller hausting av metadata og
- eintydige referansar som kan brukast når ein skal referera til data. Det inkluderer også informasjon om versjon.

«**Accessible**» inneber at det er lett å få tilgang til forskingsdata. For å oppnå det bør applikasjonen tilby

- ein veldefinert og open protokoll for data- og metadatautveksling,
- protokollen bør tilby tilstrekkeleg autentisering, og
- metadata bør vera tilgjengelege også når sjølve dataa ikkje er tilgjengelege lenger.

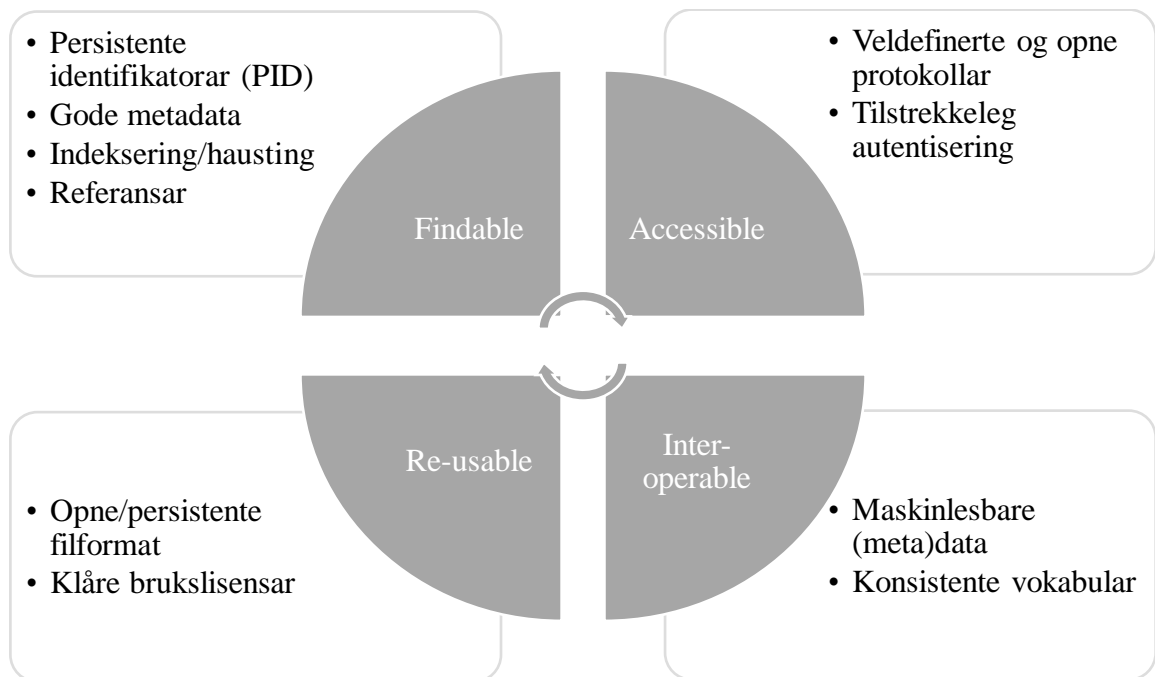
«**Interoperable**» går ut på at metadataa og sjølve dataa frå ulike prosjekt kan kommunisera med kvarandre, slik at det går an å søkja, jamføra og behandla data og metadata på tvers av arkiv. Applikasjonen bør derfor ha støtte for

- maskinlesbare data og metadata og
- konsistente metadatavokabular.

«**Re-usable**» tyder at data i størst mogleg grad skal kunna replikerast eller gjenbrukast på andre måtar. For å bidra til det bør applikasjonen gje støtte for

- opne og persistente filformat og
- klåre brukslisensar.

«FAIR»-prinsippa er oppsummerte i figur 1.



Figur 1: «FAIR»-prinsippa

I tillegg sette vi som krav at applikasjonen bør vera basert på open kjeldekode. Det er mange fordelar med slike applikasjonar. For eksempel er det ofte enklare å integrera dei med andre system, og dei blir ofte utvikla i eit internasjonalt fellesskap, noko som aukar kvaliteten og gjer det lettare å følgja internasjonale standardar. På eit meir prinsipielt plan er det også meir passande å bruka eit ikkje-proprietært, ope system for å handtera opne forskingsdata.

Vi fann ikkje nokon applikasjon som tilfredsstilte alle desse krava fullt ut. Men den applikasjonen vi landa på, Dataverse, var den einaste som oppfylte i alle fall nokre av dei viktigaste krava. Dataverse var også den einaste applikasjonen som i 2013 var tilgjengeleg som det vi kan kalla for hyllevare. Men dersom institusjonen ønskjer å lagra datasetta i Dataverse på eigne serverar, må applikasjonen installerast lokalt. Det føreset at institusjonen rår over visse IT-faglege og -tekniske ressursar.

I tillegg er Dataverse relativt fleksibelt når det gjeld strukturering av arkivet. Vi byrja som sagt med TROLLing, som seinare blei gjort om til ei tematisk samling («sub-dataverse») innanfor UiT Open Research Data. Sommaren 2017 la vi til eit nytt nivå på toppen, slik at vi no tilbyr ei fleirinstitusjonell teneste. Namnet på denne nye tenesta er DataverseNO etter mønster av den nederlandske konsortiemodellen DataverseNL. Planen er å få DataverseNO sertifisert som ei påliteleg forskingsdataarkivteneste i løpet av 2018, i første omgang gjennom sertifiseringsordninga CoreTrustSeal.

