



Digital forskningsfremtid ved NMBU

Rapport fra fagråd for e-infrastruktur



Innholdsfortegnelse

1	Mandat	3
2	Innledning.....	4
3	Databehandling i forskning.....	5
3.1	Planlegging.....	5
3.2	Innsamling.....	6
3.3	Lagring og arkivering.....	7
3.3.1	Eksisterende systemer for lagring av forskningsdata	7
3.3.2	Systemer og verktøy for arkivering og tilgjengeliggjøring av forskningsdata	8
3.3.3	Utfordringer med dagens løsninger	8
3.3.4	Anbefalinger vedrørende lagring.....	8
3.4	Analyse.....	9
3.4.1	Fornyning av Orion	10
3.5	Tilgjengeliggjøring.....	10
4	Tungregning.....	10
5	Undervisning.....	11
6	Anbefalinger	11

Versjonslogg

Versjon 1.0 — 4. mars 2021

1 Mandat og sammensetning

Forskningsutvalget ved NMBU har i sak 42/20 opprettet fagråd for e-infrastruktur og gitt rådet følgende mandat:

- Konkretisere NMBUs digitaliseringsstrategi med hensyn til e-infrastruktur for forskning og forskningsbasert undervisning
 - Behov, potensiale, risiko
 - Status dagens situasjon
- Beskrive avgrensning mellom e-infrastruktur på nasjonalt, universitetsnivå og "kontornivå"
- Beskrive oppbygging/ etablering av nødvendige ressurser
 - Maskinvare og grunnleggende systemer
 - Personalbehov for drift og avansert vitenskapelig brukerstøtte
 - Integrasjon med NMBUs øvrige systemer
- Beskrive forslag til langsiktig representasjon ved NMBU av e-infrastruktur-brukere innen forskning og forskningsbasert undervisning

Alle fakultetene har en representant til arbeidsgruppen. I tillegg er IT-avdelingen, Universitetsbiblioteket og Forskningsavdelingen representert:

Fakultet	Representant
BIOVIT	Torfinn Nome (forsker)
HH	Eirik Romstad (førsteamanuensis)
KBM	Lars Snipen (professor)
LANDSAM	Ramzi Hassan (førsteamanuensis)
MINA	Jon Gustav Kirkerud (postdoktor)
REALTEK	Hans Ekkehard Plesser (professor, leder)
VET	Ane Nødtvedt (professor) og Clara Jalland (seniorrådgiver)
IT-avdelingen	Ola-Gunnar Juelsrud (seksjonssjef IT-infrastruktur)
Universitetsbiblioteket	Linn Benjaminsen Hølvold (førstebibliotekar)
Forskningsavdelingen	Solveig Fossum-Raunehaug (seniorrådgiver)

2 Innledning

Forskning og forskningsbasert undervisning er basert på innsamling av data og prøvemateriale. Begge foreligger i dag i stor grad i digital form og forvaltes og utnyttes ved hjelp av datamaskiner. For å være en ledende forsknings- og utdanningsinstitusjon i fremtiden må NMBU ha funksjonelle IT-systemer for håndtering av forskningsdata.

NMBU skiller seg fra forskningstunge universiteter i Norge og verden ellers ved ikke å ha hatt et sentralt kompetansemiljø innen vitenskapelig IT siden FDB-sentralen ved NLH ble nedlagt på 1980-tallet. Dette har medført en fragmentert tilnærming til vitenskapelig IT, der mange grupper har utviklet sine egne løsninger. Synergiene på tvers av universitetet har vært begrenset og NMBUs slagkraft innen digitalisert forskning har i hovedsak utviklet seg innenfor faggruppene og fakultetene.

Denne rapporten synliggjør dagens situasjon og systematiserer behovene ved NMBU for å styrke ansatte og hele institusjonen for en digital forskningsframtid. Rapporten skisserer også løsninger for fremtiden.

