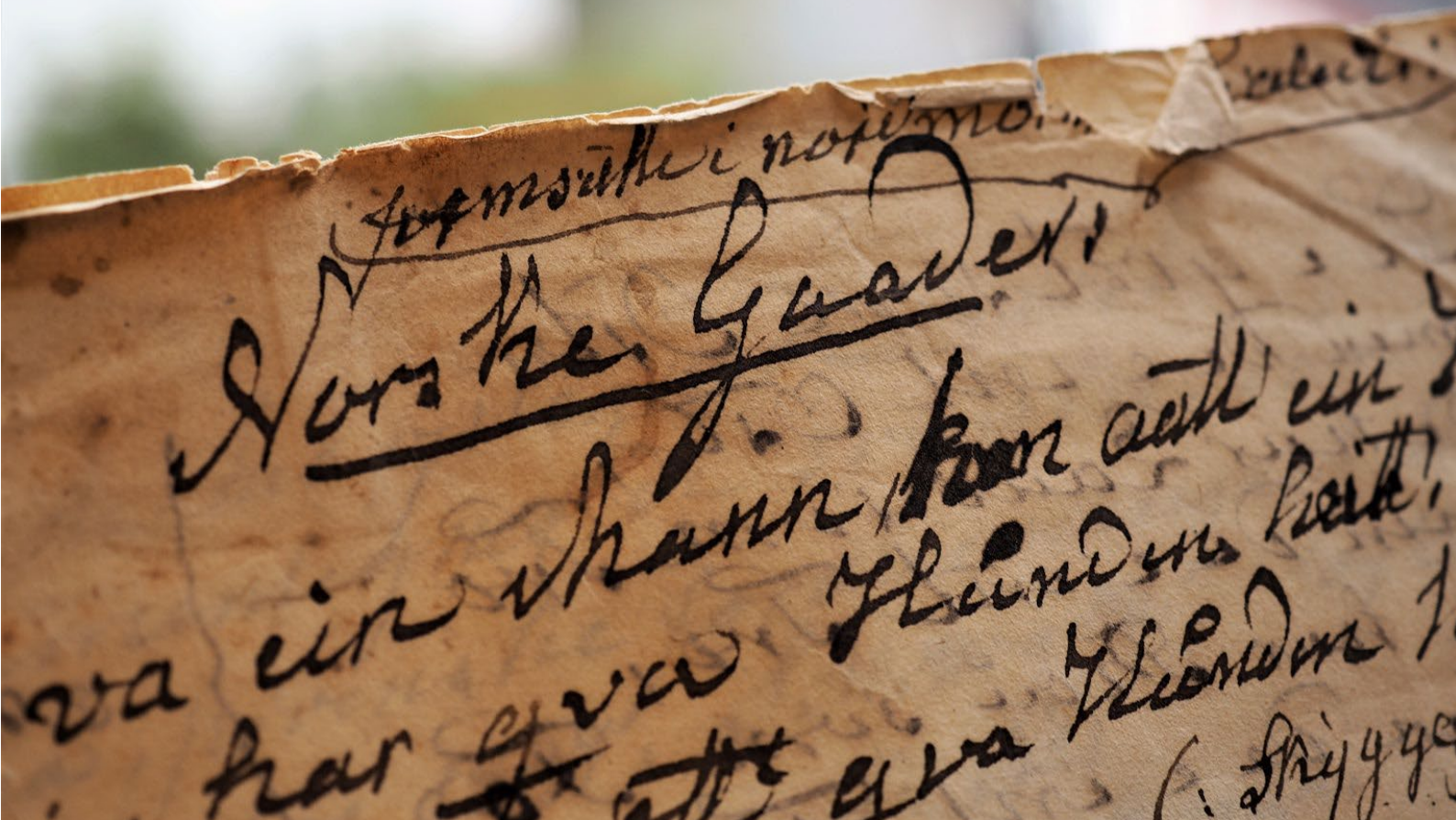




Veikart for forsknings- infrastruktur ved HF

Det humanistiske fakultet

Godkjent i Forsknings- og forskerutdanningsutvalget (FFU) 04.06.25



UNIVERSITETET I BERGEN

Innholdsfortegnelse

<i>Innledning</i>	3
<i>SAMLA</i>	3
<i>CLARINO</i>	4
<i>Strategiske behov for forskningsinfrastruktur i forbindelse med det nasjonale veikartet og Forskningsrådets utlysning av midler til forskningsinfrastruktur av nasjonal viktighet i 2025</i>	4
Koordinatorsøknad fra HF	4
Nasjonale infrastrukturer	5
<i>Veien videre</i>	6

Innledning

Det humanistiske fakultets (HF) strategi (2023-2030) har som ambisjon å stå for internasjonalt toneangivende grunnforskning. For å oppnå det skal HF blant annet bli ledende innen digital humaniora og satse videre på humanistisk forskningsinfrastruktur. HFs forskningsinfrastruktur er preget av en sammensatt fagportefølje, med ulike behov. Fellesnevneren er at infrastrukturbehovet for humanistiske fag er økende. Dette understrekes ikke minst av at fakultetet i dag er ledende på humanistisk forskningsinfrastruktur, med to nasjonale infrastrukturer (CLARINO og SAMLA) som inngår i det nasjonale veikartet for forskningsinfrastruktur for humaniora.

