

Saksdokument

Porteføljestyre for utdanning og administrasjon / forskning og kunnskapsressurser / data og infrastruktur

Til Porteføljestyre for forskning og kunnskapsressurser

Dato 01.06.2023

Saksnummer 5/23-A

Saksarkivnr. 23/00000

Sakstype O/D-sak

Saksansvarlig Vigdis Kvalheim

Saksbehandler Frode Arntsen

Presentasjon av Nasjonalt vitenarkiv (CRIS/NVA)

Forslag til vedtak

Porteføljestyret tar saken til orientering og ber om at de innspill som kom i gjennomgang av saken inngår i det videre arbeidet knyttet til risikovurdering og veien videre.

Bakgrunn

Saken presenterer prosjektet *Nasjonalt vitenarkiv (Cris/NVA)* og gir status pr medio mai 2023. Den nye tjenesten skal erstatte dagens Cristin og de institusjonelle vitenarkivene (70+) i løpet av november 2023.

Regjeringen etablerte i 2017 retningslinjer for åpen tilgang til forskning. Med utgangspunkt i retningslinjene ga Kunnskapsdepartementet i 2018 Unit som oppdrag å utrede hvordan en skal realisere et felles nasjonalt vitenarkiv. Unit leverte utredningen ved årsskiftet 2018/19 og fikk i august 2019 i oppdrag å etablere et nasjonalt vitenarkiv i tråd med anbefalingene i utredningen. Arbeidet startet høsten 2019.

Prosjektet Nasjonalt vitenarkiv (Cris/NVA) er et av virkemidlene/satsinger fra regjeringens side for å realisere ambisjonen om åpen forskning. Cris/NVA bidrar til åpen forskning ved å gjøre informasjon om og resultatene av offentlig finansiert forskning åpent tilgjengelig i en felles nasjonal løsning som samler innholdet fra 75 eksisterende institusjonelle vitenarkiv og det nasjonale forsknings-informasjonsystemet Cristin på en felles, skybasert plattform. Det vil potensielt være ca 200 brukerinstitusjoner på tjenesten inkludert UH- og instituttsektoren samt de regionale helseforetakene.

Overgangen fra institusjonelle vitenarkiv til et felles nasjonalt vitenarkiv vil i betydelig grad redusere mengden arbeid som utføres i dag. For eksempel vil arbeidet med de vitenskapelige publikasjonene som er sampublikasjoner med forfattere fra flere enn en institusjon reduseres betraktelig. Totalt tilføres dagens Cristin ca 100000 metadatatposter årlig der ca 30000 er vitenskapelige publikasjoner, hvorav



de fleste av disse er sampublikasjoner. Resterende poster i Cristin registreres oftest av de vitenskapelig ansatte selv, en prosess som vil bli enklere i den nye Cris/NVA-tjenesten enn i dagens systemer.

Gevinsten for institusjonene så vel som brukerne som benytter dagens vitenarkiv og Cristin til å finne informasjon og resultater, vil primært være at all informasjon om- og resultater fra norsk forskning er tilgjengelig på ett felles sted. Dermed legges det til rette for at den enkelte institusjon kan automatisere arbeidsprosesser og koordinere arbeidsoppgaver med andre og på den måten oppnå besparelser. Ved å ha alle data samlet på ett sted legges det også bedre til rette for å dele/gjenbruke data og evt gjøre analyser på tvers.

Gjennomgående bruk av persistente identifikatorer (ORCID/DOI/ROR) vil også lette forskernes bruk av systemet i betydelig grad ved å redusere behovet for gjentatte administrative oppgaver. Persistente Identifikatorer (PID) brukes for å koble informasjon. For eksempel kan en publikasjon eller et datasett (med DOI) kobles til en forfatter som har en ORCID og derved redusere problemer knyttet til for eksempel forskjellige måter å skrive/stave forfatternes navn på. PID-ene blir en del av metadataene til publikasjonene eller datasettene og benyttes videre for å koble sammen forfattere og data/resultater. Det benyttes også PIDer for å identifisere organisasjoner (ROR) og Sikt er i samarbeid med Forskningsrådet i ferd med å prøve ut bruk av DOIer på Forskningsrådets finansieringer. Forskningsprosjekter vil også ha en PID, og på denne måten knyttes alle relevante entiteter sammen på en enkel måte.

