

Rapport fra utredning av nasjonalt vitenarkiv

«Alle vitenskapelige artikler basert på offentlig finansiert forskning skal deponeres i et egnet vitenarkiv senest ved publiseringstidspunktet, uavhengig av publiseringskanal og uavhengig av når det er mulig å gjøre dem åpent tilgjengelige.»

Fra regjeringens nasjonale mål og retningslinjer for åpen tilgang til vitenskapelige artikler



Direktoratet
for IKT og fellestjenester
i høyere utdanning
og forskning

Sammendrag

Denne rapporten oppsummerer anbefalinger fra arbeidet med å utrede hvordan et nasjonalt vitenarkiv (NVA) kan realiseres. Anbefalingene er basert på kartlegging av problemstillinger i dagens arkivlandskap og analyse av hvilke tiltak som er mest relevante for å oppnå gevinster gjennom å etablere NVA.

Dagens problemstillinger rundt egenarkivering av forskningsresultater, herunder import (høsting) av resultater som er publisert i tidsskrifter, er i hovedsak knyttet til:

- Liten gjenbruk av metadata fra tidligere arbeidsprosesser i verdikjeden medfører unødvendig mye manuelt arbeid ved at metadata må gjenskapes manuelt.
- Dagens organisering av arbeid for rettighetsklarering, valg av lisens og kvalitetssikring av metadata for rapportering til Norsk Vitenskapsindeks (NVI) er lite effektive. Deler av årsaken til dette problemet ligger i en sub-optimal infrastruktur for å håndtere den norske modellen med et felles nasjonalt CRIS-system¹ og distribuerte institusjonsarkiv.

Disse problemene kan reduseres betydelig gjennom utredningens to anbefalte hovedtiltak:

1. Etablere nasjonale masterdatakilder (autoritetsregistre) for søknader (om finansiering), prosjekter, personer og organisasjonsheter, og gjøre data i disse tilgjengelig på tvers.
2. Etablere en felles informasjonsarkitektur og felles datagrunnlag for Cristin (CRIS-system) og institusjonenes vitenarkiv (IR²).

Ansvar for å realisere hovedtiltak 1 ligger i hovedsak utenfor NVA, men tiltaket er så sentralt og viktig for NVA at det vil bli gitt prioritet i Unit-prosjektet som jobber med å etablere 3-års handlingsplaner. Det er klare signaler fra mange aktører at opprettelsen av de nevnte autoritetsregistrene vil gi merverdi i flere andre sammenhenger.

Hovedtiltak 2 innebærer at det realiseres en ny felles plattform som ivaretar behovene som i dag dekkes av de lokale vitenarkivene. Funksjonaliteten i dagens Cristin vil også over tid måtte sees i nær sammenheng med den samme plattformen. Når funksjonaliteten fra Cristin og vitenarkivene fremstår som et hele og opererer på et felles datagrunnlag, forsvinner behov for duplisering av data slik en har i dag mellom Cristin og de lokale vitenarkivene. Dermed understøttes en viktig strategisk målsetning for digitalisering i sektoren - data lagres én gang og tilgjengeliggjøres fra én kilde³.

Hovedtiltak 2 kan realiseres på ulike måter, som omtalt i rapporten. Utredningens anbefalte tiltak er å etablere NVA med basis i en teknologisk plattform som er utviklet av Unit for å publisere digitale objekter, som f. eks. læringsobjekter og forskningsdata. Plattformen er basert på *leading edge*-teknologi og utnytter infrastruktur i skyen.

Ved anbefalt løsning, vil NVA bli et fundament for å kunne realisere betydelige gevinster i forhold til dagens situasjon. Det må poengteres at for å kunne ta ut de store gevinstene, vil det være nødvendig å endre på organisering av arbeidet (prosess og roller) som NVA skal understøtte - slik at en utnytter muligheten som NVA gir for å «jobbe smartere».

¹ Current Research Information System

² Institutional Repository

³ Hentet fra [Digitaliseringsstrategi for universitets- og høyskolesektoren](#)

Følgende personer har deltatt i utarbeidelsen av rapporten

Prosjektgruppe

Nina Andersen, NTNU
Marit Henningsen, Unit
Hanne Hole Young, Unit
Tore Vatnan, Unit
Lars Wenaas, Unit
Hege Johannesen, Unit
Jan Erik Garshol, Unit
Eirik Viken Kallset, Unit
Per Hovde, Unit (prosjektleder)

Styringsgruppe

Frode Arntsen, Unit (prosjekteier)
Jone Thingbø, Unit
Katrine Weisteen Bjerde, Unit
Rune Brandshaug, NTNU
Tomas Willebrand, Høgskolen Innlandet

Referansegruppe for prosjektet:

Jonny Edvardsen, Nasjonalbiblioteket
Trude Eikebrott, OsloMet
Laila Aksetøy, Sintef Energi AS
Hilde Trygstad, VID
Elin Holmstrøm Frøshaug, UBO
Katrine Segadal, NSD
Sonja May Amundsen, Sørlandet Sykehus HF
Leif Longva, UiT
Sigbjørn Hernes, Høgskolen Innlandet
Margrethe Storm, Oslo Universitetssykehus
Pål Richard Romundstad, NTNU
Harald Lenschow, NTNU

Innhold

Sammendrag	1
Innledning.....	5
Bakgrunn	5
Problemstilling.....	6
Definisjon og formål med Nasjonalt vitenarkiv.....	6
Hva kan arkiveres i Nasjonalt vitenarkiv?.....	6
Dagens arkivlandskap.....	8
Problemstillinger knyttet til dagens arkivlandskap.....	10
Lite veldefinerte roller og ansvar	12
Lite enhetlige arbeidsprosesser	12
Mye overflødig manuelt arbeid.....	12
Etterslep i prosessen for import av metadata fra tidsskrift	12
Dupliserte arbeidsprosesser for rettighetsklarering og valg av lisens i vitenarkivene.....	12
Heterogent systemlandskap.....	13
Mangelfull helhetlig og gjennomgående informasjonsarkitektur for verdikjeden knyttet til forskning og publisering	14
Målsetninger for Nasjonalt Vitenarkiv	16
Hovedmål 1	16
Hovedmål 2	16
Hovedmål 3	16
Hovedmål 4	17
Arbeidsflyt i nasjonalt vitenarkiv.....	18
Overordnet, generisk arbeidsflyt for NVA.....	18
Steg 3: Kvalitetssikre og lagre metadata	18
Steg 4: Godkjenne og publisere.....	19
Steg 5: Pliktavlevere til Nasjonalbiblioteket.....	19
Steg 1: Fange opp nytt materiale	19
Steg 2: Registrere / berike metadata	22
Gevinstpotensial ved realisering av et nasjonalt vitenarkiv.....	24
Eliminere dupliserte prosesser for rettighetsklarering, valg av lisens og generell forvaltning av registreringer i lokale vitenarkiv.....	24
Etablere felles, nasjonale autoritetsregistre (masterdatakilder) for sentrale entiteter i verdikjeden	26
Gjenbruk av metadata som er registrert ved utstedelse av DOI.....	27
Kanaliser forskningsdata gjennom NVA og kople disse til forskningsresultater	28

Legge til rette for større spredning av norske forskningsresultater og forskningsdata.....	28
NVA som autoritativ kilde for resultater fra norsk forskning gir bedre grunnlag for gode analyser og rapportering på all forskning som foregår i Norge på tvers av disipliner	28
NVA som kanal for pliktavlevering av resultater fra forskningen	28
Alternative tiltak.....	29
Hovedtiltak 1: Etablere nasjonale autoritetsregistre (masterdatakilder) for prosjekt, person, organisasjonsenhet og søknad (om finansiering).....	29
Gevinstpotensial.....	30
Hovedtiltak 2: Etablere en ny felles basis infrastruktur for Cristin og vitenarkiv.....	30
Overordnede arkitekturmessige valg.....	31
Hvordan kan en best mulig løsning med basis i anbefalt infrastruktur realiseres?	36
Alternative måter å realisere dette tiltaket på.....	39
Noen enkle betraktninger om det økonomiske gevinstpotensialet ved oppnådd målbilde for NVA...	43
Forslag til plan for gjennomføring.....	45
Konklusjon	47
Ordliste	49
Vedlegg 1 – Eksempler på arbeidsprosesser i NVA og hvordan disse utnytter den anbefalte informasjonsarkitekturen.....	51
Import og registrering av publikasjon fra ekstern kilde	51
Manuell registrering av resultat som ikke skal rapporteres i NVA	52
Vedlegg 2 – Kostnadsestimater og for realisering av NVA.....	53
Utvikling av NVA, trinn 1	53
Videreutvikling av NVA fra trinn 1 til trinn2	54
Integrering av NVA med nye autoritetsregistre etter hvert som disse blir etablert.....	55
Forslag til tidsplan for realisering	56

Innledning

Bakgrunn

Denne utredningen har utspring i **Nasjonale mål og retningslinjer for åpen tilgang til vitenskapelige artikler**⁴ som ble lansert av regjeringen 22. august 2017. Bakgrunnen for retningslinjene er regjeringens mål om at alle norske vitenskapelige artikler finansiert av offentlige midler skal være åpent tilgjengelige innen 2024. Retningslinjene sier at Norge skal være en pådriver for at alle vitenskapelige artikler finansiert av offentlige midler skal gjøres åpent tilgjengelig fra publiseringstidspunktet. For å nå målet om full åpen tilgang må både forskere, fagmiljøer, forskningsinstitusjoner, forskningsfinansierende institusjoner og myndigheter bidra. Fagmiljøene forventes særlig å bidra gjennom sine nasjonale og internasjonale nettverk for å fremme åpen vitenskapelig publisering, og for å endre viktige tidsskrifter på sine fagfelter fra å være lukket til å bli åpne.

Den 4. september i 2018 ble tidsplanen for overgangen til åpen publisering akselerert ytterligere gjennom Plan S. Plan S er et initiativ fra "cOAlition S", et konsortium bestående av EU-kommisjonen, Det europeiske forskningsrådet (European Research Council) og store forskningsfinansierer fra tolv europeiske land. Planen krever at forskning som er offentlig finansiert fra og med 2020 skal publiseres i åpne arkiver eller i åpne tidsskrifter. Den 27. november 2018 lanserte Coalition S sine retningslinjer for innføring av plan S. Denne sier blant annet at fra 1.1.2020 skal alle nye forskningsprosjekter følge kravene i Plan S.

Nasjonale mål og retningslinjer for åpen tilgang til vitenskapelige artikler består av fire retningslinjer som skal bidra til at alle involverte aktører arbeider mot samme mål. I tillegg er det en liste over tiltak som skal sikre at målene oppnås. Retningslinje nummer 2 er grunnlaget for å utrede et nasjonalt vitenarkiv:

Alle vitenskapelige artikler basert på offentlig finansiert forskning skal deponeres i et egnet vitenarkiv senest ved publiseringstidspunktet, uavhengig av publiseringskanal og uavhengig av når det er mulig å gjøre dem åpent tilgjengelige.

Tiltakene som skal bidra til at målene nås, er:

1. Forbedre funksjonaliteten for deponering av artikler via Cristin
2. Utrede hvordan et nasjonalt vitenarkiv skal realiseres.
3. Kreve deponering i vitenarkiv lokalt eller nasjonalt som en forutsetning for at artiklene skal telle i den resultatbaserte finansieringen. Kravet forutsetter at vitenarkiv og annen nødvendig infrastruktur for deponering er tilgjengelig for alle forskningsinstitusjonene.
4. Bidra til videreutvikling av nye og bærekraftige modeller for finansiering av åpen publisering nasjonalt og internasjonalt
5. Utvikle indikatorer og statistikk for åpen tilgang

Tiltak 2 er den direkte årsak til at denne utredningen ble startet. Unit ble tildelt arbeidet med å gjennomføre utredningen i brev fra Kunnskapsdepartementet (KD) datert 31. januar 2018. Denne rapporten er resultatet av denne utredningen.

⁴ <https://www.regjeringen.no/no/dokumenter/nasjonale-mal-og-retningslinjer-for-åpen-tilgang-til-vitenskapelige-artikler/id2567591/>

Problemstilling

I oppdragsbrevet skriver KD:

“Det er knyttet flere tiltak til retningslinjene. Kunnskapsdepartementet ber Tjenesteorganet om å gjennomføre tiltak 2 for å følge opp de nasjonale retningslinjene: utrede hvordan et nasjonalt vitenarkiv skal realiseres.

Utredningen må sees i sammenheng med tiltak 1: forbedre funksjonaliteten for deponering av artikler via Cristin og tiltak 3: kreve deponering i vitenarkiv lokalt eller nasjonalt som en forutsetning for at vitenskapelige artikler skal telle i den resultatbaserte finansieringen. Kravet forutsetter at vitenarkiv og annen nødvendig infrastruktur for deponering er tilgjengelig for alle forskningsinstitusjonene.”

Definisjon og formål med Nasjonalt vitenarkiv

Med Nasjonalt vitenarkiv menes et arkiv som samler, tar vare på, og gjør tilgjengelig digitale versjoner av resultater fra virksomheten ved de norske forskningsinstitusjonene. Dette samsvarer med definisjonen av det engelske ordet «institutional repository» (IR), men på nasjonalt nivå.

Formålet til NVA bør sees i en bred kontekst og kobles mot den overordnede oppgaven med å forvalte og framheve nasjonens forskning. Dette innebærer å sikre hensiktsmessig oppbevaring, tilgang og tilgjengeliggjøring av resultatene.

Etter [lov om avleveringsplikt for allment tilgjengelige dokument](#), den såkalte pliktavleveringsloven, skal alle dokument som er gjort allment tilgjengelig i Norge pliktavleveres til Nasjonalbiblioteket. En oppgave som kan dekkes av NVA er å samle forskningspublikasjonene slik at de blir tilgjengelig for pliktavlevering.

En viktig årsak til at vitenarkiv har fått oppmerksomhet i Norge og internasjonalt er målet om åpen tilgang til forskningsresultater. Andelen vitenskapelige artikler som publiseres med åpen tilgang er økende, men det er fortsatt mye som publiseres bak betalingsmurer. Inntil alle forskningsresultater publiseres med åpen tilgang, er egenarkivering i vitenarkiv med åpen tilgang sett på som en del av løsningen. Regjeringens mål er at innen 2024 skal alle norske vitenskapelige artikler finansiert av offentlig midler være åpent tilgjengelige. Plan S skjerper dette kravet ytterligere. Begge deler gjør det nødvendig å sikre at alle institusjoner har tilgang til å deponere artikler i et vitenarkiv. Merk også at kravet om pliktavlevering av publikasjoner til Nasjonalbiblioteket gjelder uansett, selv i en framtidig situasjon der alle vitenskapelige publikasjoner publiseres med åpen tilgang.

I tillegg til vitenskapelige artikler og andre forskningsresultater som publiseres i eksterne kanaler, er det mange resultater som kun publiseres i institusjonenes lokale vitenarkiv. Dette omfatter institusjonenes rapportserier, masteroppgaver og annet materiale produsert ved hver enkelt institusjon. For dette materialet fyller dagens vitenarkiv en viktig rolle som primær publiseringskanal, og denne funksjonen må ivaretas også i et Nasjonalt vitenarkiv.

Hva kan arkiveres i Nasjonalt vitenarkiv?

For å kunne svare på hvordan NVA bør realiseres er det nødvendig å definere hvilke typer resultater som kan inngå i NVA.

De vitenskapelige publikasjonene ved de norske forskningsinstitusjonene som bruker Cristin inngår i Norsk vitenskapsindeks (NVI) og rapporteres til myndighetene hvert år. Enkelt sett kan man som et absolutt minimum si at alle publikasjoner som inngår i NVI skal arkiveres i Nasjonalt vitenarkiv.

Dette omfatter tre kategorier:

- vitenskapelige artikler i periodika, serier og nettsteder
- vitenskapelige kapitler i antologier
- vitenskapelige monografier

Også andre resultater av forskning og faglig virksomhet må kunne gjøres åpent tilgjengelige. Et nasjonalt vitenarkiv må derfor favne relativt vidt for å dekke summen av behovene til alle relevante interessenter. Sterkt inspirert av åpen vitenskap mener prosjektgruppen bak denne rapporten at også følgende informasjon hører hjemme i et fremtidig NVA:

- Resultater i form av artikler, kapitler og bøker/rapporter som ikke inngår i rapporteringen til NVI
- Resultater av kunstnerisk utviklingsarbeid
- Studentoppgaver
- Forskningsdata (datasett)⁵
- Formidling
- Læringsobjekter



Figur 1 Innhold i Nasjonalt vitenarkiv

Det er nødvendig å skille mellom hvilke resultater som skal inngå i NVA og hvilke andre resultater som kan arkiveres i NVA, alt etter den enkelte institusjons valg og behov. For at det skal være mulig å lagre en resultattype må NVA ha støtte for denne. I dagens lokale vitenarkiv er det ingen tekniske begrensninger med hensyn på filformat. Denne utredningen tar derfor som utgangspunkt at det skal være mulig å arkivere hele spekteret av resultater.

⁵ NVA vil primært gjøre forskningsdata tilgjengelig gjennom referanse til eksterne infrastrukturer for forskningsdata hvor dataene er fysisk lagret.

Dagens arkivlandskap

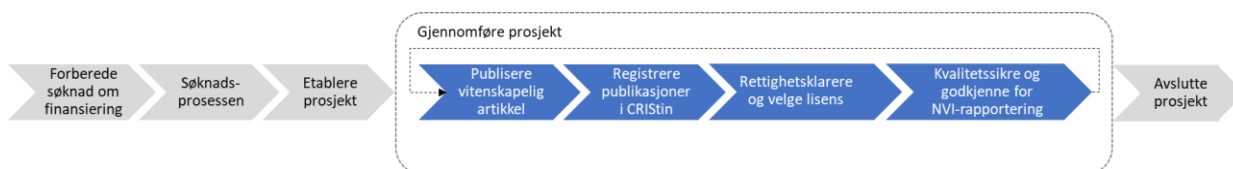
Det er i arbeidet med denne utredningen gjort en relativt grundig kartlegging av dagens situasjon for å identifisere og beskrive problemstillinger knyttet til denne. Kartleggingen har hatt fokus på arbeidsprosesser knyttet til registrering og egenarkivering av forskningsresultater. Ved å se på arbeidsflyt er det relativt enkelt å identifisere områder hvor arbeidsflyten er ineffektiv, og forklare hva som er årsaken til at det blir slik. Dette danner også et godt grunnlag for å kunne peke på hva som må endres for at arbeidsprosessene blir effektive og brukervennlige.

Dagens arkivlandskap er kartlagt hovedsakelig gjennom:

- Studier av tilgjengelig informasjon/dokumentasjon av eksisterende tjenester for lokale vitenarkiv
- Intervjuer med brukere av Cristin og brukere av lokale vitenarkiv
- Interessentseminar på Hell 28. juni 2018
- Interne diskusjoner med sentrale ressurser i Unit knyttet til dagens tjenester levert av Unit inn i dette landskapet

Kartleggingen viser at dagens arkivlandskap kjennetegnes av et sammensatt systemlandskap med lite enhetlige arbeidsprosesser, spesielt på tvers av institusjonene, men dels også innad i institusjonene. Det siste gjelder i særlig grad innenfor de store universitetene og høgskolene. Kartleggingen viser også at det er duplisering av arbeidsprosesser knyttet til en publikasjon som har medforfattere fra flere institusjoner.

Rapportering av resultater ved registrering i Cristin og egenarkivering av de samme resultatene i lokale vitenarkiv er en del av en større verdikjede knyttet til forskningsprosessen. I de tilfeller hvor forskningen er direkte finansiert av offentlige eller private aktører, er forskningen gjerne organisert gjennom et forskningsprosjekt⁶. Når dette er tilfelle, ser dagens verdikjede overordnet ut som vist i figuren nedenfor:



Figur 2 Verdikjeden for forskningsprosessen og formidling av forskningsresultater

Arbeidsprosessene markert med blått i Figur 2 omfatter publisering, registrering i Cristin, rapportering til NVI og egenarkivering i lokalt vitenarkiv, og er den delen av verdikjeden som naturlig berører NVA. I dagens verdikjede er det lite gjennomgående flyt av metainformasjon mellom de ulike arbeidsprosessene. Mange metadata som er relevante i forbindelse med publisering, rapportering og egenarkivering oppstår i de forutgående arbeidsprosessene gjennom søknadsprosessen, og under gjennomføringen av prosjektet. Metadata lagres i ulike entiteter⁷ hos flere aktører. F. eks. i godkjente søknader og prosjektinstanser hos finansøren og i prosjektinstanser i Cristin, men det er fortsatt

⁶ Mye av forskningen foregår uten at det er del av et eksternt finansiert forskningsprosjekt. Det vil med de tiltakene som anbefales i denne rapporten være mulig å definere prosjekter uavhengig av hvem som finansierer det. Prosjekt er ingen forutsetning for å registrere resultater i NVA, men prosjekt vil gi fordeler ved at metadata knyttet til prosjekt-definisjonen (og tilknyttede entiteter) kan gjenbrukes på tvers av institusjonene som deltar, for alle resultater som registreres med tilknytning til dette prosjektet.

⁷ Entitet i denne konteksten betyr en «container» for lagring av en spesifikk type dataobjekt (f. eks. prosjekt).

begrenset flyt av metadata mellom entitetene. Dette betyr at metadata er fordelt over flere kilder uten relasjoner som kan utnyttes maskinelt til å fylle ut metadata. Dette resulterer i mye manuelt arbeid for å få gjenskapt korrekte metadata, med høy risiko for feil, eller at viktige metadata mangler.