Forskere i HFs fag er avhengige av tilgang til databaser, arkiv, digitale og fysiske samlinger og laboratoriefasiliteter, samt teknologi og organisering som muliggjør bruk i forskning og undervisning. Fakultetet har i dag samarbeid om forskningsinfrastruktur med både Universitetsbiblioteket og Universitetsmuseet.

Fakultetet har en [handlingsplan for forskningsinfrastruktur](#), vedtatt i Fakultetsstyret den 22.10.2024. Handlingsplanen følger fakultetets [strategi for 2023-2030](#). Det overordnede målet i handlingsplanen er å forvalte og utvikle humanistiske forskningsinfrastrukturer som bidrar til forskning og undervisning av høy internasjonal kvalitet. Planen legger vekt på at humanistisk forskningsinfrastruktur er viktig samfunnsinfrastruktur. Derfor er et av hovedmålene å gjøre forskningsinfrastrukturene tilgjengelige for forskning og innovasjon, men også øvrige aktuelle brukere. Planen legger videre vekt på en helhetlig og langsiktig forvaltning av infrastrukturene, både når det gjelder bakenforliggende e-infrastruktur, langtidslagring og økonomi.

Fakultetets veikart bør i første rekke sees i direkte sammenheng med fakultetets prioritering til utlysningen i 2025 angående midler til forskningsinfrastruktur av nasjonal viktighet.

SAMLA

SAMLA er en nasjonal infrastruktur for tradisjonsarkiver som ble etablert i perioden 2020 til 2025. SAMLA består av en komponent for langtidslagring, en underliggende database og et søkbart nettarkiv, i samsvar med FAIR-prinsippene. Digitalarkivet samla.no ble lansert i 2024, og inneholder 477 000 manuskriptsider fra Norsk folkeminnesamling (UiO), Etno-folkloristisk arkiv (UiB) og Norsk etnologisk samling (SNF). Det gjør SAMLA til et verdens største og mest komplekse folkloristiske digitalarkiver.

SAMLA inneholder et stort mangfold av kulturelle uttrykk og praksiser, både fra majoritetsbefolkningen, minoritetsbefolkningen og urfolkskulturer i Norge. Sammen gir samlingene et godt bilde av utviklingen av hverdagsliv, forestillinger og verdisyn. Arkivet inneholder et rikt materiale som blant annet omfatter eventyr, sagn, ballader og annen sangtradisjon, svartebøker, tradisjoner knyttet til liv og helse, værtegn, juletradisjoner, ølbryggingstradisjoner, mattradisjoner, plantetradisjoner, håndverkstradisjoner og dokumentasjon av eldre driftsformer i jordbruket og mye mer.

CLARINO

CLARINO er den norske noden i ESFRI-infrastrukturen CLARIN (Common Language Resources and Technology Infrastructure) og er organisert som et nettverk av noder med tjenester som utfyller hverandre. CLARINO Bergen er nasjonal koordinator og tilbyr tjenester som lagring av data, avansert analyse, metadataredigering, og brukerstøtte til forskningsprosjekter. CLARINO har vært viktig i etableringen av Termportalen, og leverer tjenester til UiBs leksikografiske arbeid med ordbøkene. UiO og UiT tilbyr henholdsvis tilgang til egne data og datalagring for åpne språkdata, mens Nasjonalbiblioteket forvalter den nasjonale ressurskatalogen, som gjør språkdataene synlig i CLARIN internasjonalt.

Bruken av CLARINO er betydelig, og fyller en rolle nasjonalt ved at den inneholder kuraterte data av høy kvalitet som har gjennomgått kvalitetskontroll. Ved UiB betjener CLARINO fagområder som norrønt (MENOTA-arkivet), dialektforskning (Talebanken), andrespråk (ALAN og ASK), og ordbokmiljøet. Korpuskel hadde nesten 250 000 søk i 2024, som omfatter både Nynorskkorpuset og Norsk aviskorpus, og hvor ressurser som Menota og Talebanken er jevnt og mye brukt. Det kommer fortløpende nye data inn, både fra forskningsprosjekter og som del av oppdateringer, og ved UiB ble 1240 unike klient-IP-adresser registrert i 2024, noe som indikerer et stabilt behov for tilgjengelige språkdata av høy kvalitet.

Strategiske behov for forskningsinfrastruktur i forbindelse med det nasjonale veikartet og Forskningsrådets utlysning av midler til forskningsinfrastruktur av nasjonal viktighet i 2025

Koordinatorsøknad fra HF

Fase to søknaden SAMLA 2 – digitalisering av norske tradisjonsarkiv har bred intern støtte og er godt forankret i strategien og internt ved UiB samt ved partnerinstitusjonene.