Vi har valt ikkje å gjera lokale tilpassingar av Dataverse utover slike som applikasjonen sjølv legg opp til. Det gjer det mellom anna lettare å oppgradera til nye versjonar.

I tillegg til ei arkivløyning for forskingsdata såg vi i 2017 også behovet for ei teknisk løyning for å oppretta og følgja opp datahandteringsplanar. Løysinga skal stø opp om vedtaket om obligatoriske datahandteringsplanar for tilsette på UiT som tredde i kraft 1. september 2017

(sjå meir om dette i avsnittet «Retningsliner og standardar»). Våren 2017 evaluerte vi derfor aktuelle løysingar. Det er i skrivande stund særleg to løysingar som er i bruk på ulike UH-institusjonar. Den eine er DMPonline, som er mest utbreidd i Europa, medan den andre, DMPTool, er mest brukt i USA. Aktørane bak begge verktøya samarbeider om den vidare utviklinga. Vi fall ned på DMPonline ettersom det er meir tilpassa europeiske forhold.

DMPonline er eit nettbasert verktøy basert på open kjeldekode der forskarar kan oppretta datahandteringsplanar med utgangspunkt i malar som er tilpassa krava frå ulike finansørar. Som institusjon kan ein oppretta sine egne malar og brukarretteiingar i tråd med lokale reglar og retningsliner. Vidare kan ein følgja opp datahandteringsplanane til tilsette på institusjonen. Det inneber til dømes at forskingsadministrasjonen kan gje tilbakemelding på planar, og at institusjonen kan bruka opplysningane i planane som grunnlag for planlegging og vidareutvikling av lagringskapasitet og støttetenester.

I motsetnad til Dataverse har vi valt ikkje å installera DMPonline lokalt, men heller bruka installasjonen som er drifta av utviklaren, Digital Curation Centre. Også når det gjeld arkivløysingar, kan ein basera seg på eksternt drifta tenester. I skrivande stund kan ein velja mellom fleire operative tenester, til dømes BIRD, DataverseNO, Figshare, Harvard Dataverse, NSD – Norsk senter for forskningsdata, Zenodo. Eit meir uttømande oversyn finn ein i arkivkatalogen re3data.org.

Det viktigaste med tekniske løysingar for data og datahandteringsplanar er at dei tilfredsstillar grunnleggjande krav om opne og etablerte standardar, interoperabilitet og tryggleik. Det er mindre viktig kor løysingane er lokaliserte.

I tillegg til tekniske løysingar for arkivering av forskingsdata og oppfølging av datahandteringsplanar er det behov for løysingar for å handtera forskingsdata undervegs i forskingsprosjekt. Her brukar UiT per i dag tradisjonelle fillagringstenester, men skal innan 2020 tilby forskarane sine meir integrerte løysingar, også for sensitive data.

Våre tilrådingar for val av tekniske løysingar kan samanfattast slik:

- Løysingane bør stø opp om «FAIR»-prinsippa.
- Vurder eksternt drifta løysingar eller felles løysingar med andre institusjonar.
- Følg aktivt med i diskusjonsforum for brukarmiljø, og spel inn synspunkt til tilpassing og vidareutvikling.

## **Arbeids- og rollefordeling og samarbeid**

I dette kapittelet skal vi gjera greie for korleis vi internt på UiT har fordelt arbeidsoppgåvene og ansvaret for dei ulike støttetenestene knytte til forskingsdatahandtering, og korleis vi samarbeider med eksterne aktørar.

### **Arbeids- og rollefordeling på UiT**

Forskningsdatatenestene på UiT byggjer på eit samarbeid mellom ITA, AFU og UB. ITA har primæransvar for teknisk drift, medan UB er systemeigar og ansvarleg for opplæring og støttetenester. AFU jobbar med strategiske saker og institusjonell politikk på området. Samarbeidet starta med det omtalte pilotprosjektet, og har vore – og er framleis – ein føresetnad for eit vellykka tenestetilbod.

ITA har innsikt i tekniske løysingar. Infrastruktur og lagring av forskingsdata er eit viktig område i ITAs strategidokument (Avdeling for IT, 2015).

AFU er del av UiTs sentraladministrasjon og har ei rekkje forskingsadministrative støttefunksjonar. Avdelinga koordinerer også det forskingsadministrative arbeidet mellom fakulteta. AFU har hatt og har ei sentral rolle i utforminga og implementeringa av UiTs retningsliner for forvaltning av forskingsdata (sjå meir om dette i «Retningsliner og standardar»).

UB har lang erfaring med open publisering (open access). Derfor er det naturleg å utvida kompetanse og tilbod til å inkludera opne data i et meir heilskapleg arbeid med open vitskap. Forskingsstøtte, opne løysingar for deling av kunnskap og tenester for arkivering av forskingsdata er viktige punkt i UBs strategi fram mot 2020 (Universitetsbiblioteket i Tromsø, 2015).

