



DESEMBER 2023

Veien videre - NIB 2.0

Jens Aasheim, Ivar Ternsell Torgersen og Lars Wenaas



Sammendrag av rapporten

Bakgrunn og mandat

En arbeidsgruppe nedsatt av ledelsen har fått i oppdrag å vurdere tiltak for å heve kvaliteten på tjenestene i Nasjonal infrastruktur for Bibliometri - NIB. Arbeidet er begrunnet med at NIB har vært operativ i 5 år og at kilden som utgjør kjernen i NIB-basen har fått en rekke nye utfordrere som kan være bedre egnet. Evaluering av kilden er derfor på sin plass før inneværende kontrakt løper ut i 2024 og NIB går inn i en ny 3-årsperiode med dagens leverandør. Det er tidligere utarbeidet en behovs- og mulighetsstudie, bestilt av Sikt, som tar for seg både utfordringer og muligheter i NIB. En del av bildet er at rutiner knyttet til dagens tekniske løsning påfører tjenesten en uforholdsmessig ressursbelastning.

Arbeidsgruppen har fokusert på neste generasjon NIB i tråd med mandatet, med fokus på tekniske forbedringer, valg av eventuell ny kilde og hva slags tjenester morgendagens NIB må levere.

Status NIB

NIB vurderes til å levere på kjernen av oppdraget. Samtidig er det uttalte ambisjonsnivået for tjenesten høyere og har et urealisert potensial. Fagbibliometrikere har fått dekket sine behov, men administrative bibliometrikere, med oppgaver innen forskningsevaluering, benchmarking og virksomhetsstyring i egen institusjon, mangler et godt tilbud i NIB-systemet. Det er også blant disse brukerne behovet for tilrettelegging er prekært og der NIBs uforløste potensial er størst. I dag er tjenesten lite tilgjengelig og mangelfullt tilpasset denne brukergruppen.

NIB har også tekniske utfordringer med import av data, som er basert på en løsning adoptert fra Karolinska institutt. Løsningen, som i seg selv er vurdert som solid, er ikke godt nok tilpasset den norske tjenesten. Spesielt import-rutinen har stor teknisk gjeld, noe som fører til en krevende prosedyre. Import tar for lang tid og har medført en uholdbar belastning på tjenesten og tjenesteansvarlig. Dette er også poengtert i nevnte mulighetsstudie.

Valg av kilde

Valg av kilde er sentralt i NIB og bestemmende for kvalitet og hvilke muligheter for bibliometriske analyser som kan utføres. Dagens kilde er en tradisjonell og solid, men også en kostbar kilde. Likevel, det skal argumenteres godt for et eventuelt bytte. Sikt er allerede i et kontraktsforhold med en betalt og velfungerende kilde og det vil påbeløpe store transaksjonskostnader ved et evt. anbud. En ny kilde bør derfor ha store fordeler over dagens. Kilder har forskjellige egenskaper og fortrinn som f.eks. dekningsgrad, dokumenttyper og pris. Med utgangspunkt i NIFU sitt notat om bibliometriske kilder, anser arbeidsgruppen kun to som potensielle kandidater: Dimensions og OpenAlex. Begge med svært attraktiv dekningsgrad, og med dokumenttyper som muliggjør analyser utover tradisjonelle analyser med fokus kun på forskningsartikler. Dette gjelder analyser som inkluderer patenter, sosiale medier eller offentlige dokumenter.

Allikevel vurderer arbeidsgruppen OpenAlex som en spesielt interessant mulighet. Dedikasjon til åpenhet og gratis datatilgang, bidrar til et fremtidsrettet inntrykk. Dette inntrykket forsterkes ytterligere av at ledende bibliometriske aktører tar den aktivt i bruk.

Tjenestenivåer

Valg av kilde er logisk adskilt fra hvilke tjenester NIB skal levere, selv om kvaliteten på tjenestene nødvendigvis er avhengig av kvaliteten på kilden. Vi deler inn NIB i grunnmur og 4 tjeneste nivåer. Grunnmuren omhandler import, rydding og tilrettelegging av data, noe som står for anslagsvis tre fjerdedeler av det tekniske arbeidet i NIB. Betegnelsen "grunnmur" er ment å illustrere at her legges grunnlaget for hvor godt tjenestenivåene bidrar til at veien til analyse blir så kort som mulig.

Vi ser for oss følgende tjenestenivåer:

T1: SQL-tilgang og utlevering av generiske datasett

T2: Tilpasset eksport av NIB-data inn i eksisterende DUCT-struktur (arbeidsdeling med NVA)

T3: Utlevering av datasett til eksisterende bibliometrisk verktøy

T4: Analyseplattform

Arbeidsgruppens mandat presiserer at NIB ikke skal utarbeide analyseverktøy. Vi behandler derfor ikke tjenestenivå 4, men har dette med da en analyseplattform er en prioritert oppgave i mulighetsstudien. Mulighetsstudien løfter også spesielt frem tjenestenivå 2 som den viktigste tjenesten for brukergruppen som ikke har et tilbud i dag. Tjenestenivåene er anvendelige til å sette tekniske krav til grunnmuren, der et sentralt element er å omskrive dagens importrutine fra bunn av.

Arbeidsgruppens anbefalinger

Arbeidsgruppen tegner opp 5 scenarier som alternative retningsvalg for ledelsen å navigere ut fra. For å realisere ambisjonene om en fremtidsrettet NIB er det derfor ikke nok å omskrive tekniske rutiner, selv om dette løser mange problemer og kan frigi store ressurser.

Vi anbefaler det scenariet som argumenterer for å omskrive alle tekniske rutiner forbundet med WoS, og som prioriterer mer tilrettelegging av kilden. For å utnytte potensialet i NIB må Sikt levere på tjenestenivå 3, der NIB møter brukernes behov gjennom de verktøy de foretrekker. Bibliometriske verktøy kan være krevende, det er gjennom tilpassede datasett til disse verktøyene at vi «gjør veien til analyse så kort som mulig».

Scenariet innebærer videre å supplere dagens WoS-kilde med OpenAlex. Dette er et forslag som gjør NIB fremtidsrettet, samtidig som NIB utnytter en god og solid kilde som allerede er betalt. Sikt vil dermed ha god tid til å vurdere styrker og svakheter ved kildene uten hastverk og videre om WoS fortsatt skal være en del av NIB eller ikke. Fra et rent faglig ståsted er det ønskelig at WoS er med også neste tre-årsperiode, men vi anerkjenner at dette er knyttet til ressursvurderinger. Vi mener rapporten argumenterer godt for at dagens ressursmessige utfordringer ikke utgjør et hinder for å supplere NIB med OpenAlex. Kilder som kan berike NIB ytterligere blir kun overfladisk behandlet, da dette temaet fortjener en mer detaljert vurdering enn denne rapporten har hatt mulighet til.

Organisatorisk anbefaler vi at NIB knyttes sterkere til andre produktområder i Sikt. Vi ser et stort potensial i grenseflatene mellom NIB, LÅT og NVA. Vi tror også det finnes flere koblinger.

NIB har et viktig nettverk i Bibliometriforum som består av nettopp de brukerne som i dag mangler et godt tilbud. Dette nettverket ser vi som svært viktig i dialogen med NIB med en rolle som bestiller inn til NIB.

Arbeidsgruppen takker for å ha fått muligheten til å gi sitt syn på en spennende tjeneste.

Innholdsfortegnelse

Innledning	4
NIB 1.0 - Mål, status og veivalg.....	5
Status	5
Viktige valg i implementeringsfasen	5
Vurdering av valg.....	6
NIB 2.0 Realisering av potensialet for brukere	9
Brukere, bruksområder og kompetanse	10
Kildevalg – avgrensende og bestemmende for analysemulighetene	12
Vurdering av mulige nye kilder	12
Berikende kilder	14
Oppsummerende diskusjon og vurdering	15
Tjenestenivåer	17
Krav til tjenestenivåene	20
Bruk av bibliometriske verktøy	21
Krav til dokumentasjon og dialog med brukerne	22
Scenarier for å komme fra NIB 1.0 til NIB 2.0	22
0: Null-scenario (ingen endringer)	24
1: «Minimums-varianten» (Reparer import, spar tid, frigi ressurser)	25
2: «Total-renovering» (Behold WoS og bygg ut tjenestene).....	26
3: «Ny kommersiell NIB» (Bytte kilde til Dimensions)	27
4: «Gå åpent» (Bytt kilde til OpenAlex).....	28
5. «Kjør dobbelt» (Ny og forbedret NIB - Supplere WoS med OpenAlex).....	29
Anbefaling og oppsummering	30
Anbefaling.....	30
NIB i Sikt	31
Oppsummering	32

Veien videre: NIB 2.0

Rapport fra arbeidsgruppen

Innledning

Arbeidsgruppen har fått i oppdrag å vurdere tiltak for NIB i tråd med utlevert mandat sendt 06.09.2023. I mandatet er behovet for en hurtig forberedende utredning begrunnet med at NIB har vært operativ i 5 år, at valgte kilde Web of Science (WoS) har fått en rekke nye seriøse utfordrere, og at det er utarbeidet en behovs- og mulighetsstudie¹ med en rekke betraktninger på utfordringer og muligheter i NIB. Rutiner knyttet til dagens tekniske løsning belaster også tjenesten uforholdsmessig ressursmessig. Til sammen begrunner disse forholdene en gjennomgang av dagens NIB.

