

Rapport fra utredning av nasjonalt vitenarkiv

Kortversjon



Direktoratet
for IKT og fellestjenester
i høyere utdanning
og forskning

Innledning

Kunnskapsdepartementet (KD) har bestilt en utredning av hvordan et nasjonalt vitenarkiv skal realiseres. Oppdraget kommer i kjølvannet av «Nasjonale mål og retningslinjer for åpen tilgang til vitenskapelige artikler»¹ vedtatt og publisert av regjeringen 22. august 2017. Til målsetningene er det knyttet noen tiltak for at målene oppnås. Ett av tiltakene var: «Utrede hvordan et nasjonalt vitenarkiv skal realiseres».

Vitenskapelige artikler inngår i en verdikjede for forskning hvor resultater blir kvalitetssikret ved fagfelleevaluering før publisering i et vitenskapelig tidsskrift. Så langt har en vesentlig andel av de vitenskapelige artiklene kun vært tilgjengelige bak betalingsmurer. Etter sterkt press fra offentlige myndigheter og forskningsfinansiører om at resultater av offentlig finansiert forskning må gjøres åpent tilgjengelige for alle, er denne modellen under sterkt press. Spesielt etter lansering av «Plan S» høsten 2018.

Et resultat av betalingsmurene er at institusjonene som bidrar til vitenskapelige artikler også publiserer disse i sitt lokale vitenarkiv. Den versjonen av artikkelen som kan publiseres i det lokale vitenarkivet vil som regel være en annen enn den som er publisert i tidsskriftet. Det er også vanlig at tidsskriftet ikke tillater publisering i lokalt vitenarkiv før en gitt tid etter publiseringen i tidsskriftet.

I Norge er det etablert en ordning hvor offentlige midler til forskning fordeles på de institusjonene som bidrar i forskningen etter et resultatbasert prinsipp (RBO-ordningen²). Dette innebærer at institusjonene må rapportere sin vitenskapelige produksjon til Norsk vitenskapsindeks (NVI) for å få uttelling.

Et vitenarkiv omfatter mer enn vitenskapelige artikler publisert i tidsskrift. 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 følgende informasjon hører hjemme i et fremtidig NVA:

- Resultater i form av artikler, kapitler og bøker/rapporter som inngår i rapporteringen til NVI
- Resultater i form av artikler, kapitler og bøker/rapporter som ikke inngår i rapporteringen til NVI
- Resultater av kunstnerisk utviklingsarbeid
- Studentoppgaver
- Forskningsdata (datasett)³
- Formidling
- Læringsobjekter

Åpen vitenskap dekker også åpen fagfelleevaluering, åpen forskningsmetode og generell åpen formidling av forskningsresultater til allmenheten.

¹ <https://www.regjeringen.no/no/dokumenter/nasjonale-mal-og-retningslinjer-for-åpen-tilgang-til-vitenskapelige-artikler/id2567591/>

² Resultatbasert omfordeling hvor institusjonene tildeles midler basert på størrelsen på deres vitenskapelige produksjon. Dette rapporteres inn til Norsk Vitenskapsindeks (NVI).

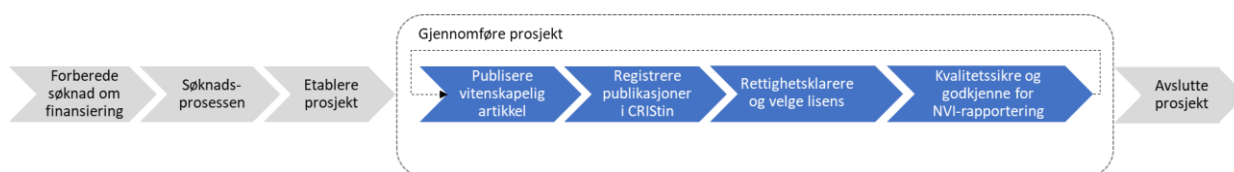
³ NVA vil primært gjøre forskningsdata tilgjengelig gjennom referanse til eksterne infrastrukturer for forskningsdata hvor dataene er fysisk lagret.