Som nemnt i innleiingskapittelet var det med utviklinga av TROLLing at arbeidet med forskingsdata for alvor skaut fart på UB og UiT. I tillegg til to til tre tilsette frå Institutt for språkvitskap bestod TROLLing-prosjektgruppa av fem UB-tilsette. UB-gruppa var sett saman, slik at ho i størst mogleg grad skulle byggja på eksisterande kompetanse og erfaring frå liknande arbeid. Tre av medlemene hadde omfattande erfaring frå arbeid med open publisering, den eine som IT-systemutviklar, og dei to andre som rådgjevarar og prosjektleiarar. IT-systemutviklaren tok seg av dei tekniske oppgåvene, som til dømes installering, konfigurering og oppgradering av programvare. Dei to rådgjevarane stod for mykje av den konseptuelle utviklinga og planlegginga av tenesta. Dei to siste medlemene i gruppa var fagansvarlege for ulike språkfag og hadde såleis med seg både disiplinfagleg kunnskap og erfaring frå undervisning i informasjonskompetanse, og nettverksbygging og kontakt med fagmiljøet. I driftsfasen er dei ansvarlege for kuratering av datasett og utvikling og gjennomføring av opplæringstilbod (sjå seinare i artikkelen).

TROLLing-arbeidsgruppa på UB danna også utgangspunkt for arbeidet med pilotprosjektet, som hadde som mål å byggja ut tenesta til eit generelt institusjonsarkiv for opne forskingsdata. Men gruppa blei no utvida. På UB blei det inkludert to nye fagansvarlege som dekte natur- og realfag. I tillegg inngjekk UB eit samarbeid med ITA og AFU. Denne arbeidsfordelinga har vi halde fast ved også etter avslutting av pilotprosjektet. På sikt er planen å involvera alle fagansvarlege på UB. Meininga er at kvar fagansvarleg skal kuratera forskingsdata innanfor sine fag. Denne breidda skal kunna sikra eit godt heilskapleg tilbod på tvers av UiTs fagportefølje. Samarbeidet mellom UB, ITA og AFU blir halde ved like gjennom ulike kontaktpunkt, til dømes gjennom ei arbeidsgruppe som jobbar med implementering av UiTs retningsliner for handtering av forskingsdata. Vidare blir representantar for UBs forskingsdatagruppe inviterte til møte når det er aktuelt i Forskingsadministrativt forum, på fagdagar i regi av ITA, og utvida leiarmøte, bestående av universitetsleiinga og ei gruppe dekanar.

Når det gjeld oppfølging av datahanderingsplanar (sjå «Kuratering og oppfølging»), så vil det truleg òg bli aktuelt å involvera fagleg tilsette på dei ulike fakulteta og institutta.

## **Samarbeid med andre aktørar**

Som det går fram av gjennomgangen av arbeidsoppgåvene i førre avsnittet, så krev god forvaltning av forskingsdata kompetanse og ressursar på ei rekkje område. På fleire område vil det vera fornuftig å samarbeida med andre aktørar, både nasjonalt og internasjonalt.

I Noreg er det ulike nasjonale aktørar som driv med forvaltning av forskingsdata, til dømes NSD - Norsk senter for forskningsdata og UNINETT Sigma2. Tenestene deira har tradisjonelt konsentrert seg om lagring av forskingsdata. Men dei siste åra har fokuset blitt retta meir mot andre aspekt ved forvaltninga av forskingsdata, som til dømes kuratering og deling. Det generelle inntrykket på mange lokale institusjonar er likevel at det per i dag ikkje finst noka klår nasjonal rollefordeling.

Trass i manglande nasjonal koordinering såg vi det som viktig å orientera oss mot nasjonale aktørar allereie i ein tidleg fase av utviklinga av støttetenestene våre. Til dømes i arbeidet vårt med TROLLing har vi frå tidleg av vore i tett dialog med CLARINO, den norske noden i CLARIN, som er eit europeisk infrastrukturnettverk for språkressursar. TROLLing og CLARINO samarbeider mellom anna om felles standardar for metadatautveksling. UB har også vore med på å stifte Linguistics Data Interest Group, som ein del av Research Data Alliance (RDA), som er eit internasjonalt nettverk som har som mål å fremja deling av forskingsdata på tvers av teknologiar, disiplinlar og land.

UB samarbeider også med NSD. På Virak-konferansen 2017 arrangerte vi i samarbeid med NSD ein workshop om forskingsdatahandtering, der også nasjonalt samarbeid var tema. Vi har også sett eit meir konkret behov for å involvera NSD for å heva kompetansen vår på handtering av persondata. Hausten 2017 arrangerte vi eit todelt heildagsseminar. I første delen fekk dei som jobbar med forskingsstøtte opplæring, og i del 2 fekk forskarar og stipendiatar opplæring, medan forskingsadministrasjonen var med som observatørar.

Val av teknisk løysing styrer delvis korleis rollefordelinga blir mellom lokal institusjon og andre aktørar. Sidan vi valde å bruka dei eksisterande tekniske løysingane som Dataverse og DMPonline tilbyr, sparar vi mykje arbeidstid som elles ville ha gått med til utvikling. Desse aktørane har hovudansvaret for velfungerande system, medan vi har gjort nokre få lokale tilpassingar som systema legg til rette for. For Dataverse gjeld det innlogging via Feide og tilpassingar av metadatataskjema. Når vi ser andre behov for lokale tilpassingar, melder vi dei inn til Harvard University, som leier utviklinga av applikasjonen. Det har dei siste åra oppstått eit aktivt brukarmiljø rundt Dataverse. Vidareutviklinga av applikasjonen skjer i stor grad etter innspel frå og i samråd med brukarmiljøet. I DMPonline har vi lagt inn ein tilpassa mal for datahandteringsplan til bruk for forskarar på UiT.