Arbeid med data i forskning har tre aspekter: produksjon, forvaltning og analyse. Produksjon av data spenner vidt, fra innsamling av data i laboratorier, klinikk og felt, via spørreundersøkelser og registerdata til simuleringer og andre beregninger. Forvaltning av data omfatter lagring og dokumentasjon ved hjelp av metadata, arkivering og tilgjengeliggjøring så vel for lokal behandling som for eksterne brukere. Analyse er alt arbeid for å ekstrahere mer informasjon og kan i seg selv omfatte krevende beregninger for å frembringe nye data.

Rapporten drøfter i sin første del databehandling i forskning utenom tunge beregninger. Disse behandles i andre del, fulgt av en diskusjon av tynge databruk i undervisningen. Rapporten avsluttes med anbefalinger.

Som del av samlokaliseringsprosjektet leverte arbeidsgruppen til delprosjektet *A5—Lagring av forskningsdata* et utkast til en rapport våren 2019 (vedlegg 1). Delprosjektet ble terminert før ferdigstilling, men utkast til rapport inneholder mye nyttig informasjon. Rapporten vår bør derfor ses i sammenheng med A5-rapporten.

3 Databehandling i forskning

Internasjonalt er det et stort fokus på at forskningsdata skal være FAIR: *findable, accessible, interoperable, reusable*, med andre ord: Forskningsdata må være tilgjengelige, gjenfinnbare og gjenbrukbare. Skal NMBU over tid fremstå som en troverdig og attraktiv samarbeidspartner i forskningen, må våre forskningsdata være FAIR. Siden det er krevende å etablere FAIRness i etterkant av en datainnsamling, er det avgjørende å legge til rette for FAIRness fra det øyeblikket datainnsamlingen planlegges: Straks data oppstår skal de registreres og lagres med nødvendig metadata i systemer som (a) gir forskerne tilgang til veldokumentert arbeid med egne data og (b) senere gir dem anledning til å dele data i henhold til internasjonale standarder. Nødvendig tilgangskontroll for sensitive data samt trygg arkivering for fremtiden er underforstått i et slikt system.

Dataenes livsløp kan deles i følgende trinn:

1. Planlegging
2. Innsamling
3. Lagring og arkivering
4. Analyse
5. Tilgjengeliggjøring (deling)

Rapporten gjennomgår disse trinn hver for seg.

3.1 Planlegging

Å bygge sett med solide vitenskapelig data som kan deles FAIR krever planlegging. Forskningsrådet, European Commission med flere har derfor innført krav om datahåndteringsplaner (Data Management Plans, DMP) for prosjekter de finansierer. En DMP beskriver type og omfang av forskningsdata som frembringes i prosjektet, hvilke metadata som registreres for å systematisere data, dokumentasjon av relevante godkjenninger for håndtering av data, hvordan data skal lagres sikkert i løpet av levetiden og hvordan de vil bli arkivert og tilgjengeliggjort i tråd med FAIR-prinsippene etter at prosjektet er avsluttet. Planen skal også oppgi hvem som er ansvarlig for behandling av data.

NMBU har inngått en egen avtale med Norsk senter for forskningsdata (NSD) knyttet til deres datahåndteringsplanverktøy DMP+ hvor NMBUs krav og retningslinjer er integrert. DMP+ verktøyet er dynamisk, og datahåndteringsplanen kan oppdateres løpende så lenge et prosjekt er operativt. DMP+ kan benyttes for alle typer av forskningsdata.

Mange forskere opplever datahåndteringsplanene som en byråkratisk byrde heller enn som et nyttig verktøy. På dette området er det stort potensiale for synergier ved å ta i bruk og videreutvikle protokoller for diverse fagfelt ved NMBU i fellesskap, ved å identifisere egnede verktøy og utvikle arbeidsflyt som lett kan overføres fra et prosjekt til neste. Arbeidet med datahåndteringsplaner kan slik bli til et produktivt verktøy.