Motivasjonen for å inkludere forskningsdata⁴ i NVA ligger i at dette er grunnlaget for den vitenskapelige publikasjonen og derfor bør gjøres tilgjengelig sammen med publikasjonen. Åpne forskningsdata vil også lette gjenbruk i videre forskning og publisering. Åpen vitenskap har også et sterkt fokus på å trekke allmennheten inn i forskningsprosessen (*Citizen science*), både ved å bidra i forskningen⁵ men primært for at befolkningen skal få sterkere tillit til forskerne og konklusjonene fra forskningen. Formidling, læringsobjekter og til dels studentoppgaver er gjerne enklere å benytte seg av for allmenheten enn mange av de vitenskapelige publikasjonene.

Vår tilnærming til arbeidet med utredningen

Da arbeidet med denne utredningen startet, innså prosjektgruppen at denne måtte se NVA i en større sammenheng, ellers kunne NVA risikere å bli enda en ny tjeneste i sektoren. Visjonen til NVA er at aktørene som inngår eller berøres av verdikjeden til NVA skal oppleve forbedringer ved at arbeidet blir enklere, at datakvaliteten blir høyere, at tilgangen til resultatene blir bedre og at gjenbruken øker. I tillegg har det selvsagt vært en klar målsetting at realisering av NVA skal gi samfunnsøkonomiske gevinster for sektoren og staten Norge.

Verdikjeden prosjektet har forholdt seg til har fokus på registrering og tilgjengeliggjøring av forskningsresultater, samt nødvendige administrative rutiner ved RBO. En prosjekt-sentrert skisse av denne verdikjeden er vist i figuren under. Hvis et forskningsresultat ikke er et resultat av en aktivitet som kan defineres som et prosjekt, vil verdikjeden vil se litt annerledes enn hvis forskningen foregår i en prosjektkontekst. Men delprosessene markert med blått i figur 1 vil være de samme.



Figur 1 Verdikjeden for forskningsprosessen og formidling av forskningsresultater

Problemstillinger knyttet til dagens arkivlandskap

Prosjektet har foretatt en relativt grundig kartlegging⁶ av dagens arkivlandskap for å identifisere problemstillingene knyttet til dette. I denne sammenhengen er det viktig å forstå hvordan kostnadskjeden for offentlig finansiert forskning ser ut: Staten betaler først for selve forskningsarbeidet, og også for det arbeidet forskerne gjør gjennom fagfelleevaluering av artikler for tidsskriftene. Deretter betaler staten for kostbare abonnementer for tilgang til de samme tidsskriftene.

En vesentlig del av problemstillingene i dagens arkivlandskap er knyttet til prosesser som skjer etter at en vitenskapelig artikkel er publisert i et tidsskrift. Publisering i åpne tidsskrift vil kun i beskjeden grad fjerne eller redusere disse problemstillingene så lenge det er nødvendig å registrere

⁴ «Regjeringen vil be det nye organet for høyere utdannings- og forskningstjenester (dvs. Unit) ta ansvar for å utrede hvordan et nasjonalt vitenarkiv skal realiseres» Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata på side 10.

⁵ Gjerne med store samtidige datainnsamlinger, som for eksempel <http://artobservasjoner.no> i regi av Artsdatabanken.

⁶ Kartleggingen er gjort gjennom en interessentsamling, dybdeintervjuer med ulike aktører og studier av dokumentasjon om de tjenestene som er de mest sentrale i dag.

vitenskapelige resultater ifm. RBO-ordningen, eller så lenge det er nødvendig å egenarkivere publikasjoner og andre resultater.

Vår kartlegging viser at prosessen med å registrere resultater i Cristin og gjøre dem åpent tilgjengelige i de lokale vitenarkivene har et stort effektiviseringspotensial.