Våre tilrådingar for arbeids- og rollefordeling kan samanfattast slik:

- Bruk organisasjonseiningar der arbeid med forskingsdata blir naturleg å knyta opp mot eksisterande arbeidsoppgåver.
- Lag ei klår ansvars- og rollefordeling, og etabler kontaktpunkt for samarbeid.
- Ta i bruk ekstern ekspertise der det finst, men sørg også for å byggja opp lokal kompetanse.

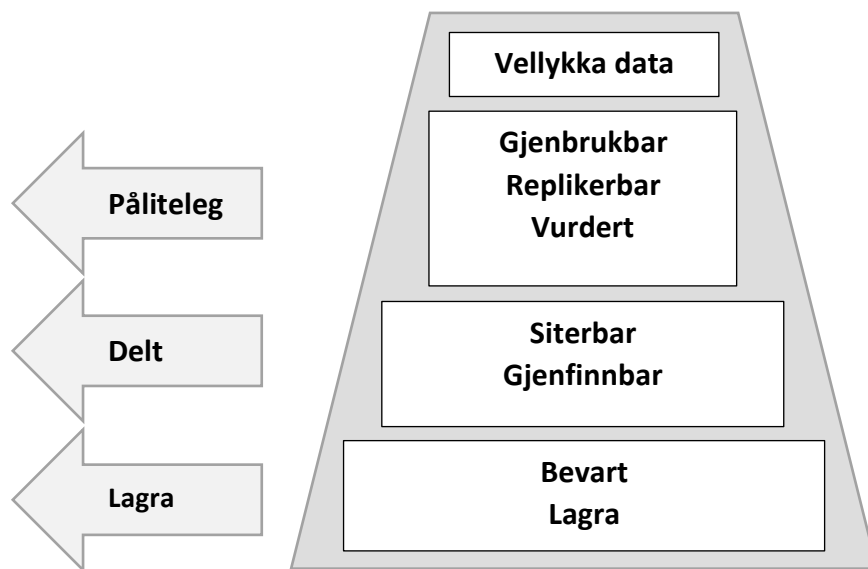
## **Kuratering og oppfølging**

Med utgangspunkt i dei nemde «FAIR»-prinsippa har de Waard, Cousijn og Aalbersberg (2015) sett opp ei liste over kriterium som beskriv kva som skal til for at forskingsdata skal



kunna gjenbrukast på best mogleg måte. Kriteria er hierarkisk plasserte i ein slags Maslows behovspyramide. Ein lett tilpassa og forenkla versjon av pyramiden er vist i figur 2. Vi har allereie sett på tekniske tiltak som er med på å gjera forskingsdata gjenbrukbare. Eit anna viktig bidrag frå UB i denne prosessen er kuratering av datasett som skal arkiverast.

Dei mest grunnleggjande kriteria, lagring og (langtids)bevaring av forskingsdata, ligg i botnen. Den mellomste bolken i pyramiden handlar om deling. Her er det viktig å velja ei arkivløysing som følgjer «FAIR»-prinsippa og som gjer at forskingsdata kan gjerast tilgjengelege, at dei er gjenfinnbare, og at dei kan siterast. Men i tillegg må ein ha gode rutinar på plass som sørgjer for at forskingsdata blir forsynte med gode metadata. Før datasett blir arkiverte i UiT Open Research Data, blir dei kuraterte av ein fagansvarleg på UB. Som ein del av kurateringa kontrollerer vi om forskaren har lagt fornuftige metadata på datasettet.



Figur 2: Vegen til vellykka forskingsdata (forenkla og tilpassa versjon av pyramiden i de Waard et al., 2015)

Lagring og deling er to viktige element i forvaltninga av opne forskingsdata. Men som framstilt i figuren, så er det fleire tiltak som må til for at forskingsdata skal kunna gjenbrukast. For det første må dei vera forståelege også for andre enn forskaren som har skaffa dei fram. Det er derfor viktig med god strukturering og dokumentasjon, inkludert konsistente filnamn. For det andre må filene som inngår i eit datasett, kunna opnast og brukast i tilgjengeleg programvare. For at dette skal vera mogleg også fleire år etter at data er genererte, bør dei arkiverast i eit persistent eller arkivverdige filformat. Strukturering, dokumentasjon og filformat er derfor tre sentrale faktorar som kuratoren sjekkar før datasettet blir publisert.

På UB brukar vi to dokument som er sentrale i kurateringsprosessen. For det første er det retningslinene for datasett som skal arkiverast i DataverseNO. Dei er mynta på forskarar og skal gje dei rettleiing på korleis dei skal leggja til rette dataa sine før dei skal arkiverast. For det andre har vi ein kuratorguide som beskriv rutinane for korleis eit datasett skal kuraterast, inkludert arbeidsflyten i arkivapplikasjonen. Begge dokumenta er baserte på forslag til beste praksis som andre institusjonar har utarbeidd, til dømes Curation Reference Manual ved Digital Curation Centre. Når det gjeld val av filformat, rådfører vi oss også med IT-avdelinga og fagmiljøa.

Ideelt sett burde forskingsdata også vurderast av fagfellar på line med det som er vanleg for forskingspublikasjonar. Først når kriteria i figur 2 er oppfylte, kan forskingsdata seiast å vera pålitelege.