Et sentralt prinsipp i utformingen av datahåndteringsplaner bør være å tenke på systematisk, maskinlesbar dokumentasjon (metadata) og langsiktig tilgjengeliggjøring fra dag én: I et noe lengre perspektiv bør data arkiveres straks data oppstår og all videre behandling bør skje på basis av disse arkiverte data (kopi av rådata). Fordelen ved dette er at data er trygget ivaretatt fra starten, at bidrag og opphav er systematisk dokumentert og at kvaliteten på dataarkivet blir løftet siden man unngår arkivering av data som en ettertanke. Umiddelbar arkivering i egnede systemer kan også bidra til å redusere behov for lagringskapasitet siden

alle brukere kan benytte seg av de arkiverte data. Per i dag må data håndteres med eksisterende teknologi og verktøy: tryggest mulig lagring under arbeid etterfulgt av arkivering i egnet dataarkiv.

Ved samarbeid med forskere fra andre institusjoner må datahåndteringsplaner også beskrive hvordan ansvar for data deles mellom partnerinstitusjonene. Her kreves en balanse mellom et ønske om å samle alle data fra et prosjekt på et sted og et ønske om å samle alle data produsert av hver partner på tvers av prosjekter hos partneren. Det er ønskelig å unngå duplisering av data innenfor prosjektet, samtidig som det er viktig å sikre langsiktig tilgjengelighet av data. En løsning her kan være å lagre data i sentrale dataarkiver for forskning som nå utvikles mange steder.

Fagrådets inntrykk er at arbeidet med datahåndtering ved NMBU er forholdsvis fragmentert i dag, der de forskjellige forskergruppene arbeider med sine respektive strategier. Det er allerede etablert et grunnleggende opplæringstilbud i datahåndtering generelt og håndtering av sensitive data samt utvikling av datahåndteringsplaner. Opplæringen gis i form av kurs organisert av Norsk senter for forskningsdata og Digitalt Liv Norge/Elixir. Sistnevnte kurs har fokus på prosjekter innen livsvitenskap spesielt. Implementering av protokoller for datahåndtering innenfor NMBUs ulike fagfelt er viktig for tiden fremover, og det er viktig at fagmiljøene selv bidrar innenfor sine fagområder.

3.2 Innsamling

Data samles inn på mange forskjellige måter og er av mange typer og størrelser. Noen typer data er sensitive og må behandles deretter, alle er verdifulle siden det vil være meget krevende å gjenskape disse ved tap.

Forskere ved NMBU møter i dag utfordringer i forbindelse med datainnsamling på forskjellige områder:

- Avansert laboratorieutstyr kan ofte bare brukes ved hjelp av datamaskiner og spesialprogramvare som følger med utstyret. Hvis disse systemene ikke holdes à jour av leverandøren, kan datamaskinene utgjøre en sikkerhetsrisiko, slik at tilkoblingen til NMBUs nettverk kan være problematisk. Foreldet programvare som ikke støtter åpne protokoller kan også gjøre det vanskelig å overføre data direkte fra måleinstrumenter til elektroniske labjournaler (LIMS, laboratory information management system). Art og omfang av disse problemer ved NMBU bør kartlegges nærmere.
- Overføringskapasitet for data kan sette begrensninger, spesielt der nettverket ikke er i samsvar med dagens standard. Et konkret eksempel er deler av Meieribygget som fremdeles bare har et 100 Mbit nettverk. Her kreves det investeringer i ny infrastruktur.
- Omfanget av data kan være såpass stort at overføring er prinsipielt krevende. Dette gjelder spesielt bilde- og videodata, som raskt når mange TB (terabyte, 1000 GB). Her kan man raskt nå en grense der den eneste praktiske løsningen er å transportere data på harddisk fra innsamlingspunktet til et sted på campus der data overføres til permanent lagring og deretter unngå mest mulig å flytte på eller duplisere data.
- Transport av data i form av disk er spesielt relevant ved innsamling i felt med ofte begrenset båndbredde på nett. I slike tilfeller er det viktig å ha gode planer og rutiner for sikring av data før de overføres til permanent lager på campus, for å unngå at tap av eller skade på en enkel disk ødelegger et forskningsprosjekt.

