

Forslag til modell for framtidige investeringer og finansiering av datainfrastrukturer

Utvalget skal anbefale hvordan organisering og investeringer i datainfrastrukturer for FAIR forskningsdata bør innrettes framover. Det er Kunnskapsdepartementet som har bedt om å få et slikt grunnlagsdokument oversendt innen 1. mai 2022 som et innspill til arbeidet med langtidsplanen for forskning og høyere utdanning. Viktige momenter i utvalgets oppdrag er å identifisere hvilke datatjenester det vil være behov for framover, innenfor hvilke områder det er behov for nye tiltak, hvilke gevinster (og eventuelle risikoer) ulike finansieringsmodeller vil kunne ha, og en generell anbefaling om ambisjonsnivå. I dette notatet er det imidlertid hovedsakelig viet plass til foreløpige anbefalinger om en finansieringsmodell for investering og drift av datainfrastrukturer.

I anbefalingene tar utvalget hensyn til at det er stor variasjon i eksisterende datainfrastrukturer, for eksempel hvilke behov de dekker, hvor godt utviklet de er, om de tilbyr generiske eller domenespesifikke datatjenester og om de tilbyr kun lokale tjenester eller er åpen for nasjonale brukere utenfor vertsinstitusjonen. Fordi behovene datainfrastrukturene dekker er så forskjellige, mener utvalget at det vil være urealistisk og lite hensiktsmessig å etablere en "superinfrastruktur" som skal dekke alle behov, men at det utvikles løsninger som gjør det mulig å dele og bruke data på tvers av infrastrukturer som betjener ulike områder. Infrastrukturenes funksjon, organisasjonsmodell, kostnadsnivå, kapasitet, brukermasse m.m. vil også påvirke hvordan kostandene knyttet til driften kan og bør dekkes. Av denne grunn vil heller ikke utvalget foreslå én måte å finansiere datainfrastrukturer på, men en finansieringsmodell (et rammeverk) med flere ulike komponenter som kan kombineres på ulike måter.

Det er viktig å understreke at det allerede brukes mye ressurser på investering og drift av infrastruktur- og datatjenester. Et av formålene med utvalgets anbefalinger er å sørge for kostnadseffektiv bruk av disse ressursene, uten at det går på bekostning av ulike brukerbehov eller begrenser forskernes tilgang til gode og relevante infrastrukturer og tjenester. Et viktig tiltak for å oppnå mer optimal bruk av ressurser er å koordinere investeringer i infrastruktur på et nasjonalt nivå for å unngå at det etableres *unødvendige* parallelle infrastrukturer og/eller tjenester rundt om på institusjonene. Det bør også vurderes i hvilken grad norske miljøer bør delta i internasjonale infrastrukturensamarbeid, enten i kombinasjon med, eller som en erstatning for investering i nasjonal infrastruktur. I tilfeller hvor det allerede finnes mekanismer for organisering, styring og finansiering som fungerer godt, vil utvalgets generelle anbefaling være at dette videreføres.

Anbefalinger om finansieringsmodell for investering og drift av datainfrastrukturer

Etablering, videreutvikling og drift av datainfrastrukturer kan innebære mange ulike kostnadselementer, som for eksempel anskaffelse og oppgradering av fysisk infrastruktur, drift av fysisk infrastruktur, utvikling og drift av digitale tjenester, brukerstøtte, utvikling av relevant kompetanse for både brukere og tilbydere og eventuelt videreutvikling av infrastrukturene. En bærekraftig finansieringsmodell for slike infrastrukturer må ta opp i seg at alle disse kostnadselementene må dekkes inn.

Inndeling i tre tjenestekategorier

Forskningsdata vil typisk gå gjennom et dynamisk livsløp i et forskningsprosjekt, fra planlegging og innsamling, via aktiv bruk med analyser og bearbeiding, til arkivering og eventuell deling, og underveis i denne prosessen vil det oppstå ulike infrastruktur- og tjenestebehov. Utvalgets

anbefalinger baserer seg i det følgende på en grov, overordnet inndeling i tre kategorier med ulike typer verktøy, tjenester eller funksjonalitet (heretter "tjeneste") som kan tilbys av datainfrastrukturer underveis i denne livssyklusen. Hvilken kategori tjenesten faller inn under vil ha innvirkning på hvilken type finansiering som er hensiktsmessig og realistisk.

En enkelt infrastruktur kan ha tilbud innen flere av kategoriene, og beskrivelsen av disse kategoriene er ikke uttømmende og vil ikke i detalj dekke alle typer funksjonalitet og varianter av tjenester tilknyttet datainfrastrukturer.