Kartleggingen viser også at arbeidsprosessen knyttet til registrering og kvalitetssikring for NVI-rapportering i Cristin og påfølgende egenarkivering i de lokale vitenarkivene har en reaktiv natur. Dette fordi registreringer i Cristin blir overført til de lokale vitenarkivene før de (i de fleste tilfeller) har vært igjennom en full kvalitetskontroll i Cristin, da denne som regel skjer på et senere tidspunkt.

En vanlig arbeidsflyt⁸ er at en metadatapost først registreres i Cristin. Hvis det blir lastet opp en fil til denne metadataposten, overføres metadata og fil til Cristins overføringsarkiv. Overføringsarkivet gjør registreringen (metadata og fil) tilgjengelig for høsting⁹ til de lokale vitenarkivene til de institusjonene som har medforfattere på publikasjonen. Denne overføringen skjer daglig, og typisk før registreringen er kvalifisert i Cristin. Når det er forfattere fra flere institusjoner vil hver institusjon få sin egen kopi av den ukvalifiserte posten. I hvert av de lokale vitenarkivene starter så en prosess for rettighetsklarering av fil, valg av lisens og generell beriking og kvalitetssikring av metadata. Dette betyr at mange utfører den samme jobben på (kopier av) de samme postene.

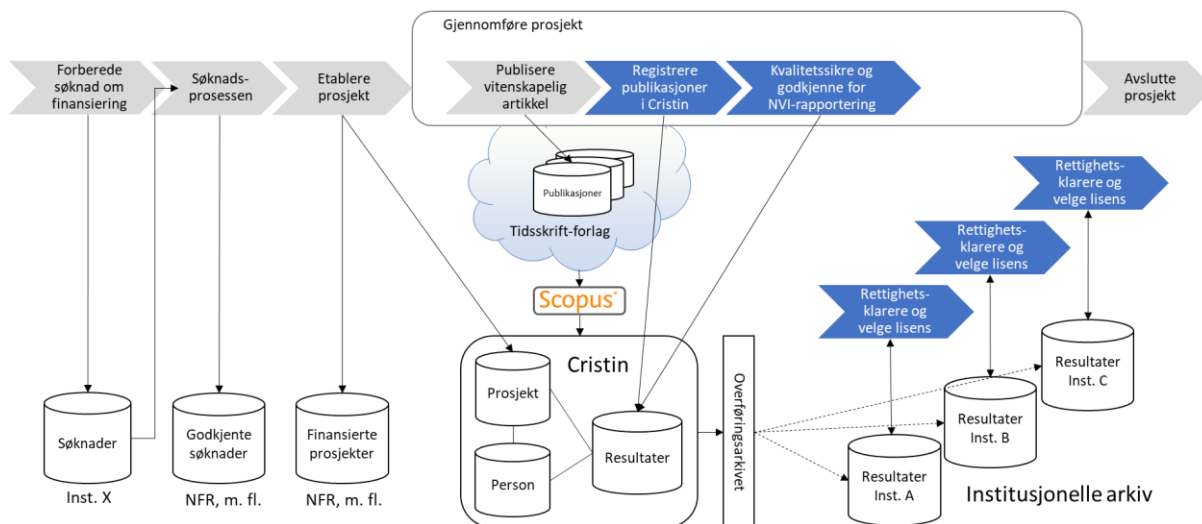
Det er et klart inntrykk at det i mange miljøer er liten støtte til forskerne i prosessen knyttet til manuell registrering av resultater og opplasting av filer i Cristin. Dette gjelder både systemstøtte og tilgang til menneskelige ressurser som jobber tett på forskerne. Dette gir i mange tilfeller dårlig kvalitet på metadata og at opplastet fil har feil versjon.

Selv om kvalitetssikring i Cristin og i lokale vitenarkiv er ulike arbeidsprosesser med ulikt fokus, vil det være overlappende data i Cristin og i de lokale vitenarkivene. Det er derfor uheldig at registreringer i Cristin distribueres til de lokale vitenarkivene før kvalitetssikring og godkjenning i Cristin. Det at samme resultat er registrert med autonome kopier flere steder, bryter med prinsippene viktige arkitekturprinsipper for sektoren.

Dagens situasjon kan illustreres med figuren under som tar utgangspunkt i et tenkt tilfelle hvor en publikasjon som registreres i Cristin har medforfattere fra tre ulike institusjoner (Inst. A, B og C) og hvor prosjektet også er registrert i Cristin.

⁸ Enten ved import fra Scopus og Norart, eller ved manuell registrering.

⁹ Overføringsarkivet omfatter en OAI-server som de lokale vitenarkivene kan bruke for å høste metadataposter og filer på en standardisert måte.



Figur 3 Eksempel på prosessflyt gjennom verdikjeden for forskning med fokus på publisering, rapportering og egenarkivering av forskningsresultater

Det er i dag mange institusjoner som ikke har egne vitenarkiv. Publikasjoner og andre resultater vil selvsagt være registrert og lagret, men ikke i det som forbindes med et vitenarkiv i denne kontekst. Det er her opp til institusjonen, eller sågar forskerne selv, hvordan dette gjøres.

Eksempler er filkataloger på felles servere, Sharepoint, etc. Dette fragmenterer landskapet ytterligere og gjør resultatene lite tilgjengelig utenfor institusjonen.

Problemstillinger knyttet til dagens arkivlandskap

Dette er en oppsummering av de mest sentrale problemstillingene slik de fremkommer i dagens arkivlandskap.

Figur 3 ovenfor kan grovt sett oppsummeres slik¹⁰:

1. Institusjonen(e) utarbeider underlag for søknad om finansiering av forskningsprosjekt.
2. Søknad sendes inn til finansjør(er) og behandles der.

Søknaden inneholder mye (meta-) informasjon om blant annet prosjektets målsetninger og innhold, samfunnsmessig relevans og deltakere i prosjektet.

3. Finansjør godkjenner søknaden og det etableres en prosjektinstans hos finansjøren.

Prosjektinstansen omfatter metadata fra søknaden og evt. nye metadata som har oppstått under behandling og godkjenning av søknaden.

4. Institusjonen(e) etablerer prosjektet i sin(e) institusjon(er), f. eks. i sitt ERP¹¹-system. Prosjektleder kan velge å registrere en prosjektinstans for prosjektet i Cristin, men dette er frivillig, og blir i mange tilfeller ikke gjort.

¹⁰ Det er her valgt å fokusere på publikasjoner som oppstår i kontekst av et prosjekt. Publikasjoner oppstår også uten at de er knyttet til et prosjekt. Og det er ikke er krav at prosjekter skal registreres i Cristin.

¹¹ Enterprise Resource Management.

Metadata fra prosjektinstansen hos finansøren er ikke tilgjengelig i Cristin. Den som registrerer prosjektet i Cristin må enten slå opp i finansørens prosjektinstans og kopiere innhold i datafelter manuelt, eller definere alle metadata på nytt.

5. Resultater fra prosjektet i form av publikasjoner sendes til et vitenskapelig tidsskrift eller en annen publiseringskanal for publisering.

Ulike publiseringskanaler håndterer innsending av manuskripter på ulike måter, fra egne opplastingsportaler for opplasting av filer og evt. skjema for å fylle ut metadata knyttet til publikasjonen, til innsending i form av vedlegg til e-post uten noen formell spesifisering av metadata.

Publiseringskanalen har ikke tilgang til metadata som allerede er definert i søknad og prosjekt og lager derfor «sitt eget» metadatasett når den utsteder en DOI eller kun lagrer metadata om publikasjonen i sin egen løsning.

6. Publikasjoner fra prosjektet registreres i Cristin

I de tilfeller hvor Cristin importerer metadataposter fra Scopus¹² og Norart¹³ blir metadata som er opprettet i disse datakildene brukt. Disse må kontrolleres manuelt og berikes ved import til Cristin. Dette er et omfattende arbeid som i dag utføres av Cristin-sekretariatet. Publikasjoner som ikke leveres fra Scopus eller Norart, må registreres manuelt.

Det er en betydelig forsinkelse fra publisering i tidsskrift og til metadataposter blir levert fra Scopus slik at Cristin-sekretariatet kan importere dem til Cristin. Denne forsinkelsen oppleves av institusjonene som et betydelig problem. Dersom institusjonene har behov for data tidligere enn for NVI-rapportering, vil de måtte registrere publikasjonene manuelt i Cristin.

Ved manuell registrering i Cristin må metadata i stor grad registreres på nytt. For den aktuelle publikasjonen stemmer ikke nødvendigvis de deltagende forskernes organisatoriske tilknytninger med det som er registrert i Cristins personregister og det må eventuelt gjøres manuelle endringer.

7. Metadataposten i Cristin høstes (via overføringsarkivet) sammen med opplastet fil til det lokale vitenarkivet til alle institusjoner som har medforfattere på publikasjonen, dersom institusjonen har et slikt. Dette skjer hvis det blir lastet opp en fil til metadataposten i Cristin.

Feil eller mangler i metadata og feil versjon av fila i Cristin vil forplantes over i de lokale vitenarkivene. Ved feil angivelse av forfatters institusjonstilhørighet vil ikke registreringen bli høstet av riktig vitenarkiv. Den opplastede fila kan være en versjon av publikasjonen som ikke kan egenarkiveres i et lokalt vitenarkiv. Ulike tidsskrift stiller ulike betingelser for egenarkivering i lokalt vitenarkiv. F.eks. tillater mange kun egenarkivering av forfatterens versjon etter fagfelleevaluering, ikke den versjonen som er publisert i tidsskriftet, og kun etter en angitt tidsperiode etter publisering i tidsskriftet (embargo).

8. Registreringen i Cristin kvalitetssikres og godkjennes for rapportering til NVI av Cristin superbruker.

¹² Scopus er Elseviers akademiske database for sammendrag, referanser og henvisninger for artikler i forskningstidsskrifter.

¹³ Norart er en referansedatabase fra Nasjonalbiblioteket over norske og nordiske tidsskriftartikler.

9. Det skjer ingen oppdatering i lokale vitenarkiv når det gjøres korrigerer av kun metadata til posten i Cristin. Tilsvarende vil ikke endringer som gjøres i lokale vitenarkiv gi oppdatering av den originale metadata-posten i Cristin. Registreringer som har blitt overført fra Cristin til de lokale vitenarkivene kvalitetssikres og godkjennes for publisering i de lokale vitenarkivene hos institusjoner som har forfattere på publikasjonen.

Når ulike kopier av samme registrering (metadata og fil) kvalitetssikres manuelt av ulike personer kan det gi inkonsistente metadata og forskjellige filer. Dette problemet gjelder for alle poster som har medforfattere fra flere institusjoner.

Hvis det blir lastet opp fil ny fil til en metadata-post i Cristin, lages det en oppdatert kopi av posten som overføres som en ny registrering til de lokale vitenarkivene som har medforfattere på publikasjonen og det oppstår en dublett i hver av disse. Dette må fanges opp og håndteres manuelt.

Ved større institusjoner er det vanligvis forskjellige personer som arbeider med kvalitetssikring i henholdsvis Cristin og det lokale vitenarkivet, mens ved mindre institusjoner kan dette være en og samme person.

Lite veldefinerte roller og ansvar

Kartleggingen viser at for noen av trinnene i den arbeidsflyten som er beskrevet overfor, mangler det veldefinerte roller. Det er også stor variasjon både på tvers av institusjonene og innenfor institusjonene mht. hvilke ressurser som utfører de ulike oppgavene. Dette gjør det vanskelig å etablere en felles beste praksis for arbeid knyttet til å understøtte forskningen ved publisering av resultater, registrering i Cristin og egenarkivering av forskningsresultater.

Lite enhetlige arbeidsprosesser

De arbeidsprosessene som understøtter publisering, manuell registrering i Cristin og egenarkivering av forskningsresultater varierer betydelig mellom institusjonene. Både med hensyn til selve arbeidsflyten og hvilken type ressurser som utfører de ulike arbeidsoppgavene. I mange tilfeller sitter støtteapparatet som skal bistå forskeren i disse arbeidsprosessene langt unna forskeren og forskningsarbeidet. På de større institusjonene er dette gjerne bibliotekarer på et sentralisert bibliotek.

Mye overflødig manuelt arbeid

Mangel på gjennomgående flyt av metadata mellom de ulike arbeidsprosessene i verdikjeden gir unødvendig mye manuelt arbeid for å gjenskape metadata som allerede er definert i en annen del av verdikjeden. Dette skyldes fragmentert lagring av data og mangel på gode integrasjoner mellom datakilder som lagrer data fra de ulike arbeidsprosessene.

Etterslep i prosessen for import av metadata fra tidsskrift

Scopus er en tjeneste som aggregerer metadata for artikler i flere tidsskrifter og det kan være en betydelig forsinkelse fra en artikkel blir publisert i et tidsskrift til metadata-posten er tilgjengelig for import fra Scopus.

Dupliserte arbeidsprosesser for rettighetsklarering og valg av lisens i vitenarkivene

Problemene beskrevet i pkt. 7 og 9 i listen ovenfor medfører at samme arbeidsflyt for rettighetsavklaring mht. hvilken fil som kan lastes opp og valg av lisens, samt generell beriking og kvalitetssikring av metadata, utføres uavhengig av hverandre på hver av de institusjonene som har

medforfattere på publikasjonen. Samproduksjon og sampublisering av forskningsresultater er mye utbredt¹⁴. Det betyr at denne problemstillingen er relativt omfattende.

Kvalitetssikring og godkjenning for NVI-rapportering i Cristin skjer ofte etter at de samme registreringene er overført til de lokale vitenarkivene. Eventuelle feil spres dermed videre til de lokale vitenarkivene og kvalitetssikring må gjøres flere ganger. I noen tilfeller av forskjellige personer, i andre tilfeller av en og samme person.

Heterogent systemlandskap

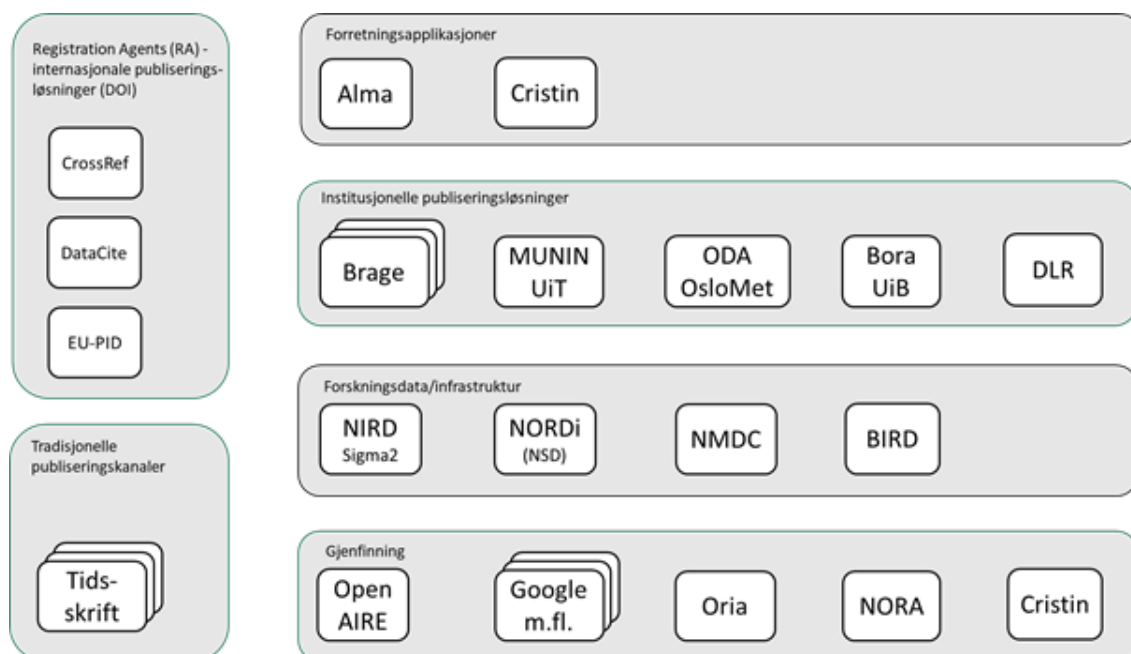
Det finnes pr. i dag 5 forskjellige utgaver av lokale vitenarkiv i Norge. Felles for alle er at de er basert på DSpace som er åpen kildekode.

Det er altså en situasjon med flere ulike systemløsninger for å understøtte en arbeidsflyt som i utgangspunktet burde være veldig lik for alle institusjoner. En vesentlig årsak til dette er at de første lokale vitenarkiv ble etablert før det fantes noen fellestjeneste i Norge.

Brage ble etablert av BIBSYS som en fellestjeneste i 2006 etter initiativ fra høgskoler som ikke ønsket å utvikle egne lokale systemer. De fleste som på det tidspunktet allerede hadde egne løsninger har valgt å videreføre disse.

Institusjonene som bruker Brage er organisert i et konsortium hvor hver deltaker betaler sin andel av de kostnadene Unit har til drift, vedlikehold og utvikling av tjenesten. Pr. i dag har Bragekonsortiet 64 deltakende institusjoner.

Det sammensatte systemlandskapet illustreres gjennom figuren under som viser sentrale systemer/tjenester som benyttes. Figuren er ikke ment å være uttømmende.



Figur 4 Sentrale tjenester/systemer i dagens arkivlandskap

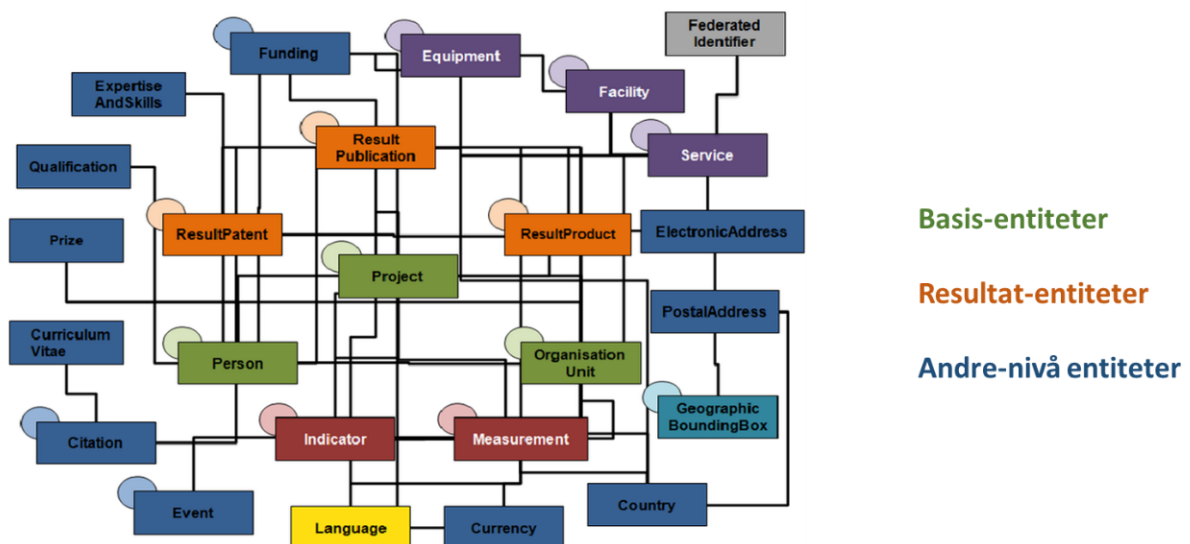
¹⁴ NTNU opplyser at ca. 40 % av deres publikasjoner er samproduksjoner med andre institusjoner.

Mangelfull helhetlig og gjennomgående informasjonsarkitektur for verdikjeden knyttet til forskning og publisering

Det eksisterer ingen helhetlig informasjonsarkitektur for denne verdikjeden. Det er mye duplisering av samme type dataobjekter flere steder uten noen synkronisering mellom dem. Mangel på en felles informasjonsarkitektur (infrastruktur) for denne verdikjeden gjør at det er vanskelig å oppnå gjennomgående flyt av metadata gjennom verdikjeden, og i hvert steg må metadataene i stor grad gjenskapes manuelt.

Det er her relevant å bruke det konseptuelle nivået av CERIF (the Common European Research Information Format) som inspirasjonskilde. Figuren nedenfor viser hvordan forskningsentiteter kan knyttes sammen i veldefinert kontekst som gir merverdi ved at sentrale metadata er definert gjennom denne kontekstinformasjonen. CERIF er en standard informasjonsarkitektur for forskningsentiteter og deres relasjoner.

På konseptuelt nivå ser modellen slik ut:



Figur 5 CERIFs internasjonale standard relasjons-datamodel for lagring og utveksling av forskningsresultater

Figur 5 viser at prosjekt, person og organisasjonsenhet er basis-entiteter i modellen, hvor relasjonene definerer hvilke personer og organisasjonsenheter som deltar i prosjektet og hvilken institusjonell tilhørighet personene har i kontekst av prosjektet.

I dag er prosjekt definert i en eller annen form hos finansjør(er) og eventuelt også i Cristin, uten noen synkronisering av data mellom de ulike prosjektdefinisjonene.

Ved etablering av Cristin ble det innført et register over personer. Dette vedlikeholdes enten ved manuell registrering eller ved import fra de institusjonene som bruker Cristin. Dette registeret benyttes kun i Cristin-kontekst, men andre områder har behov for et tilsvarende register. Med etableringen av Unit, og det mandatet Unit er tildelt, er det realistisk at det kan etableres et felles, nasjonalt autoritetsregister både for person, og entitetene nevnt over.

Nasjonalbiblioteket bygger opp et nasjonalt autoritetsregister for personer og korporasjoner, og det bør vurderes om dette kan benyttes som et forfatter-/opphavsregister.

