

# Spesialsamlinger og databaser

Til styringsgruppa etter møte 28. mai 2021

## Innhold

Oppsummering av presentasjon og diskusjon på møtet .....	1
Alvin:.....	2
Anbefaling fra prosjektgruppa .....	3
Beslutningspunkter .....	3
Vedlegg:.....	5
1. Pilotprosjektene behovsbeskrivelser med kommentarer .....	5
EMMA – Etnografisk multimediearkiv .....	5
Digital infrastruktur for historisk kildemateriale ved UiO (MUV, IPED).....	5
Bokstavinnskrifter fra vikingtid og tidlig middelalder .....	6
RITMO.....	7
Senter for Ibsen-studier .....	7
2. UBs opprinnelige tekniske kravspesifikasjoner med kommentarer .....	9
3. Innspill fra Senter for Ibsen-studier .....	11
4. Ressursbehov og tidsplan for prosjektet Bokstavinnskrifter i vikingtid og tidlig middelalder .....	13

## Oppsummering av presentasjon og diskusjon på møtet

(Se også vedlagt kopi av presentasjonen)

Prosjektets viktigste leveranser er:

- Digital(e) plattform(er)
- Ivaretagelse av behov hos fagmiljøene
- Etablert arbeidsflyt for aktuelle prosjekter
- Hub/node-organisering av infrastrukturarbeidet
- Kompetanse hos UB

Prosjektet holder så langt tidsplanen. En overingeniør har blitt rekruttert til UB i fast stilling, i to år finansiert av prosjektet. En prosjektstilling som metadatabibliotekar er rekruttert til HumSam-biblioteket for to år (innstilling hos ansettelsesrådet).

Av midler til pilotprosjektene er foreløpig 11 månedersverk (ca NOK 600 000 av de 2M avsatt til pilotene) tildelt ILN i henhold til deres opprinnelige søknad om midler (se vedlagt oversikt og tidsplan).

Det skal også settes av månedersverk til EMMA-prosjektet og IPED-prosjektet. Dette er enda ikke avklart i detalj, men vil legges fram på neste møte.

I vårsemesteret har prosjektgruppa kartlagt behov hos pilotprosjektene. Videre har de hatt møte med UBB om plattformen Marcus for å høre om status der, og de har hatt flere møter med Alvin i Uppsala, både med og uten representanter for pilotprosjektene.

Siden påske har gruppa hatt innlogging til Alvins testmiljø og testet med materiale fra EMMA- og IPED/MUV-prosjektene. I tillegg har vi testet med en samling epitafer fra Teologisk fakultet, og videoer fra RITMOs Musiclab.

Piloter med behov som i sin helhet ikke ser ut til å kunne dekkes av en felles plattform på dette stadiet, har fått dette diskutert i møter med USIT og det er lagt planer for å løse disse. Dette innebærer at det fortsatt er deler av behovene som kan dekkes i f.eks. Alvin per i dag (i hovedsak lagring og tilgjengeliggjøring) men ikke all funksjonalitet nødvendig for forskning på materialet. Dette gjelder RITMO (mer avanserte videovisningsmuligheter), ILN (oppbygning av database), Ibsen-senterets digitale plattform, og RTV-datasettet/Cerex som vil komme tilbake med materiale til plattformen på et seinere stadium. NutriFoodCalc har et eget samarbeid med USIT som videreføres.

### Alvin:

I søknadsprosessen var UB i kontakt med både UiB og NTNU for å diskutere deres erfaringer med deres respektive digitale plattformer. Ved NTNU bruker de Tind (utviklet av et Trondheim-basert selskap) som UB har vurdert tidligere, men ikke valgt. Tind har mange gode løsninger, men gir ikke de mulighetene for katalog/arkiv-strukturer prosjektet har ønsket seg. Ved UiB er UB avhengig av midlertidige prosjektmidler for å opprettholde drift av sin plattform Marcus, og har en rekke ulike systemer under den paraplyen med de utfordringer det innebærer. Vi har videre sett på den svenske plattformen Alvin og prioritert å teste i den. Plattformen er godt etablert som et samarbeid på tvers av institusjoner, har sikker drift over tid og gir tilgang til viktig kompetanse for framtidige digitaliseringsprosjekter og utvikling lokalt.