De viktigste funnene knyttet til problemstillinger er:

1. Tjenester som inngår i verdikjeden for forskning er i stor grad silo-baserte og det er lite flyt av data mellom de ulike tjenestene. For NVA handler det i stor grad om at metadata som oppstår i verdikjeden frem til registrering/tilgjengeliggjøring av resultat skal knyttes til resultatene som registreres i Cristin og gjøres tilgjengelige i vitenarkiv. Så lenge eksisterende metadata foreligger i utenforliggende tjenester uten noen integrasjoner, må disse gjenskapes manuelt i registreringsøyeblikket. Dette gir betydelig merarbeid for hver registrering. Og siden volumet er høyt vil dette både gi store kostnadsmessige konsekvenser og økt risiko for dårlig datakvalitet.
2. Arbeidsprosessene knyttet til henholdsvis registrering av resultater og prosesser for kvalitetssikring av nye registreringer mht. NVI-rapportering i Cristin, og rettighetsklarering og valg av lisens i lokale vitenarkiv, er lite enhetlige og utføres av ulike roller ved institusjonene. Vi savner en enhetlig, beste praksis for disse arbeidsprosessene.
3. Informasjonsstrømmen i dag, hvor vitenskapelige artikler registreres og lastes opp i Cristin (manuelt eller via import), og deretter distribueres til lokale vitenarkiv ved de institusjonene som har forfattere på artikkelen, gir unødig ekstraarbeid. Hver enkelt institusjon får en kopi av metadatapost og tilhørende fil og må utføre samme jobb med rettighetsklarering, valg av lisens og evt. beriking av metadata – på hver sin kopi! Det kan føre til forskjellige versjoner. I tillegg utfører også hver av de involverte institusjonene arbeid på metadataposten i Cristin for å sikre korrekt rapportering til NVI og riktig uttelling mht. RBO. Dette arbeidet utføres av administratorer i de lokale vitenarkivene og av Cristin-superbrukere som ved noen institusjoner er en og samme person.
4. Dagens arkivlandskap er svært sammensatt med mange og til dels overlappende tjenester som gir høye kostnader til drift og vedlikehold.

Målsetninger for nasjonalt vitenarkiv

Basert på innspill fra ulike aktører (roller) knyttet til den delen av forskningsprosessen som omhandler forvaltning og tilgjengeliggjøring av resultater, og de identifiserte problemstillingene beskrevet ovenfor, har vi definert fire hovedmål for NVA:

1. Oppnå brukervennlige og effektive arbeidsprosesser for registrering og tilgjengeliggjøring av forskningsresultater.
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.
3. Støtte nødvendige metadata for NVI-rapportering for alle resultater som gir uttelling i RBO-ordningen.
4. NVA skal sikre at alle registrerte resultater i NVA som faller inn under loven om pliktavlevering, gjøres tilgjengelig for avlevering til Nasjonalbiblioteket.

Det er et indirekte mål at NVA skal ha tilstrekkelig fleksibilitet til å kunne understøtte nødvendige prosessendringer og nye tjenesteleveranser som måtte vise seg nødvendige for at åpen forskning skal kunne bli realisert.

To anbefalte hovedtiltak

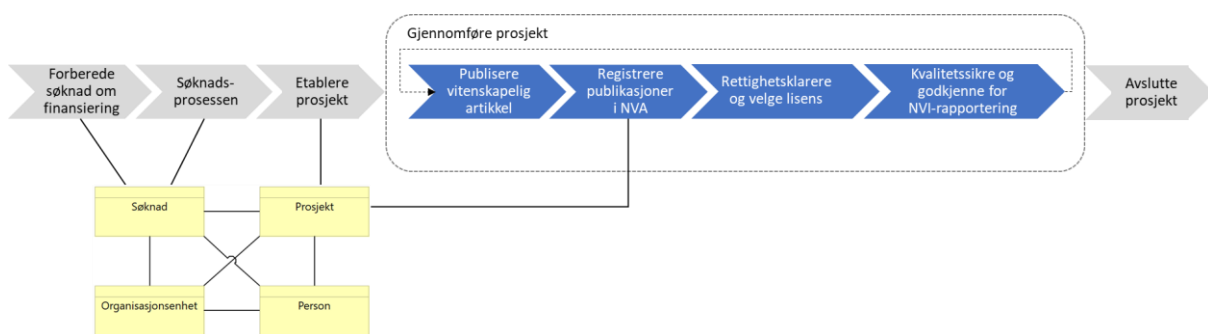
Problemstillingene beskrevet ovenfor har klare årsakssammenhenger. Det er fullt mulig å fjerne eller kraftig redusere problemene, men det vil kreve noen relativt drastiske tiltak å oppnå dette. Rapporten viser at det er et stort gevinstpotensial i å minimalisere de påpekte problemstillingene. Det er identifisert to overordnede tiltak som i sterk grad vil redusere problemene. Det er ingen avhengighet mellom disse hovedtiltakene så de kan innføres hver for seg og hver på sin måte bidra til å realisere påpekte gevinster.