Arbeidsgruppen bestående av Jens Aasheim, Ivar Ternsell Torgersen og Lars Wenaas, har tolket kjernen i mandatet til å utarbeide løsningsforslag på utfordringer knyttet til dagens tekniske drift, samt tiltak for å nå ambisjonene for prosjektet som blant annet er redegjort for i mulighetsstudien.

Innhold i rapporten

Arbeidsgruppen noterer seg at prosjektet vurderes som å ha innfridd de grunnleggende mål som ble satt ved oppstart. Vi har derfor valgt å bruke begrepet NIB 2.0 for å signalisere en utvikling av dagens løsning i tråd med ambisjonsnivået som delvis er formulert i grunnlagsdokumenter, delvis i mulighetsstudien. I dette bildet er flaskehalser knyttet til dagens drift en av flere utfordringer for å nå målene. Rapporten vil i liten grad dvele ved historien til NIB, men vil likevel kommentere strategiske valg som arbeidsgruppen vurderer som innflytelsesrike og bestemmende for videreutvikling av dagens løsning. Hovedfokus i rapporten vil i alle tilfelle bli en beskrivelse av morgendagens NIB, hvordan den kan se ut og hvordan vi kan komme dit.

Alle prosjekter av en viss størrelse må skaleres etter tilgjengelige ressurser, dette gjelder også NIB. Da rapporten skal fungere som et beslutningsgrunnlag for ledelsen når det gjelder det videre arbeid, har vi valgt en scenario-basert tilnærming. Rapporten drøfter derfor først ulike krav, muligheter og behov knyttet til NIB. Denne gjennomgangen blir fundament i vurderingen av i hvilken grad de skisserte scenariene oppfyller kravene og ambisjonene til morgendagens NIB. Selv om rapporten peker ut flere veier, så har tiden vært for kort til at dybdenivået i scenariene og det vi antar kreves av ressurser, er godt nok utredet. Målet med rapporten er at scenariene er egnet for valg av himmelretning, men påpeker at det vil være nødvendig med en egen prosjektering av valgt(e) løsning(er) i etterkant.

Rapporten er lagt opp på følgende måte

- NIB 1.0. En kort beskrivelse av bakgrunn, mål, teknisk implementering, brukergrupper og viktige veivalg.
- NIB 2.0. Tolkning av mulighetsstudien, fokus på brukergrupper og deres behov, vurdering av datakilde.
- Tjenestenivåer. Et forslag til organisering av tjenester
- Scenarier for å komme til NIB 2.0.
- Oppsummering og anbefalinger.

¹ Behovs og mulighetsstudie for Nasjonal infrastruktur for bibliometri (2022)

NIB 1.0- Mål, status og veivalg

Status

Formålet med Nasjonal infrastruktur for bibliometri (NIB) er å tilby data om norsk og internasjonal forskning til analyseformål. Målgruppen for NIB er norske forskningsinstitusjoner, andre offentlige organ innen forskning og myndigheter.

Prosjektet ble etablert 2017 med 3 mill NOK i budsjettøkning til daværende Ceres for å håndtere en komplett bibliometrisk database med verdens-data og tilrettelegge denne for bruk. Rent praktisk skulle det etableres en enkel infrastruktur for raskest mulig å gjøre disse dataene tilgjengelige for pilotinstitusjoner med kompetanse på bibliometri og senere kople data med Cristin. I 2020 ble NIB satt i ordinær drift etter at implementeringsperioden ble ansett som ferdig, da også styringsgruppen og tilhørende utvalg ble lagt ned.

En viktig del av NIB i dag er Bibliometriforum, et bruker- og kompetanse-nettverk bestående av 40 institusjoner og ca. 80 personer som jobber aktivt med bibliometri. Forumet ble etablert som en viktig arena for deling av kunnskap og kompetanse samme år NIB ble satt i drift.

Kjernen i tjenesten er NIB-basen. Dette er en database på en dedikert fysisk server som består av rundt 80 millioner poster fordelt på over 40 årganger. Brukermassen er forholdsvis lav på selve basen og består av rundt 13 delvis aktive brukere ved NIFU, NFR, UiB, UiT og Sikt. Analyse og utnyttelse av data krever høy teknisk kompetanse da data i all hovedsak må hentes via SQL.

Denne rapporten har brukt tid på å forstå valg, krav og avgjørelser i de første faser av NIB. Spesielt før og under implementeringsperioden ble det tatt en rekke valg som er viktige for å forstå hvilke muligheter NIB kan innfri og hvilke utfordringer NIB står overfor i dag. Rapporten vil først fremheve noen av disse valgene og hvilken betydning vi vurderer valgene har i dag.

Viktige valg i implementeringsfasen

I utredningsfasen (før prosjektstart)

I den innledende fasen ble det lagt vekt på å gjøre data tilgjengelig «*raskest mulig*». Prosessen var altså preget av en viss utålmodighet etter å få en database med verdensdata tilgjengelig så fort som mulig. Samtidig er underlagsdokumenter preget av å vektlegge behovet til de mer profesjonelle fagbrukerne, primært NIFU.

Valg av kilde var fokusert på tradisjonelle trygde kilder, og kanskje den viktigste av dem, Web Of Science (WoS), ble vurdert anskaffet gjennom et direktekjøp uten anbud. Helt siden 1960-tallet har denne kilden hatt status som «gull-standarden» for bibliometriske data blant fagbrukere. Scopus har vært en seriøs konkurrent siden lansering i 2004, mens andre kilder generelt var regnet som interessante, men umodne. Med i denne vurderingen er trolig også at WoS og Scopus er kuraterte databaser, mens f.eks. Dimensions i all hovedsak er basert på automatiserte rutiner.

Dette beredte grunnen for et naturlig, men også konservativt, valg av kilde som var egnet til å dekke behovet til fagbrukere med høy kompetanse. Andre typer brukere enn fagbrukere synes derimot mindre vektlagt, utover en anerkjennelse av et behov som skulle turneres gjennom analyseverktøy i den kommende anskaffelsen.

Integrasjon med Cristin og analyseløsninger

De to aktuelle leverandørene av henholdsvis WoS og Scopus, leverer også dedikerte analyseverktøy. Dette er komplette løsninger med både analyseverktøy og data, men å gi direkte tilgang til rådata. De to verktøyene, Insights og SciVal, var tiltenkt en rolle i å dekke behovet for andre brukere enn fagbrukerne. Analyser gjort i disse verktøyene er derimot avhengig av dekningsgraden på norsk publisering, noe som var årsaken til intensjonen om å forene Cristin-data med kilde-data for senere

reimport av berikede kilde-data tilbake til valgt leverandør. Ideen var å øke datakvaliteten i analyseverktøyene, hvorpå norske brukere kunne nyte godt av at Cristin-data var tilgjengelig. Den ønskede reimporten viste seg også å være vanskelig å gjennomføre, da leverandørene viste liten interesse for en feedback-løsning. Det var ikke heller ikke aktuelt med kjøp av rene løsninger som SciVal og Insights alene for NIB sitt vedkommende, dette ville ikke dekke fagbrukernes behov for tilgang til rådata.

Krav i anbudsprosessen

Anbudsprosessen ble utformet som en tredelt anskaffelse som sondret mellom 1) kjøp av rådata, 2) analyseverktøy og 3) importdata til Cristin. I ettertid fremstår disse del-elementene med krav som pekte i forskjellige retninger og forutsetninger som ikke lot seg oppfylle. Anbudet kunne derfor i prinsippet enklere blitt turnert som separate anbud.

Et krav i NIB-anbudet var eierskap til rådata. Dette var et naturlig valg begrunnet med et eventuelt senere skifte av leverandør og et ledd i å unngå klausuler om sletting av data ved opphør av kontrakt. Kravet om eierskap var i utgangspunktet ikke et sentralt element, men ble utslagsgivende for Scopus og Elsevier som ble diskvalifisert i siste forhandlingsrunde på den delen av anbudet som omhandlet kjøp av rådata.

Samtidig fortsatte separate forhandlinger med Elsevier om kjøp av Scopus-data til importformål til Cristin. Motivasjonen for at import av data til Cristin var med i anbudet, var at samme kilde skulle velges til NIB og Cristin, dvs. at Cristin-importen skulle nyte godt av NIB-avtalen. Cristin importerte tidligere WoS-data, men skiftet kilde til Scopus rundt 2013, en avtale som ble fornyet i forbindelse med NIB-anbudet.

Analyseverktøyene i anbudet var frikoblet og uavhengig av den delen som omhandlet kjøp av rådata, samtidig var rådata-kildene en del av henholdsvis SciVal og Insights fra leverandørenes side.

Analyseverktøy-delen av anbudet munnet ut i to innkjøpskonsortier for norske forskningsinstitusjoner, som da kunne kjøpe rabatterte tilgang til henholdsvis SciVal og Insights. Begge konsortiene hadde få deltagere og ble forholdsvis fort avvirket. Analyseverktøy og rådata var altså forholdsvis uavhengige elementer fra starten av, men må likevel sees i sammenheng med nevnte feedback-loop der norske data fra Cristin skulle berike original-kilden.