Det synes vanskelig å finne en løsning som møter alle behov, men Alvin dekker godt det som vanligvis snakkes om som spesialsamlinger og det vi i vår søknad refererte til som "herreløse arkiver". Under følger en oversikt over de begrensninger og muligheter som er relevante for dette prosjektet.

### Begrensninger slik plattformen fungerer i dag:

- Begrensede APIer for å hente ut data og ingen APIer for å importere og redigere data.
- Metadata registreres etter standarder utviklet i bibliotek-verdenen, primært MODS, og mangler mulighet til å tilføye egne metadatafelt og sette opp samlingsspesifikke skjemaer.
- Tilbyr ikke løsninger for navigasjon i, og presentasjon av, mer komplekse data enn det som kan representeres i MODS som f.eks. TEI-kodete transkriberte tekster, annotert med kommentarer og krysslenker innad i teksten, bibliografier og andre databaser. Kan kanskje beskrives og arkiveres (langtidslagringsperspektivet) i Alvin, men vil trenge eksterne grensesnitt for presentasjon.
- Eneste norske institusjon i Alvin

### Muligheter:

- Gode organiseringsmuligheter for samlinger og arkiver
- Kjente standarder for registrering som det finnes god kompetanse i ved UB
- Samarbeidsarena som legger til rette for kompetanseheving
- Tilgang til bred kompetanse på digitaliseringsprosjekter, og hjelp med import/konvertering av eksisterende samlinger
- Samarbeid om utvikling - Alvin vil prioritere UiOs behov i prosjektperioden

- Skal over til ny plattform som også innebærer at APlene som mangler kommer på plass
- Sikker drift over tid
- Lave kostnader frigjør midler til utvikling lokalt
- Tilgang til Alvins mangfoldige samlinger fra Sverige

Senter for Ibsen-studier har sendt et notat med innspill til et mulig valg av Alvin som plattform – det ligger i sin helhet under. Prosjektgruppa har stor forståelse for deres innvendinger, og noen av disse er også dekket av punktene over. Senter for Ibsen-studier mener bl.a. at det av strategiske grunner ikke er hensiktsmessig å bli med i Alvin-konsortiet fordi man heller bør bygge kompetansen lokalt. Fra prosjektets side ser man det imidlertid slik at samarbeid med Alvin nettopp kan bidra til å bygge kompetanse lokalt, ikke minst for digitaliseringsprosjekter, gjennom samarbeid og erfarings/kompetansedeling. Å satse på Alvin som en solid og stabil løsning for lagring og tilgjengeliggjøring kulturarvsmateriale/spesialsamlinger, frigjør ressurser i prosjektet til å jobbe med løsninger for prosjekter med mer komplekse behov som nevnt under lista med begrensninger all den tid vi ikke har funnet en løsning som kan favne alle behovene.

I diskusjonen etter presentasjonen, ble det på møtet stilt spørsmål ved at ingen av HFs pilotprosjekters behov kan løses i sin helhet i Alvin. Spesialsamlinger og kulturarvsmateriale knyttes i stor grad til humaniora, selv om den organisatoriske tilhørigheten ikke viser akkurat det hos pilotene. Fra bibliotekets og prosjektets side er arbeidet med å finne løsninger for digitale forskningsmetoder i humaniora (spesielt i arbeidet med tekst) ikke avsluttet selv om man velger å fortsette å prøve Alvin. Vi foreslår at vi parallelt jobber videre med å finne løsninger for dette. Prosjektets mandat er også åpent for at løsningen ligger i flere enn en plattform. Finner vi en plattformvariant som bedre møter pilotene sett under ett enn Alvin, ønsker vi ikke være bundet til å fortsette samarbeidet med Alvin