Cris/NVA-tjenesten benytter allerede PIDer på personer, publikasjoner, datasett og organisasjoner. Den enkelte bruker kan enkelt, dersom de ikke allerede har registrert seg for en ORCID, registrere seg for ORCID ved pålogging til Cris/NVA. Tilsvarende kan institusjonene utstede DOIer på datasett og publikasjoner som publiseres av institusjonen selv. ROR (OrganisasjonsID) registreres ved opprettelse av organisasjonen i Cris/NVA. PID på forskningsprosjekter og PID knyttet til finansieringer vil implementeres på et senere tidspunkt.

Ved å samle nødvendig funksjonalitet fra flere forskjellige løsninger til en felles nasjonal løsning, legges det også til rette for mer kostnadseffektiv forvaltning av systemene. Ved å benytte en moderne og fremtidsrettet skyløsning som plattform for utvikling av Cris/NVA, sikrer dette en kontinuerlig oppdatert og sikker løsning som legger godt til rette for videre utvikling av eksisterende og ny funksjonalitet på Cris/NVA-tjenesten

Cris/NVA-prosjektet finansieres i hovedsak med bevilgninger over statsbudsjettet. Kunnskapsdepartementet bevilget 30 MNOK fordelt over 2019-2021 øremerket til Nasjonalt vitenarkiv. I tillegg er det benyttet bevilgninger som ellers ville vært benyttet til videre planlagt utvikling av Cristin for årene 2020-2023 og ellers midler fra Sikt sin risikokapital.

Organisering

Cris/NVA-prosjektet ledes av en styringsgruppe med representanter fra UiO, UiB, UHR, FFA, NFR, og HK-Dir. Denne gruppen tar de formelle beslutningene for prosjektet, herunder beslutning om migrering av data knyttet til institusjoner som i dag benytter Cristin. Tidligere *Fagutvalg for forskning* og *Tjenesteråd for forskning* (Begge er avviklet) ble holdt jevnlig orientert om prosjektets fremdrift og relevante saker ble drøftet med spesielt Tjenesterådet. Porteføljestyret vil på samme måte som Fagutvalget og Tjenesterådet bli jevnlig involvert i forhold til prosjektets fremdrift og innføring.

Digitaliseringsstyret har vært jevnlig orientert om prosjektet og ba i sitt siste møte den 28 april Sikt komme tilbake til møtet i juni med en risikovurdering knyttet til tid, kostnad, opplæringsplan, innføring og funksjonalitet.

Prosjektet har en omfattende og løpende brukerinvolvering gjennom referansegruppe og flere arbeidsgrupper som behandler spesifikke funksjonsområder. I tillegg ble det etter dialog med UHR/FFA opprettet en overordnet arbeidsgruppe sommeren 2022. Spørsmålet om hva som er nødvendig funksjonalitet ble utdypende behandlet av UHR/FFA/Sikt-arbeidsgruppen. Arbeidsgruppens [rapport](#) legger til grunn at de leveranser som er planlagt, sikrer at samtlige Cristin-institusjoner kan gå over til det nye systemet når nødvendig funksjonalitet er på plass og det er gjennomført ende til ende test av systemet.