De to delene i anbudet som omhandlet innkjøp av importkilde og analyseverktøy fremstår i ettertid som mindre relevante for NIB-basen slik den er i dag.

Krav til teknisk løsning

Da Clarivate ble valgt som leverandør, var neste skritt en teknisk løsning for import av rådata (som ble levert som XML-filer fra leverandør) og å bygge opp NIB-basen fra grunnen av. På dette tidspunkt var arbeidet med datamodell allerede startet. Som et ledd i nettverksbygging søkte NIB samarbeid med blant annet Karolinska Institutet (KI). KI hadde implementert løsninger basert på tilsvarende behov og kilde og hadde rutiner som etter grundig vurdering ble ansett som solide og attraktive. Dette var også en nødvendig snarvei for å komme fort i gang med import av data, noe som naturlig innebar en viss tilpassing til KI's valg av teknologi, datamodell og metode for import.

Bibliometriutvalget, som satt tett på prosessen med å bygge NIB, poengterte spesielt at det fremdeles var behov for analyseløsninger for brukere uten tung teknisk kompetanse. Det ble derfor initiert en dialog med flere aktører om analyseløsninger.

Vurdering av valg

Prosjektet har tatt en rekke valg i implementeringsfasen som både er naturlige og fornuftige, situasjonen og ressursene tatt i betraktning. I ettertid kan man si at en del av forutsetningene og

premissene som ble lagt til grunn, ikke viste seg å være helt som først antatt. Overordnet kan premissene fremstå som i overkant fokusert på hurtig anskaffelse av *data* snarere enn brukere, bruksbehov og *bruk av data*. Et bilde på dette er at styringsgruppen avviste et forslag om å utarbeide en prosjektplan (basert på prosjektveviseren til Difi) med den konsekvens at det ikke ble utarbeidet en interessentanalyse som kunne bidratt til å identifisere og håndtere brukerbehovene bedre.

De tematiske kjernepunktene arbeidsgruppen spesielt vil fokusere på er følgende:

Brukerne

Relativt sett, har fagbrukerne hatt stor prioritet og samtidig som 'andre brukere' har vært en viktig motivasjonsfaktor for å opprette en nasjonal infrastruktur, har de ikke fått tilstrekkelig tilgang til tjenester som dekker deres behov. 'Andre brukere' er også uklart definert og behovene deres svakt formulert. Hverken grunnlagsdokumenter eller tidlige prosjektdokumenter gjør grundige undersøkelser av brukerbehov. Det er først i anbudsprosessen at brukerscenarier og user cases blir utarbeidet. Her er brukerbehovene skissert, men uten at det har satt et tydelig preg på dagens NIB. Produktet ansees i dag som lite brukervennlig og med begrenset mulighet for flere brukere og virksomheter². Bibliometriforum ble opprettet for å gi disse en stemme, noe som understreker at den store forskjellen mellom fagbibliometrikere og andre, er kompetansenivået på både bibliometri som fag, analyse av data og teknisk prosessering. Det er derfor kun fagbibliometrikere med høy kompetanse som kunne utnytte NIB-ressursen i dag.

Tilgang og analyseverktøy

Dagens tilgangsnivåer er en konsekvens av at fagbrukere har blitt prioritert over brukerbehovene til andre brukere. NIB-basen er tilrettelagt for fagbibliometrikere gjennom direkte tilgang til basen via SQL. Dette krever høy tekniske kompetanse og kjennskap til datamodellen. Et viktig unntak er grepet med å eksportere data fra NIB (utvalgte tall og indikatorer) til Cristin/NVA-basen. Dette er et godt valg som tilgjengeliggjør NIB-data i en eksisterende infrastruktur for datautlevering og analyse, som dekker mange behov for administrative brukere. Likevel fremstår denne ordningen som uferdig og usikker, blant annet fordi DUCT og den underliggende programvaren Tableau har en uavklart fremtid i Sikt. Det er uansett et spørsmål om hvor langt brukere med lav teknisk og bibliometrifaglig kompetanse må hjelpes på veien til analyse.

Kildevalg

Valget av WoS som kilde skjedde på et tidspunkt da andre tilbydere enn WoS og Scopus ble regnet som umodne. I dag er situasjonen en annen. Både andre kommersielle tilbydere (f.eks. Dimensions) og ikke-kommersielle tilbydere (f.eks. OpenAlex), regnes nå for å være seriøse utfordrere som kan konkurrere på dekningsgrad og etter hvert datakvalitet. Det er et poeng at enhver datakilde setter naturlige begrensninger både for hva som kan analyseres og hvor god analysene kan bli. Dette alene er grunn nok til å kontinuerlig vurdere hvilken kilde som er best egnet. Dersom gjennomslagskraften til forskning i offentlige dokumenter og patenter, eller oppslag i aviser og sosiale medier, er et viktig interesseområde for brukerne av NIB, kan andre kilder være bedre egnet enn dagens. Da må WoS enten suppleres eller erstattes. Kildevalg er tett knyttet til brukernes behov og hvilke spørsmål de ønsker besvart i bibliometriske analyser.

Teknisk standard

Teknologisk infrastruktur (operativsystem, databasemotor etc.) i NIB er basert på solide løsninger basert på åpen kilde kode og med godt renommé. Teknologisk infrastruktur vil ikke bli vurdert nærmere i rapporten, men bør likevel berøres i en senere fase.

² <https://sikt.no/tiltak/utvikling-av-nasjonal-infrastruktur-bibliometri-nib>

Generelt har åpen kildekode-løsninger ofte en lav pris for inntreden, men kan likevel ha høye kostnader i tilpassing og adaptasjon. Dette gjelder spesielt applikasjonslaget og andre egenutviklede rutiner, som i motsetning til programpakkene i den teknologiske infrastrukturen (database, operativsystem etc.) ikke er gjennomtestede og har sunne bærekraftige økosystemer tilknyttet.

I NIB sitt tilfelle har en automatisk velfungerende løsning tilpasset KI, ikke blitt tilstrekkelig tilrettelagt og tilpasset norske data. NIB inneholder større datamengder og andre kildetyper som sammen med manglende utviklingsressurser, har skapt en stor belastning på systemet. Den norske implementasjonen mangler tilpasset automatisering og feilhåndtering og har dermed teknisk gjeld. Spesielt import-rutinen oppleves som uholdbar med en svært lang periode fra mottak av rådata til de er importert, som er medvirkende til en uakseptabel belastning på tjenesteansvarlig. Det bemerkes også at NIB ble knyttet tett til Cristin og Sikt, delvis med den begrunnelse at det fantes kompetanse på import av bibliografiske data fra både WoS og Scopus. Dette har ikke blitt utnyttet i prosjektet.

Prinsipper knyttet til åpenhet og transparens

I nyere tid har verdigrunnlaget som ligger i bruk av åpen kildekode og åpne data (og dermed åpen forskning) fått fotfeste innen bibliometri-faget. Prinsippet om åpenhet og deling skal også gjelde bibliografiske poster, siteringer og andre bibliometriske data.

Åpenhetsprinsippet ser man blant annet i tjenester som Crossref, Open Citations, Unpaywall og OpenAlex, der data skal være fritt tilgjengelige for alle mulige interessenter, langt på vei med samme begrunnelse som med andre åpne data. Dette gjenspeiler seg også i praksis, der en rekke prosjekter integrerer åpne kilder. CWTS i Leiden, ansett for å være verdensledende på bibliometri, er et eksempel på et miljø som er i ferd med å implementere OpenAlex, delvis av ideelle grunner. Blant annet vil en kjernetjeneste fra CWTS, den såkalte Leiden-rankingen, også bli basert på OpenAlex i 2024³.

Det er altså et slektskap mellom verdigrunnlaget til moderne bibliometri og til åpne data som har direkte relevans for NIB. Denne åpenhets- og delings-kulturen vil være en av flere komponenter i vår vurdering av kilder.

Selv om NIB vurderes som å ha innfridd de mest grunnleggende målene, strekker ambisjonene seg lenger enn der prosjektet er i dag. Det gjelder spesielt «institusjonene i UH-sektoren [som] har behov for å kunne gjøre benchmarking av fagmiljøer, enheter og enkeltforskere mot relevante aktører nasjonalt og internasjonalt⁴». Dette behovet er helt i tråd med senere tids administrative bruk av bibliometri i virksomhetsstyring utover ren forskning og bestillingsoppdrag. Tilgang til data som støtter denne type bruk har lenge «vært savnet og har begrenset mulighetene for bibliometrisk analyser i Norge»⁵. Den voksende interessen for bibliometri ved norske forskningsinstitusjoner er altså en viktig motivasjon for en «nasjonal infrastruktur», noe som også reflekteres i formålsbeskrivelsen.

Sikt formulerer selv forventningene til NIB slik: Med «tilrettelegging av data og brukergrensesnitt, som senker brukerterskelen, forventes det at bruken av produktet kan øke på hver av de mer enn 60 institusjonene som hittil har vist interesse for NIB-data» [med gevinster som] «mindre behov for bruk av lokale ressurser og individuelle innkjøp av kommersielle systemer»⁶.