Målsetninger for Nasjonalt Vitenarkiv

Problembeskrivelsen i forrige kapittel peker på flere områder hvor det er åpenbare potensial for forbedring. Med utgangspunkt i denne og gjennom intervjuer med ulike aktører i den delen av verdikjeden som NVA skal betjene, er det definert et sett med hovedmål for Nasjonalt vitenarkiv som vil redusere, eller eventuelt fjerne dagens problemstillinger. Hvert hovedmål er brutt opp i mer konkrete delmål eller tiltak. Disse vil ligge til grunn ved vurdering av ulike løsningsalternativer mht. grad av måloppnåelse. Målene vil også danne grunnlag for å utarbeide en kravspesifikasjon til NVA når det etableres et prosjekt for realisering av NVA.

Hovedmål 1

Oppnå brukervennlige og effektive arbeidsprosesser for registrering og tilgjengeliggjøring av forskningsresultater.

Delmål:

- a. Etablere en enhetlig, beste praksis arbeidsflyt for registrering av forskningsresultater inkludert forskningsdata, herunder registrering av nødvendige metadata for NVI-rapportering.
- b. Rendyrke roller og hva deres oppgaver og ansvar i denne arbeidsflyten er.
- c. Oppnå størst mulig grad av automatisert utfylling av metadata ved registrering (manuelt eller ved import) i NVA.
- d. Manuelle oppgaver utføres kun én gang for en registrering.
- e. Fjerne (eller minimalisere) behov for kopiering av metadata mellom kilder for metadata.
- f. Kvalitetssikring og godkjenning må være proaktiv, dvs. at en registrering må kvalitetssikres og godkjennes før den «publiseres» (gjøres tilgjengelig) for distribusjon (høsting) til andre arkiver/systemer.

Hovedmål 2

Gjøre alle metadata som registreres i NVA tilgjengelig for indeksering/høsting i søkemotorer og søkeportaler som til enhver tid er de mest relevante. En differensiering basert på materialtype mht. hva som skal eksponeres må vurderes.

Delmål:

- a. NVA skal tilby tjenester som gjør det mulig å indeksere innholdet i NVA.
- b. NVA skal tilby tjenester som gjør det mulig høste metadata i NVA på etablerte formater som benyttes av de tjenestene hvor en ønsker å eksponere innhold i NVA.
- c. NVA må støtte OpenAIREs retningslinjer for *Data Archive, Literature Repository and CRIS Managers*, slik at innhold i NVA blir tilgjengelig gjennom OpenAIRE.

Hovedmål 3

Støtte nødvendige metadata for NVI-rapportering for alle resultater som gir uttelling i RBO-ordningen.

Delmål:

- a. Ved registrering i NVA skal metadata som er nødvendig for å oppfylle gjeldende krav til NVI-rapportering registreres og kvalitetssikres slik at NVA danner grunnlag for NVI-rapportering.

Hovedmål 4

NVA skal sikre at alle registrerte resultater i NVA som faller inn under loven om pliktavlevering, gjøres tilgjengelig for avlevering til Nasjonalbiblioteket.

Delmål:

- a. NVA tilgjengeliggjør materiale som faller inn under pliktavleveringsloven på et omforent format slik at Nasjonalbiblioteket kan hente dette for langtidslagring i sin infrastruktur (nasjonens hukommelse).
- b. NVA må håndtere en elektronisk kvittering fra Nasjonalbibliotekets tjeneste for høsting av data for å verifisere at registrering (metadata og fil) er lagret i Nasjonalbibliotekets arkiv.

Arbeidsflyt i nasjonalt vitenarkiv

NVA skal være et arkiv som samler resultater fra norsk forskning og vil bestå av mange ulike typer materiale som har ulike behov mht. hvordan de registreres, og videre behandling etter at de er registrert. Det synes hensiktsmessig å definere en overordnet, generisk arbeidsflyt som vil dekke alle typer materiale i NVA. Denne består av fem hovedsteg som vist i figuren nedenfor. Noen av disse vil variere avhengig av hva slags type materiale som behandles og dette beskrives gjennom spesialiseringer av steget for en type materiale, eller grupper av materiale som følger samme arbeidsflyt.

Overordnet, generisk arbeidsflyt for NVA

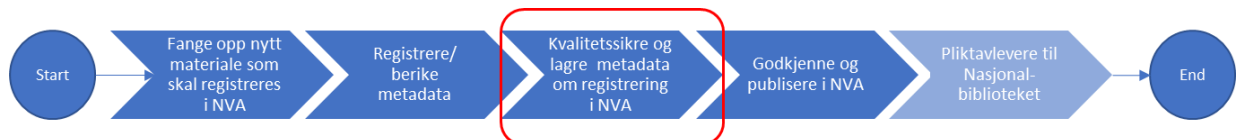
Den overordnede arbeidsflyten i NVA består av fem hovedsteg på overordnet nivå som skissert i Figur 6 nedenfor.



Figur 6 Overordnet, generisk arbeidsflyt i NVA

De to første stegene vil variere avhengig av hvordan det resultatet som skal inkluderes i NVA gjøres tilgjengelig for NVA og hvordan metadata defineres eller gjenskapes. De tre siste stegene vil kun variere mht. hva slags metadata som skal kvalitetssikres og lagres avhengig av type materiale, f. eks. vil dette steget omfatte rettighetsklarering og valg av lisens for filer som er publisert i tidsskrift tidligere. De tre siste stegene er derfor beskrevet først, mens varianter av de to første stegene er beskrevet i de etterfølgende kapitlene.

Steg 3: Kvalitetssikre og lagre metadata



Figur 7 Kvalitetssikre og lagre metadata om registrering i NVA

Startbetingelsen for dette steget er at metadata for resultatet som skal registreres i NVA er klare for kvalitetssikring og lagring. Steget omfatter validering av at dataene er konsistente og at det resultatet som beskrives, ikke fra før er registrert i NVA (dublettkontroll). Når metadataposten validerer, blir den lagret.

Hvis resultatet metadataposten beskriver ikke er publisert tidligere, kan også filen(e) som utgjør selve resultatet lastes opp i NVA hvis dette ikke er gjort i foregående steg. Hvis resultatet allerede er publisert med åpen tilgang og hvis NVA har en referanse i form av en lenke som gir tilgang til filen(e), er det ikke nødvendig å laste opp filen(e) i NVA. I stedet vil metadataposten da kun referere til resultatet i den kilden der det er publisert, samt at det vil bli overført en kopi til nasjonalbiblioteket dersom resultatet faller innenfor loven om pliktavlevering.

Forskningsdata kan også være omfattet her, men da typisk i form av metadata som refererer til datasett lagret i en etablert, ekstern infrastruktur for forskningsdata. Prosjektet har vært i dialog med Nasjonalbiblioteket mht. hvorvidt forskningsdata skal pliktavleveres, men dette er ikke i skrivende stund avklart.

Steg 4: Godkjenne og publisere



Figur 8 Godkjenne og publisere i NVA

I steg 4 skjer det en godkjenning som er en betingelse for publisering¹⁵ i NVA. Dette kan være et automatisert steg, men vil i mange tilfeller kreve en menneskelig kontroll. Dette steget omfatter rettighetsklarering av det resultatet som skal gjøres tilgjengelig i form av fil(er) eller lenker. Med rettighetsklarering menes å avklare at en fil som er lastet opp eller referert til, kan arkiveres i NVA (jfr. betingelser gitt av den som har opphavsretten). Et resultat i form av en fil som skal gjøres åpent tilgjengelig gjennom NVA må også kobles til en lisens som regulerer bruken (f.eks. en CC-lisens). For resultater som skal rapporteres til NVI vil dette steget omfatte godkjenning av at resultatet kvalifiserer for å gi uttelling i RBO. Når det er medforfattere fra flere institusjoner må alle de involverte institusjonene godkjenne resultatet før det kan rapporteres til NVI.

Steg 5: Pliktavlevere til Nasjonalbiblioteket



Figur 9 Pliktavlevere til Nasjonalbiblioteket

Pliktavlevering til Nasjonalbiblioteket foregår ved høsting av data fra NVA til pliktavleverings-tjenesten ved Nasjonalbiblioteket. Det betyr at NVA har gjort sin del av jobben når et resultat er gjort tilgjengelig for høsting. For å validere pliktavlevering og arkivering ved Nasjonalbiblioteket, bør det innføres en elektronisk kvitteringsmekanisme fra Nasjonalbiblioteket til NVA. Når kvittering er mottatt av NVA, kan eventuell temporær informasjon utover det som skal ligge i NVA, f.eks. filer, slettes i NVA.

Steg 1: Fange opp nytt materiale

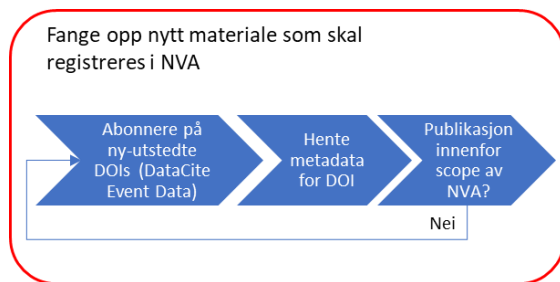


Figur 10 Fange opp nytt materiale som skal registreres i NVA

Dette steget vil ha flere varianter (spesialiseringer) avhengig av hvordan resultatet som skal inkluderes i NVA gjøres tilgjengelig for NVA. De ulike variantene er beskrevet nedenfor.

¹⁵ Publisering i denne sammenheng innebærer at resultatet gjøres tilgjengelig for eksterne kilder (blir søkbart/kan høstes).

Resultater publisert med DOI



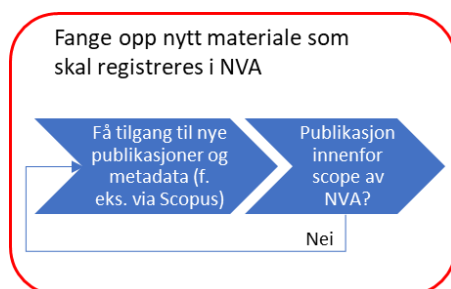
Figur 11 Resultater publisert med DOI

Publiseringskanaler som tildeler en Document Object Identifier (DOI) til sine publikasjoner må publisere metadata for å få utstedt en DOI. Det er et mindre antall "Registration Agents" (RA) som har lov til å dele ut DOI, for eksempel CrossRef, DataCite og EU.

DataCite og CrossRef har etablert en tjeneste (DataCite Event Data) som beriker hverandres metadata. Selv om begge tildeler DOI til alle typer dokumenter, er CrossRef favorisert av tradisjonelle forlag, mens DataCite er favorisert for DOI på forskningsdata.

Alle metadata for DOI er åpne data, publisert for gjenbruk. NVA kan, ved å høste metadata fra DataCite og CrossRef, fortløpende finne relevante resultater (basert på forfatter, institusjon og andre metadata) og automatisk opprette disse i NVA. Disse kanalene vil gjøre resultater tilgjengelig når DOI er utstedt og vil ikke ha en forsinkelse slik en opplever med Scopus som aggregerer informasjon fra mange kilder. Gitt at både høstede metadata og prosjekt i autoritetsregister for prosjekt inneholder sammenfallende finansieringsinformasjon (for eksempel en "Grant-ID"), kan resultatet automatisk kobles til rett prosjekt.

Resultater publisert uten DOI



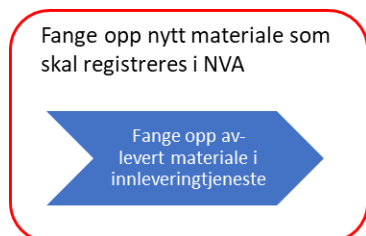
Figur 12 Resultater publisert uten DOI

NVA kan overvåke andre kilder (f. eks. Scopus) som er viktige for å fange opp relevante resultater og opprette metadatatposter etter samme mal som beskrevet for resultater publisert med DOI. Scopus vil være kilde for resultater både med og uten DOI.

Ulike datakilder vil måtte håndteres individuelt, men overordnet er prosessen lik; identifisere relevante metadata, opprette metadatatpost i NVA, og forsøke å knytte den til et prosjekt i autoritetsregister for prosjekt for å få fylt inn (beriket) metadata. Det må avveies om automatisert import fra kilder med dårlig kvalitet på metadata-kvalitet gir gevinst, da det kan medføre behov for økt manuell etterkontroll.

I dette steget bør det vurderes å ta i bruk maskinlæring og kunstig intelligens for å etablere en metadata-kontekst for publikasjonen basert på ustrukturerte data i publikasjonen og kople dette mot autoritetsregistre for person, organisasjonsenhet og evt. prosjekt for å finne en entydig kontekst. På denne måten kan en oppnå automatisering av import.

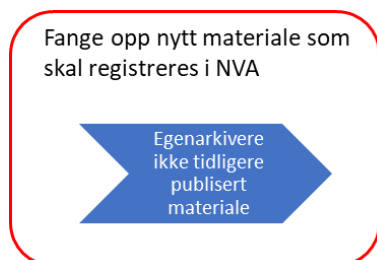
Resultater som ikke er publisert utenfor NVA, men som er tilgjengelig i en ekstern kilde



Figur 13 Resultater som ikke er publisert utenfor NVA, men tilgjengelig i en ekstern kilde

Utdanningsinstitusjonene har, i varierende grad, etablert løsninger for innlevering av studentarbeid for distribusjon til det lokale vitenarkivet. NTNU har et egenutviklet system hvor studenter registrerer metadata for masteroppgaven og laster opp filen(e) som utgjør denne. Når oppgaven er sensurert, klagetiden utløpt og kandidaten har samtykket til publisering, blir metadata og tilhørende fil(er) som er ferdig kvalitetssikret høstet til det lokale vitenarkivet og publisert automatisk. Dette systemet er nå i ferd med å bli erstattet av et kommersielt digitalt eksamenssystem som vil bli høstet på samme måte av det lokale vitenarkivet. Det er stor grunn til å anta at flere institusjoner vil følge etter NTNU. NVA kan overta rollen til det lokale vitenarkivet som primær publiseringskanal for studentoppgaver. NVA vil generelt kunne settes opp til å høste fra alle kilder som tilbyr resultater som er relevante å registrere i NVA.

Resultater som ikke er publisert utenfor NVA



Figur 14 Resultater som ikke er publisert utenfor NVA

For resultater som ikke er publisert noe sted tidligere og hvor NVA vil være primær publiseringskanal, vil registrering i NVA gjøres manuelt. Dvs. at steget trigges av at noen, f. eks. en forsker, ønsker å publisere et resultat i NVA som ikke er tilgjengelig fra noen ekstern kilde. Dette blir som når slikt materiale registreres i Cristin i dag. Et eksempel er institusjonenes rapportserier.

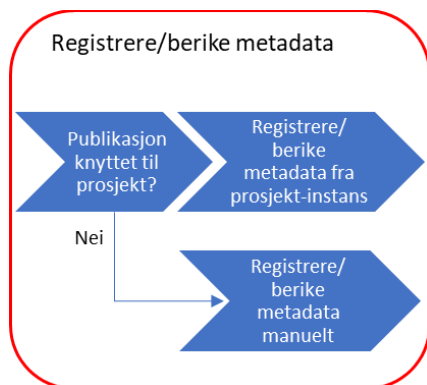
Steg 2: Registrere / berike metadata



Figur 15 Registrere og berike metadata

I likhet med steg 1 vil dette steget ha ulike varianter avhengig av hvordan det resultatet som skal inkluderes gjøres tilgjengelig for NVA. De ulike variantene er beskrevet nedenfor.

Resultater publisert med eller uten DOI



Figur 16 Resultater publisert med eller uten DOI

Startsiden til en forsker i NVA vil inneholde alle prosjekt i autoritetsregisteret for prosjekt der forskeren har en aktiv rolle. Alle registrerte resultater i NVA blir listet opp, fordelt på prosjekt. Resultater uten prosjekt-tilknytning, blir listet til slutt. Forskeren får slik oversikt over hva som finnes i NVA, før forskeren har tilført eller godkjent noe selv.

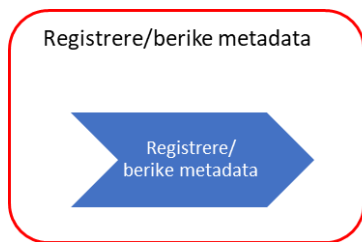
Forskeren kan åpne alle registreringer og godkjenne dem, med eller uten endringer. Forskeren kan også laste opp filer eller legge inn referanser i form av en lenke. Om forskeren savner et prosjekt, så kan forskeren søke det opp eller opprette det. Om forskeren knytter et resultat til et prosjekt, vil metadata fra prosjektet automatisk bli lagt til.

Om en publikasjon skulle mangle i NVA, åpner forskeren et tomt registreringsskjema. Ved å fylle inn en URI (f. eks. en DOI eller lenke) til en publikasjon, hentes relevante metadata inn i registreringsskjemaet, før forskeren ferdigstiller skjemaet og lagrer.

Forskeren må bekrefte alle endringer av metadata i NVA som stammer fra en DOI innhøsting, da slike endringer har konsekvenser ut over NVA. Andre personer og tjenester vil sitere dette resultatet basert på metadata fra masterdata-kilden¹⁶. Det er derfor ikke ønskelig at slike metadata blir endret kun i NVA. I stedet bør NVA initiere en endringsmelding til masterdata-kilden. Eier av masterdatakilden er forpliktet til å rette feil i metadata gjennom sin kontrakt med Registration Agent (RA) når feil blir identifisert.

¹⁶ Det er "Registration Agent" som utsteder DOI, basert på metadata som blir levert inn fra publiseringskilden. Google, NVA og andre som høster metadata, høster disse fra RA. Det er ikke vanlig å høste alle masterdata-kilder direkte.

Resultater som ikke er publisert utenfor NVA, men som er tilgjengelig i en ekstern kilde

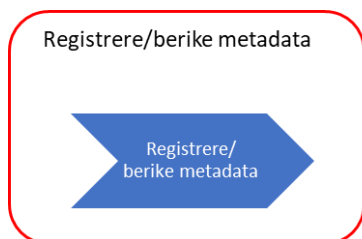


Figur 17 Resultater som ikke er publisert utenfor NVA, men som er tilgjengelig i en ekstern kilde

Som beskrevet for steg 1, har noen institusjoner løsninger for innlevering av studentoppgaver. Ved å etablere nasjonale fellesløsninger for innlevering av studentarbeid, kan sektoren spare mye manuelt arbeid. En annen sannsynlig gevinst vil være at dekkningen av slikt materiale i NVA blir større, og at metadatakvaliteten blir bedre.

Ved høsting fra slike innleveringsløsninger vil all registrering i NVA kunne gjøres automatisk hvis det er gode rutiner for definisjon og kvalitetssikring av metadata i innleveringsløsningene.

Resultater som ikke er publisert utenfor NVA



Figur 18 Resultater som ikke er publisert utenfor NVA

Resultater som skal førstegangs-publiseres i NVA må som hovedregel registreres manuelt. Registreringen vil forenkles betydelig hvis resultatet som skal registreres kan knyttes til en kontekst gitt av én eller flere av entitetene prosjekt, person, organisasjonsenhet ved at metadata i registreringsøyeblikket vil kunne hentes fra disse.

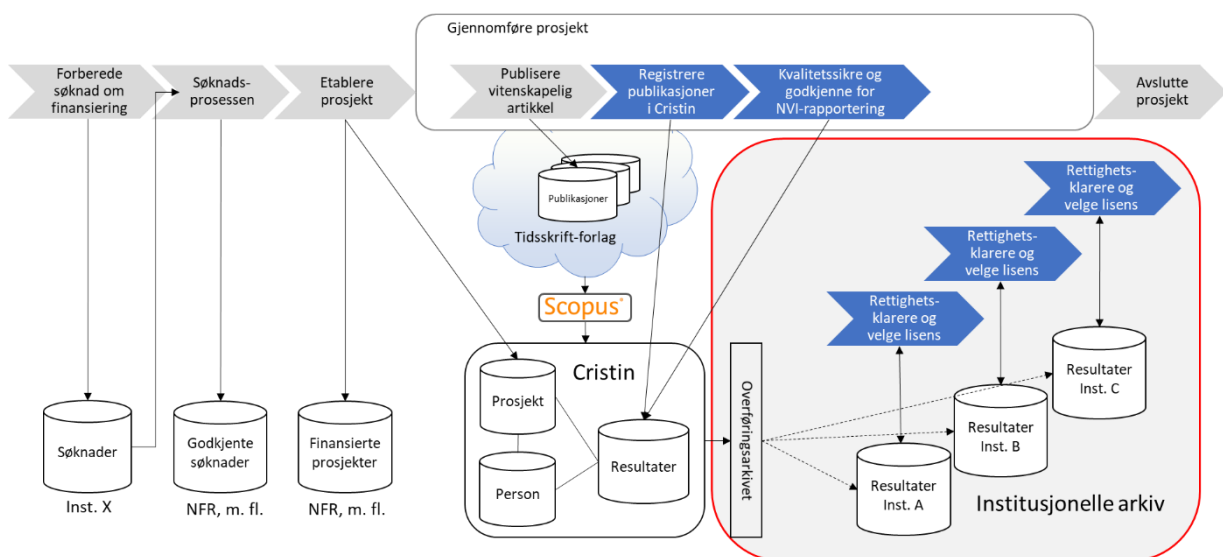
Gevinstpotensial ved realisering av et nasjonalt vitenarkiv

Med basis i dagens situasjonsbeskrivelse og overordnet arbeidsflyt beskrevet foran, vises her hvor det er mulig å hente ut de største gevinstene ved å innføre endringer. Dette vil gi føringer for hvilke tiltak som er relevante, og hvilke gevinster de muliggjør.