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.

Et nasjonalt autoritetsregister er masterdatakilde på nasjonalt nivå⁷ for en gitt entitet⁸ som f. eks. person eller organisasjon.

I Figur 2 nedenfor illustreres hvordan disse autoritetsregistrene (symbolene i gult) lagrer metadata som genereres i arbeidsprosessen og muliggjør gjenbruk av metadata fra arbeidsprosesser senere i verdikjeden. Linjene i figuren illustrerer hvordan data utveksles.



Figur 2 Samspill mellom autoritetsregistrene og arbeidsflyt i forskningsprosessen

⁷ Kan være begrenset til en spesifikk sektor, eller på tvers av sektorer.

⁸ Entitet i denne konteksten betyr en «container» for lagring av en spesifikk type dataobjekt (f. eks. prosjekt).

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.
- Omfanget av arbeid med retting av metadata vil reduseres.

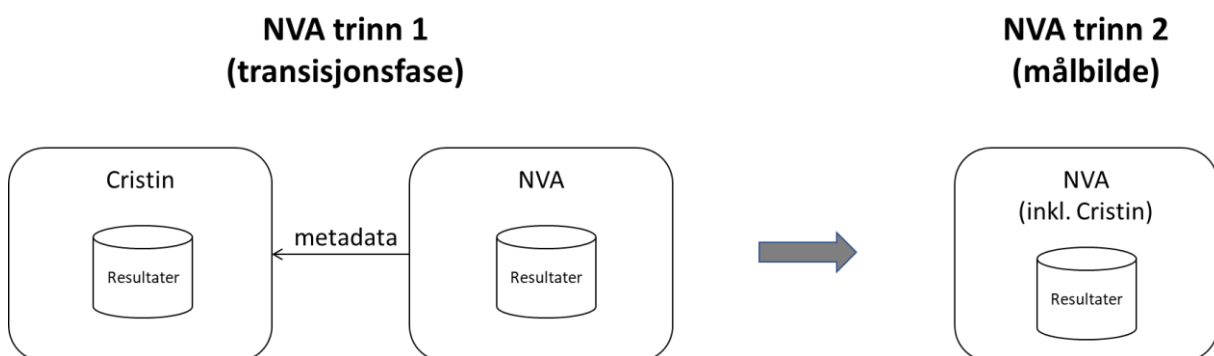
Kvalitative gevinster:

- 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.

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

Prosjektet har vurdert ulike alternativer for en ny infrastruktur som kan sikre best mulig koordinering av aktiviteter og data mellom Cristin og de lokale vitenarkivene når disse erstattes av NVA. Prosjektets anbefalte målbilde for NVA er at NVA blir en felles plattform for Cristin og vitenarkivene som opererer på ett felles datagrunnlag (referert til som NVA trinn 2 i Figur 3 nedenfor). Dette målbildet forutsetter at det etableres en informasjonsarkitektur for NVA 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.