³ <https://www.researchprofessionalnews.com/rr-news-europe-universities-2023-9-leiden-rankings-to-add-open-source-version-in-2024/>

⁴ «Nasjonal infrastruktur for bibliometri» - Forslag fra en arbeidsgruppe nedsatt av Norges forskningsråd

⁵ «Nasjonal infrastruktur for bibliometriske data» - NFR

⁶ <https://sikt.no/tiltak/utvikling-av-nasjonal-infrastruktur-bibliometri-nib>

Vår tolkning er at ambisjonsnivået i stor grad dreier seg om administrative brukere til virksomhetsstyring ved norske forskningsinstitusjoner, snarere enn den profesjonelle fagbrukeren med høy bibliometrisk og teknisk kompetanse. Sistnevnte gruppe er i alle tilfelle operative i dagens løsning.

En arbeidsgruppe oppnevnt av Tjenesterådet for forskningstjenester og med base i medlemmer i Bibliometriforum, leverte en rapport i 2022 om potensialet i NIB. Mulighetsstudien både kompletterer og presiserer ambisjonsnivået for NIB og er et viktig dokument for å beskrive essensen i NIB 2.0. Arbeidsgruppen vil fokusere nærmere på de nevnte tematiske kjernepunktene (brukeren, analyseverktøy og tilgangsnivå, kildevalg og teknisk standard) ved vår oppsummering og tolkning av rapporten.

NIB 2.0 Realisering av potensialet for brukere

Mulighetsstudien plasserer NIB i senere tids forskningspolitiske kontekst preget av åpen forskning, som igjen har forårsaket store omveltninger i forskningsevaluering. Spesielt løftes NOR-CAM-profilen frem. Dette er den norske implementasjonen av et europeisk initiativ for nye former for forskningsevaluering som ikke hviler på kvantitative bibliometriske indikatorer alene. NOR-CAM betoner spesielt viktigheten av å meritte andre former for forsknings-output enn kun forskningspublisering. Mulighetsstudien anerkjenner dermed også at bruk av NIB ikke er uproblematisk sett i lys av Leiden-manifestet, DORA-deklarasjonen og annen kritikk rettet mot bruk av bibliometriske indikatorer i evalueringsprosesser. Dette er et viktig poeng som må gjennomsyre arbeidet med NIB, som åpenbart må legge til rette for vettig bruk av indikatorer gjennom økt kompetanse om bibliometri.

Mulighetsstudien bekrefter også utfordringer og sårbarheter i dagens drift av NIB og manglende tilgang til basen for andre enn fagbrukere. Dermed har også ressursituasjonen fått stor oppmerksomhet; i konklusjonen hevdes det at selv en forholdsvis liten ressursøkning vil kunne realisere store deler av potensialet til NIB. Mulighetsstudien påpeker videre at dagens løsning ikke er bærekraftig og er personavhengig, det anbefales derfor å vurdere annen organisering av NIB, eventuelt utenfor Sikt. Mulighetsstudien har et noe defensivt utgangspunkt da en full utvikling av NIB ikke vurderes «da det ansees som urealistisk å få gjennomført på nåværende tidspunkt.»

Mulighetsstudien har godt samsvar med de underliggende dokumenter som arbeidsgruppen har lest, men det observeres også noen forskjeller og endringer i prioriteringer. I de opprinnelige dokumentene er norske forskningsinstitusjoners behov for en bibliometrisk database hovedargumentasjonen for å tilby en nasjonal løsning, i mulighetsstudien er det (mer) åpenbart at det også er stort behov for både bibliometrisk og teknisk kompetanse samt ferdige analyseløsninger. Dette kan også leses som en erkjennelse av at institusjons-brukeres behov ikke har fått nok oppmerksomhet i implementerings-fasen.

Ifølge studien er dagens hovedutfordring den høye brukerterskelen for ikke-avanserte brukere og spesielt knyttet til at tilgang til databasen krever kompetanse i SQL. Noe av potensialet kan derfor realiseres «gjennom enklere tilgang til data, med lavere brukerterskel.». Videre mener studien at det uansett er behov for analyseløsninger som må bygges fra grunnen av, fortrinnsvis gjennom bruk av open source-verktøy. Mulighetsstudien ser for seg en brukervennlig analyseplattform som kan utvikles i R/Shiny og som kan utnytte det som finnes av ferdige programbiblioteker for bibliometriske analyser. En slik løsning bør ifølge studien suppleres med API-tilgang til NIB. Omfanget av disse oppgavene er ikke prosjektert.

En helt spesiell integrasjon er den mellom NIB og Cristin/NVA. Mulighetsstudien «anser koblingen mellom NIB og Cristin (CRIS/NVA - MIME) som den viktigste tilretteleggingen av dataene, på nåværende tidspunkt.» Denne må «prioriteres fremfor all annen tilrettelegging, uavhengig av nivå på videreutvikling».

Viktigheten av denne integrasjonen ligger i primært to forhold.

- 1) Brukergruppen med dårligste tilbud i dag har størst interesse av den norske delen av verdens publisering (avgrenset av Cristin/NVA) for analyser til virksomhetsstyring.
- 2) Tilgang til dette datasettet (Cristin/NVA + NIB-data) gis gjennom eksisterende utleveringsplattformer (DUCT) som også gir tilgang til et generisk analyseverktøy (Tableau).

Denne ordningen er også en god illustrasjon på mulig samarbeid og arbeidsdeling mellom NIB, NVA og andre tilstøtende produktområder, der ansvaret for NIB naturlig stopper ved datautlevering.

Mulighetsstudien er forholdsvis summarisk i beskrivelse av brukere og bruksbehov, likevel tegnes det et noe klarere bilde av bruksbehov hvis man samtidig skuer til *user cases*-dokumentet som ble utarbeidet i forbindelse med anbudet i 2018. Neste avsnitt er derfor et forsøk på å sammenstille informasjon om brukeren, bruksområder, kompetansebehov og data-behov.

Brukere, bruksområder og kompetanse

Mulighetsstudien lister en rekke forskjellige interessenter og brukere av NIB. Brukere av NIB er primært ansatt i institusjoner i UH-sektoren, instituttsektoren og helsesektoren, der en forutsetning er at institusjonen holder seg med et minimum av bibliometrisk kompetanse. Det kan likevel ikke forventes et høyt kompetansenivå blant administrative bibliometrikere helt uten videre, noe som krever en forankring av NIB i miljøer med tung forskningskompetanse. I Norge er det i praksis NIFU som representerer denne kompetansen og som bidrar gjennom NIB-nettverket. NIB har også samarbeid på nordisk nivå, i tillegg til tilgang til internasjonal kontakt gjennom NIFU.

I vår rapport skiller vi mellom to primærbrukere og en sekundærbruker som har forskjellig grad av kompetanse og behov for tilgang til NIB-tjenester og analyseverktøy.

- *Fagbibliometrikere*: Fagpersonell i NIFU, samt forskere og studenter innen relevante kvantitative disipliner og andre dedikerte miljøer. Sikt kan også plasseres i denne gruppen, gjennom avansert bruk av data til understøttelse av forhandlingsoppdraget samt statistikk og rapporteringsansvar for åpen tilgang. Brukergruppen karakteriseres av å ha, og ha behov for, tung teknisk og bibliometrifaglig kompetanse.
- *Administrative bibliometrikere*: Lokalt administrativt personell i forskningsinstitusjoner involvert i virksomhetsstyring og rapportering. Brukergruppen karakteriseres av å ha mindre teknisk og bibliometrifaglig kompetanse og krever mer tilrettelegging av data og enklere tilgang til data enn på SQL-nivå. De har også behov for analyseverktøy og mest mulig ferdige løsninger. Databehovet er sentrert rundt data om egen institusjon. Brukergruppen representerer også fronten mot de mer indirekte brukerne i neste brukergruppe, bestillere.
- *Bestillere*: Institusjonsledere på alle nivåer, nasjonale evalueringsorganser, Norge forskningsråd, norsk publiseringsutvalg, departement og andre statlige og offentlige institusjoner. Brukergruppen har ikke behov for teknisk kompetanse, men må ha forståelse for bibliometri og ansvarlig bruk av indikatorer. Vil ikke være direkte brukere av NIB.

For å adressere den minst kompliserte brukergruppen først, så er *bestillere* brukere som må kjenne til bibliometriens begrensninger for å kunne bestille og tolke bibliometriske analyser. Brukergruppen nevnes spesielt da den er viktigste målgruppen for Leiden-manifestets og NOR-CAM profilens advarsler om bruk av bibliometriske indikatorer; et kvantitativt uttrykk for en forsker, forskergruppe, institutt eller institusjon er aldri godt nok alene i evalueringer.

Advarselen har konsekvenser for NIB: hvis NIB-tjenestene bygges ut til å enkelt levere ferdige tall og rangeringer uten at det bygges inn mekanismer som påpeker svakhetene ved bruk av indikatorer alene, så gjør NIB forskningen en bjørnetjeneste.

Det kan finnes andre 'nytere' enn bestillere. Det er vist en viss interesse fra representanter fra norske kommuner som kan ha nytteverdi av informasjon om forskning som vedrører egen kommune i næringssammenheng. Allmenheten kan også både ha interesse av, og krav på, informasjon og statistikk om forskning. En tilsvarende tjeneste som dekker dette behovet leveres av Sikt gjennom statistikk-sidene til Cristin/NVA. Det er ikke foretatt noen systematisk undersøkelse av disse behovene i mulighetsstudien, men det antas at potensialet til resultater fra NIB kan vise seg å være større enn antatt.