I analysen er gevinstene delt opp i tre typer som beskrevet nedenfor:

Gevinsttype	Forklaring / eksempler
Direkte budsjettmessige gevinster	<ul style="list-style-type: none"> • Reduserte driftskostnader, f.eks. rimeligere avtale med en IKT-driftsleverandør. • Reduserte utgifter, f.eks. til porto og papir som følge av elektronisk utsendelse av dokumenter. • Redusert bruk av innleide/vikarer.
Indirekte budsjettmessige gevinster	<ul style="list-style-type: none"> • Effektivisering som følge av endrede arbeidsprosesser eller potensielle stillingsreduksjoner (ikke hele stillinger), som følge av elektronisk dialog med tjenestemottakere. • Denne type gevinster kan kostnadsberegnes, men er vanligvis vanskeligere å ta ut i budsjettene fordi disse ikke har direkte utgiftsposter i regnskapet.
Kvalitative gevinster	<ul style="list-style-type: none"> • Kvalitative gevinster kan være utfordrende å sette et tall på, men er likevel viktige å dokumentere. Eksempler: <ul style="list-style-type: none"> - Alle institusjoner som utfører forskning får et sted å deponere sine forskningsresultater. - Økt brukertilfredshet, mer tilgjengelige tjenester, økt datakvalitet, økt omdømme, økt evne til omstilling o.l.

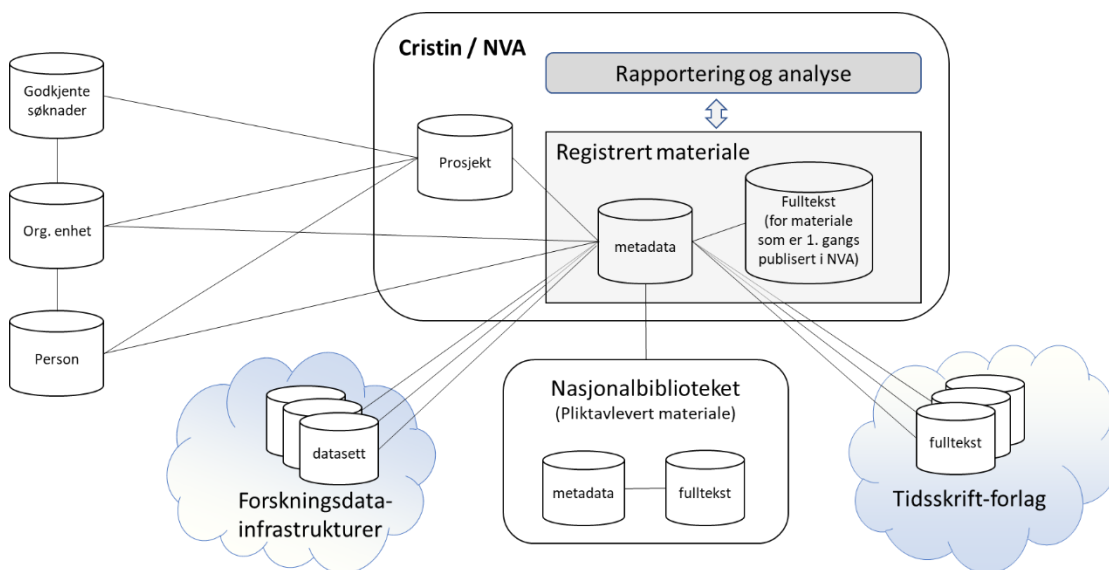
Eliminere dupliserte prosesser for rettighetsklarering, valg av lisens og generell forvaltning av registreringer i lokale vitenarkiv



Figur 19 Samme arbeidsoppgaver utføres flere ganger på ulike kopier av samme registrering

Figur 19 viser den delen av dagens verdikjede som omfatter kvalitetssikring og godkjenning av forskningsresultater som er overført fra Cristin til lokale vitenarkiv. Denne problemstillingen er beskrevet tidligere.

Ideelt sett bør NVA samle alle data om et resultat ett sted, slik at rettighetsklarering og valg av lisens gjøres én gang på én instans. Da vil det ikke være behov for å distribuere kopier av metadata eller fil ut av NVA til lokale vitenarkiv. I stedet vil hver institusjon som er «medeier» i en registrering i NVA få denne som en del av sin produksjon gjennom metadata som identifiserer institusjonen. Med et institusjonelt perspektiv i NVA vil institusjon A kun se registrert materiale som har en eller flere medforfattere fra institusjon A. Konsekvensene av dette er at kopiering av informasjon mellom systemer slik som det gjøres i dag, blir overflødig, og behovet for overføringsarkivet forsvinner.



Figur 20 Felles datagrunnlag for Cristin og vitenarkivene

Ved å utføre alle arbeidsprosesser knyttet til kvalitetssikring av metadata på én og samme instans, vil dette kun gjøres én gang. Dette innebærer også at ansvaret for kvalitetssikringen kan delegeres til én person/rolle, eller til flere personer/roller som er ansvarlig for kvalitetssikring av hver sine deler av registreringen. For eksempel kan en person/rolle ha ansvar for rettighetsklarering og valg av lisens, mens en annen person/rolle har ansvaret for å kvalitetssikre metadata for publikasjonen. Poenget er at hver av disse oppgavene utføres én gang, og på vegne av alle institusjoner som har eierskap til registreringen. Men godkjenning for NVI-rapportering kan ikke i samme grad forenkles, da denne godkjenningen er knyttet til finansiering og er et institusjonsansvar. Det vil fortsatt være nødvendig at hver enkelt institusjon godkjenner sine (bidrag i) vitenskapelige publikasjoner.

Åpenbare gevinster som gjøres mulig med denne arkitekturen/arbeidsflyten er beskrevet nedenfor.

Direkte budsjettmessige gevinster:

- Med denne arkitekturen forsvinner dagens lokale vitenarkiv¹⁷ og driftskostnadene antas å reduseres betydelig for den enkelte institusjon. NVA kan f. eks. finansieres som et spleiselag hvor kostnadene til drift og vedlikehold fordeles på institusjonene, f. eks. gjennom et konsortium.

¹⁷ Består av til sammen ca. 80 DSpace installasjoner.

Indirekte budsjettmessige gevinster:

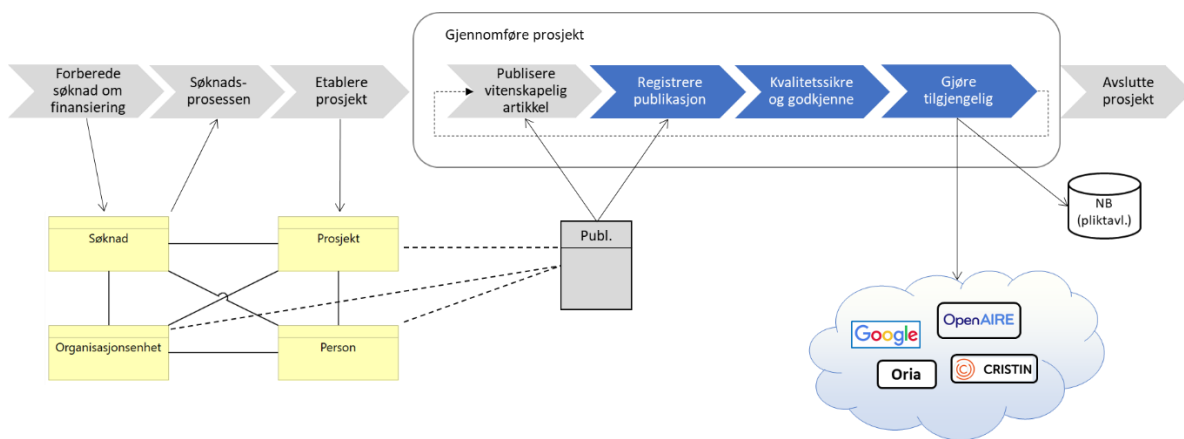
- Det vil samlet sett frigjøres mye tid i institusjonene ved at kvalitetssikring foregår én gang pr. registrering, uavhengig av hvor mange institusjoner som har bidratt i arbeidet med det som registres. For de store universitetene vil dette gi en innsparing på flere årsverk pr. institusjon.

Kvalitative gevinster:

- Det er overveiende sannsynlig at ved å sentralisere og profesjonalisere kvalitetssikring og godkjenning av det som registreres i NVA, vil kvaliteten bli høyere og det blir mindre behov for justeringer i etterkant av registreringen.

Etablere felles, nasjonale autoritetsregistre (masterdatakilder) for sentrale entiteter i verdikjeden

Figuren nedenfor er en forenklet illustrasjon av hvordan autoritetsregistre for søknad, prosjekt, personer og organisasjonsenheter kan sikre flyt av sentrale metadata gjennom verdikjeden slik at disse kan gjenbrukes.



Figur 21 Autoritetsregistre gir mulighet for gjenbruk av metadata

Symbolene i gult i figuren over viser fire entiteter som er svært sentrale i forskningsprosessen. I dag er disse entitetene fragmentert og/eller duplisert i flere kilder med lite eller ingen synkronisering mellom dem. Ved å etablere en felles nasjonal informasjonsarkitektur hvor disse entitetene blir autoritetsregistre (masterdatakilder) som hver for seg gjøres tilgjengelig fra én kilde, og det opprettes koplinger mellom dem via API'er, vil en kunne hente ut registrerte metadata for en gitt kontekst definert av koplingen mellom disse registrene. Dette vil gi potensial for betydelige gevinster ved at metadata for et resultat som skal registreres i NVA i stor grad kan fylles inn automatisk. Dette vil også kunne benyttes ved import av resultater fra eksterne kilder, dersom metadataene som følger med importen er tilstrekkelige til å identifisere en kontekst som f. eks. en *grant-id* eller forfatteres ORCID. I denne prosessen kan det også være et stort potensial i å ta i bruk maskinlæring for tekstklassifisering og på denne måten skape koplinger som definerer en kontekst ved import.

Ved import av resultater fra Scopus og Norart til Cristin i dag, kreves det mye manuelt arbeid av Cristin-sekretariatet for å definere slike kontekstuelle metadata. Faktainnhenting viser at det er et betydelig etterslep i denne kjeden for å få resultatene inn i Cristin, noe som da tydelig tilsier at det vil være en betydelig gevinst i å effektivisere denne prosessen. Som tidligere nevnt vil bruk av maskinlæring og kunstig intelligens ifm. importprosessen kunne øke effektiviteten ved import. En høyere grad av automatisering er også høyt prioritert i den pågående videreutviklingen av Cristin.

Gevinster som muliggjøres ved å innføre denne informasjonsarkitekturen er beskrevet nedenfor.

Indirekte budsjettmessige gevinster:

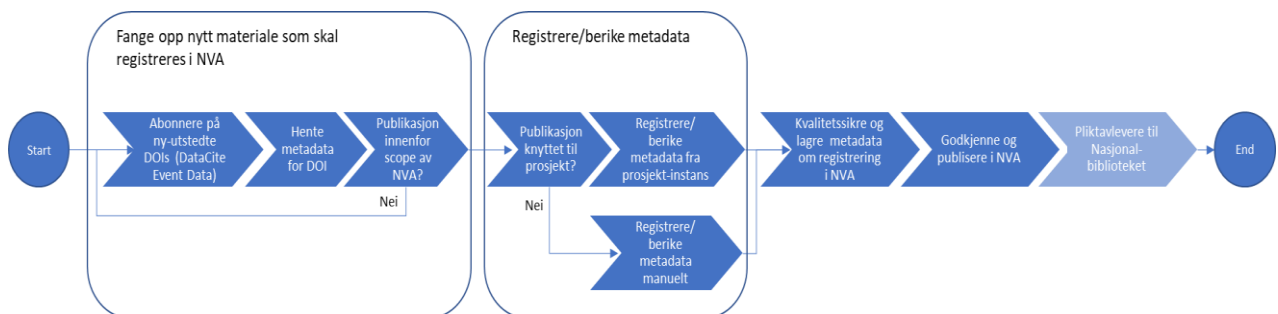
- Når arbeidsoppgaver knyttet til rettighetsklarering, valg av lisens og generell kvalitetssikring av metadata utføres én gang i én kilde på vegne alle institusjoner som har medforfattere på publikasjonen, gir dette stor arbeidsbesparelse ved sampublisering.
- Mengden manuelt arbeid for å hente ut og registrere metadata vil reduseres betydelig.
- Omfanget av arbeid med retting av metadata vil reduseres.

Kvalitative gevinster:

- Kvaliteten på metadataene blir høyere siden metadata som er etablert underveis i forskningsprosessen allerede er kvalitetssikret gjennom søknadsprosessen som ender opp i en godkjent søknad og etablering av en prosjektinstans.
- Etterslep på import av metadataposter fra tidsskrift kan fjernes slik at de blir tilgjengelige i NVA umiddelbart etter at tidsskriftet har gjort dem tilgjengelige for import.
- Det blir ikke lenger nødvendig å registrere publikasjoner i tidsskrift manuelt pga. etterslep i importprosessen.

Gjenbruk av metadata som er registrert ved utstedelse av DOI

I kapittelet Arbeidsflyt i nasjonalt vitenarkiv, ble det i underkapittelet Resultater publisert med DOI, vist eksempel på høsting av metadata fra DataCite/CrossRef for å fange opp nye publikasjoner som er tildelt DOI. På denne måten er det potensial for å automatisere arbeidsflyten som fanger opp nytt materiale som skal inn i NVA og registrere det med gode metadata. Se figuren nedenfor som er hentet fra kapittelet om arbeidsflyt tidligere i rapporten.



Figur 22 Import av resultater som er publisert med DOI

Som beskrevet om import foran, vil det i mange tilfeller være mulig å knytte en publikasjon til en entydig kontekst som gir automatisk innfylling av metadata.

Indirekte budsjettmessige gevinster:

- Spart tid ved registrering av publikasjoner.

Potensiell kvalitativ gevinst:

- Gjennom høsting fra kilder som CrossRef, DataCite og tilsvarende tjenester, er det sannsynlig at alle relevante publikasjoner blir fanget opp av NVA, og med kvalifiserte metadata.

Kanaliserer forskningsdata gjennom NVA og kople disse til forskningsresultater

I NVA legges det opp til en knytning mellom forskningsresultater (f. eks. publikasjoner) og tilhørende forskningsdata. Dette gir kvalitative gevinster som:

- Økt sporbarhet og etterprøvbarhet av forskningen gjennom tilgang til forskningsdata som ligger til grunn for resultatene.
- Gjenbruk av forskningsdata.

Legge til rette for større spredning av norske forskningsresultater og forskningsdata

Et mål for NVA er å gjøre resultater fra norsk forskning lettere tilgjengelig ved spredning i alle relevante kanaler nasjonalt og internasjonalt¹⁸. Dette kombinert med pågående initiativ mot åpen publisering gir potensial for økt innovasjon i samfunnet ved at forskningsresultater fra offentlig finansiert forskning blir lettere tilgjengelig. Dette gir grunnlag for at forskningsresultater i større grad blir utnyttet av kommersielle aktører for å operasjonalisere de mulighetene for fremskritt som forskningen legger grunnlag for. Det er primært kvalitative gevinster som følge av dette som igjen vil kunne gi en samfunnsøkonomisk gevinst.

NVA som autoritativ kilde for resultater fra norsk forskning gir bedre grunnlag for gode analyser og rapportering på all forskning som foregår i Norge på tvers av disipliner

På samme måte som man i dag samler metadata om alle resultater fra forskningen på ett sted i Cristin, vil NVA gi mulighet for brede analyser og rapportering fra den norske forskningen. Selv om innholdet ligger samlet nasjonalt, kan det også analyseres ned på den enkelte institusjon, ned til et nivå tilsvarende antall nivåer som støttes av autoritetsregisteret for organisasjonsenhet. Dette muliggjør analyse og rapportering innenfor ulike avdelinger/seksjoner innenfor én institusjon (institusjonelt perspektiv), eller aggregert over all norsk forskning, ved å se på Norges samlede forskning under ett. Både for den enkelte institusjon og for staten Norge ligger det et kvalitativt gevinstpotensial i å kunne analysere produksjonen på ulike aggregeringsnivåer. Dette vil igjen gi bedre grunnlag for kunnskapsbasert styring på departementsnivå.

NVA som kanal for pliktavlevering av resultater fra forskningen

Gjennom å sikre at alt som skapes av resultater fra forskningen i Norge blir registrert i NVA, vil dette kunne bidra til å forenkle, og sikre pliktavlevering til Nasjonalbiblioteket av dette materialet. Det vil kunne gi direkte budsjettmessige gevinster ved at alle forskningsresultater kanaliseres gjennom én kilde. Dette gir også mindre sårbarhet for fremtidige endringer i pliktavleveringsloven ved at alle forskningsresultater kanaliseres gjennom én kilde og lovendringer som påvirker den tekniske løsningen håndteres ett sted.

Kvalitative gevinster oppnås gjennom sikring av at alle forskningsresultater som registreres i NVA, og som faller inn under pliktavleveringsloven, blir håndtert i tråd med gjeldende regelverk.

¹⁸ Unit er medlem av DataCite og kan utstede Digital Object Identifiers (DOI) på dokumenter.

Alternative tiltak

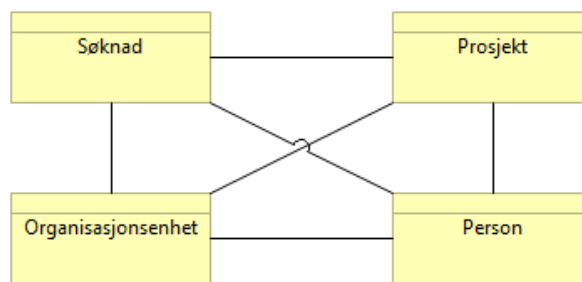
Gjennom problembeskrivelsen og analysen av hvor en kan hente ut de største gevinstene for NVA vises her forslag til konkrete tiltak for realisering av NVA.

I vurderingen av tiltak, er disse delt i to nivåer der det først anbefales to hovedtiltak som hver for seg vil gi betydelige gevinster. Disse hovedtiltakene er ikke gjensidig avhengig av hverandre og kan innføres hver for seg, og til forskjellig tid.

Hovedtiltak 1: Etablere nasjonale autoritetsregistre (masterdatakilder) for prosjekt, person, organisasjonsenhet og søknad (om finansiering)

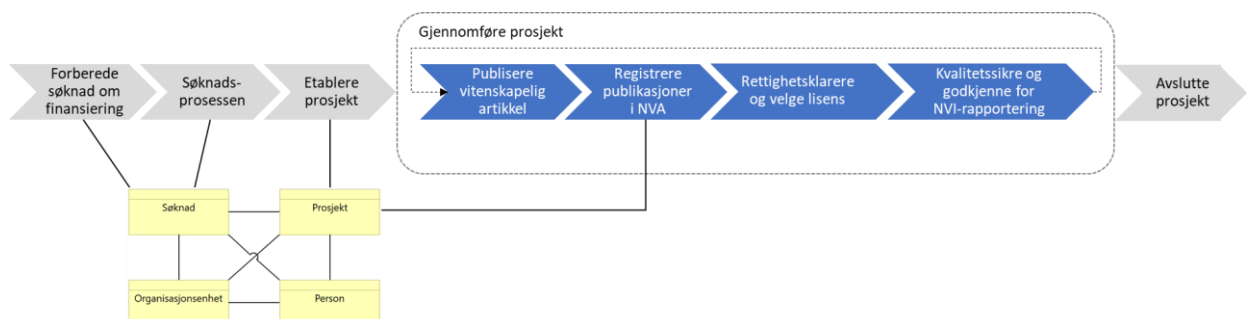
Dette er et tiltak som i hovedsak ligger utenfor nasjonalt vitenarkiv, men det er dog hensiktsmessig å anbefale dette tiltaket da det vil gi store gevinster for NVA, men ikke bare for NVA. Det er klare signaler fra mange aktører at opprettelsen av de nevnte autoritetsregistrene vil gi merverdi i flere andre sammenhenger.

Skissen nedenfor er en konseptuell informasjonsmodell som definerer disse fire autoritetsregistrene og sammenhengene mellom dem.



Figur 23 Autoritetsregistre

Et nasjonalt autoritetsregister er en autoritativ kilde på nasjonalt nivå¹⁹ som er masterdatakilde for en gitt entitet som f. eks. person eller organisasjon. Linjene mellom boksene i figuren viser hvordan data i en kilde er tilgjengelig fra en annen kilde og omvendt gjennom persistente identifikatorer (f. eks. prosjekt-id, person-id, etc.). Ved å relatere verdikjeden for forskningsprosessen til denne informasjonsmodellen, ser en hvordan metadata som oppstår i verdikjeden og lagres i de enkelte autoritetsregistrene gjøres tilgjengelig på tvers av disse når det er tilgang til en eller flere identifikatorer som gir unik kopling inn mot disse.



Figur 24 Samspill mellom autoritetsregistrene og arbeidsflyt i forskningsprosessen

¹⁹ Kan være begrenset til en spesifikk sektor, eller på tvers av sektorer.

I arbeidsprosessen «Forberede søknad om finansiering» vil en institusjon, eller flere institusjoner i fellesskap, forberede en søknad om finansiering av et forskningsprosjekt.

I denne prosessen oppstår det metadata som beskriver fagområde og tema for forskningen, personer og institusjoner som skal bidra, etc. Ved å opprette en instans av Søknad i Søknads-registeret med en unik, persistent id, vil metadata som er relevante for søknaden registreres der. Når alle data relatert til søknaden er definert kan søknaden gjøres tilgjengelig for finansiør(er) det søkes midler fra.