Å generera pålitelege forskingsdata krev god planlegging. Ein datahandteringsplan femner om alle fasar og aspekt av datahandteringa i eit forskingsprosjekt. Her skal forskaren beskriva korleis forskingsdata skal samlast inn eller genererast, og korleis dei skal forvaltast både i og etter prosjektperioden. UB tilbyr ikkje berre kuratering av datasett i sluttfasen av eit forskingsprosjekt, men også rettleiing om og oppfølging av datahandteringsplanar før eller i ein tidleg fase av prosjektet. For opne data inneber det at planen legg opp til ei forvaltning som er i tråd med «FAIR»-prinsippa og kriteria i figur 2, men også at planen oppfyller andre spesifikke krav frå finansiørar. Arbeidet med oppfølging av datahandteringsplanar er enno i startfasen. Konkrete rutinar vil bli etablerte når vi tek i bruk DMPonline (sjå «Tekniske løysingar og standarder»).

Vi har følgjande gode-praksis-råd til kuratering og oppfølging av forskingsdata:

- Lag klare retningsliner for datasett. Skaff deg oversikt over tilsvarande dokument på andre institusjonar, og rådfør deg med IT-avdelinga og fagmiljøa om naudsynt.
- Sørg for grunnleggjande kuratering av datasett som forskarar på din institusjon skal arkivera.
- Set opp klare kurateringsrutinar.
- Set deg inn i grunnleggjande krav til datahandteringsplanar, og tilby hjelp for forskarar.

Gode rutinar for søk, sitering, planlegging, strukturering og dokumentasjon av forskingsdata er også ein viktig del av opplæringstilbodet vårt (sjå neste kapittelet). Vi forventar derfor at UB på sikt treng å leggja ned mindre kurateringsinnsats per datasett.

## **Kompetanseheving, opplæring og formidling**

### **Intern kompetanseheving**

Sjølv om fleire tilsette på UB har lang erfaring med Open Access, så er opne data eit ganske nytt felt både i bibliotekverda og for andre tilsette i UH-sektoren. Derfor har vi sett til arkiv og tilknyttate tenester på andre institusjonar for å læra av dei. Dataverse er utvikla på Harvard University, og det har vore naturleg å setja seg inn i tenesta lokalt der. Fleire institusjonar i Storbritannia har også erfaring med eigne arkiv. I september 2016 arrangerte vi ein studietur til Bristol. Universitetet der er om lag på same storleiken som UiT, men dei er komne lenger i utviklinga av forskingsdatatenester. Forskingsdatagruppa på UB har også lese seg opp på og held seg oppdatert på internasjonale standardar for handtering av forskingsdata, til dømes om kuratering og metadatahausting. Vi har delteke på fleire konferansar, til dømes International Digital Curation Conference.

Elles har vi prøvd oss fram og gradvis bygd opp tenestetilbodet etter kvart. Vi har hatt nyttige diskusjonar med AFU og ITA, forskarar og andre tilsette på UiT, noko som har resultert i justeringar undervegs. Det er viktig for oss å heile tida utvikla tilbodet vidare for å kunna tilby

ei mest mogleg aktuell teneste. Vi prøver å spreia kompetansen til fleire tilsette på UB ved å fortelja om forskingsdata i ulike interne forum, og oppmoda kollegaer til å delta på opne kurs.

### **Kurs og rettleiing for ulike målgrupper**

Vi ønskjer å nå ut til fleire typar grupper, og tilpassa tilbodet etter deira behov. Dei første kursa i datahandtering blei gjevne som del av seminarserien *Take control of your PhD journey*, som er eit kurstilbod frå UB retta mot doktorgradsstudentar frå alle fagområde. Seminarserien blei gjeven første gongen våren 2015 og bestod då av seminar om akademisk ærlegdom, litteratursøk, open publisering og referansehandtering. Våren 2016 blei datahandtering inkludert. Kwart seminar varer i tre timar og har fokus på diskusjon og refleksjon. I seminaret om datahandtering snakkar vi om motivasjon for å dela data, korleis ein best kan strukturera og dokumentera data, om kriterium for val av arkiv, og om korleis ein lagar ein datahandteringsplan.

Frå hausten 2016 har vi gjeve opne kurs med introduksjon til tenesta UiT Open Research Data som varer i to timar. Her er fokuserer vi på bruk av arkivet. Våren 2017 starta vi med fleire opne kurs, der vi sette av ein heil dag til ulike tema, med tre kvarter per tema. Deltakarane kan velja om dei vil vera med på eitt eller fleire temakurs, men temaa høyrer saman og blir underviste i ei naturleg rekkefølge. Så langt tilbyr vi kurs i følgjande tema og i denne rekkjefølgja: søking og sitering av forskingsdata, strukturering og dokumentering av forskingsdata, deling og lisensar, og datahandteringsplanar. Sjølv om ein datahandteringsplan bør skrivast tidleg i eit prosjekt, skal den innehalda delar om dei andre temaa, så deltakarane bør ha ei grunnleggjande forståing av kva dette inneber. Difor kjem delen om datahandteringsplanar til slutt.

UB er også involvert i eit kurs om forskingsfinansiering arrangert av High North Academy. Her underviser vi i skriving av datahandteringsplanar og open publisering i lys av retningsliner frå ERC/H2020 (European Commission Directorate-General for Research & Innovation, 2016). Kurset blei arrangert første gongen i 2016, men er no planlagt som eit årleg tilbod.