- Pasientjournalssystemet ved dyresykehuset er sentralt for sykehusdriften. Hvis data fra journalssystemet skal brukes i forskning vil det være behov for løsninger som støtter uthenting av forskningsdata fra som ivaretar sikker klinikkdrift.
- Når sensitive data innhentes i felt, f.eks. i form av dybdeintervjuer med informanter, videoopptak i samhandlingssituasjoner, o.l., må disse data behandles trygt fra det øyeblikket de oppstår og før de overføres til f.eks. TSD (Tjenester for sensitive data ved UiO). Dette krever ryddige rutiner i form av definerte protokoller. Gode og etablerte protokoller vil dessuten gjøre det enklere å veilede studenter innen datahåndtering av sensitive data. Disse bør omfatte klare regler for sletting av midlertidig lagrede data etter overføring til permanent lager.

3.3 Lagring og arkivering

Det er krevende å produsere forskningsdata og dataene kan tape verdi dersom det ikke er godt dokumentert hvordan dataene ble til og hvordan man har sikret at dataene er uforandret. Det er derfor vesentlig at alle forskningsdata beskrives med metadata, og at data og metadata lagres sammen. Det er også viktig at man i denne prosessen legger til rette for bruk av data i videre analyser. Tilgang til visse type data må dessuten kontrolleres nøye, enten fordi data er sensitive (personvern) eller konfidensielle (forretningshemmeligheter).

Ansvarlig lagring av NMBUs forskningsdata har flere komponenter: For det første selve "rålageret" bestående av hardware og systemprogramvare, inkludert systemer for data-sikring og tilstrekkelig backup. Derest systemer for trygg forvaltning av data inkludert metadata; disse systemene må legge til rette for bruk av data i analyser og for deling av data, så vel internt som eksternt. Det øverste sjiktet er klientprogramvare som gir forskerne tilgang til sine data, støtter overføring av data fra instrumenter til lageret og åpen deling av data i form av forskningsdataarkiv.

3.3.1 Eksisterende systemer for lagring av forskningsdata

NMBU har i all hovedsak to dedikerte systemer for lagring av forskningsdata: (a) W: (også kjent som LargeFile) som benyttes fra Windows-systemer og (b) disksystemet til Orion clusteret. Selv om systemene er logisk adskilt fra hverandre kjører de på samme serverhardware. Den totale lagringskapasiteten for systemene inkludert backup er i dag 3,2 PB (petabyte). For å sikre data også i tilfelle brann eller liknende, lagres data på fysisk adskilte steder.

I tillegg benytter NMBU Microsoft OneDrive som verktøy for lagring og deling av data i sky. OneDrive er ikke ment til lagring av forskningsdata og en rekke forskermiljøer rapporterer om tekniske problemer med OneDrive i forbindelse med store datasett. OneDrive kan allikevel benyttes som et samhandlingsgrom for samarbeid om data ved NMBU og med eksterne partnere. Rådata skal da lagres på W: og kopieres til OneDrive. Dette er den foretrukne løsningen for datadeling i pågående prosjekter og det vises til at bruk av Dropbox eller Google Drive ikke er tillatt. NMBU tar backup av data som befinner seg i OneDrive skyen til lokale servere.

For store datasett (10 TB og mer) kan den nasjonale lagringsløsningen Sigma2/NIRD benyttes. For lagring av sensitive forskningsdata har NMBU avtale med UiO om tilgang til deres lagringsplattform for sensitive data (TSD). Hvordan kostnadene for bruk av NIRD og TSD vil utvikle seg over tid er usikker og utenfor NMBUs kontroll.