Hvis søknaden godkjennes, oppretter noen en prosjektinstans i Prosjekt-registeret med referanse til søknaden(e) som var grunnlaget for opprettelsen av prosjektinstansen. I henhold til informasjonsmodellen i gult vil prosjektinstansen da ha tilgang til metadata i søknadsinstansen som igjen gir kopling til hvilke personer som inngår i prosjektet og hvilken organisasjonstilhørighet de har i denne konteksten. Koplingene sikres gjennom at alle instanser av disse autoritetene har unike, persistente identifikatorer. Opprettelse av prosjektinstans har potensial for automatisering, men må trigges på en eller annen måte, f. eks. når søknad om finansiering av prosjektet blir godkjent. Det er fanget opp behov for å integrere institusjonenes ERP-systemer med prosjektinstanser i prosjektregisteret. Dette bør et realiseringsprosjekt se nærmere på.

Når et resultat fra dette prosjektet skal registreres i NVA, kan dette koples til den aktuelle prosjektinstansen gjennom en prosjekt-id. Denne koplingen vil gi tilgang til alle metadata i prosjektinstansen, den assosierte instansen av Søknad og dennes referanser til personer og organisasjonsheter i henholdsvis Personregisteret og Organisasjonsheterregisteret. Disse metadataene vil da kunne koples automatisk til resultatet som registreres. I dag må metadata i stor grad gjenskapes manuelt.

Gevinstpotensial

Gevinster som muliggjøres ved å innføre denne informasjonsarkitekturen er:

Indirekte budsjettmessige gevinster:

- Mengden manuelt arbeid for å hente ut og registrere metadata vil reduseres betydelig ved registrering i NVA.
- Det er overveiende sannsynlig at omfanget av arbeid med retting av metadata vil reduseres pga. høyere datakvalitet.
- Automatisert opprettelse av prosjektinstans i prosjektregisteret, vil gi spart tid.

Kvalitative gevinster:

- Det forventes at kvaliteten på metadata blir høyere siden de som er etablert underveis i forskningsprosessen allerede er kvalitetssikret gjennom søknadsprosessen som ender opp i en godkjent søknad og etablering av en prosjektinstans.

Hovedtiltak 2: Etablere en ny felles basis infrastruktur for Cristin og vitenarkiv

Med basis i dagens problemstillinger og målsetningene for NVA beskrevet foran, vurderes her noen konkrete alternative tiltak som beskriver hvordan NVA kan realiseres. De ulike alternativene vil ha ulik grad av dekning av målsetningene, og vil også variere i omfang og kostnad. Basert på hvilket tiltak som sannsynligvis gir størst gevinstpotensial sett i forhold til kostnadene, vil dette være det alternativet som anbefales å videreføre i et prosjekt som skal realisere NVA.

Overordnede arkitekturmessige valg

Gjennom arbeidet med alternative tiltak fremkom det som hensiktsmessig å plassere tiltakene i to nivåer. Øverste nivå knytter seg til hvordan den overordnede arkitekturen (infrastrukturen) for økosystemet for registrering og publisering av forskningsresultater bør utformes.

Dette knytter seg til hvordan NVA som et CRIS-system og de lokale vitenarkivene som IR-systemer er organisert for å sikre optimal interoperabilitet mellom dem. Gjennom kartleggingen fremkommer det at mange av dagens problemstillinger kan knyttes til denne infrastrukturen.

Av denne årsak presenterer og vurderer utredningen noen alternative løsninger på dette nivået med basis i hvordan de ivaretar interoperabilitet mellom NVA (inkludert Cristin) og de lokale vitenarkivene for å oppnå effektiv arbeidsflyt i alle ledd.

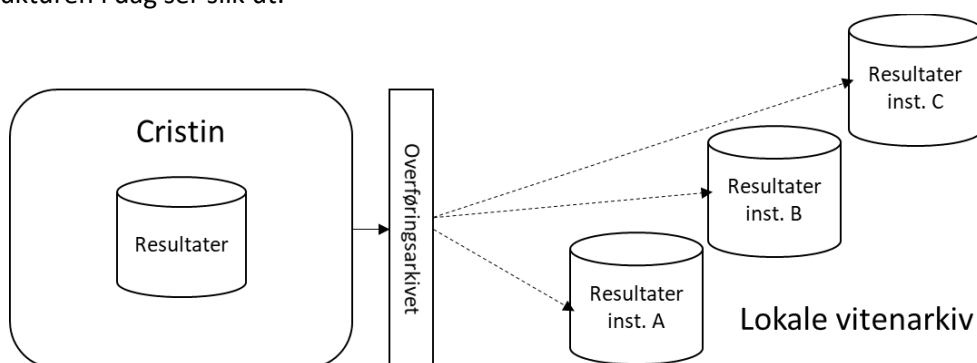
Alternativene tar utgangspunkt i dagens infrastruktur (0-alternativet) og fire andre, nye måter å realisere infrastrukturen på.

Infrastruktur-alternativ 0: Videreføring av dagens infrastruktur

Dagens infrastruktur er basert på at et stort antall institusjoner som utfører forskning (men ikke alle) har et lokalt vitenarkiv (IR) for egenarkivering av resultater fra forskningen. Cristin ble etablert i 2010 med formål om «å samle og gjøre informasjon om norsk forskning tilgjengelig gjennom forskningsinformasjonssystemet Cristin, inkludert Norsk Vitenskapsindeks (NVI)» og «forenkle forskningsadministrative prosesser ved å legge til rette for gjenbruk av informasjon». Dette innebærer at det er overlapp mellom informasjonen som lagres i de lokale vitenarkivene og i Cristin og det er behov for dataflyt mellom dem.

Pr. i dag flyter informasjonen kun én vei – gjennom høsting av Cristin til vitenarkivene via overføringsarkivet som tilbyr en OAI-server²⁰ som gjør det mulig for vitenarkivene og høste sine data. Kopiering av data over i overføringsarkivet skjer kun dersom det lastes opp en fil til metadataposten i Cristin. Det betyr at dersom det kun defineres metadata i Cristin, skjer det ingen overføring til lokale vitenarkiv. Dagens infrastruktur innebærer at data som først er definert i lokalt vitenarkiv, og som også ønskes registrert i Cristin, må defineres manuelt i Cristin. Da vil normalt ikke filen lastes opp i Cristin for å unngå overføring av data som allerede finnes i vitenarkivet.

Infrastrukturen i dag ser slik ut:

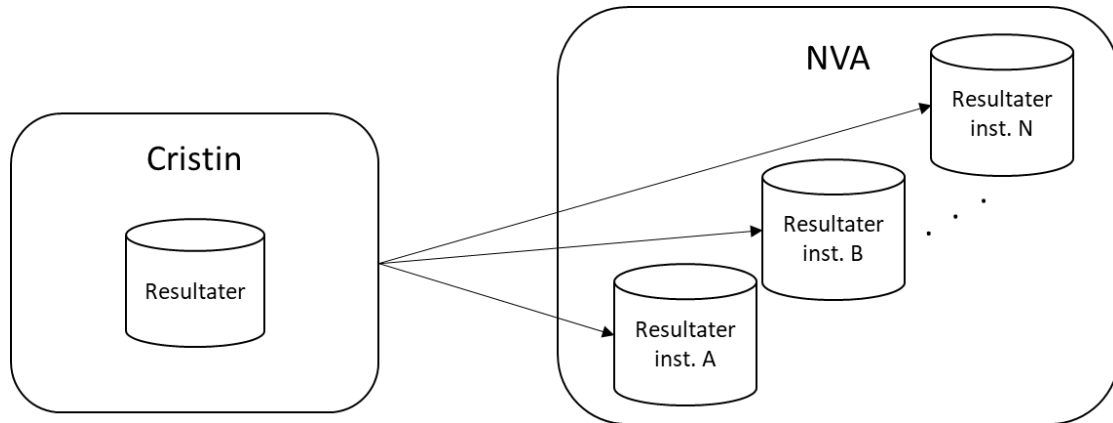


Figur 25 Dagens infrastruktur

²⁰ En OAI-server er en tjeneste som tilbyr høsting av metadata fra en kilde vha. OAI-PMH-protokollen. I en ny løsning er det behov for tilsvarende funksjon, dog ikke denne teknologiske løsningen.

Infrastruktur-alternativ 1: Videreføring av dagens infrastruktur med en liten utvidelse

Overføringsarkivet er nødvendig i dag fordi Cristin ikke har en OAI-server som gjør at de enkelte vitenarkivene kan høste informasjon som de har eierskap til direkte fra Cristin. Det vil være relativt enkelt å tilby en OAI-server i Cristin slik at overføringsarkivet blir unødvendig. Infrastrukturen vil da se slik ut:



Figur 26 NVA realisert med distribuerte vitenarkiv med direkte høsting av Cristin

Cristin omfatter her en OAI-server som gjør at vitenarkivene kan høste direkte uten behov for overføringsarkivet. Videre bør prosessen endres slik at nye registreringer ikke gjøres tilgjengelig for høsting før de er kvalitetssikret i Cristin.

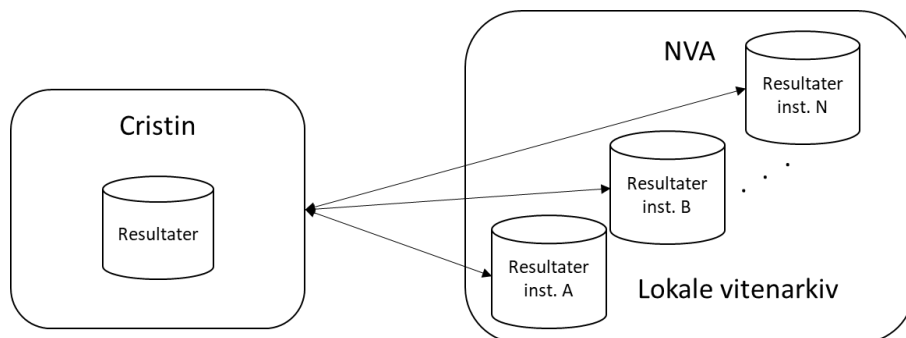
Denne infrastrukturen har åpenbare mangler fordi det ikke flyter noen informasjon fra vitenarkivene til Cristin. Det betyr at alt som skal inngå i NVA må defineres i Cristin. Hvis den enkelte institusjon vil ha en lokal kopi, må den overføres til vitenarkiv hos institusjoner som har eierskap til registreringen, eller så må den defineres manuelt i vitenarkivene.

Dette alternativet løser ikke problemstillingene knyttet til dupliserte arbeidsprosesser og vil ha begrensede gevinster utover at overføringsarkivet blir unødvendig. Høsting direkte fra Cristin har imidlertid noen fordeler ved at kvaliteten på metadata blir bedre, og problemet med dubletter i vitenarkivene når det lastes opp ny fil til en allerede eksisterende metadatatpost i Cristin kan reduseres.

De begrensede gevinstene, og i tillegg at dette alternativet ikke understøtter målsetningen om at data lagres én gang og tilgjengeliggjøres fra én kilde, gjør at dette alternativet ikke anbefales.

Infrastruktur-alternativ 2: Toveis kommunikasjon mellom NVA og vitenarkivene

I dette alternativet flyter informasjon begge veier mellom NVA og de lokale vitenarkivene som skissert i figuren nedenfor.



Figur 27 NVA realisert med distribuerte vitenarkiv og toveis synkronisering med Cristin

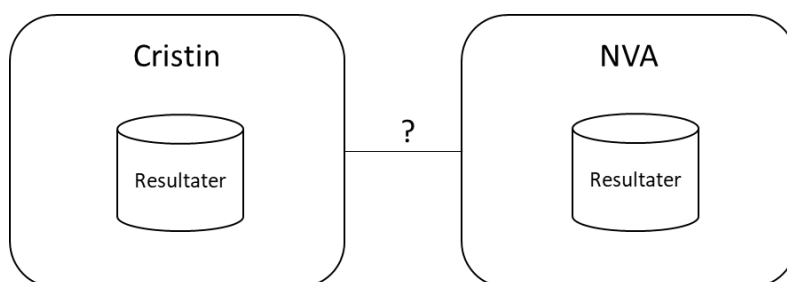
Denne infrastrukturen muliggjør full synkronisering mellom vitenarkivene og Cristin, og mellom vitenarkivene via NVA. I utgangspunktet høres det ut som en attraktiv løsning, men den har noen åpenbare utfordringer og svakheter:

- Hvor ligger masterdata for en registrering (metadata og filer)?
Modellen åpner for et «anarki» mht. at en endring ett sted medfører oppdatering av alle andre versjoner i andre arkiv. Det vil kreves rutiner, sannsynligvis i stor grad manuelle, hvis en skal kunne unngå et slikt anarki. En løsning kan være å definere hvor masterdatakilden er. Siden NVA er en fellestjeneste, er det et naturlig valg at dette også er masterdatakilden. Eller en kan også se for seg at masterdatakilde kan variere avhengig av hvilken type resultat det er snakk om. Uansett vil en slik løsning kreve mye forretningslogikk som vil være utfordrende å håndtere.
- Målsetningen om at «data lagres én gang og tilgjengeliggjøres fra én kilde» understøttes ikke med denne arkitekturen.

Disse utfordringene gjør at denne arkitekturen ikke anbefales.

Infrastruktur-alternativ 3: Felles instans for vitenarkiv

I dette alternativet samles de forskjellige lokale vitenarkivene i én felles instans, men skillet mellom Cristin og vitenarkivet opprettholdes. Dette er vist i figuren under.



Figur 28 NVA realisert som felles instans for vitenarkiv

I dette alternativet flyttes funksjonaliteten som i dag tilbys gjennom de lokale vitenarkivene²¹ over i NVA som blir en fellestjeneste for alle forskningsinstitusjoner, også for de som i dag ikke har egne lokale vitenarkiv. Alternativet baserer seg på at det etableres en felles informasjonsarkitektur og et felles datagrunnlag i NVA for alle institusjonsarkivene. Informasjonsarkitekturen må ivareta logiske

²¹ Dagens lokale vitenarkiv er basert på instanser av DSpace.

skiller mellom registreringer (metadata og filer) basert på eierskap, slik at en institusjon kun har tilgang til data den har eierskap til. For resultater med medforfattere fra flere institusjoner, vil de involverte institusjonene ha delt eierskap.

Det er noen åpenbare fordeler med en slik arkitektur:

- Gjør det mulig å fjerne problemet med dupliserte arbeidsprosesser ifm. rettighetsklarering og valg av lisens for filer i vitenarkivene. Dette gir store effektiviseringsgevinster.
- Alle dagens lokale vitenarkivene blir unødvendige og erstattes av én felles tjeneste. Det forventes å gi betydelig reduksjon i drifts- og vedlikeholdskostnader.

Denne arkitekturen krever også en synkronisering av data mellom Cristin og NVA. I figuren over har er det satt et spørsmålstegn på koplingen mellom NVA og Cristin fordi det er ulike måter å realisere synkroniseringen på mht. retning på dataflyten som beskrevet nedenfor:

- Toveis synkronisering
Det som er beskrevet i forrige alternativ mht. hva som er masterdatakilde ved toveis-kommunikasjon mellom Cristin og NVA gjelder også her, selv om det ikke vil være behov for synkronisering innad i NVA da det kun finnes én instans av en registrering i NVA.
- Synkronisering fra Cristin (master) til NVA
Resultat registreres (manuelt eller ved import) i Cristin og fil lastes opp. En kopi overføres til NVA. Rettighetsklarering og valg av lisens gjøres i NVA. Kvalitetssikring av metadata for NVI-rapportering gjøres i Cristin.
NVA kan tilby mer omfattende metadata enn det som kommer fra Cristin dersom det er behov for det, men disse må da registreres manuelt i NVA. Nye poster som registreres i NVA, eller endringer på eksisterende poster, vil ikke bli synlig i Cristin uten at det samme gjøres manuelt der.
Utfordring: Skal endringer som gjøres på en allerede synkronisert registrering i Cristin overføres til NVA på nytt?
- Synkronisering fra NVA (master) til Cristin
Resultat registreres (manuelt eller ved import) i NVA. Metadatapost overføres fra NVA til Cristin når registreringen i NVA er kvalitetssikret, dvs. rettighetsklarering og valg av lisens er gjennomført. Alternativt kan også kvalitetssikring av metadata for NVI-rapportering gjøres i NVA før overføring til Cristin slik at posten er klar for rapportering når den kommer inn i Cristin.
Utfordring: Skal endringer som gjøres på en allerede synkronisert registrering i NVA overføres til Cristin? Svar på dette er at hvis NVA tilbyr en OAI-server som Cristin høster metadata fra, vil endringer av en metadatapost i NVA føre til en automatisk oppdatering av metadataposten i Cristin. Alternativt kan Cristin tilby en API for å opprette nye og endre eksisterende metadataposter i Cristin som NVA kan benytte.

Noen ulemper med denne løsningen er:

- Målsetningen om at data lagres én gang og tilgjengeliggjøres fra én kilde understøttes heller ikke med denne arkitekturen, selv om det nå kun er to kopier fordelt på to kilder uavhengig av om flere institusjoner har eierskap til resultatet.
- Avhengig av hvilket av synkroniseringsalternativene ovenfor som benyttes, vil det være utfordringer med at en post som i utgangspunktet er identisk i Cristin og NVA vil kunne endres forskjellig.

Det konseptuelt mest ryddige alternativet vil være å definere posten i NVA som master og la alt vedlikehold av posten skje der, inklusive kvalitetssikring for NVI-rapportering, og replikere endringer på posten til Cristin. Dette er en aktuell løsning i en utviklingsfase for å redusere avhengigheten mellom utvikling av Cristin 2 og NVA (kalt NVA trinn 1 senere i rapporten). Cristin må med denne løsningen kun utvides til å håndtere replikering av metadatatpotser fra NVA. NVA vil trolig kunne gjenbruke en del funksjonalitet i Cristin for å håndtere prosessen for kvalitetssikring av metadata for NVI-rapportering.

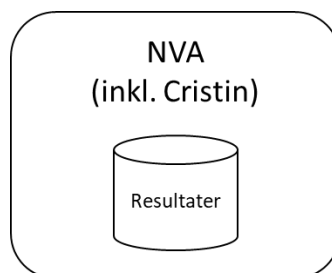
Infrastruktur-alternativ 4: Cristin og vitenarkivene ses i sammenheng og realiseres med felles informasjonsarkitektur og felles datagrunnlag

Vi ser av tilgjengelig informasjon om CRIS- og IR-systemer på ulike nettsteder hos (i hovedsak) europeiske aktører innenfor dette området, at det er sterkt fokus på problemstillinger knyttet til interoperabilitet mellom disse systemene. Dette krever standardiserte formater for overføring av data. Synkronisering av data mellom ulike kilder er utfordrende og vil fungere best når det er en definert masterdatakilde for dataene.

Dette er utfordrende i et regime med mange IR-systemer som kommuniserer med en felles CRIS-løsning slik situasjonen er i Norge. Replikering av data bryter også med et viktig digitaliseringsprinsipp om at data lagres én gang og tilgjengeliggjøres fra én kilde.

Ved å etablere en felles informasjonsarkitektur og felles datagrunnlag for CRIS og IR, dvs. at NVA rommer både funksjonalitet som ligger i Cristin og i de lokale vitenarkivene i dag, vil behov for synkronisering av data mellom dem bli unødvendig og behovet for overføringsarkivet vil forsvinne.

Denne arkitekturen er skissert i figuren nedenfor.



Figur 29 NVA realisert med Cristin og vitenarkivene på felles datagrunnlag i én og samme kilde

I dette alternativet flyttes funksjonaliteten som i dag tilbys gjennom de lokale vitenarkivene over i NVA og blir en fellestjeneste for alle forskningsinstitusjoner, også de som i dag ikke har egne lokale vitenarkiv. Behov for flyt av data mellom vitenarkiv og Cristin, og mellom vitenarkiv, vil forsvinne, da dataene befinner seg kun ett sted. Dette krever at det etableres en informasjonsarkitektur som ivaretar logiske skiller mellom registreringer (metadata og filer) basert på eierskap, slik at en institusjon kun har tilgang til de den har eierskap til. For resultater med forfattere fra flere institusjoner, vil de involverte institusjonene ha delt eierskap.

Denne løsningen krever også at datagrunnlaget er tilstrekkelig i omfang og detaljeringsgrad for å understøtte behovene sett fra både CRIS- og IR-perspektivet. Det er behov for ulike arbeidsprosesser som opererer på ulike subsett av metadata i de to perspektivene. Dette må informasjonsarkitekturen designes for å understøtte.

Omtale av CRIS- og IR-systemer har mye fokus på metadata-formater, og at det i liten grad skiller mellom utvekslingsformat og lagringsformat i disse to kategoriene av systemer. Eksisterende

systemer ser ut til å sette likhetstegn mellom lagringsformat og utvekslingsformat. Dette synes å være et lite gunstig valg. Det er allment akseptert at Dublin Core (DC) som typisk benyttes i IR-systemer er mangelfullt mht. å dekke behov for rapportering i CRIS-systemer.

En mer fleksibel modell er å definere et lagringsformat med omfang og granularitet som er tilstrekkelig for å dekke alle metadatabehov og som benytter standard *state of the art* lagringsteknologi for å oppnå en effektiv administrasjon av data. Hvilke utvekslingsformater som er aktuelle vil endre seg over tid og det er derfor lite bærekraftig å gjenspeile dette i datamodellen. Utvekslingsformat handler om å utveksle data med andre tjenester på veldefinerte formater, som f. eks. DC og CERIF. Et skille mellom lagringsformat og utvekslingsformat gir en langt mer fleksibel arkitektur.