Status fremdrift

Prosjektets opprinnelige plan fra 2018/19 skisserte et overordnet forløp som ville gjennomføre prosjektet frem til ferdigstilling sommeren 2022. Med et mål om at dagens løsninger erstattes av den nye tjenesten i løpet av november 2023 er det forståelse for at det er noe uro for forsinkelser og hva det kan medføre for omlegging ved hver institusjon. NVA vil treffe forskerne direkte, om dette ikke fungerer er fallhøyden stor. Prosjektet har i store deler av sitt forløp hatt utfordringer med tilgang på IT-utviklere og samtidig en skjev balanse mellom senior- og juniorutviklere. Det har vært et svært anstrengt arbeidsmarked og betydelige utfordringer med både å rekruttere IT-utviklere og å finne egnede konsulenter som er tilgjengelige. Resultatet av denne situasjonen er at det har tatt lenger tid enn forutsett å realisere den funksjonaliteten som må være på plass før en kan bytte fra dagens løsninger. De siste månedene har prosjektet klart å innlemme nye utviklere samtidig som eksisterende utviklere kontinuerlig blir mer kompetente. Ved utgangen av 1. tertial 2023 er prosjektet tilfredsstillende bemannet med nødvendig kapasitet på IT-utvikling. Prosjektet er imidlertid også avhengig av flere andre kompetanseområder som fortsatt gir høy risiko i forhold til fremdrift ved for eksempel sykdom hos prosjektmedarbeidere.

Store deler av planlagt funksjonalitet er nå på plass og det som ikke er testet tidligere vil bli gjenstand for strukturert brukertesting i første del av juni mnd. De to områdene som ikke er klar for testing i denne omgang er import av publikasjonsdata (Også omtalt som sentralimport) og NVI Kontroll og godkjenning. Begge områder er imidlertid langt på vei i utviklingen og for import av publikasjonsdata vil det meste være på plass før sommerferien i demo/test. Til forskjell fra dagens importfunksjonalitet, som er en separat tjeneste integrert med Cristin, vil den nye importtjenesten være tett integrert med øvrig funksjonalitet i Cris/NVA og vil som en konsekvens gi raskere import og bedre kvalitet på publikasjonsdata i forhold til dagens løsninger.

Utviklingen av funksjonalitet for NVI Kontroll og godkjenning vil i hovedsak replisere dagens løsninger i Cristin, med enkelte planlagte forbedringer. Utviklingen av denne funksjonaliteten er godt i gang og kan muligens foreligge i demo/testversjon før sommerferien.

Det er imidlertid viktig å huske at den nye Cris/NVA-tjenesten (<https://nva.sikt.no>) allerede benyttes av 17 institusjoner, i hovedsak institusjoner som ikke har hatt vitenarkiv tidligere. Den første institusjonen (NVE) er allerede flyttet fra et eksisterende vitenarkiv (Brage) og over til det nye nasjonale vitenarkivet. Cris/NVA fungerer nå godt som et felles nasjonalt vitenarkiv.

Det gjennomføres jevnlig prøvemigreringer av alt innhold i dagens Cristin til en utviklingsversjon av Cris/NVA. Det er avdekket behov for å justere/endre importfunksjonaliteten for å sikre at den «endelige» importen forløper feilfritt. Delvis handler dette om hvordan vi skal håndtere poster som foreligger både i vitenarkivene og i Cristin for de ca 70 institusjonene som i dag benytter begge tjenestene, delvis også justeringer av funksjonaliteten for å redusere tiden det tar å migrere ca 2 mill poster fra Cristin.

Status innføring

Høsten 2022 ble det etablert et delprosjekt for innføringsaktiviteter som omfatter både testing og opplæring i tillegg til løpende kommunikasjon rundt prosjektet i form av nyhetsbrev og webinarer.

Testing av Cris/NVA-tjenesten skjer delvis ustrukturert ved at ca mange institusjoner har tilgang til testversjonen av Cris/NVA, hvor flere gir løpende tilbakemeldinger på funksjonalitet. Så langt er det gjennomført en omfattende strukturert brukertest på et avgrenset funksjonsområde som fikk gode tilbakemeldinger. Det er nylig gjennomført webinar som hadde hovedfokus på funksjonalitet for arbeidsflyt i Cris/NVA. Denne funksjonaliteten vil være klar for strukturert brukertesting i første del av juni mnd. Det planlegges ytterligere 1-2 webinarer før sommerferien og opptak fra tidligere webinarer er tilgjengelig på Sikt sine nettsider under [Webinarer og svar på spørsmål](#) ved å følge lenken.