I tillegg til løsningene beskrevet hittil lagres forskningsdata også på ymse lokale harddisker, interne eller eksterne. Sistnevnte har ingen backupløsning og er derfor svært dårlige lagringsalternativer.

3.3.2 Systemer og verktøy for arkivering og tilgjengeliggjøring av forskningsdata

NMBU har et eget dataarkiv, NMBU Open Research Data (NMBU ORD), for arkivering av forskningsdata ved prosjektslutt eller ved publisering. NMBU ORD kan også brukes til å dele data samt metadata gjennom et webgrensesnitt. NMBU ORD er dermed egnet til å gjøre data FAIR etter at arbeidet med data er avsluttet. Samtidig er NMBU ORD egnet for arbeid med data mens et prosjekt pågår og er heller ikke tilpasset sensitive data eller store datasett.

Mange forskningsmiljøer benytter seg av fagspesifikke dataarkiv. Dette er spesielt utbredt innenfor systembiologi, men er også voksende innenfor andre fagområder.

3.3.3 Utfordringer med dagens løsninger

Dagens lagringsløsning nærmer seg kapasitetsgrensen og slutten av sin tekniske levetid.

Det er en stor utfordring at dagens lagringsløsning W: ikke tilgjengelig fra Orion clusteret og at filsystemet til Orion omvendt ikke enkelt er tilgjengelig fra Windows. Et fremtidig system må sikre gjensidig tilgang og vil da sannsynligvis også sikre tilgang til W: fra arbeidsstasjoner og personlige datamaskiner som ikke benytter Windows. Det er også utfordringer med å dele store datasett via OneDrive.

Lagring av vitenskapelig data ved NMBU skjer i dag ofte i form av egenutviklede mappe-strukturer som representerer metadata. I noen tilfeller har forskere benyttet seg av dato for lagring av filer som et implisitt organiseringskriterium. Sistnevnte kan være et problem ved overføring mellom ulike systemer, da metadata lett kan gå tapt når filer flyttes.

I dataarkivet NMBU ORD finnes i dag bare 12 datasett, noe som tyder på at det ikke er tatt i bruk i særlig grad av NMBUs forskere. Det er viktig at NMBU ORD gjøres bedre kjent blant forskere og studenter, og at det legges til rette for kompetanseheving. For at universitetsbiblioteket skal kunne bistå forskere og studenter er det viktig at det settes av nok ressurser både til generell opplæring og til bistand ved arkivering av data og eventuell tilgjengeliggjøring.

3.3.4 Anbefalinger vedrørende lagring

Viktige punkter for videreutvikling av lagringsløsningen for forskningsdata ved NMBU er:

- Data må lagres trygt, med backup på fysisk adskilt sted.
- Alle typer data som forskere ved NMBU behandler må kunne håndteres (mulige unntak for sensitive data som må lagres ved TSD og ekstremt store datamengder må lagres ved Sigma2/ NIRD).
- Data må være direkte tilgjengelig for analyse både fra arbeidsstasjoner og lokale servere (Orion), helst også fra nasjonale og europeisk tungregneanlegg.
- Data må dokumenteres med metadata og underlegges versjonskontroll. Dette krever styrkingen av "datakulturen" ved NMBU. Løsninger for dokumentering av data som ligger f.eks. ved NIRD må vurderes.
- Kildekode til programmer og skripter som brukes i databehandling er i seg selv en form for data som må arkiveres underlagt versjonskontroll.

- Dagens lagringsløsning må erstattes når den når slutten av sin tekniske levetid og supportperiode i Q4/2021. Investeringsvolumet er på cirka 9 millioner kr. IT-avdelingen har kommunisert dette i langtidsprognoser for budsjett siden 2017.
- Det må planlegges med tilstrekkelig kapasitet for betydelig vekst i volum av forskningsdata. Basert på tidligere vekst forventes at volumet vil vokse fra 3,2 PB i dag til 4,5 PB innen 2023 og 7 PB innen 2026.
- I henhold til standarder i bransjen må vi planlegge med og budsjettere for en totalfornying av lagringshardware hvert femte år.