Dette alternativet har mange fordeler:

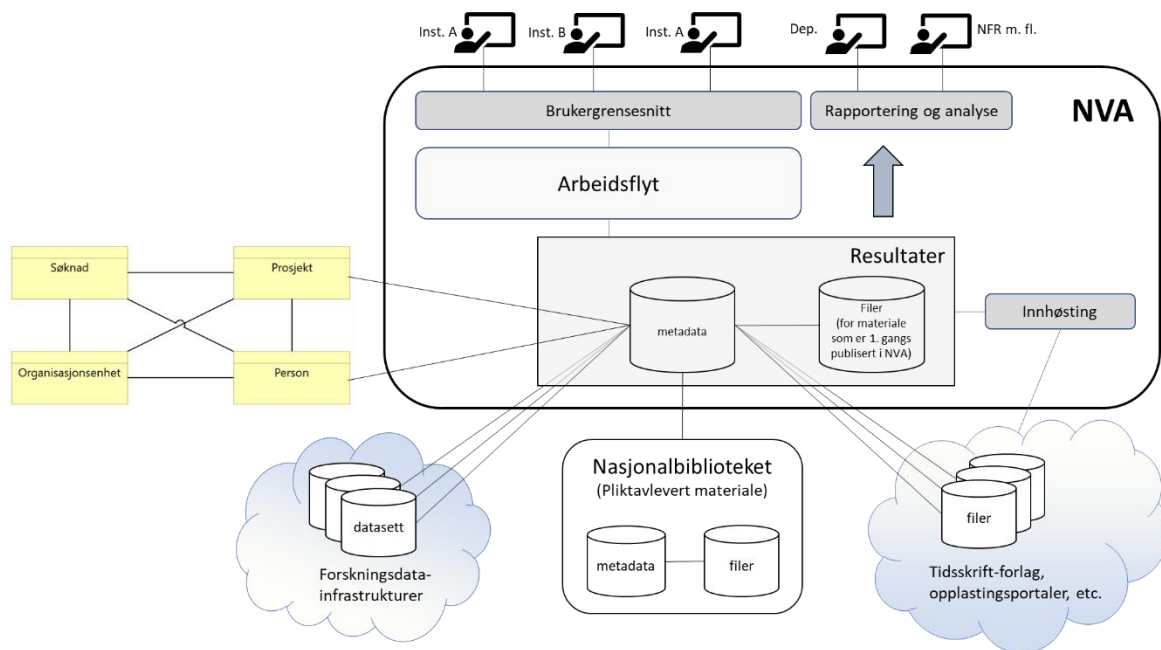
- Det oppfyller flere av de overordnede målsetningene nedfelt i styrende dokumenter for digitalisering:
 - o Data lagres én gang og tilgjengeliggjøres fra én kilde
 - o Fellesløsninger som åpner for mange forretningsmodeller
 - o Felles arkitektur
 - o Effektiv dataflyt
 - o Skille data fra funksjoner
- Disse gir generelt grunnlag for både kvalitative og kvantitative gevinster.
- Gjør det mulig å fjerne problemet med dupliserte arbeidsprosesser ifm. rettighetsklarering og valg av lisens for filer. Dette gir store effektiviseringsgevinster.
 - Alle de lokale vitenarkivene blir unødvendige. Det vil gi betydelig reduksjon i drifts- og vedlikeholdskostnader.
 - Problemstillinger knyttet til dagens infrastruktur hvor en ny opplasting av en fil i Cristin medfører at det oppstår dubletter i de lokale vitenarkivene, med tilhørende arbeid for å rydde opp, forsvinner.
 - Det å etablere en løsning basert på denne infrastrukturen vurderes å være kostnadseffektiv. Det er en løsning som fjerner problemer ifm. synkronisering av data som kjennetegner alle de øvrige alternativene beskrevet foran.

Dette alternativet vurderes som å ha det klart største gevinstpotensialet og er derfor det alternativet for infrastruktur som anbefales som målbilde for NVA (kalt NVA trinn 2 senere i rapporten).

I det videre vurderes ulike måter å realisere en løsning med basis i det anbefalte infrastrukturalternativ 4.

Hvordan kan en best mulig løsning med basis i anbefalt infrastruktur realiseres?

Det er gitt en anbefaling om at NVA og de lokale vitenarkivene ses i sammenheng og realiseres med felles informasjonsarkitektur og felles datagrunnlag. Figur 30 nedenfor viser en skisse av hvordan dette kan se ut. I figuren er det også tatt med relasjoner mellom NVA og autoritetsregistre (masterdatakilder) for prosjekt, søknad, person og organisasjonsenhet. Det er tidligere vist til gevinstpotensial for NVA ved å etablere slike registre.



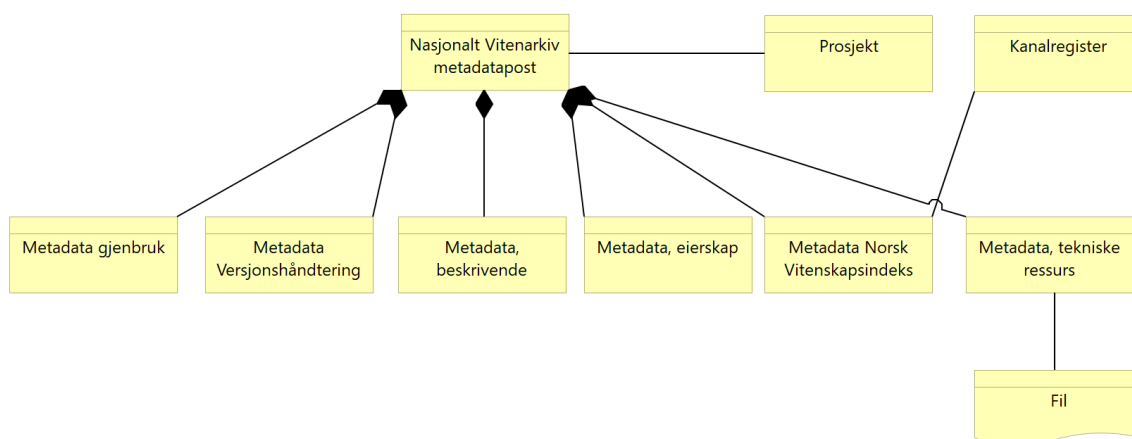
Figur 30 Endelig målbilde for NVA

Dette tiltaket innebærer at NVA vil omfatte både funksjonalitet som i dag ligger i Cristin og funksjonalitet som ligger i de lokale vitenarkivene. Det betyr at de lokale vitenarkivene som selvstendige tjenester basert på DSpace blir unødvendig. Institusjonene får sitt vitenarkiv levert som en fellestjeneste i NVA, hvor innholdet skilles gjennom metadata for eierskap. Resultater som har medforfattere fra flere institusjoner vil ha delt eierskap.

NVA vil i de tilfeller hvor filen allerede er publisert med åpen tilgang i en annen kilde, f. eks. i et åpent tidsskrift, i utgangspunktet ikke lagre en kopi av filen. I tilfeller hvor det vurderes at det nødvendig å laste opp en kopi av fil, vil dette kunne gjøres. For vitenskapelige artikler hvor tidsskriftet ikke tillater egenarkivering av original versjon, kan en annen/tidligere versjon lagres og gjøres tilgjengelig i NVA.

Dette tiltaket baserer seg på at det etableres en informasjonsarkitektur som dekker behovene sett både fra Cristin-perspektivet og fra de lokale vitenarkivene. Det betyr at resultater i NVA vil bestå av en metadatatpost satt sammen av metadata-subsett for ulike formål, og eventuelt en eller flere filer som inneholder selve resultatet. En fil kan inneholde en artikkel, en rapport, en masteroppgave, en video, en presentasjon med innleste kommentarer, etc. Referanse til filen ligger i et eget metadata-subsett og kan peke til en fil internt i NVA (typisk for materiale som publiseres for første gang i NVA) eller eksternt til f. eks. et tidsskrift og eventuelt til Nasjonalbibliotekets infrastruktur for pliktavlevert materiale dersom resultatet faller inn under pliktavleveringsloven. Referanse til et evt. forskningsdata vil også ligge som et metadata-subsett.

Figur 31 nedenfor viser en skisse av hvordan en registrering i NVA kan bygges opp (figuren er ikke uttømmende).



Figur 31 Konseptuell informasjonsarkitektur for registreringer i NVA

Figuren illustrerer hvordan en metadapost settes sammen (aggregeres) av et sett av metadata-subsett og referanse til fil(er). Hvilke subsett som inngår vil være avhengig av type resultat. F. eks. vil et resultat som ikke gir uttelling i RBO-ordningen og dermed ikke skal rapporteres til Norsk Vitenskapsindeks (NVI) ikke ha metadata for dette.

I vedlegg 1 er det vist noen eksempler på hvordan ulike arbeidsprosesser forholder seg til og utnytter denne informasjonsarkitekturen.

Dette hovedtiltaket innebærer at en unngår duplisering på tvers av vitenarkiv og mellom CRIS-delen av NVA og vitenarkivene. Dette gjør det også mulig å unngå at samme arbeidsprosess utføres parallelt på kopier av samme resultat hos alle institusjoner som har forfattere på resultatet. Dette er beskrevet under avsnittet Dupliserte arbeidsprosesser for i kapitlet Problemstillinger knyttet til dagens arkivlandskap. Tiltaket vil drastisk redusere antall tjenester (eller instanser av disse) som i dag dekker samme funksjonelle behov, ref. avsnittet Heterogent systemlandskap i kapitlet Problemstillinger knyttet til dagens arkivlandskap.

Det er viktig å påpeke at det vil være en del investeringskostnader knyttet til å etablere den felles informasjonsarkitekturen som er beskrevet her, og å tilpasse arbeidsprosessene til den. Totalt sett er det sannsynlig at de gevinstene denne løsningen muliggjør, vil bety at investeringen er spart inn i løpet av et tidsperspektiv på to til fire år.

Gevinstpotensial

Gevinster som muliggjøres med dette tiltaket er:

Direkte budsjettmessige gevinster:

- Med denne arkitekturen forsvinner dagens lokale vitenarkiv og driftskostnadene antas å reduseres drastisk for den enkelte institusjon.
- Institusjoner som ønsker å benytte NVA kan ta dette i bruk uten investeringskostnader, men vil måtte betale for bruk på en eller annen måte.

Indirekte budsjettmessige gevinster:

- Det vil samlet sett frigjøres mye tid i institusjonene ved at kvalitetssikring gjøres på én felles instans av det registrerte resultatet, uavhengig av hvor mange institusjoner som har bidratt til resultatet. Totalt for sektorene som benytter NVA vil dette gi en innsparing på mange årsverk hvert år.

Kvalitative gevinster:

- Det er overveiende sannsynlig at ved å sentralisere og profesjonalisere kvalitetssikring og godkjenning av det som registreres i NVA, vil kvaliteten bli høyere og behovet for justeringer i etterkant vil bli mindre.

Alternative måter å realisere dette tiltaket på

Realisering av en løsning med basis i en felles informasjonsarkitektur kan skje på en av flere måter:

1. Egenutvikle en løsning med basis i beste praksis arbeidsprosesser. Et realiseringsprosjekt bør omfatte en forprosjektfase som kartlegger dagens arbeidsflyt og ser på prosessforbedring med basis i ny, felles informasjonsarkitektur som åpenbart vil kunne gi store gevinster²².
2. Ta utgangspunkt i eksisterende egenutviklede komponenter og tilpasse disse. Et aktuelt alternativ er en eksisterende teknisk plattform som håndterer alle typer digitale ressurser, utviklet av Unit. Den er realisert med basis i moderne teknologi som understøtter en viktig målsetning nedfelt i stortingsmelding 27 (2015-2016); «Digital agenda for Norge» om å skille data og funksjoner. Denne plattformen er grunnlaget for tjenestene DLR/Bird og har en løs kopling mellom data og funksjoner og understøtter dermed denne målsetningen. Plattformen er klargjort for å kjøre som en sky-tjeneste. Siden den er realisert med et skille mellom data og funksjoner vil det også være enklere å gjenbruke brukergrensesnittene og forretningslogikken i Cristin.
3. Benytte en åpen kildekode-løsning med nødvendige tilpasninger (f. eks. DSpace eller Fedora)
4. Anskaffe et kommersielt produkt.

Vurdering av alternativene

Alternativ 1 – Egenutvikling av en ny tjeneste for NVA fra bunnen av

Dette alternativet innebærer at en kan starte på nytt og kvitte seg med eventuell teknisk gjeld i de tjenestene som blir erstattet av en ny løsning, samtidig som en kan utnytte erfaringer fra eksisterende tjenester for å spesifisere en best mulig løsning. Frihet i valg av teknologi for den nye tjenesten gjør også at en kan velge en mest mulig bærekraftig teknologisk plattform for å redusere risikoen for å ende opp med ny teknisk gjeld over tid. Den teknologiske utviklingen går imidlertid svært raskt, så det er en risiko for at ingen av dagens tilgjengelige teknologier vil ha svært lang levetid. Egenutvikling fra bunnen av tar gjerne lang tid og dermed blir en mer sårbar for teknologiske endringer.

En fordel med å kunne utvikle noe helt fra bunnen av vil gjøre det enklere å operasjonalisere førende strategier og målsetninger fra myndighetene, som pekt på tidligere. Det kan være svært utfordrende å endre eksisterende systemer for å understøtte nye strategier, da det kan medføre totalt redesign av tjenesten. Dette alternativet krever at det settes av tid og ressurser i en forprosjektfase for å velge teknologisk plattform. Valget vil kunne medføre at det må bygges kompetanse på den nye teknologien, og eventuell innleie av konsulenter med spesialkompetanse på den valgte teknologien.

²² Dette bør gjøres uavhengig av hvilken løsning som velges.

Vi har oppsummert egenskapene til dette alternativet i tabellen nedenfor.

Styrker	Svakheter
<ul style="list-style-type: none"> - Kan kvitte seg med teknisk gjeld i eksisterende tjeneste(r) ved at disse erstattes. - Enklere å operasjonalisere førende strategier og målsetninger fra myndighetene når dette tas hensyn til fra grunnen av. - Fleksibilitet mht. å oppnå brukervennlige og effektive arbeidsprosesser gjennom at en har full kontroll på tjenesten. - Eventuelle endringer i lover eller forskrifter, eller andre faktorer som påvirker arbeidsprosessene vil lettere kunne håndteres fordi en enkelt kan tilpasse tjenesten ift. dette. - God intern kompetansebygging på forretningsforståelse fordi man selv utvikler tjenesten. 	<ul style="list-style-type: none"> - Behov for kompetansebygging ved valg av «leading edge» teknologi. - Vanskelig å forutsi omfang av arbeidet - «Time-to-market» er lengre enn ved anskaffelse av eksisterende tjenester eller ved å ta utgangspunkt en allerede etablert teknisk plattform som det kan bygges tjenester på toppen av. - Egenutvikling er i utgangspunktet dyrere enn ferdige løsninger og kan ha større risiko for feil. - Forventes å bli mer kostbart enn alternativ 2.
Muligheter	Risiki
<ul style="list-style-type: none"> - Ta i bruk sky-teknologi. - Utnytte erfaringer fra eksisterende tjenester for å forstå behov og utfordringer. - Frihet i valg av teknologi for tjenesten kan gi en mer bærekraftig teknologisk plattform. - Endringsbehov kan realiseres raskt uten avhengighet til en leverandørs planer og strategi for sine kommersielle tjenester. - Konsulenter med spesialkompetanse på teknologiplattformen kan gi effektiv kompetansebygging og god fremdrift i utviklingen. - Bruk av skyteknologi gir tilgang til infrastruktur som vil begrense egenutviklingen til «forretnings-funksjoner» (f. eks. vil bruk av en hendelses-drevet arkitektur med mikrotjenester i skyen redusere behovet for å utvikle infrastrukturkomponenter og fokus blir å realisere forretningslogikk). 	<ul style="list-style-type: none"> - Den teknologiske utviklingen går veldig raskt og store teknologiske milepæler kan gi ny teknisk gjeld. - Sårbarhet knyttet til et lite team av utviklere. - Begrenset erfaring med å drifte skyløsninger, men det er et satsningsområde.

Alternativ 2 – videreutvikling av eksisterende tjeneste/plattform

Dette alternativet innebærer gjenbruk av en nylig etablert plattform som baserer seg på beste praksis-teknologi. Plattformen er utviklet av Unit som mulig grunnlag for levere fremtidens Bragearkiv samt til å levere to eksisterende tjenester; Digitale læringsressurser (DLR) og løsning for å lagre, dele og tilgjengeliggjøre forskningsdata (Bird).

Med dette alternativet kan mye av funksjonaliteten i Cristin videreføres. I første omgang kan NVA realiseres som et felles, frittstående vitenarkiv som beskrevet i Infrastruktur-alternativ 3: Felles

instans for vitenarkiv, med basis i informasjonsarkitekturen beskrevet i Figur 31. Metadata for poster i NVA må i denne fasen overføres fra NVA til Cristin. Det endelige målbildet (se Infrastrukturalternativ 4: Cristin og vitenarkivene ses i sammenheng og realiseres med felles informasjonsarkitektur og felles datagrunnlag) er at Cristin integreres inn i NVA-plattformen og opererer på samme datagrunnlag som NVA gjennom tilpasning til den samme informasjonsarkitekturen.

Med dette alternativet vil NVA gradvis kunne etableres som en felles løsning for sektoren som også kan dekke behov for forvaltning og tilgjengeliggjøring av læringsobjekter og forskningsdata der det er hensiktsmessig sett i sammenheng med øvrige løsninger.

Styrker	Svakheter
<ul style="list-style-type: none"> - Benytter beste praksis teknologi basert på event-baserte mikrotjenester som kjører i skyen. - Utnytter velprøvd infrastruktur og basistjenester levert som en skytjeneste fra en seriøs skyleverandør. - Mange behov for integrasjon er allerede dekket i plattformen og vil gjenbrukes, eksempelvis lenke sammen registreringer, tredjeparts autentisering og granulert autorisering. - God intern kompetansebygging på forretningsforståelse fordi man selv utvikler tjenesten. - Har lavere risiko enn alternativ 1 	<ul style="list-style-type: none"> - Behov for kompetansebygging ved valg av «leading edge» teknologi. - Vanskelig å forutsi omfang av arbeidet, men mindre omfattende enn alternativ 1. - Egenutvikling er i utgangspunktet dyrere enn ferdige løsninger og kan ha større risiko for feil.
Muligheter	Risiki
<ul style="list-style-type: none"> - Billigere å etablere enn alternativ 1 - Eksisterende funksjonalitet i Cristin 2 kan gjenbrukes ved tilpasning til felles informasjonsarkitektur. - Videre utvikling av funksjonalitet som i dag ligger i Cristin flyttes over på ny plattform. - Tjenester kan utvikles (som mikrotjenester) i parallell uten tette koplinger, med minimal utviklingstid. - Mulighet for å lage en fellestjeneste for sektoren som kan benyttes for læringsobjekter og forskningsdata. 	<ul style="list-style-type: none"> - Sårbarhet knyttet til et lite team av utviklere. - Begrenset erfaring med å drifte skyløsninger, men det er et satsningsområde.

Alternativ 3 - Benytte en løsning basert på åpen kildekode med nødvendige tilpasninger

Det vil være utfordrende å tilpasse en åpen kildekode-løsning fullstendig til informasjonsarkitekturen som er anbefalt. Det er dermed en risiko for at dette alternativet ikke klarer å ta ut hele gevinstpotensialet som ligger i en slik løsning. Samtidig vil dette alternativet trolig også gi mindre fleksibilitet med hensyn til vedlikehold på grunn av at den opprinnelig er designet for et snevrere perspektiv enn hva det er behov for i NVA.

De løsningene som i dag finnes som åpen kildekode-løsninger er i stor grad dedikert til å håndtere behov sett kun fra institusjonelle vitenarkiv (IR) perspektivet eller kun fra CRIS-perspektivet.

I tråd med tidligere argumentasjon under valg av overordnet arkitektur (infrastruktur), er det mange fordeler med en felles informasjonsarkitektur og felles datagrunnlag for CRIS og IR, og at arbeidsprosesser for de to perspektivene forholder seg til samme datagrunnlag. Dét medfører utfordringer med en åpen kildekode-løsning som ikke er tilrettelagt for å håndtere både IR- og CRIS-perspektivene.

DSpace er en åpen kildekode-løsning som i dag benyttes for å realisere alle de lokale vitenarkivene, men anses ikke som egnet til formålet da den er designet og skreddersydd for å dekke enkeltinstitusjoners behov. Det er lite hensiktsmessig å forsøke å tilpasse denne til å betjene mange institusjoner i én instans. Dette har vært forsøkt med Brage og har vist seg å være problematisk. Det har medført at det må innføres en ny infrastruktur hvor hver institusjon i Brage-konsortiet får sin egen instans av DSpace for lettere å kunne rendyrke IR-perspektivet.

Fedora er på samme måte som DSpace basert på åpen kildekode, men er mer omfattende og i større grad mulig å tilpasse, men det vil også bety at den blir mer kostbar å tilpasse og drifte.

DSpace-CRIS er en utvidet versjon av DSpace som er laget for å kunne håndtere data både i CRIS- og IR-perspektivet ved utvidelse av datamodellen i DSpace. Det faktum at det er laget en slik løsning, underbygger argumentasjonen om at samling av data for CRIS og IR er en god idé. DSpace-CRIS er i likhet med DSpace designet med utgangspunkt i å betjene enkeltinstitusjoner og har derfor samme utfordring som DSpace i den norske modellen hvor vi har ett sentralisert CRIS-system.

Det vil være mange behov for integrasjoner mellom NVA og andre tjenester. Ved å realisere NVA vha. en standard åpen kildekode-løsning vil det kunne bli betydelige utfordringer for å realisere disse integrasjonene uten at det må lages spesialfunksjonalitet for dette. Når det er nødvendig å lage mange spesialtilpasninger til et standard-system for å dekke alle behov, er dette alternativet erfaringsmessig ikke et smart valg. Derfor frarådes dette alternativet.