Mange, kanskje alle, bibliometriske analyser vil være av interesse for begge de to primære brukerne. Litt forenklet kan man si at der fagbibliometrikeren undersøker forskningens påvirkning og oppførsel i en rekke ulike kontekster, vil den administrative bibliometrikeren ha interesse av å undersøke det samme spørsmålet hovedsakelig avgrenset på egen institusjon. Det er altså først og fremst et kompetanseskiller mellom de to brukergruppene, bruksområder og analysemetoder er derimot overlappende.

Bruksområder

Tradisjonelle analyser dekker samarbeid, nettverk, siteringer, tematiske undersøkelser og evaluering av forskere, forskergrupper og institusjoner. Dette er analyser som er interessant både for rene forskningsformål og i virksomhetsstyring og som strategisk styringsinformasjon i nasjonal kontekst. Analysene kan skje regelmessig for å overvåke forskningsaktivitet over tid eller som enkelt-stående analyser basert på skreddersøm.

I tillegg til de mer tradisjonelle analysene, har bibliometri den senere tid fokusert på *utvidete analyser* som undersøker påvirkning gjennom siteringer i patenter (gjennomslagskraft i innovasjonssammenheng), sosiale medier (oppmerksomhet i offentligheten), offentlige dokumenter (policy-påvirkning) og datasett. Dette krever kilder i NIB med dokumenttyper som støtter oppunder slik bruk, men er ellers basert på de samme statistiske metodene som i tradisjonell bibliometri.

Mulighetsstudien peker også på mer *avanserte metoder* der NIB kan brukes til maskinlæring og metoder som Natural Language Processing (NLP) på større mengder tekst-korpus (typisk abstracts).

Rent konseptuelt adresserer NIB 2.0 det store kompetanses skillet mellom brukertypene. Dette kompetanses skillet setter store krav til hvordan data tilrettelegges og utleveres og dermed møter brukernes analyseverktøy og analysebehov.

Kompetanses skillet gjør seg kanskje spesielt gjeldende i behovet for tilrettelegging av indikatorer i NIB og krever i alle tilfeller koordinering. NIB består i utgangspunktet av rådata som danner beregningsgrunnlaget for diverse bibliometriske indikatorer (persentiler, normaliserte siteringsgjennomsnitt etc.). Dette er krevende beregninger og det er følgelig fornuftig og ofte nødvendig med sentral beregning av indikatorer en gang, i stedet for at beregninger gjøres av brukerne selv på rådata. Indikatorene må kvalitetssikres av NIFU (noe som også gjøres i dag og som understrekes av mulighetsstudien) og det er et stort behov for å utvide antall indikatorer for at

administrative bibliometrikere skal kunne utnytte NIB bedre. Ferdig beregnede indikatorer bidrar også til standardisering av analyser og sikrer lik bruk ved institusjonene. Tilrettelegging i NIB er følgelig et innsatsområde som gir store gevinster for brukerne.

Kildevalg – avgrensende og bestemmende for analysemulighetene

Kilden som ligger til grunn for NIB vil sette rammene for alle tjenestene i hele systemet. Kildevalg er derfor helt sentralt, noe mandatet til dette arbeidet også påpeker. I tillegg har det vært en hypotese at den tunge tekniske belastningen på import er eksklusivt knyttet til dagens kilde. Belastningen er åpenbart de facto knyttet til WoS, men arbeidsgruppen vil likevel påpeke at tekniske utfordringer ikke nødvendigvis er knyttet til WoS per se. Tekniske utfordringer med import må derfor håndteres separat og vil i prinsippet kunne gjelde alle kilder. Det betyr også at diskusjonen om kildevalg ikke heftes av tekniske utfordringer som uansett må løses.

Før vår diskusjon av kilder, så noteres det at arbeidsgruppen bak mulighetsstudien stiller seg åpen for endring av dataleverandør, men vil prioritere ressursene til videreutvikling av NIB som tjeneste, fremfor en ny utredning av dataleverandør og ny anbudsprosess. Mulighetsstudien fant det (derfor) lite hensiktsmessig å vurdere andre datakilder for dette formålet.

Vurdering av mulige nye kilder

Valg av kilde kan like gjerne bety valg av «kilder» i flertall. Det er ikke gitt at en enkelt kilde kan levere data på alle de bruksbehovene som er formulert eller som kan bli aktuelle i fremtiden. Det er fornuftig å skille mellom *supplerende* kilder og *berikende* kilder, der supplerende kilder øker antall bibliografiske poster, mens berikende kilder supplerer mer informasjon på eksisterende poster. Hovedkilden må være dekkende nok uten at det er behov for supplerende kilder, og bør også helst minimere behovet for berikende kilder, hvis mulig.

Vurderingen av mulige hovedkilder starter heller ikke i et vakuum. NIB er basert på en eksisterende kilde og Sikt er i et kontraktsforhold med en kommersiell leverandør som løper ut 2024 med mulighet for forlengelse nye tre år. Da vi uansett betaler for nåværende kilde i minst et år til, skal et akutt bytte av kilde i tilfelle begrunnes svært godt. Transaksjonskostnadene vil bli betydelige, og omfatte mer enn bare et nytt anbud, det vil også være endringer for brukerne som skal forholde seg til full ny implementasjon av NIB. Bibliografiske poster er i all hovedsak lik fra kilde til kilde og forholdsvis bestandig, men antall siteringer vil variere relativt mye pga. ulike konvensjoner for hva som er en relevant sitering i de ulike kildene. Med andre ord; en kilde kan ikke ta over der en annen slutter, tidsserier må typisk gjenskaper fra bunn. Andre forhold enn kun de rene kilde-spesifikke forhold vil derfor spille inn i vurderingen om dagens skal skiftes ut. Vi har delt inn vurderingskriteriene i følgende hovedgrupper:

- *Ressursmessige hensyn*
Årlige lisensutgifter, oppstartskostnad, krav om anbudsprosess, omfang av arbeid med data
- *Egnethet for brukerne*
Dekningsgrad og innholdstyper, dekningsgrad på norsk publisering, datakvalitet, lisens- og bruksbetingelser, egnethet for henholdsvis administrative brukere og fagbrukere
- *Åpenhet og samspill*
Bruk av åpne standarder, samspill med andre kilder og tjenester

I vår vurdering av hovedkilder vil vi i stor grad låne informasjon og vurderinger fra annet hold. Det er også begrenset hvor detaljert det har vært hensiktsmessig å dykke ned i kildene. Vi har heller ikke vurdert det juridiske i det å håndtere kilder som inneholder personnavn med tanke på GDPR og evt. krav om databehandleravtale. Vi tar derfor utgangspunkt i NIFU-arbeidsnotatet «Nye kilder til data

om forskningsaktivitet⁷» publisert i 2023 for å vurdere kandidater til hovedkilde som vi supplerer med forskning og andre rapporter vi har søkt frem. Kandidater til hovedkilde er WoS, Scopus, Dimension, Crossref, OpenAlex, Semantic Scholar, Scite, og OpenCitations

Initialt finner vi at Semantic Scholar og Scite bør begrenses til situasjoner hvor det er ønskelig med kvantitative vurderinger av semantisk innhold i tekst, og ikke tradisjonelle bibliometriske analyser. OpenCitations har også en litt annen innretning enn de andre og egner seg mer som komplementær (berikende) kilde enn som separat hovedkilde. Google Scholar er heller ikke en aktuell kilde da tjenesten ikke utleverer data eller tilbyr et api.

Dette gir oss i utgangspunktet en liste på fem kilder: er WoS, Scopus, Dimension, Crossref, OpenAlex

Ressursmessige hensyn

De fem aktuelle kildene kan deles i kommersielle og åpne kilder:

Kommersielle: WoS, Scopus, Dimension

Åpne: Crossref, OpenAlex.

Hovedanliggende med å skille mellom kommersielle og åpne kilder ligger på tre plan, der åpne kilder har en rekke fordeler:

- Åpne kilder er gratis å anskaffe, de har ikke lisensbetingelser som begrenser bruksområder eller antall (evt. samtidige) brukere, fortløpende oppdateringer av datagrunnlaget er også gratis. En åpen kilde har derfor et langt større potensial for mer bruk for lavere kostnad enn de kommersielle.
- Åpne kilder er transparente. Sammenlignet med proprietære kilder muliggjør åpne kilder mer ansvarlig og etterprøvbar bruk av bibliometri i forskningsevaluering og i politikktutforming, spesielt gjennom at analysene blir mer gjennomsiktige og reproducerbare. Hvem som helst kan etterprøve analyser da data er åpent tilgjengelig.
- Rent merkantilt er ikke behov for en anbudsprosess ved anskaffelse av en åpen kilde. En anbudsprosess vil innebære betydelige transaksjonskostnader som i beste fall vil øke belastningen på NIB. Likevel bør anskaffelsesreglementet konsulteres så man ikke trår feil f.eks. i regler for statsstøtte.