3.4 Analyse

Forskningsdata vil typisk gå gjennom mange behandlingstrinn før publiserbare resultater er klare. Dette strekker fra initial kontroll av data for registreringsfeil og liknende problemer til avanserte analyser. Arbeidet med data bør dokumenteres systematisk, slik at det foreligger en gjennomgående "audit trail" fra registrering av rådata til endelig resultat.

Dette betyr at hvert trinn i behandlingsprosessen må bygge på data som er registrert med metadata i datalageret, utføre operasjoner på disse data og lagre resultatene i datalageret med metadata som viser til inngangsdata og behandlingsmåter brukt. Data blir i en slik prosess aldri forandret i datalageret, det blir bare lagt til ny data¹.

En slik prosess krever at data er lett tilgjengelig fra datalageret på alle maskiner som skal brukes til analyser. For enklere analyser vil dette som oftest være brukernes PC, for tyngre analyser av f.eks. omfattende omics-, bilde- og videodata dedikerte servere som Sigma2 NIRD eller Orion clusteret. Prosessen krever dessuten programvare som gjør det så enkelt som mulig for forskere å arbeide med data fra datalageret innenfor sine egne arbeidsprosesser.

Avansert analyse krever avanserte datasystemer. Forskning pågår ofte på grensen av det teknisk mulige og krever tilpasset utstyr. Orionclusteret ble f.eks. anskaffet fordi avanserte genomanalyser krever maskiner med meget stor minnekapasitet. Den siste utvidelsen av Orion clusteret i 2020 la til regneservere med kraftige GPUer (grafikkort til beregningsformål) dedikert til deep learning metoder, spesielt konvolverende nevrale nettverk.

Behovene i forskningsfronten oppstår gjerne betydelig raskere enn investeringssyklusen og tilpasningsfarten til de nasjonale tungregnesystemene. Sigma2 har f.eks. nylig oppgradert sine tungregnesystemer, men har allikevel bare 40 GPUer tilgjengelig på alle sine systemer til sammen (Orion har p.t. 12 GPUer). For at NMBU skal være med i forskningsfronten, må universitetet kunne reagere raskere og også tilby verktøy som følger våre prioriteringer på grunnlag av vår forskning.

Det er nå en betydelig økning i behov for regnekraft som kan støtte avansert visualisering og behandling av grafikkdata som CAD (Computer Aided Design), BIM (Building Information Modeling), VR/AR (Virtual/Augmented Reality), AI (Artificial Intelligence), 3D modellering og rekonstruksjon, punktskymodeller og fotogrammetriske analyser. Disse behov må kartlegges

¹ Behandles data ved hjelp av skripter, Jupyter notebooks o.l., kan benhandlingstrinn dokumenteres ved å lenke til versjonskontrollert kode. Dokumentasjon av databehandling via grafiske brukergrensesnitt er mer krevende. Også i slike tilfeller kan inngangsdata, benyttet programvare og utgangsdata registreres; ideelt sett skulle programvaren generere en logg av behandlingstrinn som kan lagres sammen med data. Regneark (Excel) er kanskje den vanskeligste formen siden den blander data, programkode (formlene i cellene) og arbeidsskritt gjennom GUI. Her må man nok lagre en ny versjon av regnearket etter hver arbeidsøkt.

nærmere for å legge føringer for videre utvikling av Orion clusteret og server for virtuelle maskiner.

3.4.1 Fornyng av Orion

Orion består i dag av 18 servere. Åtte av disse ble anskaffet i 2019 og 2020. Resterende servere er utenfor garantiperioden, og bør erstattes snarest. Estimert kostnad for oppgradering, før en eventuell utvidelse, er 2–3 millioner kroner.