Styrker	Svakheter
- Så lenge åpen kildekode-løsningen dekker alle behov når den benyttes som tiltenkt (ift. hvordan den er designet), og det ikke er nødvendig å gjøre spesialtilpasninger, vil dette være et godt alternativ.	- Behov som ikke dekkes av standard-funksjonaliteten må lages spesielt, og da forsvinner mye av styrken til en åpen kildekode-løsning. Hvis ingen andre har de samme behovene faller mye av grunnen til å velge åpen kildekode bort.
Muligheter	Risiki
- Hvis åpen kildekode-løsningen dekker behovene til NVA, vil dette alternativet være et godt valg	- Det må bygges så mye spesialfunksjonalitet for å dekke alle behovene til NVA at gevinstene ved valg av åpen kildekode faller bort.

Alternativ 4 - Anskaffe et kommersielt produkt

Dersom det finnes kommersielle produkter som dekker de behovene som ligger til grunn for NVA, er dette et interessant alternativ som kan ha klare fordeler. Men, dersom produktet i mindre grad dekker behovene, vil dette kreve spesialtilpasninger som enten kan bli veldig kostbare, eller som leverandøren av produktet ikke vil levere fordi det ikke faller inn under deres produktstrategi. Det blir derfor en sterk avhengighet til leverandørens vilje og evne til å tilby slike spesialtilpasninger.

Med dette alternativet vil arbeidsprosessene i stor grad bli låst til det aktuelle produktet som er valgt – og vil kunne medføre betydelige endringer i arbeidsprosessene ved seinere skifte av system. Unit har blandet erfaring med bruk av store kommersielle systemer. I stor grad viderefører disse silo-

problematikken som er påpekt tidligere i rapporten, og er ofte ikke enkle å tilpasse til andre tjenester.

De medfører ofte også lock-in overfor det valgte systemet som gjør det kostbart å skifte ut systemet (byttekostnader), blant annet fordi alle integrasjoner må skapes på nytt og arbeidsprosesser må endres. Det betyr også at muligheten (og viljen) til å ta i bruk ny og mer innovativ teknologi blir mindre.

Flere av disse kommersielle produktene er store og omfattende, ofte med langt mer funksjonalitet enn det en strengt tatt trenger. Dette gjør produktet kostbart. Ofte er slike produkter realisert med én felles kodebase for alle kunder og da blir det vanskelig for leverandøren av slike produkter å lage spesialtilpasninger som ingen andre, eller svært få, andre kunder har behov for. Av samme grunn tilsier all erfaring at det å få inn ny funksjonalitet i slike produkter kan ta veldig lang tid.

Styrker	Svakheter
<ul style="list-style-type: none"> - Velprøvde produkter siden antall anvendelser er stort. - Kort «time-to-market». 	<ul style="list-style-type: none"> - Behov for spesialløsninger kan være vanskelig å oppnå. - Kommersielle systemer/tjenester kan være lite fleksible.
Muligheter	Risiki
<ul style="list-style-type: none"> - Hvis produktet dekker alle behov er dette et godt alternativ. 	<ul style="list-style-type: none"> - Betaler for funksjonalitet som ikke tas i bruk. - Mange integrasjoner og spesialløsninger kan gi høye byttekostnader.

Noen enkle betraktninger om det økonomiske gevinstpotensialet ved oppnådd målbilde for NVA

For å danne et bilde av hva innsparingspotensialet for den anbefalte løsningen kan dreie seg om er det gjennomført en forenklet analyse. Det er viktig å poengtere at tallene er beheftet med betydelig usikkerhet. Prosjektet anbefaler at det gjøres en grundigere kostnadsanalyse når det er besluttet hvordan NVA skal realiseres. Dette kan være en innledende fase av et realiseringsprosjekt hvor en nærmere detaljerer det vedtatte løsningsalternativet og samtidig gjør en kartlegging av hvordan arbeidsprosessene kan effektiviseres med denne løsningen. Det vil gi et langt bedre grunnlag for en økonomisk analyse.

Dagens kostnadsbilde kjennetegnes av drift, utvikling og administrasjon av Cristin og de lokale vitenarkivene (både Brage-konsortiet og lokale løsninger), registreringen og kvalitetssikringen som gjøres sentralt av Cristin-sekretariatet og det som gjøres lokalt i Cristin og i vitenarkiv av institusjonene. Estimer for disse kostnadene er vist i tabellen under. Kostnadene for Cristin og Brage er budsjetttall, mens øvrige tall er kvalifiserte estimer. For estimatene er det tatt utgangspunkt i antall brukerinstitusjoner inndelt i størrelseskategorier og estimert antall årsverk for én institusjon pr. størrelseskategori. Dagens kostnadsbilde estimeres til 108 mill. kroner årlig.

Dagens kostnadsbilde (årlig)	
Drift, utvikling og administrasjon:	
- Cristin	11 mill. kr
- Brage	4 mill. kr
- Øvrige vitenarkiv	1 mill. kr
Registrering og kvalitetssikring:	

- Sentralt av Cristin-sekretariatet	8 mill. kr
- Lokalt i Cristin	50 mill. kr
- Lokalt i vitenarkiv	34 mill. kr
Totalt:	108 mill. kr

Ved å realisere målbildet for det anbefalte tiltaket vil det være kun én felles tjeneste som dekker funksjonaliteten i Cristin og dagens vitenarkiv. Det er sannsynlig at en felles løsning basert på den anbefalte plattformen kan forvaltes mer effektivt enn hva situasjonen er i dag for eksisterende løsninger. Effektiviseringspotensialet skyldes at det er færre løsninger som skal forvaltes, at teknologien baserer seg på en moderne plattform med mindre teknisk gjeld, og at en del funksjonalitet knyttet til at personregisteret i Cristin blir overtatt av autoritetregister for person som vil ligge utenfor NVA.

Selv om det er mulig med en mer effektivt forvaltning, betyr ikke det nødvendigvis at utviklingskostnadene i absolutte tall vil bli lavere. Kostnadsbildet vil påvirkes av ambisjonsnivået og de framtidige ønskene om ny funksjonalitet. Dersom NVA blir en suksess kan dette gi økte forventninger om videreutvikling²³. For et gitt ambisjonsnivå er det sannsynlig at et felles NVA vil være mer effektivt enn dagens situasjon anslagsvis i størrelsesorden 20 %. Med et forvaltningsbudsjett som i dag er på 16 mill. kroner, tilsvarer dette ca. 3 mill. kroner.

Potensialet for å redusere kostnadene knyttet til registrering og kvalitetssikring er betydelig. Det er flere tiltak som gjennomføres uavhengig av NVA for å utnytte dette potensialet, og det er viktig å ikke telle gevinster dobbelt. To pågående forbedringstiltak er automatisk overføring av studentoppgaver til vitenarkivene og arbeidet med Cristin 2.

Arbeidet med Cristin 2 vil forenkle oppgavene som gjøres av Cristin-sekretariatet og det som gjøres lokalt i Cristin av institusjonene. Deler av tiltakene i denne utredningen sammenfaller med det som allerede er tenkt for Cristin 2, eksempelvis å knytte personregisteret sammen med ORCID. I denne utredningen kvantifiseres ikke gevinstpotensialet knyttet til dette utover å konstatere at dagens kostandsnivå for disse aktivitetene estimeres til 58 millioner og at det trolig er et vesentlig forbedringspotensial.

For arbeidet som i dag gjøres i vitenarkivene er anslagsvis halvparten av arbeidsmengden knyttet til studentoppgaver. Med den planlagte løsningen for automatisk overføring fra digitale eksamenssystem til vitenarkivene, vil denne arbeidsmengden forsvinne i sin helhet. I henhold til estimatet over tilsvarer dette en effektivisering på 17 mill. kr årlig.

Det resterende ressursforbruket er da knyttet til det som i dag gjøres i de lokale vitenarkivene som ikke er knyttet til studentoppgaver. Dersom NVA realiseres som anbefalt estimeres det at disse arbeidsprosessene vil kunne effektiviseres med 25 %. I henhold til estimatet utgjør dette 3 mill. kroner årlig. Denne effektiviseringen skyldes hovedsakelig at ved sampubliseringer så trenger ikke hver enkelt institusjon å avklare rettigheter og valg av lisens og en kan unngå dobbeltarbeid.

Inntil den årlige besparelsen kan realiseres vil det være engangskostnader i mellomperioden før NVA er fullt oppe å gå. I en periode vil NVA, Cristin, Brage og andre lokale vitenarkiv eksistere side om side og det vil være kostnader knyttet til migrering og omstilling.

²³ Disse kan også gi opphav til nye gevinster som forsvaret investering i ny funksjonalitet.

Forslag til plan for gjennomføring

En forutsetning for en gradvis utvikling er at arbeidet med Cristin 2 går videre uten for mye påvirkning og ekstraarbeid. Cristin-applikasjonen kjører i dag på en utdatert teknisk plattform med mye teknisk gjeld. Inntil Cristin blir klar for videre utvikling utover plattformoppgradering slik at Cristin 1 kan termineres, må arbeidet med Cristin 2 få gå som planlagt inntil det kan skje. Utviklingsløpet for NVA må derfor planlegges med hensyn til dette.

En fornuftig måte å oppnå minst mulig avhengighet mellom utviklingsløpene for Cristin 2 og NVA slik at disse kan foregå parallelt, er å starte utviklingen av NVA med basis i Infrastruktur-alternativ 3: Felles instans for vitenarkiv. Da kan den anbefalte informasjonsarkitekturen som er beskrevet under Infrastruktur-alternativ 4 etableres uten at Cristin-arbeidet blir påvirket. På denne informasjonsarkitekturen vil det etableres en NVA-kjerne på den anbefalte tekniske plattformen som både DLR og BiRD er basert på. Det vil være naturlig å først tilby de institusjonene som i dag ikke har noe eget vitenarkiv å ta i bruk den nye NVA-kjernen.

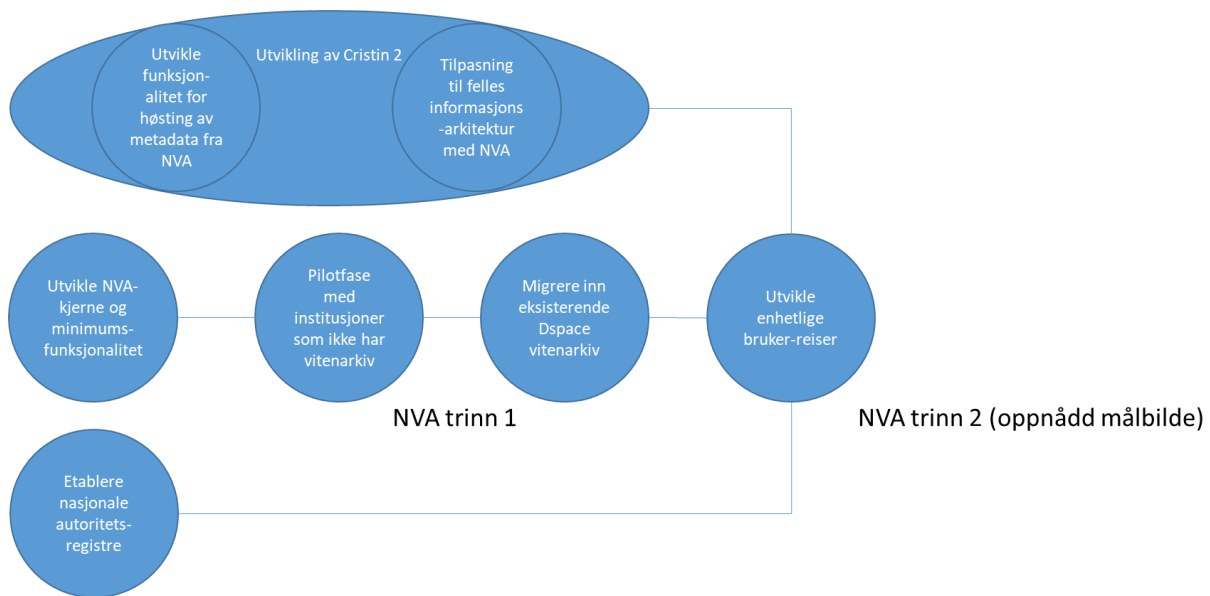
Relativt tidlig i løpet med utvikling av NVA, må det gjøres mulig å replikere metadata fra NVA-kjernen over i Cristin. Dette gjøres trolig enklest ved at Cristin tilbyr en API for å opprette nye, endre og slette eksisterende poster (som er opprettet fra NVA) i Cristin.

Når NVA-kjernen er ferdig utviklet og testet ut gjennom en pilotfase, vil neste steg være å migrere data fra de eksisterende, lokale vitenarkivene over i NVA-kjernen. Deretter vil NVA-kjernen erstatte de lokale vitenarkivene som da kan slås av. Da er NVA trinn 1 oppnådd.

Et tredje parallelt løp kan være å utvikle de nasjonale autoritetsregistrene som skal brukes på tvers av tjenester og systemer. Dette vil, med sannsynlig unntak av autoritetsregister for prosjekt, bli ivaretatt gjennom andre prosjekter. Cristin vil som en del av dette løpet måtte tilpasses den nye virkeligheten hvor prosjekt blir en autoritet som også benyttes av andre tjenester utenfor Cristin/NVA. I tillegg må det gjøres tilpasninger som følge av at Cristins personregister erstattes med et nasjonalt autoritetsregister på utsiden av Cristin/NVA.

Når tiden er moden, bør Cristin tilpasses informasjonsarkitekturen i NVA-kjernen slik at NVA og Cristin opererer på samme datagrunnlag for å oppnå målsetningen om at data lagres én gang og tilgjengeliggjøres fra én kilde, NVA trinn 2.

Denne fremdriften i illustrert i figuren under.



Figur 32 Gradvis utvikling av NVA

Konklusjon

Oppdraget som var utgangspunkt for arbeidet oppsummert i denne rapporten var å utrede hvordan et nasjonalt vitenarkiv skal realiseres. I stedet for å se NVA i et vakuum, valgte prosjektgruppen å se NVA i et større perspektiv som en sentral del av verdikjeden for forskningsprosessen.

Rapporten anbefaler at det innføres to hovedtiltak:

1. Etablere nasjonale autoritetsregistre (masterdatakilder) for prosjekt, person, organisasjonenhet og søknad (om finansiering).
2. Etablere en ny felles basis infrastruktur for Cristin og vitenarkivene.

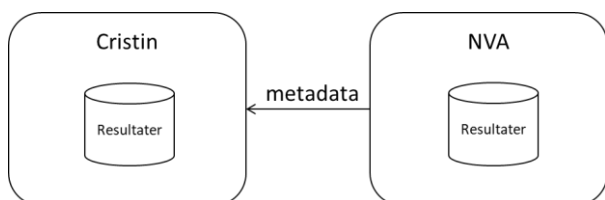
For hovedtiltak 1 er det kun etablering av autoritetsregister for prosjekt som er kandidat for å bli realisert i NVA/Cristin. Cristin har i dag et prosjektregister som kan utvides til å bli et felles, nasjonalt prosjektregister med persistent id for prosjekt-instanser. De øvrige autoritetsregistrene som dette tiltaket peker på, vil bli realisert gjennom egne initiativ/prosjekter. Rapporten beskriver hvilket gevinstpotensial det ligger i å gjenbruke metadata som alt er registrert tidligere i verdikjeden.

Hovedtiltak 2 medfører å endre dagens infrastruktur, ved å gå bort fra separate løsninger for de lokale vitenarkivene, til en felles tjeneste med felles datagrunnlag (NVA-kjernen). Målbildet er at Cristin og NVA integreres i plattformen, slik at de deler et felles datagrunnlag. Da forsvinner behovet for å kopiere data mellom Cristin og vitenarkivene og dermed fjernes behovet for alt dobbeltarbeid som foregår på kopier av samme post fra Cristin. Det ligger et stort gevinstpotensial i dette tiltaket.

For å realisere den nye basis infrastrukturen som hovedtiltak 2 beskriver, anbefaler prosjektgruppen at denne bygges på en teknologisk plattform som er utviklet av Unit for å publisere digitale objekter, som f. eks. læringsobjekter og forskningsdata. Plattformen er basert på *leading edge*-teknologi og utnytter infrastruktur i skyen. Tjenestene DLR og BiRD kjører i dag på denne plattformen.

Det anbefales at hovedtiltak 2 innføres i to trinn som beskrevet nedenfor.

NVA trinn1: Vitenarkivene samles i NVA i en egen tjeneste adskilt fra Cristin:



Figur 33 Målbilde for trinn 1 av NVA

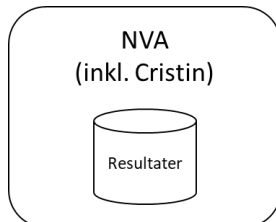
I dette trinnet etableres det som er kalt NVA-kjernen (boksen til høyre i figuren). NVA er her en fellestjeneste for vitenarkiv med ett felles datagrunnlag for alle institusjoner. Dette forutsetter en informasjonsarkitektur som ivaretar behovene for å kunne skille registreringer mellom institusjoner basert på eierskap. For resultater som har medforfattere fra flere institusjoner vil disse ha delt eierskap mellom de involverte institusjonene. I rapporten er det beskrevet en konseptuell informasjonsarkitektur som ivaretar disse behovene og som danner grunnlag for en effektiv administrasjon av forskningsresultater knyttet til import, registrering, rapportering og tilgjengeliggjøring.

I dette trinnet vil Cristin og det arbeidet som foregår under prosjektet Cristin 2, bli minimalt påvirket. Som vist i figuren over blir metadata fra registreringer som blir publisert i NVA overført til Cristin.

Det betinger at Cristin i dette trinnet utvides med funksjonalitet for å replikere metadata-poster fra NVA i Cristin.

Dette trinnet vil bidra til å realisere gevinster knyttet til at dupliserte arbeidsprosesser i institusjonene for rettighetsklarering og valg av lisens fjernes. Sammenslåing av alle de lokale instansene for vitenarkiv på en felles plattform vil medføre sparte driftskostnader.

NVA trinn 2: Cristin og NVA samles i samme plattform og opererer på samme datagrunnlag:



Figur 34 Endelig målbilde for NVA

I dette trinnet kan Cristin tilpasses informasjonsarkitekturen som NVA benytter, og flyttes over på NVA-plattformen. Da oppnås det en løsning hvor det blir unødvendig å ha kopier av samme registrering i NVA og i Cristin. Informasjonsarkitekturen er designet for å støtte behov både sett fra NVA-perspektivet (egenarkivering og tilgjengeliggjøring) og Cristin-perspektivet (NVI-rapportering).

Dette trinnet vil bidra til å utløse nye gevinster som ligger i ett felles datagrunnlag.

Ordliste

BiRD

BiRD er en tjeneste for lagring og deling av forskningsdata, levert av Unit. BiRD støtter datahåndteringsplaner vha. integrasjon med easyDMP (Uninett Sigma2).

BORA

BORA er Universitetet i Bergen sitt åpne digitale forskningsarkiv. BORA inneholder masteroppgaver, doktoravhandlinger, forskningsdata og andre vitenskapelige publikasjoner i fulltekst.

Brage

Brage er en løsning for åpne arkiv som benyttes av norske universiteter, høyskoler, forskningsinstitutter og forvaltningsinstitusjoner som er medlem av Bragekonsortiet. Den enkelte institusjon registrerer og vedlikeholder selv innholdet i sitt eget arkiv for å gjøre dette åpent tilgjengelig. Innholdet består av artikler, studentoppgaver, skriftserier og annet materiale som produseres ved institusjonen. Brage forvaltes og driftes av Unit.

Bragekonsortiet

Gjennom Bragekonsortiet sikres brukerinstitusjonene medvirkning og innflytelse på drift og utvikling av plattformen BIBSYS Brage. Konsortiets øverste organ er det årlige konsortiemøtet som oppnevner styret for konsortiet og vedtar kommende års budsjett og leveranseplan. Konsortiet har også en faglig arbeidsgruppe som skal gi råd om prioriteringer og valg av løsninger. Brage er basert på DSpace som er åpen kildekode.

CERIF (Common European Research Information Format)

CERIF definerer en formell, konseptuell modell for forvaltning av forskningsresultater, inkludert interoperabilitet mellom sysetemer som forvalter forskningsresultater.

Cristin

Cristin er et nasjonalt forskningsinformasjonssystem som samler og tilgjengeliggjør informasjon om norsk forskning og forenkler forskningsadministrative oppgaver ved å legge til rette for gjenbruk av forskningsinformasjon. Videre gir Cristin støtte for rapportering av vitenskapelige publikasjoner til Kunnskapsdepartementet og Helse- og omsorgsdepartementet (NVI-rapportering).

Crossref (tidl. CrossRef)

Crossref er et offisielt registreringsorgan for den internasjonale DOI ordningen og utsteder digitale objekt identiteter (DOI). Det drives av Publishers International Linking Association Inc. (PILA) og ble lansert tidlig i 2000 som et samarbeid mellom utgivere for å muliggjøre persistent sitering av vitenskapelige artikler på tvers av utgivere av nettbaserte akademiske tidsskrifter.

DataCite

Internasjonal, ideell (non-profit) organisasjon, som Unit er medlem av, og som har som mål å forbedre sitering for forskningsdata gjennom å:

- etablere lettere tilgang til forskningsdata på internett
- øke aksept av forskningsdata som legitime, siterbare bidrag til vitenskapelig materiale
- støtte dataregistrering som gjør at resultatene kan verifiseres og gjenopptas for fremtidig studie.

DLR - Digitale Læringsressurser

DLR er en skybasert tjeneste for digitale læringsressurser som gjør det enkelt å benytte de samme læringsressursene flere steder, på tvers av emner, programmer, systemer og institusjoner. DLR støtter alle filformater, og de enkelte ressursene kan enkelt søkes frem og gjøres tilgjengelig der det er behov for dem. DLR forvaltes og driftes av Unit.