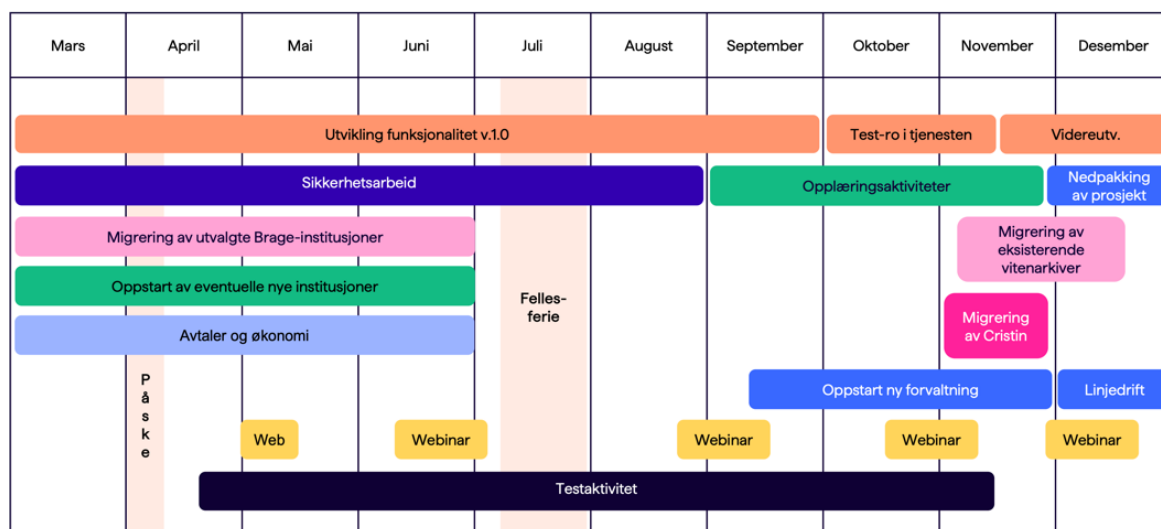


Før det besluttes at tjenesten skal settes i full drift skal det gjennomføres ende-til-ende testing av all tilgjengelig funksjonalitet, dette er planlagt i september/oktober 23.

Basert på brukernes innspill til opplæringsbehov og tidligere erfaringer utarbeides det opplæringsopplegg som skal gjennomføres høsten 2023. Utgangspunktet for opplæringen er at de institusjonene som har over et visst antall ansatte administrative brukere selv gjennomfører opplæring av de fleste ansatte, både administrative og vitenskapelige ansatte. Sikt vil gjennomføre nettkurs via Teams/Zoom for de som skal gjennomføre lokale kurs.

Opplæringen vil bli en kombinasjon av webinarer (åpent for alle), nettkurs (Teams/Zoom med begrenset deltagelse som muliggjør dialog, korte oversiktsvideoer og trinn-for-trinn veiledninger som er tilgjengelig på nett og i pdf-versjoner. Det vil gjennomføres et webinar i siste halvdel av juni der vi presenterer eksempel på utvalgte opplæringsopplegg for utvalgt funksjonalitet og roller. Basert på tilbakemeldinger på disse vil vi utarbeide endelig opplæringsmaterieell som også vil være tilgjengelig for den enkelte institusjon til eget bruk.

Nedenfor er vist en oversikt over de aktiviteter som er planlagt i prosjektet frem mot full overgang til ny løsning. Det vil foreligge en mer konkret plan i løpet av juni mnd, der det oppgis konkrete uker og i noen tilfeller spesifikke datoer for de forskjellige aktivitetene.