På lengre sikt (2023–2026) bør åtte Orion-servere, som har ekstra stor minne-kapasitet, erstattes til en estimert kostnad på 2–4 millioner kroner.

3.5 Tilgjengeliggjøring

Forskningsdata skal håndteres i tråd med FAIR-prinsippene og tilgjengeliggjøres så raskt som mulig, senest med publisasjon av resultatene. Noen fagfelt som systembiologi har lang erfaring med dette og godt innarbeidete internasjonale løsninger. Andre fagfelt har forskningsdata som ikke kan tilgjengeliggjøres grunnet sensitivt innhold (personvern) eller andre faktorer som mulige kommersialiseringer og/eller forretningshemmeligheter. Det er allikevel viktig at forskningsfellesskapet får informasjon om hva slags type data som er innsamlet og under hvilke forhold. Dette beskrives gjerne publisasjonene og i datahåndteringsplanen.

Det er viktig å planlegge for tilgjengeliggjøring av data allerede ved innsamlingen. Dersom tilrettelegging for tilgjengeliggjøring av data skjer ved prosjektslutt eller ved publisering, kan det bli krevende og oppleves som belastende ekstraarbeid. Det vil derfor være et overordnet mål at forskere og studenter har tilgang på systemer og verktøy som gjør det mulig å tilrettelegge for tilgjengeliggjøre data fra starten av. I et slikt system vil tilgjengeliggjøring være mulig på publiseringstidspunktet ved kun å endre innstillingene for tilgang til dataene.

NMBU ORD er i prinsipp et slikt system, men arbeidet med å tilrettelegge filer må gjøres av forskeren på forhånd. Universitetsbiblioteket bistår forskere og studenter for å sikre at alle nødvendige rettigheter, lisenser og krediteringer er på plass og kvalitetssikre at nødvendig metadata i henhold til FAIR er på plass på publiseringstidspunktet. I tillegg kan de bistå forskere med å etablere egne metadata-templater i tilfeller med store prosjekter der det skal arkiveres mye over tid.

4 Tungregning

Tunge beregninger gjennomføres både som simuleringer som produserer data og analyser som bearbeider data. For de tyngste beregningene vil fagmiljøene ofte benytte nasjonale (Sigma2 og TSD) eller internasjonale ressurser (for eksempel PRACE og EuroHPC). For spesialiserte beregninger og analyser kan det allikevel være nødvendig med egnede lokale ressurser ved NMBU, som Orion (jf. avsnitt 3.4). Dette gjelder spesielt tilfeller der store datavolum er berørt og databehandling må foregå nær datalageret, og der spesialisert hardware som GPU og FPGA (programmerbar hardware for spesielle analyser) kreves. Lokale ressurser er også en stor fordel ved utvikling av nye metoder og opplæring av studenter og unge forskere.

På lik linje med vitenskapelig utstyr må utstyr (servere) for tungregning skiftes ut omtrent hvert femte år, men her er det større rom for fleksibilitet: Forskjellige deler av systemet kan fornyes i forskjellige år og til en hvis grad kan man ta sjansen på å kjøre systemet videre

etter at serviceavtalen er utløpt. Konkrete planer, både i henhold til utforming av systemet, langsiktig finansiering og innføring av leiestedsbetaling må diskuteres med relevante miljøene ved NMBU.

5 Undervisning

En rekke emner ved NMBU krever spesialisert programvare og tynge beregninger. Tidligere har man stort sett benyttet datasaler i disse emner. COVID-restriksjonene har medført betydelig overføring av programvarebruk til studentenes egne datamaskiner og dels til virtuelle maskiner.

Bruk av studentenes egne maskiner medfører en del problemer, spesielt med hensyn til lisenser og datakraft: ikke alle studenter har maskiner som er kraftige nok til å kjøre for eksempel tunge DAK programmer. Det er derfor ønskelig å legge til rette for utstrakt bruk av virtuelle maskiner. LANDSAM har gjennomført et større emne i januarblokk med utstrakt bruk av virtuelle maskiner, som kan være et godt utgangspunkt for videre utvikling. Viktige punkter her er å sikre at nødvendig programvare kan kjøre på virtuelle løsninger (både teknisk og med hensyn til lisenser) og at serverne som de virtuelle maskinene kjører på er tilstrekkelig kraftig til å kunne støtte store kurs.