DOI (Digital Object Identifier)

Digital object identifier (eller DOI) er en standard (ISO 26324) for pålitelig identifisering av et elektronisk dokument i et nettverk og med tilhørende metadata, på en strukturert utvidbar måte.

DUO

DUO er universitetet i Oslo sitt vitenarkiv som omfatter elektroniske utgaver av masteroppgaver, doktoravhandlinger og forskningsartikler.

EasyDMP

Verktøy for å lage datahåndteringsplaner, etablert av Uninett Sigma2.

metadata

Metadata (fra gresk meta "om" og latin data "opplysninger") er data som tjener til å definere eller beskrive andre data. Eksempel på metadata er navn på forfattere av en artikkel.

Munin

Munin er UiTs åpne vitenarkiv. Munins formål er å gjøre åpent tilgjengelig så mye som råd av vitenskapelige og faglige dokumenter produsert av UiTs forskere og høyere grads studenter.

Norart

Referansedatabase fra Nasjonalbiblioteket over norske og nordiske tidsskriftartikler.

Norwegian Marine Data Centre (NMDC)

Ny nasjonal infrastruktur for marine data som skal levere sømløs tilgang til dokumenterte marine datasett over viktige havområder for Norge til den marine forskningsverdenen.

For mer informasjon, se: <https://nmdc.no/om-prosjektet/norwegian-marine-data-centre-nmdc->

NORDi

NORDi tilbyr tjenester for forskere og studenter for å finne, lagre og dele forskningsdata.

Forskningsinstitusjonene tilbys tjenester for dokumentasjon og oversikt over forskningsdata i egen portefølje.

ODA

ODA er det digitale vitenarkivet for OsloMet - storbyuniversitetet. Arkivet inneholder vitenskapelige artikler, bokkapitler, doktorgradsavhandlinger og masteroppgaver.

Scopus

Elseviers akademiske database for sammendrag, referanser og henvisninger til artikler i forskningstidsskrifter.

UNINETT Sigma2 Research Data Archive

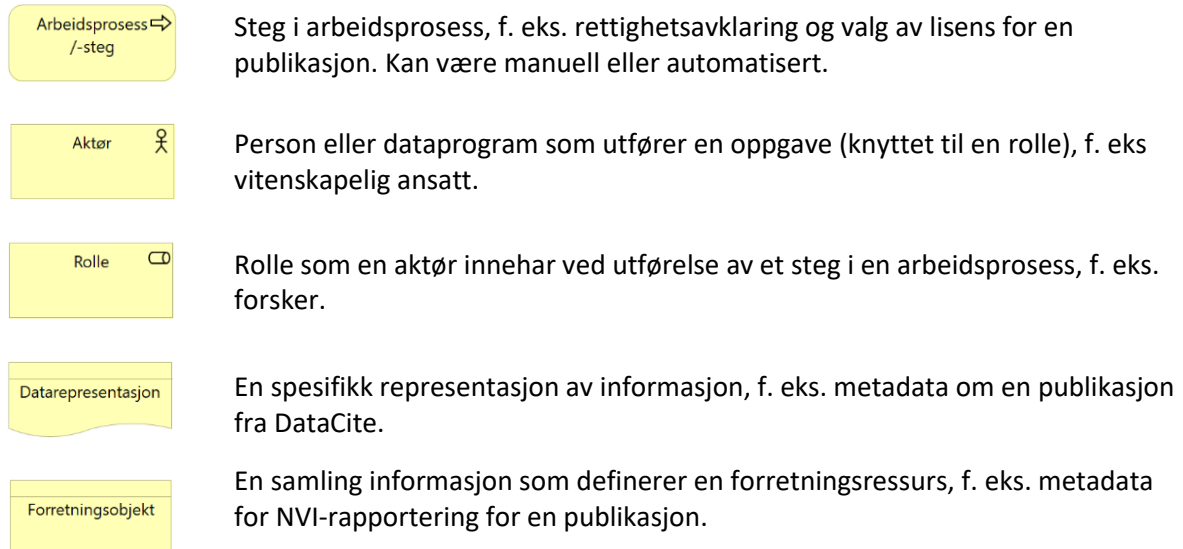
Tjeneste som gjør det mulig for brukere å publisere offentlig finansierte, akademiske forskningsdata.

Se mer om tjenesten her: <https://www.sigma2.no/content/archive-user-guide#Introduction>

Vedlegg 1 – Eksempler på arbeidsprosesser i NVA og hvordan disse utnytter den anbefalte informasjonsarkitekturen

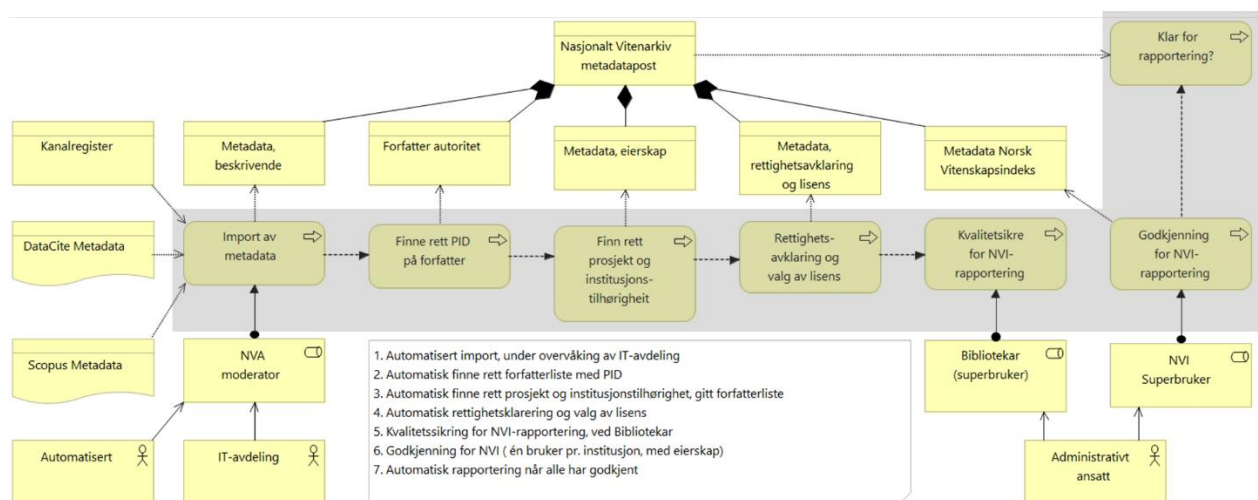
Her er vist to eksempler på arbeidsprosesser i NVA som viser hvordan disse forholder seg til og utnytter informasjonsarkitekturen beskrevet i Figur 31 Konseptuell informasjonsarkitektur for registreringer i NVA.

Eksemplene er laget vha. Archimate som er en standardisert notasjon for modellering og analyse av ulike aspekter innenfor virksomhetsarkitektur. Symbolene benyttet i figurene nedenfor har følgende betydning:



Import og registrering av publikasjon fra ekstern kilde

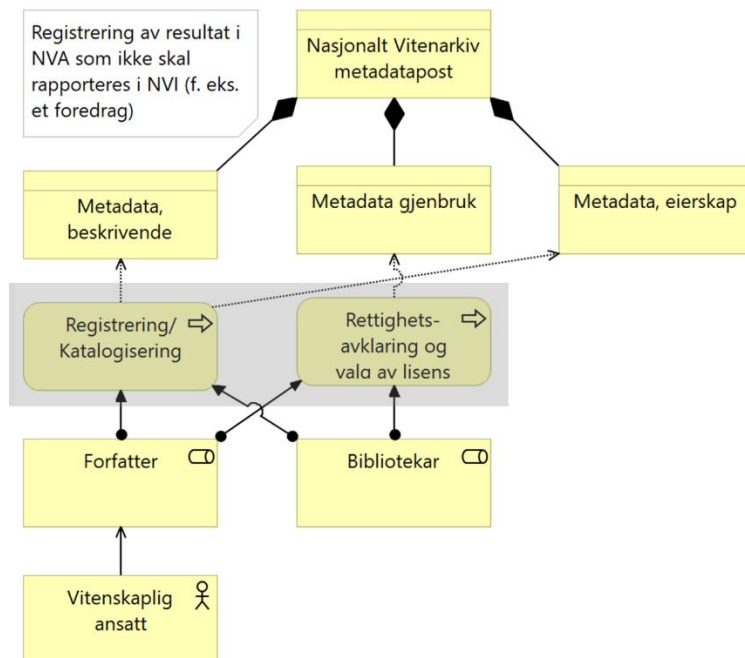
Skissen i Figur 35 nedenfor viser prosess-stegene (markert med grå skygge) som er involvert i forbindelse med import av data fra ekstern kilde, her eksemplifisert med DataCite og Scopus. Her ser en hvordan ulike aktører som innehar en bestemt rolle utfører ulike steg i prosessen, og hvilke metadata-subsett som det opereres på i de ulike stegene.



Figur 35 Arbeidsprosess for import og registrering av publikasjon i NVA

Manuell registrering av resultat som ikke skal rapporteres i NVI

Figur 36 nedenfor viser eksempel på manuell registrering av et resultat som ikke skal rapporteres i NVI, f. eks. et foredrag.



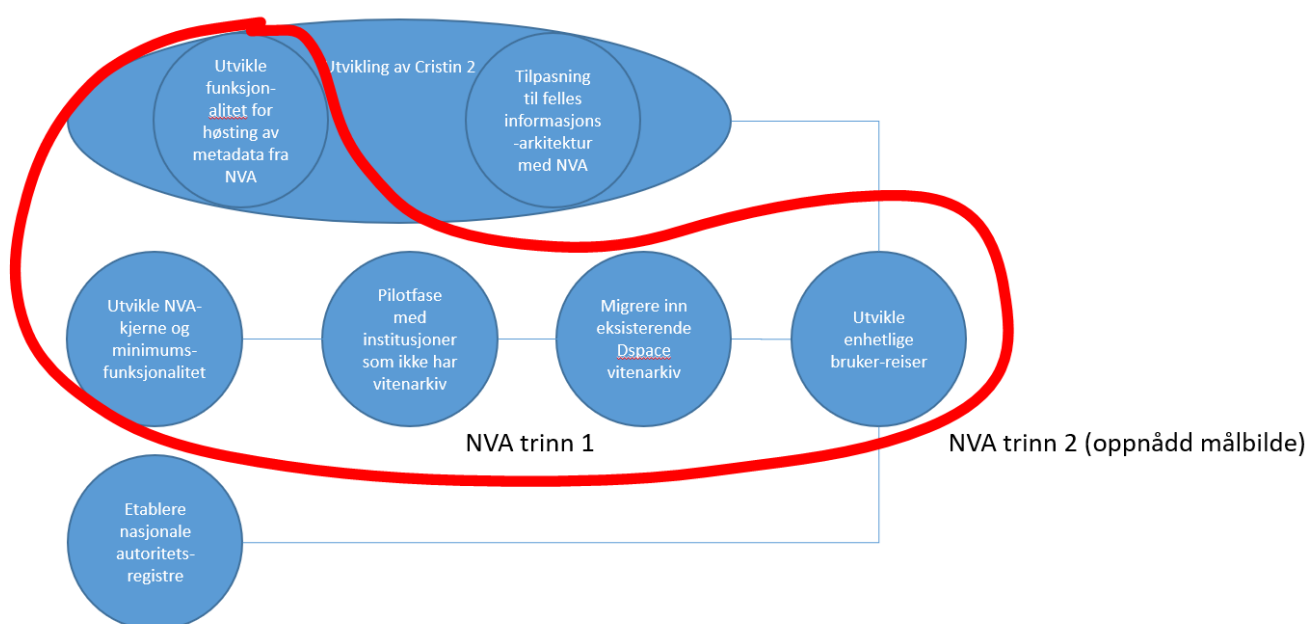
Figur 36 Manuell registrering av resultat i NVA som ikke skal rapporteres i NVI

Vedlegg 2 – Kostnadsestimater og for realisering av NVA

I figur 32 i kapittelet som beskriver Forslag til plan for gjennomføring, er det skissert en plan for hvordan NVA kan realiseres, basert på anbefalingene i rapporten. Dette vedlegget oppsummerer estimerte kostnader for å realisere NVA i ulike faser med ulike deloppgaver iht. inndelingen i figur 32. I estimatene er figuren benyttet for å vise ulike faser av utviklingen, og hva kostnadene til disse er estimert til. Det må poengteres at estimatene har betydelig usikkerhet på nåværende stadium, men vil være en indikasjon. Usikkerheten er synliggjort ved at alle kostnadsestimater er gruppert i henhold til et optimistisk, et sannsynlig, og et pessimistisk estimat.

Utvikling av NVA, trinn 1

NVA trinn 1 består av deloppgavene innenfor det rødmargerte området i figuren nedenfor. I estimatene har vi også tatt med drifts- og vedlikeholdskostnader for NVA trinn 1, inklusive lisenskostnader til applikasjonsdrift i skyen.

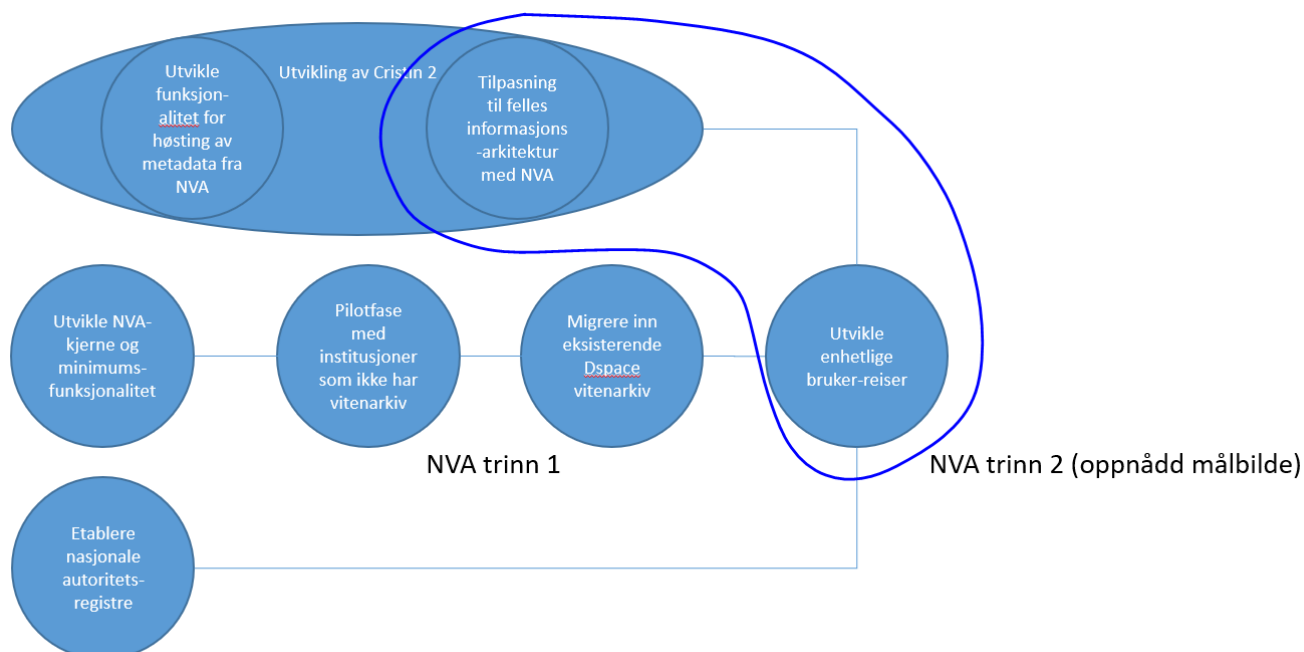


Estimerte kostnader for utvikling av NVA, trinn 1	Optimistisk	Sannsynlig	Pessimistisk
Utvikling av NVA, trinn 1, inklusive: - integrasjon med Cristin for replikering av metadata fra NVA til Cristin - prosessanalyse av arbeidsflyt knyttet til lokale vitenarkiv og etablering av en ny, beste praksis arbeidsflyt tilpasset NVA, trinn 1.	5,5 MNOK	9,5 MNOK	18 MNOK
Årlige drifts- og vedlikeholdskostnader for NVA, trinn 1	0,25 MNOK	0,5 MNOK	1 MNOK
Opsjon:			
Ta i bruk maskinlæring/kunstig intelligens ifm. med høsting og registrering i NVA	0,85 MNOK	1,5 MNOK	2,8 MNOK

Videreutvikling av NVA fra trinn 1 til trinn2

Dette innebærer i hovedsak endringer i Cristin ved at Cristin må tilpasse seg informasjonsarkitekturen/datamodellen som er realisert i NVA, trinn 1, og flyttes over på den samme plattformen som NVA kjører på. Lisenskostnader i NVA etter flytting av Cristin over i NVA forventes å opprettholdes på omtrent samme nivå som for NVA, trinn 1. Dette fordi datamengden ikke vil øke nevneverdig ift. NVA, trinn 1 da Cristin benytter de samme (meta-) dataene som allerede eksisterer i NVA. Dagens driftskostnader for Cristin hos USIT vil forsvinne når Cristin flyttes over på NVA-plattformen.

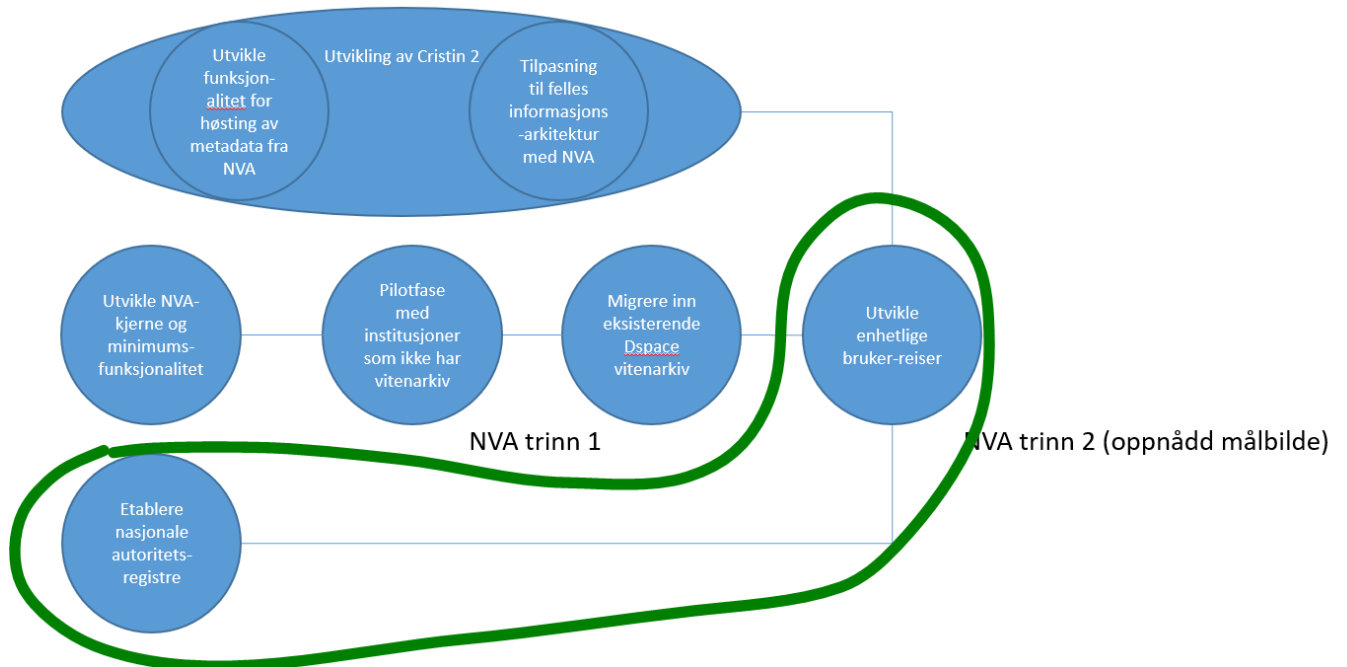
Deloppgavene som inngår er markert med blått i figuren nedenfor.



Estimerte kostnader for utvikling av NVA fra trinn 1 til trinn 2	Optimistisk	Sannsynlig	Pessimistisk
Utviklingskostnader (Cristin), inkl. prosessanalyse og etablering av en ny, beste praksis arbeidsflyt tilpasset NVA, trinn 2.	5 MNOK	8 MNOK	15 MNOK

Integrering av NVA med nye autoritetsregistre etter hvert som disse blir etablert

Når autoritetsregistrene som er omtalt i denne rapporten blir etablert gjennom andre prosjekter, vil NVA integreres mot disse. Integrasjon med disse registrene vil kunne bidra til å realisere de gevinstene som er beskrevet i rapporten ved å etablere og ta i bruk disse registrene. I figuren er de deloppgavene som inngår markert med grønt.



Estimerte kostnader for integrasjon med nye autoritetsregistre	Optimistisk	Sannsynlig	Pessimistisk
Integrasjon med autoritetsregister for Prosjekt	200 000 NOK	350 000 NOK	1 MNOK
Integrasjon med autoritetsregister for Søknad	150 000 NOK	200 000 NOK	400 000 NOK
Integrasjon med autoritetsregister for Organisasjonsenhet	150 000 NOK	200 000 NOK	400 000 NOK
Integrasjon med autoritetsregister for Person	150 000 NOK	200 000 NOK	400 000 NOK

Forslag til tidsplan for realisering

Figuren nedenfor viser forslag til tidsplan for realisering av de ulike deloppgavene beskrevet ovenfor, med antatt oppstart fom. tredje kvartal 2019. Integrasjoner med autoritetsregistrene vil utvikles etter hvert som disse blir etablert og tilbyr integrasjonsgrensesnitt.