Tidslinje for sentrale aktiviteter i Cris/NVA-prosjektet

Rapportering av Norsk vitenskapsindeks (NVI)

Kunnskapsdepartementet har foreslått at Publiseringsindikatoren som er nært knyttet til, og baserer seg på den årlige NVI rapporteringen, skal avvikles for universitet og høyskoler. For institutt- og helsesektoren planlegges det fortsatt med at NVI-rapportering og publiseringsindikatoren er grunnlag for finansiering. Kunnskapsdepartementet uttaler samtidig at det fortsatt skal gjennomføres årlige rapportering av vitenskapelige publikasjoner. Disse mulige endringene forventes ikke å påvirke utvikling og ferdigstilling av Cris/NVA.

Finansiering av Cris/NVA-tjenesten

Utvikling av Cris/NVA-tjenesten finansieres av Kunnskapsdepartementets bevilgninger og Sikt sin risikokapital. Kunnskapsdepartementet har besluttet at Cris/NVA tjenesten skal finansieres gjennom en kombinasjon av bevilgninger og brukerbetaling når den er i produksjon. Dette vil være med virkning fra 2024, og er i tidligere saker til Tjenesteråd og Digitaliseringsstyret foreslått gjort etter prinsippene som ligger i rammeverket for tjenestestyring (Vedtatt i Digitaliseringsstyret). Det er tidligere presentert en modell basert på institusjonenes FOU-årsverk i kombinasjon med fastpris/minstepris. Konkret planlegges det med en minstepris på størrelsesorden 25000 NOK for institusjoner som har 5 eller



færre FOU-årsverk og en glidende modell der prisen pr årsverk vil være avtagende med antall årsverk. Sikt har tidligere indikert at de institusjonene som i dag har vitenarkiv levert av Sikt vil få en kostnadsøkning på gjennomsnittlig 30% fra 2023 til 2024. For den største institusjonen forventes dette å utgjøre en økning på i underkant av 100.000 kroner.

Det vurderes alternative løsninger for deler av helse- og museumssektoren der en finner et stort antall institusjoner som har få FOU-årsverk. Sikt vil ta kontakt med henholdsvis Helse- og omsorgsdepartementet og Kulturdirektoratet for å avklare hensiktsmessige betalingsmodeller for disse. Vedlagte notat om finansiering av Cris/NVA er tidligere lagt frem for Tjenesterådet og Digitaliseringsstyret.

Risiki

Som det tidligere er gjort rede for har prosjektet over tid hatt utfordringer rundt tilstrekkelig kapasitet og begrenset seniorkompetanse blant IT-utviklere. Dette er essensielle kapasiteter for å lykkes med prosjektet. På nåværende tidspunkt vurderes det dithen at vi har tilstrekkelig antall personer, men fortsatt ikke en optimal balanse mellom junior og senior kompetanse. Det er imidlertid en god og positiv utvikling som synes å være leveransedyktig. Det er imidlertid fortsatt en betydelig sårbarhet dersom en eller flere prosjektmedarbeidere skulle bli utsatt for sykdom eller andre hendelser. Dette gjelder ikke kun IT-utviklere, men også andre deler av bemanningen.

Dette er prosjektets største risiko, og den følges nøye på daglig basis. En bredere risikoanalyse presenteres i Porteføljestyrets møte.

Vedlegg - Betalingsmodell



Vedlegg til Sak 5-23A Presentasjon av Cris/NVA - Betalingsmodell

Produktet/Tjenesten skal benyttes av institusjoner som utøver offentlig finansiert forskning, konkret innen universitets- og høyskolesektoren, instituttsektoren og helsesektoren. Det er anslagsvis ca 200 slike institusjoner totalt, der de fleste er svært små med et lite antall forskere. De data som legges inn i produktet handler i hovedsak om forskning og er i de fleste tilfeller kun beskrivende data (metadata). Mellom 20 - 25% av alle registreringer er vitenskapelige publikasjoner og inngår i Norsk vitenskapsindeks (NVI).