Det pågående arbeidet med Cristin 2 er kritisk for å kunne få en fullstendig erstatning for Cristin 1. Det gjør at det er vanskelig å gå rett på dette målbildet. For ikke å forstyrre arbeidet med Cristin 2 i særlig grad, anbefales det i første omgang å utvikle NVA i henhold til NVA trinn 1 i Figur 3 nedenfor. I dette trinnet etableres det som er kalt NVA-kjernen (boksen til høyre i figuren for NVA Trinn 1). NVA er her en fellestjeneste for vitenarkiv med ett felles datagrunnlag for alle institusjoner basert på den anbefalte informasjonsarkitekturen.



Figur 3 Trinnvis overgang til nytt målbilde

Dette hovedtiltaket innebærer at en unngår duplisering av registreringer på tvers av vitenarkiv (trinn 1) og mellom Cristin og vitenarkiv (trinn 2). 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. Tiltaket gir grunnlag for store gevinster i form av spart tidsbruk tilsvarende mange årsverk. Tiltaket vil også drastisk redusere antall tjenester (eller instanser av disse) som i dag dekker samme funksjonelle behov.

Gevinstpotensial

Gevinster som muliggjøres med dette tiltaket er:

Direkte budsjettmessige gevinster:

- Med denne arkitekturen forsvinner dagens lokale vitenarkiv 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.

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.

Konsekvenser av hovedtiltak 2

Dette hovedtiltaket medfører at det etableres en plattform for NVA med basis i den anbefalte informasjonsarkitekturen og at det utvikles funksjonalitet som tilfredsstillende behøver til vitenarkiv. Prosjektet har vurdert følgende fire alternative måter å etablere NVA på:

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 gevinster⁹.
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.

⁹ Dette bør gjøres uavhengig av hvilken løsning som velges.

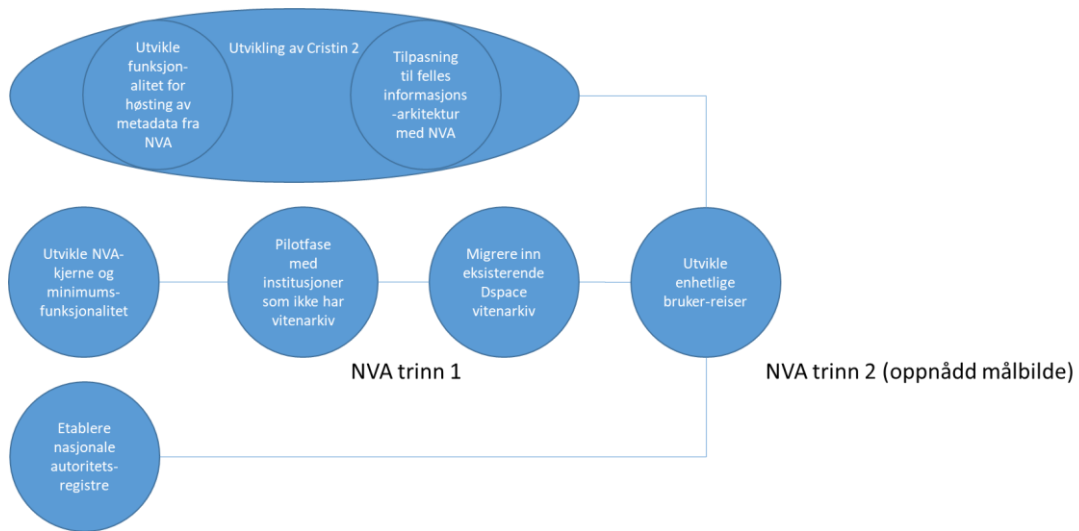
Prosjektet har i rapporten vurdert disse alternativene opp mot hverandre med basis i styrker og muligheter kontra svakheter og risiki. Anbefalt løsning er alternativ 2. Prosjektets vurderinger av dette alternativet er oppsummert i tabellen nedenfor.

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 i Unit.

Veien videre

Dersom anbefalingene i rapporten vedtas, bør utviklingen av NVA startes raskt.

En tentativ plan for realisering og innføring av NVA er skissert i figuren nedenfor.



Figur 4 Tentativ plan for realisering og innføring av NVA