1. Grunnleggende, generisk e-infrastruktur som er en forutsetning for datadrevet forskning. Dette vil inkludere fysisk infrastruktur (for eksempel høyhastighetsnett og datalagring), tungregning, samt generisk (interdisiplinær) programvare for analyse og prosessering av stordata. En del av tjenestene til Sigma 2 vil falle inn kategorien grunnleggende, generisk infrastruktur.
2. Verktøy og tjenester knyttet til aktiv bruk, deling og gjenbruk av data. Dette kan være tilknyttet både generelle og fagspesifikke infrastrukturer, men vil i betydelig grad være fagspesifikke og disiplintilpassede verktøy og tjenester. ELIXIR, TSD og microdata.no er eksempler på infrastrukturer som tilbyr slike tjenester for bruk, deling og gjenbruk av forskningsdata.
 - 2.1. Tjenester som tilgjengeliggjør data for gjenbruk i forskning. Dette inkluderer publisering av data, databaser, datakataloger for søk, plattformer og tjenester for utlevering/tilgang til data etc.
 - 2.2. Tjenester og verktøy for aktiv bruk, håndtering, analyse og prosessering av data i forskningsprosjekt. Dette inkluderer samhandlingsplattformer, verktøy for metadata, ontologier og annotasjoner, "data management systems" og elektronisk labjournal, programmer og tjenester (pipelines/workflows) for reproducerbar analyse, visualisering og prosessering.
3. Tjenester for langsiktig forvaltning av data. Dette gjøres gjerne i konteksten av et dataarkiv (repository) som ofte tilbyr mange ulike tjenester, men i denne kategoriseringen skiller selve den langsiktige forvaltningen ut som en egen, grunnleggende tjeneste, separat fra tilgjengeliggjøringen av data (kategori 2.1). Infrastruktur som tilbyr slik langtidsforvaltning kan både være generelle, tverrfaglige eller fagspesifikke. Helseregistrene, NSDs arkiv og Nasjonalbibliotekets samlinger er eksempler på slike tjenester for langsiktig forvaltning av data.

I tillegg er vellykket datadrevet og delingsbasert forskning avhengig av gode datatjenester og systemer som sørger for blant annet standardisering, gjenfinnbarhet og kreditering (PID-utstedelse, CRIS-systemer, siteringsverktøy, lisenser m.m.) som er godt integrert i datainfrastrukturene som brukes. Mange av disse tjenestene dekkes gjennom nasjonale og internasjonale fellestjenester kombinert med kompetanse og bistand i støttesystemene ved institusjonene og er derfor ikke like relevante for utvalgets anbefalinger om investering i infrastruktur.

Tre hovedformer for finansiering

Infrastrukturer som ligger til grunn for tjenestene som er beskrevet over kan finansieres på flere ulike måter. Utvalget har delt inn finansieringen i tre hovedformer, som vi vil bruke i anbefalingene under. I mange sammenhenger vil finansieringen være en kombinasjon av disse typene.

1. **Langsiktig finansiering** – stabil og forutsigbar finansiering over tid. Finansieringen kan for eksempel komme i form av grunnfinansiering fra ett relevant departement eller fra flere departementer etter en fordelingsnøkkel basert på hvilke forvaltningsområder infrastrukturen leverer forskning og forvaltning til. Slik finansiering kan også komme i form av spesifikke forvaltningsoppdrag eller gjennom forpliktende medlemskap/partnerskap fra institusjoner som bruker infrastrukturen over tid.
2. **Konkurransesatt finansiering** – finansiering som et resultat av konkurranse mellom flere aktører som vurderes etter gitte kriterier, som for eksempel Forskningsrådets infrastrukturutlysning.
3. **Brukerfinansiering/brukerbetaling** – finansiering som kommer fra de som bruker tjenestene. Dette kan enten være i form av abonnement eller stykkpris- eller forbruksfinansiering, eller en kombinasjon av disse.

Utvalgets anbefalinger

Basert på tidligere utredninger¹, diskusjoner i utvalget, innspill som utvalget har fått og foreløpige funn som er presentert i [FAIR-utredningens rapport](#), har utvalget kommet fram til følgende foreløpige hovedanbefalinger om finansiering av ulike datainfrastrukturer som er viktig for å oppnå FAIR forskningsdata:

1. Finansieringen av grunnleggende, generisk e-infrastruktur må økes og være forutsigbar i årene som kommer for å imøtekomme behovene for slik infrastruktur på tvers av hele forskningssektoren. Fra flere hold, blant annet fra Forskningsrådet og den nylige [evalueringen](#) av Forskningsrådets INFRASTRUKTUR-ordning, er det slått fast at det er behov for å øke størrelsen på finansieringen av denne typen infrastruktur. Dette er etablerte infrastrukturer som tilbyr tjenester det er behov for nå og i fremtiden.
 - a. **En større andel av finansieringen av grunnleggende, generisk infrastruktur bør komme gjennom langsiktig finansiering.** Det er uheldig at over halvparten (60%) av finansieringen av slik infrastruktur er konkurransesatt. Det skaper usikkerhet og risiko for tap av viktig kompetanse og en reell fare for at viktige forsknings- og forvaltningsoppgaver ikke kan gjennomføres. Denne typen infrastruktur betjener alle fagområder og prosjekter innenfor både UH-sektor og instituttsektor, og i takt med at forskningen innen alle fag blir mer og mer dataintensiv og datadrevet, vil behovet øke. For å oppfylle nasjonale føringer om langsiktig lagring av forsknings- og forvaltningsoppgaver vil det kreve en langsiktig og forutsigbar finansiering.
 - b. **Tjenester som spesielt tilbyr lagringsressurser underveis i et forskningsprosjekt kan helt eller delvis brukerfinansieres.** Midler til slik betaling for bruk må da legges inn som del av budsjettene til forskningsprosjekter, både i prosjekter med ekstern finansiering og i prosjekter som institusjonene selv dekker. For større institusjoner som har mange prosjekter som genererer store mengder data, kan slik betaling med fordel skje gjennom institusjonsbidrag til infrastrukturene, jf. ny Sigma2-modell. Dette vil bidra til både mindre administrasjon og mer stabil og forutsigbar finansiering av nasjonale e-infrastrukturer som tilbyr slike tjenester.

¹ Blant annet [Behov og finansieringsstrategi for nasjonal e-infrastruktur for forskning for perioden 2020 – 2030](#) og [Tilgjengeliggjøring av forskningsdata og data for forskning](#)

2. **Datainfrastrukturer for aktiv bruk, deling og gjenbruk av forskningsdata anbefales finansiert i en kombinasjon av langsiktig finansiering, konkurranse- og brukerbetaling.**
 - a. **Dersom infrastrukturen dekker et nasjonalt behov, kan det vurderes om deler av driften skal dekkes gjennom en langsiktig finansiering.** For å få grunnfinansiering, må datainfrastrukturer oppfylle bestemte krav til hvilke tjenester de skal levere. Dette kan for eksempel være krav om å gjøre data FAIR, gjøre data komplementære til andre data, at data må gi faglig merverdi ut over eksisterende grunnfinansierte tjenester og godkjente internasjonale datainfrastrukturer, og det må stilles tekniske krav til tjenesten. Dette gjør at vi kan bygge på dagens eksisterende tjenester, og med det ivareta domene- og institusjonsheterogenitet, samtidig som vi sørger for at ikke alle institusjoner eller enkeltprosjekter lager egne datainfrastrukturer og -tjenester. Utvalget tar ikke stilling til hvem som skal vurdere om en tjeneste oppfyller disse kravene, men vil peke på at dette kan sammenlignes med den jobben Forskningsrådet i dag gjør med å vurdere om forskningsinstitutter oppfyller kriteriene for å få grunnfinansiering, og delvis også vurderingen av hvorvidt en forskningsinfrastruktur fyller kravet om å være nasjonal. Det finnes også flere internasjonale sertifiseringsstandarder, for eksempel gjennom Core Trust Seal, GoFair og EOSC som vil være relevante i denne sammenheng.
 - b. **Nyetablering av datainfrastrukturer eller videreutvikling av tilbudet til en datainfrastruktur for deling og gjenbruk av forskningsdata konkurranseutsettes.** Dette kan gjennomføres gjennom dagens Infrastrukturprogram i Forskningsrådet eller en egen konkurransearena. Det er viktig å få en kvalifisert vurdering av hvilken merverdi det vil ha å utvikle tjenestetilbudet og derfor vil konkurranse inkludert faglige og strategiske vurderinger egne seg. For å oppnå varig drift av en nyetablert infrastruktur, kan finansieringen komme fra ulike kilder avhengig av hvilket behov infrastrukturen skal imøtekomme. En vurdering av modell for framtidig drift må vurderes før det fattes en beslutning om å tildele midler til nyetablering.
 - c. **Tjenester som er rettet spesielt mot enkeltprosjekter, eller er tett knyttet til fysiske infrastrukturer, kan brukerfinansieres.** Dette kan for eksempel være tilpassede tjenester som gir forskerne tilgang til behandlede eller sammenstilte data og som det vil være naturlig at prosjektet betaler for. Det kan også være tjenester knyttet til avlevering og publisering av forskningsdata i tråd med FAIR-prinsippene. Ved ekstern finansiering må slike kostnader legges inn som en del av prosjektets budsjett. Institusjoner som stiller krav om FAIR forskningsdata fra sine forskere bør da også forventes å legge dette inn i sine budsjetter, enten gjennom å "belaste" enkeltprosjekter for kostnadene, eller gjennom institusjonsbidrag til relevante datainfrastrukturer. Det vil være naturlig at infrastrukturene legger en overhead på tjenestene for å dekke inn deler av driftskostnadene gjennom slik brukerbetaling.
3. Finansiering av dataarkiver for langtidslagring og kuratering av verdifulle forskningsdata og forvaltningsdata som er relevant for forskning må være et nasjonalt ansvar, slik man til en viss grad gjør innenfor ABM-sektoren i dag (ABM – arkiv, bibliotek og museum). Dette både fordi langtidslagring ikke kan dekkes av enkeltprosjekter med begrenset budsjettperiode og fordi mange typer data kan ha verdi utover et enkelt forskningsprosjekt, hvis de har gode og maskinlesbare metadata. For å sikre langsiktige arkivtjenester for forskning, næringsliv og