For de fleste av disse publikasjonene vil det også lastes opp tilhørende innholdsdata, normalt en versjon av en vitenskapelig artikkel. Innholdsdata kan også være for eksempel datasett eller kunstneriske utviklingsarbeider i form av lyd-, bilde- eller videofiler. I tillegg legger UH-sektoren også studentoppgaver (Bachelor og Masteroppgaver) inn i dagens vitenarkiv, noe som vil videreføres i Cris/NVA. Tilsvarende legger forskningsinstitutter normalt sine rapporter inn i vitenarkivet på tilsvarende måte. Innholdet i Cris/NVA vil normalt være produsert av forskere (eller studenter), og ofte vil ansatte i bibliotekene eller forskningsadministrasjonen forvalte metadata samt lisenser og publisering av innholdet.

For å avklare hvordan den delen av kostnadene knyttet til Cris/NVA som skal brukerfinansieres bør fordeles, er det naturlig å ta utgangspunkt i hvilke kostnadsdrivere som foreligger. I tillegg vil de relevante fordelingsmodellene som tidligere er behandlet av Digitaliseringsstyret og vist i rammeverksdokumentet [Forvaltning og videreutvikling av fellestjenester i høyere utdanning og forskning](#) som sist ble revidert i mai 2021, være et relevant grunnlag for hvordan kostnader skal fordeles. Dokumentets vedlegg 2 - *Prinsipper for enhetlige betalingsmodeller for fellestjenester* gjør rede for prinsippene for prising av fellestjenester samt kriterier for enhetlige betalingsmodeller. Generelt diskuteres omkring bruk av en fast kostnad uavhengig av bruk samt et eller flere element som enkelt kan relateres til faktisk bruk. For Cris/NVA-tjenesten vil de tydeligste kostnadsdriverne være faktisk bruk som reflekteres i prosessortid hos AWS samt hvor store mengder data den enkelte institusjon lagrer hos AWS. Det vil være vesentlig høyere kostnader knyttet til lagring av lyd- bilde- og videofiler sammenlignet med PDF-filer som er det vanligste for vitenskapelige artikler, rapporter og studentoppgaver.

Modellen for nye fellestjenester kan ha et fast element og følgende fire fordelingsparametere:

- Antall studenter (registrerte studenter som definert i DBH)
- Antall tilsatte (registrerte årsverk i DBH)
- Antall FoU-ansatte (registrerte årsverk innenfor undervisnings- og forskerstillinger som definert i DBH)
- Størrelse på bevilgning (inntekt fra bevilgninger som definert i DBH)

Med utgangspunkt i disse mulige fordelingsparameterene synes det mest naturlig å benytte antall FOU-ansatte, da de øvrige parametrene vanskelig lar seg sammenligne på tvers av de forskjellige forskningsutøvende sektorene. Det finnes offentlige, kvalitetssikrede tall for FOU-ansatte i de forskjellige sektorene.

Bruk av FOU-årsverk som grunnlag for kostnadsfordeling

Som et utgangspunkt er alle UH-institusjonene sammenlignbare i den forstand at de ansatte normalt er pålagt å forske samtidig som de underviser studenter som skriver oppgaver som skal arkiveres. Det varierer hvor stor andel av deres årsverk som benyttes til forskning og til undervisning, men en gjennomsnittlig 50/50 fordeling er et naturlig utgangspunkt. Når en sammenligner med forskere som er ansatt ved forskningsinstitutter der de i all hovedsak produserer vitenskapelige artikler og rapporter kan det stilles spørsmål ved om et FOU-årsverk i UH-sektoren er sammenlignbart med et forskerårsverk ved et forskningsinstitutt.

Det kan argumenteres med at UH-sektoren skal betale en relativt høyere andel av kostnadene fordi de også lagrer studentoppgaver, men samtidig er situasjonen at et FOU-årsverk ved en UH-institusjon kun benytter ca 50% av tiden til forskning og publisering, mens en ansatt ved et institutt benytter hele årsverket til forskning, utredninger og publiseringer. Den andelen en UH-ansatt benytter til



undervisning vil også telles med som FOU-årsverk og kan slik anses å ivareta studentoppgavens plass i Cris/NVA tjenesten.