## Anbefaling fra prosjektgruppa

Prosjektgruppa foreslår at

- prosjektet tester Alvin videre gjennom et begrenset medlemskap i prosjektperioden og velger om vi skal bli fulle medlemmer i konsortiet først ved prosjektets slutt (årsskiftet 2022/23). Prosjektgruppa jobber tett med Alvin om UiOs behov i prosjektperioden for å få reell innsikt i hva man kan få til i samarbeid og hva gevinstene av konsortium-deltagelse vil være
- vi utforsker alternativer til Alvin parallelt – både med blikk på om det finnes andre løsninger som dekker alle de identifiserte behovene i en plattform, eller som et parallelt «spor» i tillegg til Alvin
- det settes av ressurser (kan spesifiseres hvor mye til neste styringsgruppemøte) til å lage portaler/innganger til enkeltsamlinger i Alvin fra UiOs nettsider. Selv om prosjektet skal jobbe med back end, trenger pilotprosjektene å se hva som er mulig å få til av front end-løsninger for å kunne vurdere løsningen (bildevisning, samlingspresentasjoner mm.)
- vurderingsgrunnlaget utvides til å omfatte andre tilsvarende utfordringer og en kartlegging av hvor omfattende behovet for plattform som Alvin er ved UiO (kan legges fram ved neste styringsgruppemøte)

## Beslutningspunkter

- Fortsette med testing i Alvin gjennom et begrenset medlemskap i prosjektperioden
- Utforske alternative/parallele løsninger
- Inkludere testing av front end-muligheter i prosjektet

- Prosjektet ber om at tildeling av 11 månedsverk til ILN, 4 måneder til IPED/MUV og inntil 3 måneder til EMMA fra de frie midlene tiltenkt pilotprosjektene tas til etterretning

**Til neste møte:**

- Vurderingsgrunnlag videre

## Vedlegg:

### 1. Pilotprosjektene behovsbeskrivelser med kommentarer

(Unntatt RTV-datasettet og NutriFoodCalc som har blitt ivaretatt av USIT)

#### EMMA – Etnografisk multimediearkiv

**Ressurspersoner:** Maria Kartveit, Arne Perminow, Rune Flikke

**Kort om bakgrunn og behov:** Dataproduksjon knyttet til Sosialantropologisk forskning ved UiO har siden 1950-tallet foregått gjennom feltarbeid der deler av forskningsmaterialet eksisterer som foto-, lyd- og (siden 1980 tallet) videoopptak. Det finnes i dag ingen rutiner ved UiO for ivaretagelse og deling av slikt multimediemateriale. Et digitalt arkiv med en delingsplattform utviklet i regi av UB vil bidra til å sikre et rikt arkivmateriale med stor forsknings- og samfunnsrelevans for ettertiden og muliggjøre deling av et unikt kildemateriale.

#### Ønskeliste:

- Lagring av allerede digitalisert materiale – (Fredrik Barth, Unni Wikan, Arve Sørnum, Signe Howell, totalt ca 3200 bilder, noe audiovisuelt materiale)
- Grensesnitt for søk/tilgjengeliggjøring
- Viktig med avklaringer i forhold til personvern og opphavsrett, juridisk ekspertise
- Digitalisering og lagring av materiale etter Johs.Falkenberg (bilder, tekster)

**Kommentarer mai-21:** Alvin fungerer fint som svar på disse behovene. En god front end-inngang fra UiOs nettsider hvor bildene kan presenteres mer direkte blir viktig. Prosjektet har fått med juridisk ekspertise i prosjektgruppa og er i gang med å undersøke de relevante juridiske begrensningene på dette materialet og hvordan opphavsrettsforholdene ser ut. EMMA er et bra testprosjekt i så måte.