forvaltning er det nødvendig med forutsigbar finansiering. **Det betyr at slike arkivtjenester i all hovedsak bør ha en langsiktig finansiering.** Om en slik langsiktig finansiering skal komme fra én eller flere kilder må vurderes. Aktuelle finansieringskilder er alle departementer som har et sektoransvar for forskning. Kunnskapsdepartementet, som har et hovedansvar for forskning og utdanning (herunder infrastruktur), må ta ansvaret for å samordne og koordinere departementene på en måte som understøtter samordning av den langsiktige finansieringen. En slik sentral finansiering kan også kombineres med et krav om at alle datagenererende institusjoner selv bidrar med midler til infrastrukturen.

Et eksempel

For å vise hvordan en slik modell vil slå ut i praksis har vi tatt utgangspunkt i Statistisk sentralbyrås funksjon som dataleverandør for forskere. I tabellen under har vi kategorisert ulike tjenester som SSB har i den forbindelse og beskrevet hvordan finansieringen av de ulike tjenestene kan finansieres ut fra den foreslåtte modellen.

Tjenestekategori/aktivitet	Innsamling og forvaltning av statistikkdata (forvaltningsoppdrag)	Tradisjonell tilgjengeliggjøring av tilrettelagte data for forskere	Etablering av microdata.no	Drift og videreutvikling av mikrodata.no
Finansieringsform	Langsiktig finansiering	Brukerfinansiering (stykkepris)	Konkurransebasert finansiering	Brukerfinansiering ("abonnement")

Finansieringsmekanismer som sikrer utvikling og modernisering av teknologi, tjenester og kompetanse

I tillegg til investering i selve infrastrukturene (utstyr og personell) som skal bidra til mer FAIR forskningsdata og relevante forvaltningsdata, ser utvalget et stort behov for å øke investeringene til utvikling av ny teknologi og nye tjenester for å gjøre infrastrukturene relevante og internasjonalt konkurransedyktige for framtidige forskningsbehov. Finansiering av slike utviklingsprosjekter må sees i tett sammenheng med investering i infrastrukturene. Disse prosjekter faller gjerne utenfor etablerte finansieringsordninger fordi vi ikke har gode nok mekanismer til å identifisere, vurdere og kvalitetssikre dem. De kan være vanskelige å definere inn under etablerte kategorier, fordi de ofte befinner seg i skjæringspunktet mellom forskning, metodeutvikling, tjenesteutvikling og infrastrukturetablering. Det er ofte forbundet mye usikkerhet og risiko til slike prosjekter fordi de søker å finne løsninger på delvis ukjente, framtidige behov og kan være eksperimentelle i sin natur. De taper da ofte i konkurranse med andre, mindre risikofylte, prosjekter. Eksempler på slike prosjekter kan være utvikling av nye beregningsteknologier som f.eks. neuromorphic computing, analysemetoder for store og komplekse datamengder som baserer seg på maskinlæring og kunstig intelligens, nettverkstjenester i takt med f.eks. 5G/6G nettverksutvikling, eller nye FAIR-løsninger for arkiver. Utvalget mener derfor at det er behov for å utvide etablerte konkurransearenaer til også å omfatte utviklingsprosjekter, eller at det etableres egne arenaer med kriterier som gjør det mulig for risikoprosjekter å nå opp i konkurransen.