Institutt	Tittel	Kontaktpunkt
AHKR	SAMLA 2.0. National Infrastructure for Tradition Archives in the age of AI	Kyrre Kverndokk kyrre.kverndokk@uib.no

SAMLA 2 vil være en oppgradering og en videreføring av SAMLA. Målet er tredelt:

1. Å digitalt sikre viktig og unikt kulturhistorisk kildemateriale.
2. Å gjøre samla.no mer mangfoldig når det gjelder kulturuttrykk og hvilke kulturelle grupper som er representert
3. Å gjøre det kulturhistoriske kildematerialet søkbart og digitalt tilgjengelig, blant annet ved hjelp av HTR-lesing og maskinlært annotering.

Partnere i SAMLA 2:

- Norsk Folkemинnesamling, IKOS, (NFS) Universitetet i Oslo
- Stiftelsen Norsk Folkemuseum (SNF)

- Universitetsbiblioteket ved Universitetet i Bergen (UBB)
- Institutet för språk och folkminnen (ISOF)
- Institutt for kulturvitenskap, Norges arktiske universitetsmuseum (UiT)

Norges arktiske universitetsmuseum (UiT) har ikke vært partner i SAMLA1. Som ny partner vil UiT tilføre samla.no kulturhistorisk sett, betydningsfulle samlinger av samiske og kvenske kulturuttrykk og praksiser, samt nordnorske folkeminner. I tillegg vil mye av det gjenværende kildematerialet fra de øvrige partnerne bli inkludert, bla. vanlige folks selvbiografier, skillingsviser, barnetradisjoner, samt sjømenns-, arbeider- og husmannsminner. Dermed vil SAMLA spenner over minoritets- og urfolkskultur, barnetradisjoner, og kulturuttrykk knyttet til ulike sosial klasse og yrker. Det gjør at samla.no vil inneholde et sosial representativt kildemateriale til en mangfoldig kulturhistorie i Norge over 500 år, med hovedvekt på 1800- til 2000-tallet. Totalt sett vil samla.no inneholde over 1 million manuskriptsider. Noe av materialet er etisk sensitivt og vil kun være tilgjengelig gjennom en tilgangskontroll utviklet i SAMLA,1 hvor tilgang gis etter søknad. Det samiske og kvenske materialet krever også at SAMLA2 samarbeider med minoritets- og urfolkrepresentanter for å sikre disse gruppenes rettigheter.

Noen viktige tekniske oppgraderinger vil bli gjort: Annotering utføres nå manuelt, mens SAMLA2 vil prøve ut maskinlæringsbaserte annoteringsmetoder. Det vil gi viktige synergieffekter for andre digitalarkiver. HTR-teknologi vil bli brukt til å produsere maskin-genererte transkripsjoner av skannede faksimiler. I SAMLA 1 er én HTR-modell utviklet og to nye er under utvikling. SAMLA2 vil bygge videre på dette for å bruke HTR-teknologi til transkribering i stor skala.

En delmålsetning er i tillegg å utvikle et fellesnordisk digitalarkiv for tradisjonssamlinger (NORDSAMLA). Et nordisk digitalarkiv vil gjøre det mulig å avdekke kulturmønstre på tvers av landegrenser, både ved hjelp av kvalitative analyser og stordatasøk.

Nasjonale infrastrukturer

Navn/kortnavn på infrastruktur	Omtale av sist tildelte prosjekt i prosjektbanken	Status	Nettsted
SAMLA Koordinert av UiB	SAMLA: National Infrastructure for Cultural History and Tradition Archives	I driftsfasen	https://samla.w.uib.no/
CLARINO Koordinert av UiB	Common Language Resources Infrastructure Norway Upgrade	I driftsfasen ESFRI Landmark	https://clarin.w.uib.no/
NORCHRON Koordinert av UiO	Norwegian Diachronic Corpus 200-1814	I oppstartsfasen	

Veien videre

Forskningsinfrastrukturen CLARINO er i driftsfasen fra 2024. Det planlegges å sende inn en oppgraderingssøknad til Forskningsrådet om to år. Infrastrukturen CLARINO gir tilgang til store datasett, og ulike eksternfinansierte prosjekter bruker infrastrukturen, noe som gir mye å bygge på. Blant prosjekter og miljøer som deponerer, analyserer og deler data gjennom CLARINO (SSI-et MediaFutures, NFR-prosjektene ALAN, Norwegian Across the Americas, Senter for norsk som fagspråk, ordbokmiljøet i Bergen, Menota og Talebanken ved UiB).

Det er betydelige endringer i det nasjonale landskapet for språk og språkteknologi siden den forrige oppgraderingssøknaden, og det vil blant annet være naturlig å integrere KI også i videreutviklingen av CLARINO.

Høsten 2025 etableres det to sentre for norsk som fagspråk i Bergen og Oslo, som vil medføre mye brukerinvolvering og kompetanseheving på språk, data og KI. Fagpråkssenteret har en egen arbeidspakke dedikert til Språkteknologi med fokus på evaluering av språkmodeller). Ordbokmiljøet i Bergen, som har vokst og er en stor bruker av CLARINO, har omfattende orddata og omtales som en egen infrastruktur. Det er også opprettet en Stadnamnportal ved UiB, samt Termportalen, delvis finansiert av CLARINO. Det er naturlig å tenke helhetlig rundt disse i forbindelse med en ny oppgraderingssøknad.