#### Digital infrastruktur for historisk kildemateriale ved UiO (MUV, IPED)

**Ressurspersoner:** Jon Røyne Kyllingstad, Anne Vaalund, Kari-Anne Ulfsnes, Sverre Tveit

**Kort om bakgrunn og behov:** Museum for universitets- og vitenskapshistorie (MUV) ser et prekært behov for å kartlegge, ivareta, forvalte og formidle historisk arkivmateriale og annet kildemateriale ved UiO. En digital infrastruktur i regi av UB er nødvendig for å muliggjøre en kartlegging, sikring og synliggjøring av materialet. Det er viktig at den digitale plattformen kan håndtere komplekst kildemateriale med berøringspunkt til både arkiv, bibliotek og museum. Forbindelsen mellom de ulike typene kilder må ivaretas, også der materialet må splittes opp. I årenes løp er det gjennomført eller påbegynt en rekke kartleggings-, transkriberings- og formidlingsprosjekter lokalt ved fagmiljøene ved UiO. Prosjektet bør også romme allerede gjennomførte prosjekter som står i fare for å bli teknisk utdatert. Materialet i pilotprosjektet er fysisk arkivmateriale, skriftlig materiell og gjenstander for psykometrisk testing i kjelleren på Helga Eng's hus.

#### Ønskeliste:

- Ivareta, formidle og tilgjengeliggjøre kildemateriale for historisk forskning
- Fotograferer fysisk materiale, strukturere og sette i sammenheng
- Registrering og katalogisering av materiale
- Registrering må gjøre det mulig å spore materialet tilbake til «funnkonteksten» uavhengig av hvordan det oppbevares i framtiden

**Kommentarer mai-21:** Alvin fungerer fint som svar på disse behovene, spesielt de fleksible løsningene for samlingsposter og arkiver. Egne portaler til ulike samlinger fra UiOs web vil være lurt å teste i prosjektperioden. Digitaliseringsarbeidet ivaretas når digitaliseringsverksted er etablert, og kan også arbeides med før det er plass med utstyr som allerede finnes.

## Bokstavinnskrifter fra vikingtid og tidlig middelalder

**Ressurpersoner:** Elise Kleivane og Alessandro Palumbo

**Kort om bakgrunn og behov:** Databasen skal inneholde innskriftenes tekster, lingvistisk, historisk og geografisk metadata, samt informasjon om de tekstbærende gjenstandene. Innholdet i databasen hentes primært gjennom forskningsprosjektet "Bokstavar mellom runer og manuskript" (BRuM), ledet av Elise Kleivane. Videre vil databasen bli supplert med materiale hentet frem gjennom ytterligere to prosjekt ved ILN, nemlig Marie Curie-prosjektet Language switching and script mixing, ledet av Alessandro Palumbo, og forskningsprogrammet Modes of Modification, der Kleivane er en del av kjernegruppa og Palumbo virksom som affilert forsker. Ytterligere én epigrafiker og fem filologer er knyttet til dette forskningsprogrammet. Databasen vil være et essensielt verktøy i det videre arbeidet i disse forskningsprosjektene. Datamaterialet som innlemmes i infrastrukturen skal kobles til andre relevante databaser (Samnordisk Runtextdatabas, Universitetsmuseenes gjenstandsdatabase, bl.a.), noe som dessuten vil gjøre den foreslåtte infrastrukturen til et viktig verktøy i tverrvitenskapelige prosjekt og maksimere mulig utbytte.

Forslag fra prosjektgruppa v/Jarle E:

- ILN bør starte med å lage en XML-representasjon av datamaterialet med utgangspunkt i TEI-standard.
- Vi må avklare om Alvin kan importere / tilgjengeliggjøre den aktuelle TEI-utgaven,
- Det kan også publiseres på nett for google-indeksering.
- Mulig andre søkeverktøy kan benyttes (Glossa og andre).
- Vi må kartlegge behovet på tvers av piloter når det gjelder bruk av TEI.
- Undersøke kompetanse og behov ved UiO i forhold til TEI, ikke minst når det gjelder bruk av Alvin som standard lagringsløsning.