6 anbefalinger

Den digitale forskningsfremtiden ved NMBU krever en betydelig videreutvikling av kulturen i organisasjonen. Dette er en krevende oppgave som må planlegges nøye for å få alle med på fremtiden: Forskerne må oppleve at det digitale hamskiftet gir deres arbeid et løft. Alle videre planer må derfor ta utgangspunkt i hvordan de vil påvirke og kan forbedre forskernes arbeidssituasjon. Samtidig må essensielle behov ivaretas løpende.

Nesten alle forskere ved NMBU arbeider med data. Data og datahåndtering er dermed et felles tema for hele campus og derfor godt egnet som fokus på veien mot den digitale forskningsfremtiden.

NMBU bør i nærmeste fremtid opprette et prosjekt med tilstrekkelige ressurser for å kartlegge dagens status i databehandling ved NMBU i detalj og utvikle konkrete forslag for systematisering av arbeid med data, både i form av tekniske løsninger og protokoller. Arbeidet kan bygge på A5-rapporten og denne rapporten, men bør gå i betydelig større detalj. Et punkt vil være å utforske og dokumentere "brukerreiser", dvs følge enkelte forskere/ forskergrupper for å dokumentere hvordan disse håndterer data i dag. Dessuten bør informasjon systematiseres og detaljeres om hva slags typer data som finnes og frembringes ved NMBU, hvilket volum det har, hvordan den kommer til å vokse, og om lagring ved NMBU eller på fagspesifikke eksterne ressurser vil være mest nærliggende.

Basert på denne informasjonen kan man så utvikle konkrete anbefalinger for en felles tilnærming til databehandling ved NMBU. For utprøving og innfasing av ny teknologi og nye rutiner vil det være en fordel å velge noen sterke forskningsgrupper som piloter som både kan gi tilbakemeldinger om nødvendige justeringer og være lokomotivene som drar de andre med seg når systemet rulles ut bredt.

Dette vil være et omfattende arbeid og det bør antakelig settes av minst ett årsverk for et slikt analyse- og utviklingsprosjekt. Prosjektet koordineres av IT-avdeling med bistand fra Forskningsavdelingen og fagråd for e-infrastruktur kan gjerne stille som et prosjektstyre.

Når prosjektet blir klar til bredere utrulling, er det vesentlig å sette av nok tid og ressurser til innfasing av ny teknologi og nye rutiner. Her må spesielt fakultetene motiveres til å skape tid hos sine ansatte for å kunne sette seg inn og bli godt kjent med de nye verktøyene. Dette gjelder kanskje spesielt doktorgradsstudenter, postdoktorer, forskere og teknikere som håndterer data daglig. Samtidig bør alle emneansvarlige, spesielt de med emner for laboratoriekurs, oppfordres til å integrere NMBUs metoder for datahåndtering i sine kurs, slik at studentene er godt opplært i metodene når de begynner på sine masteroppgaver.

Mens dette prosjektet pågår, må løpende drift av sentralt utstyr for vitenskapelig data-behandling ivaretas. **Det er essensielt at maskinvaren for lagringsløsningen fornyes** når servicekontrakten utløper i slutten av 2021 (ca 9 mill kr). Dette bør dekkes av NMBU sentralt. Videre bør de komponentene av **Orion** som ikke lenger er dekket av servicekontrakt **fornyes** så raskt som mulig. Inntil en langsiktig finansiering gjennom leiestedsmodell er på plass bør man her se på hva om kan realiseres i et spleiselag mellom de fakultetene som benytter Orion mest og NMBU sentralt.