FOU-ansatte i helsesektoren vil ofte også være både ansatt ved en UH-institusjon samtidig som de også er utøvende «praktikere» i helsesektoren. Det er erfaringsmessig krevende å gjøre en slik eksakt fordeling av hvor stor andel som er relevant i denne sammenhengen. Sikt har historisk benyttet tall for FOU-årsverk basert på det som er publisert i DBH samt av Forskningsrådet og for Helsesektoren det som er publisert i [SSBs rapport om Ressursbruk til forskning i helseforetakene](#) i det enkelte år.

Med dette som utgangspunkt anbefales at et FOU-årsverk teller likt uavhengig av hvilken institusjon og sektor en tar utgangspunkt i.

Antallet brukerinstitusjoner er i størrelsesorden 200, men det vil avhenge av hvordan de mindre institusjonene i både museumssektoren og helsesektoren regnes:

- Om samtlige blir ansett å være frittstående institusjoner, så er sannsynlig kundemasse i 2024 cirka 220 institusjoner.
- Om institusjonene som i dag er sammensatt i klynger (museene (40 stk), Private ideelle Helse Vest (8 stk) og Private ideelle Helse Øst (6 stk)) skal regnes som klynger, så er sannsynlig kundemasse i 2024 cirka 169 kunder.

Hvorvidt dette vil ha noen økonomisk konsekvens avhenger av hvilken direktefinansiering man eventuelt ender opp med for de ulike klyngene. Dette vil være avhengig av en dialog med de regionale helseforetakene (for helse) og Kulturdepartementet (for museene). Det antas å være uforholdsmessig kostbart om disse små skal faktureres som frittstående.

Brage-tjenesten (Institusjonelle vitenarkiv) har fram til i dag benyttet en trappetrinnsmodell bestående av seks trinn, basert på samme prinsipp hvor størrelse gir rabatter, hvor man basert på sine FoU-årsverk (og da til dels grove estimer av dette) plasseres på et trappetrinn med en gitt årlig kostnad. Dette oppleves å ha fungert fint i en årrekke, men det har gitt til dels uheldige utslag for spesielt de nesten største institusjonene som har plassert seg nært inntil grensene for forrige/neste trappetrinn. Dette vil gi både heldige og mindre heldige utslag, med ulik grad av hopp i kostnad, ved overgang til en glidende modell uten skarpe overganger mellom trinn.

Med grunnlag i en antagelse om at bruk av tjenesten øker lineært med antall FOU-ansatte og at lagring vil være relatert til antall FOU-årsverk foreslås følgende betalingsmodell for den brukerfinansierte delen av Cris/NVA-tjenesten.

- En andel i form av en fastpris som vil være en minstepris uavhengig av institusjonens størrelse, slik at institusjoner med et lavt antall FOU-ansatte må betale en høyere pris enn hva antall FOU-årsverk tilsier.
- En andel som baseres på antall FOU-årsverk fordelt slik at prisen pr årsverk er avtagende med økende antall.
- En andel som baserer seg på lagring av data slik at hver institusjon får en lagringskvote pr FOU-årsverk, tilsvarende x antall forskningsartikler på y MB inkludert i årlig avgift. Lagring ut over dette vil telles ved årets utløp og faktureres den enkelte institusjon etter forbruk. (Dette innebærer at kostnaden for lagring vil øke hvert år uavhengig av om institusjonen benytter mer lagringsplass enn sin kvote eller ikke.

Denne modellen hensyntar små og store institusjoner ved at de små betaler relativt sett en noe høyere pris enn de store da alle bør bidra med et minimum og de små erfaringsmessig har et større behov for brukerstøtte/support enn de store institusjonene.