### Ønskeliste

- Database med innskrifter/tekst som gjengis på minst 3 detaljeringsnivåer
- Må kunne søke på fulltekst og på ulike detaljeringsnivåer
- Tilrettelagt for søk som er relevante for forskning innenfor epigrafi, skriftkultur, historie
- Må kunne søke på metadatafelter enkeltvis og kombinert
- Avgrense søk på grunnlag av innledende søk
- Folk til å registrere data
- Mulighet for å legge inn kommentarer (hvor?)
- Skal være intuitivt i bruk, lav brukerterskel
- Databasen må kunne bygges ut over en lang periode, må kunne oppdateres med nytt materiale
- Mulighet for å oppgradere databasen med mer detaljert annotering (kvantum og kvalitet)

**Kommentarer mai-21:** Det er vanskelig å se at Alvin kan være en løsning for dette prosjektet for annet enn lagring og tilgjengeliggjøring, og piloten fortsetter det opprinnelig planlagte samarbeidet med Tekstlaboratoriet for å lage databasen.

## RITMO

**Ressursperson:** Alexander Refsum Jensenius

**Kort om bakgrunn og behov:** Datasettene som samles inn i fourMs-lab'en ved RITMO/IMV er av musikkrelaterte kroppsbevegelser. Det er to hovedtyper av studier: (1) musikere som spiller solo eller i mindre grupper, (2) folk som beveger seg til musikk på ulike måter (dans, luftinstrumentspilling, m.m.). Dataene kjennetegnes ved å være komplekse og multimodale. Et typisk datasett består av lyd- og videoopptak ( gjerne fra flere ulike kameraer), data fra bevegelsessporing og fysiologiske målinger, spørreskjemaer med psykologiske tester og fritekst, samt intervjumateriale. Målet er å kunne lagre, arkivere, laste ned, og vise frem datasettene på en helhetlig måte. Vi har et konkret case som kan brukes til uttesting: opptak av en cellist som spiller ornamenten.

### Ønskeliste:

- Strukturert lagring/arkivering av lyd- og videomateriale
- Teknisk løsning for avspilling av lyd og video
- Tilknyttet tekstlig materiale og sensordata
- Kvalitativ informasjon, f.eks. intervjudata
- Kobling mellom datasett og publikasjon
- Maskinlesbar koding av juridisk informasjon (personvern og opphavsrett)

**Kommentarer mai-21:** RITMO samarbeider framover med USIT for å finne gode tekniske løsninger for visning i VORTEX, mens vi fortsetter å se på Alvin som en løsning for de behovene som kan løses der.

## Senter for Ibsen-studier

**Ressurspersoner:** Asgeir Nesøen, Damir Nedic, Nina Marie Evensen

**Kort om bakgrunn og behov:** Senter for Ibsen-studier ved ILN er en av mange enheter ved UiO som har behov for en sentralisert lagringsløsning for digitalt materiale og metadata. Ibsen-ressursene er tilgjengelige fra nettstedet Det Virtuelle Ibsen-senteret <https://www.uio.no/ibsen> og består av nettutgaven av Henrik Ibsens skrifter, databasen IbsenStage, Ibsen-arkivet, Ibsen-bibliografien, The Multilingual Ibsen og en 3D-modell av Komediehuset i Bergen (under arbeid).