Egnethet for brukerne

En grovkornet oversikt over dekningsgrad og bruksbetingelser ser slik ut:

Faktor/kilde	Dekningsgrad (Dokumenter)	Dekningsgrad (Siteringer)	Dekning av norsk publisering	Innholdstyper	Egnet for fagbrukere	Egnet for adm-bruker	Bruksbegrensning er (lisens)
WoS (core)	87M	1.8B	69 %	Forskning (mulighet for patenter)	ja	Ja	Ja
Scopus	87M	1.8B	72 %	Forskning	Ja	Ja	Ja
Dimension	140M	1.7B	?	Forskning, patenter, offentlige dokumenter, datasett, grants	Ja (?)	Ja (?)	Ja
Crossref	145M	1.45B	?	Forskning	?	?	Ingen
OpenAlex	243M	1.9B	?	Forskning, datasett, grants	Ja(?)	Ja(?)	Ingen

Dekningsgraden favoriserer OpenAlex og Dimensions med klar margin, men det hersker også noe usikkerhet om dekningsgraden på norsk publisering i kildene. Det er likevel grunn til å være optimistisk da en av styrkene til både OpenAlex og Dimensions er nettopp ikke-engelskspråklig

⁷ <https://nifu.brage.unit.no/nifu-xmlui/bitstream/handle/11250/3074742/NIFUarbeidsnotat2023-10.pdf?sequence=1&isAllowed=y>

innhold. Dette vil naturlig påvirke i hvilken grad kildene er egnet for spesielt administrative bibliometrikere. Dimension har i tillegg flere dokument-typer enn andre kilder, noe som muliggjør flere type utvidete analyser. Patenter som dokumenttype er også planlagt i OpenAlex-sammenheng og er også en opsjon i WoS. Kildene vurderes som like på siteringer, men eventuelle svakheter i dekningsgraden av det norske domenet, vil naturlig smitte over på bruk av siteringsdata også. Vi har ikke vurdert dekningsgraden på de enkelte metadatafeltene, eller hvordan felter varierer mellom kildene.

Når det gjelder bruksbetingelser er de åpne kildene i en egen klasse. Det påpekes likevel at strenge bruksbetingelsene for kommersielle kilder først og fremst rammer fagbibliometrikere som skal gjøre analyser i kommersiell sammenheng f.eks. ved oppdrag. Disse må da betale en separat avgift (royalties) for bruk av kilden til dette formålet. Oppdrag som NIFUs analyser til Indikatorrapporten er dekket av basisavgiften.

Åpenhet og samspill

Bruk av åpne standarder er helt i tråd med senere tids utvikling av bibliometriens ideelle grunnlag som igjen deles med åpen forskning. Dette kan ha konsekvenser for integrering med andre kilder, spesielt med andre åpne og supplerende kilder med samme delingskultur. Likevel er det vanskelig å konkludere i noen retning mellom kildene når det gjelder kompatibilitet. Vi tror at alle de aktuelle kildene har høy prioritet på bruk av standarder og PID'er og det er lite trolig at kildene vil skille vesentlig på dette området. OpenAlex er likevel kommet lengst i implementering av åpne registre som rygggrad i systemet og har et naturlig fortrinn som en åpen kilde.

I dagens NIB er det begrensninger på å berike WoS-data med informasjon fra andre kilder, det betyr f.eks. skranker for å berike WoS med Cristin/NVA-data, noe vi antar også vil gjelde andre berikende kilder. Det er lisensmessige begrensninger som ligger til grunn, ikke praktiske integrasjonsmuligheter. Vi antar likevel at dette i praksis ikke utgjør et stort problem, da det finnes enkle tekniske løsninger for å unngå «ulovlig» sammenkobling, dette er også noe som er gjort NIB. Dette vil likevel kunne spille inn hvis NIB skal berikes med mange datakilder.

Berikende kilder

Dagens kilde tilfredsstiller de fleste fundamentale krav og muliggjør dermed tradisjonelle bibliometriske analyser. Likevel tror vi mye av den fremtidige økte verdien i NIB vil ligge i berikende kilder og det vi har kalt utvidete analyser. Vi tenker da i første omgang på informasjon om, og referanser mellom forskning og henholdsvis patenter, offentlige dokumenter, sosiale medier og datasett. Dette er informasjonstyper, som når koblet til hovedkilden, muliggjør nye former for analyser utover tradisjonelle siterings-baserte impact-undersøkelser. Det muliggjør at Universitetet i Tromsø f.eks. kan finne tall på i hvilken grad egen forskning refereres til i sosiale medier eller i hvor stor grad norsk medisinsk forskning resulterer i innovasjon og patenter eller policydokumenter. Denne type bruk er i liten grad behandlet i mulighetsstudien, som i alle tilfeller vil «prioritere vedlikehold av eksisterende data, fremfor nye datakilder». Denne rapporten foretar ikke en vurdering av berikende kilder, men refererer til mulige kandidater i NIFU-arbeidsnotatet⁸, spesielt avsnittet om «kilder til data om annen forskningsaktivitet». Dette gir et godt utgangspunkt for å utvide og tilpasse NIB til behovene vi tror vil komme.

Vi vil likevel nevne OpenAIRE som en berikende kilde, dels fordi det er en tjeneste som Sikt er involvert i og som flere produktområder har interesser i; LÅT, NVA, og forskningsdataarkivene har alle forbindelseslinjer til OpenAIRE. Denne tjenesten har et like stort omfang av artikler som WoS og

⁸ <https://nifu.brage.unit.no/nifu-xmlui/bitstream/handle/11250/3074742/NIFUarbeidsnotat2023-10.pdf?sequence=1&isAllowed=y>

Scopus, men er likevel ikke egnet som hovedkilde, spesielt grunnet usikker kvalitet på siteringer. Tjenesten er derimot sterk på datasett f.eks. benytter Norges forskningsråd seg av OpenAIRE-tjenesten Monitor som bidrar med å koble forsknings output (datasett, artikler etc.) til NFR-finansierte prosjekter. OpenAIRE er derfor en mulig berikende kilde på datasett også for NIB.

Oppsummerende diskusjon og vurdering

Kildene våre starter ikke likt ut fra startstreken. WoS har et godt forsprang som hovedkilde i NIB med godt renommé. Bevisbyrden for endringer i hovedkilder må derfor forankres i sterke faglige eller ressursmessige argumenter. I så måte mener vi flere av kildene i listen kan ekskluderes som aktuelle kandidater. Det er få, om noen gevinster i å bytte WoS med Scopus. Begge kilder scorer ganske likt på alle parametere, men et bytte vil medføre store transaksjonskostnader ved en ny anbudsrunde. Videre er Crossref allerede innlemmet i OpenAlex (og også mange av de andre kildene), så denne kan også utgå. Reelt står vi dermed igjen med Dimensions og OpenAlex som mulige kandidater. Disse kildene regnes også for å ha et mer moderne design og en mer tjenlig og enklere innretning i datamodellen. Begreper som innovativ og moderne hektes altså på Dimensions og OpenAlex.

Dimensions:

På pluss-siden kan Dimensions skilte med svært god dekning og dokumenttyper som muliggjør flere analyser enn de tradisjonelle forskning-til-forskning-analysene. De har datasett (12 mill), patenter, (157 mill), online mentions (260 mill), policy dokumenter (2 mill) i tillegg stipend/grants og kliniske undersøkelser. Dimensions vil derfor støtte oppunder mye av behovet som vil komme i tillegg til tradisjonelle analyser og vil redusere behovet for berikende kilder.

På minus-siden, er Dimensions en kommersiell kilde som vil medføre begrensninger i bruk. I en komparativ studie med WoS og Scopus, kommer Dimensions ut med en høyere feilrate på klassifisering av institusjoner og land⁹, noe som vil ha betydning for analyser på dette nivået. Samme studie regner Dimensions for å være en reel konkurrent til WoS/Scopus på siteringer mens WoS/Scopus anses som sterkere på kvaliteten til siteringer i andre studier¹⁰. Dimensions har nå såpass lang fartstid at den kan regnes som trygg.

OpenAlex:

OpenAlex er basert på Microsoft sin tjeneste Microsoft academic graph (MAG), som Microsoft i sin tid sluttet å vedlikeholde. I perioden etter at OurResearch overtok MAG er datagrunnlaget forbedret på en rekke områder¹¹. På pluss-siden har OpenAlex omtrent tre ganger så mange dokumenter som WoS og Scopus, og 50 % flere enn Dimensions. Dekningsgraden er altså svært høy, selv om avstanden er mindre på det som regnes for å være vitenskapelig publisering. OpenAlex har datasett (3 mill) og finansiør-informasjon, men har ellers ikke informasjon på andre dokument-typer (patenter, offentlige dokumenter etc.) som ville gjort behovet for berikende kilder mindre. Tjenesten planlegger å legge til patenter som dokumenttype. OpenAlex er såpass ny at det er foreløpig få studier av datakvaliteten, men det finnes derimot undersøkelser og studier der MAG kommer godt ut i sammenligninger med andre kilder. MAG finner flere siteringer enn både WoS, Scopus og Dimensions¹² og er like egnet som Scopus for å bygge indikatorer¹³. Som med Dimensions er bildet likevel uklart på siteringer: en studie

⁹ <https://www.frontiersin.org/articles/10.3389/frma.2020.593494/full>

¹⁰ <https://direct.mit.edu/qss/article/2/1/20/97574/Large-scale-comparison-of-bibliographic-data>

¹¹ <https://arxiv.org/abs/2206.14168>

¹² <https://link.springer.com/article/10.1007/s11192-020-03690-4>

¹³ <https://link.springer.com/article/10.1007/s11192-017-2247-8>

konkluderer med at MAG er sterkere enn WoS, Scopus og Dimension¹⁴ mens WoS/Scopus anses som sterkere på kvaliteten til siteringer i andre studier¹⁵.