Vi kan vidare tilby kursopplegg og presentasjonar på bestilling. Så langt har nokre få bede om presentasjon og vidare oppfølging etter deltaking på kurs.

### **Nettressursar**

Vi har laga fleire generelle brukarrettleiingar på nett:

- Kom i gang – med fokus på innlogging, registrering av metadata, opplasting av filer, og publisering
- Retningslinjer for datasett – med fokus på namngjeving av filer, filformat, filkonvertering, dokumentasjon av data, og filstorleik.

TROLLing har tilpassa rettleiingar, og i tillegg

- instruksjonsvideoar baserte på eldre versjon av arkivapplikasjonen.

Det er også laga ein promoteringsvideo om TROLLing, samt eit intervju med Laura Janda,

professor i russisk språk, og ein av pådrivarane for TROLLing, om opne data i lingvistik.

For å gjera det enklare for forskarane å orientera seg har UiT laga ein eigen nettportal for informasjon og rettleiing om forskingsdatahandtering, Forskingsdataportalen UiT.

### **Vidare planar**

Vi jobbar kontinuerleg med å utvikla undervisningstilbodet og inkluderer nye emne relaterte til forskingsdata etter behov. Vi planlegg òg å integrera forskingsdata i informasjonskompetansekurs for studentar i større grad.

Våre tilrådingar for tiltak for kompetanseheving, opplæring og formidling kan samanfatast slik:

- Fokuser på nokre få grupper av gongen, og prioriter dei som allereie er interesserte i/har eit behov for å læra om datahandtering.
- Ha dialogbasert undervisning for å få tilbakemeldingar på kva forskarar og ev. andre treng meir informasjon om.
- Sørg for å heva intern kompetanse slik at fleire kan undervisa, rettleia og diskutera. Fagansvarlege har ei naturleg rolle i kontakten med forskarar som har forskingsdata.

### **Retningsliner og standardar**

Sjølv om arbeidet med forskingsdatahandtering byrja som eit grasrottiltak i form av eit samarbeid mellom UB og eit fagmiljø, så har vi på UB heile tida vore klare over at arbeidet måtte forankrast breiare dersom det skulle lykkast på heile institusjonen. Etter avslutta pilotprosjekt heldt eit utval av prosjektgruppa derfor fram med å utarbeida retningsliner for forvaltning av forskingsdata på heile UiT. Arbeidsgruppa blei komplettert av to prodekanar for forskning. Gruppa la fram eit utkast for Forskingsstrategisk utval, og det blei seinare sendt på høyring til fakulteta og einingane. Etter høyringa blei forslaget lagt fram som sak for Utvida leiarmøte, og saka blei seinare handsama i Universitetsstyret. I styremøtet 9. mars 2017 vedtok universitetsstyret på UiT prinsipp og retningsliner for forvaltning av forskingsdata på UiT gjeldande frå 1. september 2017. Så vidt vi veit, er det det første styringsdokumentet av sitt slag på ein norsk UH-institusjon.

Retningslinene legg til rette for effektiv, forsvarleg og framtidsretta forvaltning av forskingsdata. Dei bestemmer kva typar data som skal forvaltast, kven som skal forvalta dei, og korleis dei skal forvaltast. Retningslinene byggjer på Noregs forskingsråd (NFR) sin policy for forskingsdata (Norges forskningsråd, 2014). Hovudprinsippet til NFR er «åpen som standard» når det gjeld tilgang til forskingsdata, og dei oppmodar institusjonane til å utforma eigne retningsliner som skal fremja tilgjengeleggjering av forskingsdata. Både NFRs og UiTs retningsliner tek omsyn til føringane som EU sitt forskingsråd har sett fram i sitt forskingsprogram Horizon 2020. Hovudprinsippet for forskingsdata som er skaffa fram i Horizon 2020-finansierte prosjekt, er «as open as possible, as closed as necessary», og frå 2017 er det krav om datahandteringsplan i alle nye prosjekt (European Commission Directorate-General for Research & Innovation, 2017, s. 8). UiT har også sett til retningsliner som er tekne i bruk på andre UH-institusjonar. Til dømes fekk vi på det tidlegare omtalte studiebesøket nærare innblikk i retningslinene som gjeld på University of Bristol (University of Bristol, 2015).

Ein nyttig lærdom vi tok med oss frå Bristol, er at ein slik policy ikkje bør utformast for

vag. Dersom policyen skal føra til konkrete endringar i måten forskingsdata blir forvalta på, så bør ein heller bruka *skal* i staden for *bør*, særleg når det gjeld delingsaspektet. Dette rådet har UiT-styret lytta til. To av dei viktigaste prinsippa i retningslinene våre er at forskingsdata som er genererte på UiT, *skal* gjerast ope tilgjengelege, og at UiT-forskarar *skal* setja opp ein datahandteringsplan ved oppstart av forskingsprosjekt. Hausten 2017 laga UiT ein handlingsplan som skisserer konkrete tiltak og reglar for korleis desse prinsippa skal følgjast opp. På tampen av 2017 kom Kunnskapsdepartementet med sin nasjonale strategi for tilgjengeleggjering og deling av forskingsdata (Kunnskapsdepartementet, 2017). Desse retningslinene vil vera sentrale for UiTs vidare arbeid med forskingsdatahandtering.