### Ønskeliste

- Etablere mediebase hvor vi kan lagre råmateriale i form av bilder, video, lyd, tekst (XML (TEI), HTML, tekst i databasefelt) og metadata på norsk og engelsk
- Må støtte ulike nivåer av metadata, f.eks. for metadata som skal skjules eller bare vises for enkelte brukergrupper (pga. rettighetsproblematikk)
- Søkegrensesnitt som muliggjør søk på tvers i alt materiale
- Bruker- og admingrensesnitt som gjør det mulig å legge inn data
- Må ha APIer som gjør det mulig å hente ut / søke i data fra eksterne applikasjoner
- Samhandlingsplan for utvikling og drift
- Eksplisitte lisenser på visning av bilder og tekst – Må tas som en del av FAIR-prosjektet
- Mulighet for å vise bilder i sammenheng/relasjon, f.eks. bøker / manuskripter, og i 3 ulike størrelser
- Tilrettelegge for bildegjenkjenning (generering av metadata)

**Kommentarer mai-21:** Senter for Ibsen-studiers behov møtes ikke i Alvin per i dag. Vi tar opp igjen tråden etter sommeren sammen med USIT som forslår å bringe inn sitt UX-team for å få bedre forståelse av designbehovene for senterets plattform. Prosjektet jobber videre med å se finne løsninger for mediefilene.



## 2. UBs opprinnelige tekniske kravspesifikasjoner med kommentarer

Gul utheving er ikke møtt i Alvin, grønn utheving er oppfylt

### Digitaliseringsplattform: tekniske krav

Nedenfor følger en beskrivelse av tekniske krav til systemer i forskningsinfrastrukturen. Listen er ikke uttømmende og beskriver et sett med minimumskrav til teknisk funksjonalitet gruppert etter system, kuratering, tilgjengeliggjøring og autentisering og sikkerhet.

Må ha-krav:

#### System:

1. Grensesnitt for eksport/-import av data, metadata og datamodeller i bulk (Dette skal være basert på REST-API).
2. Grensesnitt for opplasting av data med validering mot metadataschema.
3. Skalerbarhet (årlig dobling av datamengde).
4. Robust teknologi for backup og gjenoppretting.
5. API for redigering av brukere og data.
6. Robust teknologi for indeksering.
7. Støtte for lagring og fremvisning av bilde- og videofiler uten begrensning på filstørrelse.

**Kommentarer etter testing:** Krav 1, 2 og 3 vil møtes når Alvin flyttes til Cora, senest 2025.

#### Kuratering:

1. **Fleksible datamodeller for metadata og støtte for kobling til eksisterende vokabularer:**
  - a. Som administrator må en kunne definere metadataschemaer/sett av egenskaper som kan brukes for bestemte objekter/samlinger. Egenskapene bør være flerspråklige.
  - b. Kobling til etablerte metadataschemaer: Egenskaper i metadataschemaene må kunne knyttes til eksisterende metadataschemaer/ vokabularer via URIs (som Dublin Core).
2. **Permanente identifikatorer:**
  - a. Identifikatorene må enten være URI-er, eller kunne danne utgangspunkt for URI-er, som kan brukes som globale, permanente identifikatorer i en lenket data-kontekst.
3. **Mulighet for å knytte metadataverdier til kontrollerte vokabularer:**
  - a. Verdier må kunne knyttes til kontrollerte vokabularer, både lokale i systemet og eksterne.
4. **Støtte for lisensiering av data, helst maskintolkbar lisensinformasjon i form av URI-er.**
5. **Gode arbeidsflyter i alle ledd av kuratering og tilgjengeliggjøring.**
6. **Støtte for endringer av metadata i batch.**
7. **Web analytics og statistikk.**
  - a. Anonymisering av data før oversending til skytjenester for analytics og statistikk (for eksempel Google Analytics).
8. **Dokumentasjon pr samling (intern beskrivelse og ekstern). Denne informasjonen må være synlig og lett tilgjengelig både internt og for alle sluttbrukere.**
9. **Versjonering eller historikk over endringer for hvert objekt.**