På minus-siden er det avdekket inkonsistens mellom OpenAlex og Unpaywall¹⁶. Dette er noe overraskende, da det er samme organisasjon bak begge tjenestene, men vi antar at det er knyttet til gjenstående arbeid i samkjøring av tjenestene.

OpenAlex regnes av mange for å være en gavepakke til det bibliometriske miljøet, men det kan likevel knyttes usikkerhet til organiseringen rundt kilden. Microsoft hadde store økonomiske muskler til å vedlikeholde MAG, mens OurResearch er en liten ideell organisasjon avhengig av økonomisk støtte og donasjoner. De har sikret drift gjennom sponsorer (primært Arcadia) i lang tid fremover, og har bevist sin kompetanse gjennom f.eks. Unpaywall og UnSub, der spesielt Unpaywall er blitt en helt sentral tjeneste som blant annet Sikt er avhengig av.

Internasjonal bevegelse mot åpne kilder

Det er stor interesse i å anvende OpenAlex i bibliometrisk forskning. Kilden har på svært kort tid etablert seg som viktig, ikke minst på grunn av størst dekning og åpen tilgjengelighet. Vi har allerede nevnt CWTS-miljøet ved Leiden og en OpenAlex-versjon av Leiden-rankingen, men også en rekke andre institusjoner enten forlater WoS til fordel for OpenAlex, eller supplerer WoS med OpenAlex. Det gjelder blant annet svenske institusjoner som Karolinska Institutet (som NIB arvet importrutiner fra) og Kungliga Tekniska Høgskolan. Det er også samtaler på gang om et nordisk bibliometrisk samarbeid basert på OpenAlex. I UK har JISC forfattet en rapport vi har fått tilgang til, der konklusjonen er en anbefaling om å bytte ut Dimensions med OpenAlex som datakilde. Dette skjer i en annen kontekst enn en rent bibliometrifaglig og som en del av oppdraget for å understøtte forhandlinger med forlag, likt det Norge gjør. Data og analyser er likevel overlappende og rapporten tillegges derfor noe vekt.

Oppsummering

Den ukjente faktoren for både Dimensions og OpenAlex er dekning av norsk publisering. Denne er kartlagt av NIFU for både WoS og Scopus¹⁷. Det er likevel grunn til å være optimistisk på vegne av dekningsgraden for begge kildene.

Av de kommersielle kildene i vårt utgangspunkt er det kun Dimensions som bør vurderes som alternativ til WoS, men til tross for langt større dekning og merverdi i form av patenter og andre dokumenttyper, har Dimensions begrensninger på land og institusjoner som må hensyntas. Det er også liten tvil om at åpne kilder har store fordeler foran de kommersielle på anskaffelse, bruksbetingelser og dekning, noe som taler i favør OpenAlex. Med andre ord; Hvis det skal skje endringer i valg av hovedkilde, er OpenAlex det alternativet som totalt sett peker seg ut.

Vi vil likevel understreke poenget med at sentrale miljøer innen bibliometri snarere supplerer, enn erstatter WoS med OpenAlex. Vi tolker det som at aktørene er i en vurderingsfase, der de kjører dobbelt med to hovedkilder. Dette er en mulighet som NIB også har, da vi allikevel er bundet til WoS ut 2024 og noe som diskuteres nærmere i scenario-delen av rapporten.

¹⁴ <https://link.springer.com/article/10.1007/s11192-019-03114-y>

¹⁵ <https://direct.mit.edu/qss/article/2/1/20/97574/Large-scale-comparison-of-bibliographic-data>

¹⁶ https://subugoe.github.io/scholcomm_analytics/posts/oalex_oa_status/

¹⁷ <https://sciendo.com/article/10.2478/jdis-2019-0001>

Tjenestenivåer

Da vi nå tror NIB 2.0 er tilstrekkelig definert, brukere, behov og kildegrunnlag er diskutert, kan vi se dette i relasjon til *tjenestenivå*. Disse vil bli definert etter mønster fra blant annet analyseplattformen i Sikt, der inndelingen skjer etter i hvilken grad det leveres tilfredsstillende løsninger for tilgang til data og senere analyse. Tjenestenivåene har altså fokus på det som faktisk hjelper brukerne med å utnytte NIB. Nivåene er også egnet til å sette krav til enkelt-elementer og tjenester, både de som er planlagt og påtenkt, men også de som leveres i dag.

Vi har notert oss følgende formulering i mandatet:

«Sikt skal sørge for at avstanden fra data til analyse er så kort som mulig av data. Investeringer i analysemiljøene ligger naturlig nok utenfor oppdraget til arbeidsgruppen, jf. også samfunnsoppdraget til Sikt.»

Vi tolker dette som at utvikling av dedikerte analyseløsninger ikke skal være en del av det Sikt tilbyr gjennom NIB. Samtidig peker mulighetsstudien på at å «Utrede konsept for Dashboard/analyseplattform, lage prosjektplan og jobbe med finansiering av en større utvikling av NIB» er en prioritert oppgave i det å møte behovet ved mindre institusjoner. Her finnes det potensielt en spenning mellom hva behovet faktisk er og hva Sikt skal levere. Vi tilnærmer oss dette via spørsmålet «Hvordan kan veien til analyser bli kortest mulig?». Spørsmålet har som premiss at Sikt ikke skal utvikle dedikerte analyseløsninger, men først og fremst sørge for en solid, velordnet, vasket og ryddig, beriket kilde i det norske bibliometriske økosystemet med ansvar for utlevering av data. Denne utleveringen må skje tilpasset de bibliometriske verktøy som brukeren benytter.

Vi ser for oss en grunnmur supplert med fire tjenestenivåer som i stor grad bygger på hverandre. Grunnmuren består av import av kilde, datamodell, datarydding og beregning av indikatorer og vil følgelig sette en maks-grense for hvor gode analysene kan bli i den andre enden, spesielt for administrative bibliometrikere. Vi anslår at tre fjerdedeler av å «gjøre veien til analyse så kort som mulig» skjer i grunnmuren. Tjenestenivåene vil videre være definert av hvor godt data er tilrettelagt for bruk og analyse.

Vi har også synliggjort mulighetsstudiens ønske om en analyseplattform på tjenestenivå 4 selv om dette per definisjon ligger utenfor Sikt sitt ansvarsområde. Dette kan i tilfelle finansieres av institusjonene selv, hvis de f.eks. ønsker å rette en slik bestilling til Sikt. Vi stopper derfor ved tjenestenivå 3, der vi ser at mye/mesteparten av NIB-potensialet kan realiseres ved å tilpasse NIB-data til eksisterende bibliometriske verktøy og løsninger. En vurdering av bibliometriske verktøy blir derfor gitt senere.

Grunnmuren og de fire tjenestenivåene har følgende karakteristika:

Grunnmur: *Import av og datarydding i datakilde(r), generering av indikatorer.*

Nivået karakteriseres av en velorganisert og fungerende base med data.

- Krever velfungerende basis-import
- Krever rutiner for å håndtere endringer/oppdateringsfiler
- Krever rutiner for datarydding
- Krever rutiner for tilrettelegging og beregning av indikatorer

Tjenestenivå 1: Tilgang til rådata enten ved SQL eller via datauttrekk i flate generiske filer på filserver. Nivået karakteriseres av rudimentær tilgang (SQL) der brukere kan ta ut egendefinerte datasett til analyser via spørringer eller laste ned datasett som filer. Disse vil per definisjon være på en slik form at det vil kreves stor innsats fra brukeren for å klargjøres for analyse.

Teknisk kompetansekrav: Høy. Bibliometrisk kompetansekrav: Høy

- Data tilbys gjennom et rudimentært tilgangsnivå (SQL)
- Data tilbys gjennom nedlasting av generiske datasett.
- Krever kjennskap til datamodell
- Verktøyvalg er opp til den enkelte bruker.

Tjenestenivå 2: Tilpassing/speiling av NIB-data i NVA gjennom eksisterende infrastruktur. Datautvalg er i dag naturlig limitert til data om norske publikasjoner.

Nivået karakteriseres av at utvalgte kolonner og indikatorer i NIB eksporteres og speiles i NVA. NIB-data blir dermed tilgjengelig for institusjonelle brukere gjennom eksisterende infrastruktur i Tableau (DUCT).

Teknisk kompetansekrav: middels. Bibliometrisk kompetansekrav: Høy

- Krever integrering med NVA (NVA-basen berikes med NIB-data).
- Utlevering av NIB-data skjer via eksisterende infrastruktur, dvs. DUCT/evt. Sikt sin analyseplattform.
- Avhengighet til NVA. Sikt tilbyr analysemuligheter gjennom generiske analyseverktøy som allerede er klar til bruk i andre sammenhenger (Tableau/evt. PowerBI). Ansvar for NIB stopper ved utlevering av data til NVA.
- Opsjon: Konfigurere NIB til å kommunisere med Tableau evt. annen infrastruktur som ordinær kilde.
- Opsjon: Utrede bedre utnyttelse av NVA-data i NIB (dypkobling på mer enn post-nivå, dvs. på forfatter og institusjon.)