I tillegg til overordna retningsliner og politiske føringar er det viktig å følgja med på standardar på dei ulike områda som inngår i forvaltninga av forskingsdata. Her får ein mykje med på kjøpet ved å velja rett teknisk løysing og/eller tenestetilbydar. Det er særleg når det gjeld kuratering og opplæring, at ein er nøydd til å halda seg orientert om kva som er tilrådd praksis. Nyttige inngangar til relevant informasjon finn ein til dømes på nettsidene til Digital Curation Centre og Research Data Alliance.

Forskningsdatahandtering må forankrast i retningsliner og standardar. Våre råd på dette området er:

- Utform klare retningsliner for forvaltning av forskingsdata på din institusjon.
- Ta omsyn til nasjonale og internasjonale føringar.
- Sjå gjerne til andre institusjonar for inspirasjon og tips.
- Ver tydeleg i ansvarsfordelinga, og formuler retningslinene som plikter heller enn som ønske.
- Lag ein handlingsplan for oppfølging av retningslinene.
- Orienter deg om og følg med på tilrådingar om beste praksis for handtering av forskingsdata.

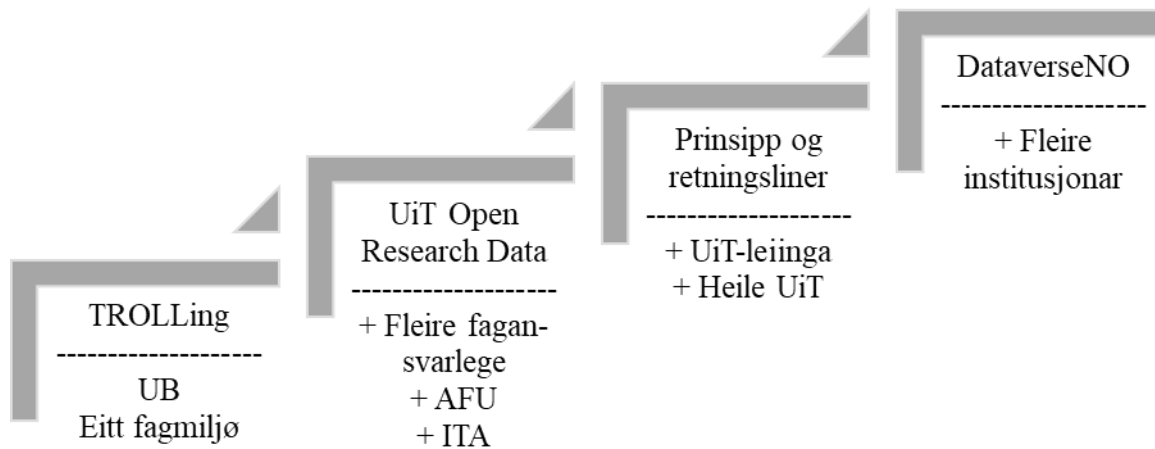
## Oppsummering

I denne artikkelen har vi gjort greie for korleis vi har utvikla og drifta støttetenester for forvaltning av forskingsdata på UiT Noregs arktiske universitet. Vi har sett på ulike aspekt ved dette arbeidet, og lagt fram konkrete forslag til beste praksis på dei ulike områda. Avslutningsvis kjem vi her med fire generelle råd for arbeidet med å utvikla og drifta støttetenester av denne typen:

1. Slagordet «adopt, adapt, develop» har vist seg som ei god rettesnor i arbeidet vårt: Bruk eksisterande løysingar, tilpass dei der det er naudsynt, og utvikle eigne løysingar berre som siste utveg.
2. Hald deg til internasjonale standardar og forslag til beste praksis.
3. Gje rom for lokal tilpassing og personleg handlingsrom, men innanfor klare føringar (jf. pkt. 2).
4. Ikkje vent med å lansera støttetenesta til den perfekte løysinga er på plass.

Eit gjennomgåande trekk ved utviklinga av støttetenestetilbodet for forskingsdatahandtering på UiT er at arbeidet har hatt preg av nedanifrå-og-opp-tiltak. Den gradvise utviklinga er illustrert i figur 3. Det byrja med eit initiativ frå eit fagmiljø, som så sette i gang eit lokalt arbeid på UB for å laga TROLLing. Seinare blei TROLLing utvida til ei generell teneste for opne forskingsdata på UiT. Då blei også fleire fagansvarlege og fagmiljø involverte, og UB inngjekk eit meir systematisk samarbeid med andre einingar på UiT. Med vedtaket og implementeringa av forskingsdatapolicyen blei UiT-leiinga og i prinsippet heile

UiT involvert i arbeidet med forvaltning av forskingsdata. Til slutt kan vi seia at utvidinga av UiT Open Research Data til ein fleirinstitusjonsmodell er endå eit steg mot breiare forankring og samarbeid.



Figur 3: Gradvis utvikling av tenester

Vegen fram til tenestetilbodet vi har i dag, har i stor grad blitt til medan vi har gått. Slik sett er det ikkje heilt rett å snakka om ein medviten nedanfrå-og-opp-modell, men vegen skil seg frå meir topptungne tilnærmingar som ofte er i bruk ved utviklinga av større tenestetilbod.