**Kommentarer etter testing:** Krav 1a, 7 og 8 vil møtes når Alvin flyttes til Cora, senest 2025.

### Tilgjengeliggjøring:

1. Inngang til samlinger (og undersamlinger), der hver samling er et sett av objekter, men også introduksjon/bakgrunnsinformasjon i et egnet markeringsspråk (Markdown, HTML e.a.):
  - a. Objektene i en samling bør kunne filtreres etter fasetter som baserer seg på metadataskjemaet som er i bruk i samlingen.
  - b. Samlinger må kunne opprettes også der man delvis er basert på eksterne datasett, for eksempel med lenking både til rådata, til kuraterte data og til publikasjoner.
2. Moderne og fleksibel discovery-funksjonalitet.
3. Søkefunksjonalitet både i back-office og i sluttbrukergrensesnittet.
4. Støtte for IIIF for fremvisning av bildefiler:
  - a. God kontroll over IIIF manifestene er viktig, for eksempel for å definere annoteringer på en viss posisjon i et bilde.
5. Videospiller for fremvisning av videofiler:
  - a. Bør/må ha støtte for undertekster WebVTT eller andre undertekstformater.
  - b. Bør/må ha støtte lenking til bestemt tidspunkt.
  - c. Bør ha støtte for kapittelmarkeringer fra WebVTT eller annet format.
  - d. Bør kunne levere en video i ulike kvaliteter (adaptiv bitrate), slik at vi kan lagre originalfilene i høy kvalitet, men levere lavere bitrate til brukere.
6. Delingsfunksjonalitet via APIer og standarder for høsting av data (OAI-PMH).
7. Sluttbrukergrensesnitt som etterkommer krav til universell utforming.
8. Støtte for flere språk i brukergrensesnittet.

**Kommentarer etter testing:** Å møte krav 4 vil prioriteres av Alvin i prosjektperioden. Krav 7 er antagelig ikke relevant.

### Autentisering og sikkerhet:

1. Integrasjon med institusjonelle systemer for autentisering og autorisering.
  - a. Moderne grensesnitt mot IDP (LDAP/AD/openID connect).
2. Støtte for tildeling av roller i back-end.
3. Følge gjeldene retningslinjer for informasjonssikkerhet og personvern.

### Bør ha-krav:

1. Tilgang til lagringsstrukturen på filnivå for å kunne kjøre analyser av data.
2. Støtte for brukergenerert innhold.
3. "Did you mean"-funksjonalitet .
4. Støtte for parallelle brukergrensesnitt med egne utforminger.

### 3. Innspill fra Senter for Ibsen-studier

Fra: Senter for Ibsen-studier

Til: Styringsgruppe for Prosjekt 1 (Spesialsamlinger og databaser) ved Astrid Anderson og Cecilia Ekström

Dato 26.05.2021

På vegne av Senter for Ibsen-studier har Nina Evensen vært med på å evaluere ALVIN som mulig løsning for behovet for å lage et system for drift, utvikling og bevaring av spesialsamlinger og databaser som kan betjene særlig HumSam-miljøene ved UiO. Selv om ALVIN ser ut til å fungere bra i en svensk kontekst med hovedfokus på kulturmateriale (og ikke hovedsakelig forskningsdata), er vår sterke anbefaling at UiO ikke velger denne løsningen, av flere grunner:

1. Av strategiske grunner er det veldig lite hensiktsmessig å inngå i ALVIN-konsortiet. Et hovedmål med dette prosjektet er å bygge kompetanse internt ved UiO/UB. Ved å «outsourc» til Sverige vil man miste denne muligheten til å styrke DH-kompetansen internt. Valg av mediebase bør være første skritt på veien mot å etablere en infrastruktur for langsiktig drift og vedlikehold av DH-materiale ved UiO, og det oppnår man ikke ved å velge ALVIN.
2. En viktig praktisk begrensning i ALVIN er at den ikke godtar alle typer formater, f.eks. to av de aller viktigste og mest utbredte innen bilde- og tekstdata, nemlig RAW-filer og XML-filer. Den tilbyr heller ingen løsning for drift og lagring av databaser. ALVIN vil dermed ikke kunne håndtere en stor del av datasettene som produseres i forbindelse med særlig humanistisk forskning.
3. Det er ikke mulig å hente ut materiale via API-er i ALVIN. Det er vilje til å utvikle dette, men vi vet ikke når det kommer, og det er en funksjonalitet som er viktig å ha på plass helt fra starten av.
4. ALVIN er en relasjonell database med hovedfokus på digitalisert svensk kulturmateriale. Det betyr at forskningsdataene fra UiOs HumSam-miljøer vil bli koblet inn i denne konteksten, noe som ikke vil være relevant i de aller fleste tilfeller. Merverdien i å inngå i en relasjonell database blir dermed sterkt redusert.
5. ALVIN er et veletablert konsortium som allerede har mange underinstanser fra ulike arenaer. Dette vil høyst sannsynlig gjøre det vanskelig for UiO – med en helt annen profil enn de fleste andre etablerte partnere – å nå gjennom med egne behov.
6. ALVINs brukergrensesnitt er tospråklig svensk og engelsk, men er ikke konsekvent. Velger man svensk visning, får man en blanding av svensk og engelsk språk i systeminfo og innhold. Dette vil bli ytterligere komplisert og uryddig ved å legge til norsk som språkvalg.

Vi mener det vil være mer hensiktsmessig å inngå samarbeid med NTNU eller UiB, som har utviklet løsninger som UiO kunne kloner, videreutvikle og tilpasse til egne behov som ledd i den ønskede kompetansebyggingen. I et slikt samarbeid vil vi kunne lære av erfaringer som er direkte relevante for UiO og bidra til å bygge et nasjonalt kompetansenettverk. Det ville være et skritt i retning av å

etablere en helhetlig infrastruktur som kan håndtere alle typer DH-materiale ved UB/UiO på lengre sikt.

## 4. Ressursbehov og tidsplan for prosjektet Bokstavinnskrifter i vikingetid og tidlig middelalder

Kostnadene for prosjektet fordeler seg på utviklingen av databases og grensesnittet og innholdet. Etter samtale mellom forskerne og Tekstlaboratoriet ved ILN anslått at det samlet sett vil kreve inntil fire månedsverk av en senioringeniørs arbeidstid. Blant disse er instituttet villige til å stå for en av disse mnd. som egenandel til søknaden.

Prisanslag utvikling av databasen med søkegrensesnitt:

Senioringeniør ltr. 68, 3 mndsværk (lønn og sosiale kostnader)	
407 timer à 494 kr	= 201020 kr.
Egenandel, Senioringeniør ltr. 68, 1 mndsværk (lønn og sosiale kostnader)	
135,67 timer à 494 kr	= 67007 kr.
Samlet sum for utviklingstjenesten	= <u>268026 kr.</u>

En vesentlig del av kostnadene vil innebære en manuell innsats fra vitenskapelig assistenter som vil utføre arbeidet med datahåndtering, mengde og type data som genereres, lagring, tilgjengeliggjøring, arkivering, og som krever tett oppfølging fra forskerne.

Vitenskapelig assistanse ltr. 51,9 mndsværk (lønn og sosiale kostnader)	
1085 timer à 366 kr	= <u>397 231 kr.</u>
Samlet prisanslag for utviklingstjenesten og vitenskapelig assistanse	= <u>665257,00kr</u>
Egenandel(10%) - Inten frikjøp	- 66525,70
<b>Søknadssum</b>	= <u><b>598731,30kr</b></u>

### Tidsplan:

**Høst 2021:** Tekstlab starter å bygge strukturen høsten 2021, og Palumbo og Kleivane leverer materiale som er hensiktsmessig å legge inn som del av utviklingen av strukturen. Rekrutterer assistenter som skal jobbe videre med prosjektet.

**Vår 2022:** De vitenskapelige assistentene jobber med prosjektet, og stipendiaten på prosjektet har levert avhandlingen og vil veilede assistentene i dette arbeidet.

Midlene til vitenskapelige assistenter vil mao ikke bli brukt før våren 2022.