## Verktøy og nettstader

BIRD: <https://bird.bibsys.no/>  
CLARINO: <http://clarin.b.uib.no/>  
CoreTrustSeal: <https://www.coretrustseal.org/>  
Curation Reference Manual: <http://www.dcc.ac.uk/resources/curation-reference-manual>  
Dataverse: <https://dataverse.org/>  
DataverseNL: <https://dataverse.nl/>  
DataverseNO: <https://dataverse.no/> og <https://info.dataverse.no>  
Digital Curation Centre: <http://www.dcc.ac.uk/>  
DMPOne: <https://dmponline.dcc.ac.uk/>  
DMPTool: <https://dmptool.org/>  
DSpace: <http://www.dspace.org/>  
EPrints: <http://www.eprints.org/>  
Fedora: <http://fedorarepository.org/>  
Figshare: <https://figshare.com/>  
Forskingsdataportalen UiT: <https://uit.no/forskingsdata>  
Harvard Dataverse: <https://dataverse.harvard.edu/>  
High North Academy: [http://site.uit.no/hna/?page\\_id=114](http://site.uit.no/hna/?page_id=114)  
International Digital Curation Conference: <http://www.dcc.ac.uk/events/international-digital-curation-conference-idcc>  
Linguistics Data Interest Group: <https://www.rd-alliance.org/groups/linguistics-data-ig>  
Munin: <http://munin.uit.no/>  
Norsk senter for forskningdata: <http://www.nsd.uib.no/>

Prinsipp og retningslinjer for forvaltning av forskningsdata ved UiT:

[https://intranett.uit.no/Content/523436/cache=20172206150343/Policy\\_forskningsdata\\_UiT\\_090317.pdf](https://intranett.uit.no/Content/523436/cache=20172206150343/Policy_forskningsdata_UiT_090317.pdf)

re3data.org: <http://www.re3data.org/>

Research Data Alliance (RDA): <https://www.rd-alliance.org/>

Take control of your PhD journey: <https://uit.no/ub/laringsstotte#linje2>

TROLLing (The Tromsø Repository of Language and Linguistics):

<https://dataverse.no/dataverse/trolling>, <http://site.uit.no/trolling/about/>) og  
<https://www.youtube.com/watch?v=8FLQwJVM-VA>

UiT Open Research Data: <https://opendata.uit.no>

UNINETT Sigma2: <https://www.uninett.no/sigma>

Virak-konferansen 2017: <http://www.virak-konferansen.no/>

Zenodo: <https://zenodo.org/>

## Referansar

Ashley, K. (2016). *Research data: value and responsibilities*. Foredrag presentert på

Introduksjonsseminar om forskningsdata, Gardermoen. Retrieved from

<http://uhrbforskertjenester.b.uib.no/files/2016/11/Ashley.pdf>

Avdeling for IT, UiT Norges arktiske universitet. (2015). Veikart for Avdeling for IT 2013-2017. Retrieved from <https://uit.no/Content/411476/Veikart-for-Avdeling-for-IT-2013-2017.pdf>

de Waard, A., Cousijn, H., & Aalbersberg, Ij. J. (2015). 10 aspects of highly effective research data. Retrieved from <https://www.elsevier.com/connect/10-aspects-of-highly-effective-research-data>

European Commission Directorate-General for Research & Innovation. (2016). *Guidelines on FAIR Data Management in Horizon 2020*. Retrieved from

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

European Commission Directorate-General for Research & Innovation. (2017). *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*. Retrieved from

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

FORCE11 (2014). *Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing*. Version b1.0. Retrieved from <https://www.force11.org/fairprinciples>

Johnston, L. R. (2014). A Workflow Model for Curating Research Data in the University of Minnesota Libraries: Report from the 2013 Data Curation Pilot. *University Libraries Staff Publications*. Retrieved from <http://hdl.handle.net/11299/162338>

Kunnskapsdepartementet. (2017). Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata. Retrieved from <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-tilgjengeliggjoring-og-deling-av-forskningsdata/id2582412/>

Longva, L., Amundsen, E. T., Dragseth, R., Østhus, R., & Høydalsvik, S. (2016).

Pilotprosjekt infrastruktur for forskningsdata ved UiT. *Septentrio Reports*, 0(4).

<http://dx.doi.org/10.7557/sr.2016.4>

Norges forskningsråd. (2014). Tilgjengeliggjøring av forskningsdata. Policy for Norges forskningsråd. Retrieved from

<https://www.forskningsradet.no/servlet/Satellite?blobcol=urldata&blobheader=applicati on%2Fpdf&blobheadername1=Content-Disposition&blobheadervalue1=+attachment%3B+filename%3D%22Tilgjengeliggj%C>

3%B8ringForskningsdataWEB.pdf%22&blobkey=id&blobtable=MungoBlobs&blobwhere=1274505424271&ssbinary=true

University of Bristol. (2015). Research Data Management and Open Data Policy. Retrieved from [http://www.bristol.ac.uk/media-library/sites/university/documents/governance/UOB\\_RDM\\_Policy.pdf](http://www.bristol.ac.uk/media-library/sites/university/documents/governance/UOB_RDM_Policy.pdf)

Universitetsbiblioteket i Tromsø. (2015). Strategisk plan for Universitetsbiblioteket 2015-2020. Retrieved from [https://uit.no/Content/424232/Strategi\\_2015-2020.pdf](https://uit.no/Content/424232/Strategi_2015-2020.pdf)