Tjenestenivå 3: Utlevering av pre-genererte datasett tilpasset eksisterende bibliometriske verktøy, applikasjoner og rutinebibliotek.

Nivået karakteriseres av pre-genererte datasett som må regenereres jevnlig (i takt med oppdatering av basen) for import til lokale bibliometriske verktøy på egne pc'er. Analyser kjøres av den enkelte bruker på egen maskin. Datasett kan genereres og utleveres som flate filer eller via API.

Teknisk kompetansekrav: Middels. For egenutviklede rutiner: Høy. Bibliometrisk kompetansekrav: Høy

- Krever automatiske tekniske rutiner for generering av skreddersydde og tilpassede datasett som er etterspurt av brukerne og koordinert via Bibliometriforum.
- Krever et analysemiljø i eget rammeverk (R/Python eller lignende) på den enkelte brukers datamaskin.
- Krever en viss grad av koordinering av analysemiljø på den enkelte brukers datamaskin (bruk av eksisterende bibliometriske programpakker og tilpassing/opplæring i disse)
- Opsjon: Miljø for utveksling/validering av skript som er tilpasset datasettene
- Opsjon: Flere berikende kilder inn i NIB, utvikling av API
- Muliggjør avansert bruk av maskinlæring/KI, men analyser vil være begrenset av datamengde.

Tjenestenivå 4: Ferdig genererte rapporter og analyser i en dedikert analyseplattform.

Nivået karakteriseres av en web-tjeneste som tilbyr ferdig-genererte bibliometriske rapporter og analyser etter «trykk på en knapp»-prinsippet. Kan tilpasses autorisering, vil ikke være begrenset av datamengde.

Teknisk kompetansekrav: Lav. Bibliometrisk kompetansekrav: Middels

- Utenfor scope og Sikt sitt ansvarsområde.
Krever et større utviklingsløp av en webbasert tjeneste. Rammeverk på server kan velges forholdsvis fritt.
- Krever utarbeidelse av de rapporter og analyser som institusjonene bestiller og knyttes til aktiviteter på tjenestenivå 3.
- Krever autoriseringsrutiner
- Opsjon: Automatisk klassifisering av tekstkorpus ved bruk av maskinlæring. (Bør sees på som eget del-prosjekt).

Får å nå ambisjonene i 2.0-analogien må NIB opp på tjenestenivå 3, der vi mener «veien til analyser er blitt kortest mulig». Det er likevel en del handlingsrom innenfor de enkelte tjenestenivåene.

Illustrasjonen viser hvordan tjenestenivåene henger direkte sammen med forbedringer i grunnmuren, datarydding, koblinger og generering av indikatorer:



Neste avsnitt vil omhandle krav til de enkelte komponentene i tjenestenivåene, noe som også vil fortelle noe om hva som må gjøres med de eksisterende tjenestene.

Krav til tjenestenivåene

Hvis vi dykker ned i tjenestenivåene og ser på de enkelte komponentene, så har også de forskjellige nivåer av kompleksitet. Det kanskje viktigste valget er kilde som må velges ut fra egnethet, men dette er også i stor grad et separat valg i forhold til de mer tekniske kravene vi setter til tjenestene under. Med andre ord, uansett hvilken kilde som velges, så må denne bearbeides og tilrettelegges. Dette er et omfattende arbeid, som blir illustrert i kulepunkter under og noe som krever egen prosjektering i senere faser.

Krav til import (Vurderes som lik for alle hovedkilder. Primært krav til grunnmur)

- Retningsgivende: Det skal ta maks tre uker på å importere data
- Automatisert import med god feil-håndtering
- Ukesoppdatering skal i prinsippet importeres fortløpende
- God og tilgjengelig dokumentasjon av datamodell og rutiner
- Redundans av personell
- Teknisk driftssikkerhet

Krav til datarydding (Primært krav til grunnmur)

- All datarydding må skje sentralt og ikke som et ledd i brukers klargjøring av datasett for analyse.
- Institusjonsnavn desambigueres og ensrettes
- Landkoder desambigueres og ensrettes
- Forfatternavn desambigueres
- Samkjøring av andre ryddetjenester (primært NVA og rydding i institusjonsnavn, forfatternavn etc.)
- Kobling mot Cris/NVA
- Evt. berikende kilder vil ha tilsvarende prosedyrer tilknyttet.
- God og tilgjengelig dokumentasjon av rutiner

Krav til indikator-generering (Primært krav til grunnmur og tjenestenivå 1)

- Indikatorer spesifiseres og implementeres sentralt etter bestilling av brukerne. Kvalitetssikring av NIFU.
- Indikatorer regenereres fortløpende ved hver oppdatering av basen.
- God og tilgjengelig dokumentasjon av indikatorer

Utlevering av enkle generiske datasett for lokal bruk (Primært krav til tjenestenivå 1)

- Generiske datasett basert på utvalg bestilles fra brukerne, samordnes og genereres.
- Regenereres fortløpende.
- God og tilgjengelig dokumentasjon av rutiner og datasett

Speiling av data fra NIB til Cristin/NVA (Krav til tjenestenivå 2)

- Matching av poster i Cristin/NVA med NIB.
- Utvalgte felter eksporteres fra NIB til Cristin/NVA og importeres.
 - Omvendt eksport av data fra Cristin/NVA til NIB
- God og tilgjengelig dokumentasjon av kolonner, felter (datamodell)

*Utlevering av tilpassede datasett tilrettelagt for eksisterende bibliometriske tjenester/bibliotek.
(Primært krav til tjenestenivå 3)*

- Datasett som er tilpasset utvalgte analyseverktøy genereres fortløpende
 - Datasett kan i utgangspunktet være varianter av generiske datasett
- Vedlikehold av datasett skjer automatisk
- God og tilgjengelig dokumentasjon av datasett
- Kan utvikles til en api-tjeneste

Bruk av bibliometriske verktøy

Selv om NIB sitt ansvar i utgangspunktet ikke inkluderer analyseverktøy, er kunnskap om hva brukeren foretrekker essensielt for å levere gode tjenester. Eksisterende verktøy er møtepunktet mellom tilpassede datasett og brukernes behov på tjenestenivå 3. (En god oversikt over bibliometriske verktøy er publisert i reviewen «Software tools for conducting bibliometric analysis in science: An up-to-date review»¹⁸).

I prinsippet kan brukerne utføre analyser i Excel eller andre generiske analyseverktøy som PowerBI og Tableau. Likevel er det mer tilpassede verktøy som peker seg ut. Da bibliometri er en disiplin basert på kvantitative analyser og statistikk, har det vokst frem et miljø innen «R». Dette er et open source rammeverk og script-programmeringsspråk der ferdig utviklede spesialtilpassede rutinebibliotek kan lastes inn og anvendes. R regnes for å ha meget høy brukerterskel. R kan brukes uten å laste inn separate bibliotek, det er likevel meget stor nytteverdi i spesialtilpassede bibliotek for bibliometriske analyser. Det mest fremskredne prosjektet er kanskje BibliometriX¹⁹, som også har en webbasert analyseløsning. Det finnes en rekke lignende bibliotek for Python, som i likhet med R også krever høy teknisk kompetanse og tilsvarende ferdigheter i programmering. De fleste av de bibliometriske bibliotekene i R og Python er tilpasset de vanligste bibliografiske kildene, bibliotekene forutsetter altså at data er eksportert på det formatet kilde-uttrekk kommer på.

Både R og Python er også gode rammeverk for mer avanserte former for bibliometri som krever tekstprosessering^{20,21} av fulltekst/abstract (NLP). Det finnes svært mange relevante bibliotek som kan benyttes, selv om disse oftest er av en generell art og ikke spesialtilpasset bibliometriske oppgaver.

I tillegg til disse rammeverkene, finnes det også nedlastbare separate programmer som lar brukeren koble seg rett på et kilde-api eller importere lokale datasett. Selv om programmene i utgangspunktet ikke krever programmeringskompetanse, så krever de som minimum tilrettelagte datasett. Her er VOSviewer og CiteSpace to av de mest vanlige programmene som også er gratis å bruke. (CiteSpace finnes i varianter som krever betaling, men har en gratis basis-versjon).

Det er både betydelige muligheter og betydelige utfordringer ved disse løsningene, de setter store krav både til NIB og til brukeren. Fagbibliometrikere vil trolig navigere i dette landskapet forholdsvis lett, men som også mulighetsstudien påpeker, behovet for kompetanse, løsninger og støtte ligger hos de administrative bibliometrikerne. Felles for alle de nevnte løsningene, er at de oftest krever et bestemt format på datasettet som skal analyseres og det er denne tilpassingen som må skje på tjenestenivå 3.

¹⁸ <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2020.ene.03/47883>

¹⁹ <https://www.bibliometrix.org/home/index.php/blog/135-main-software-to-analyze-bibliometric-data>

²⁰ https://eprints.lse.ac.uk/86659/1/Benoit_Text%20analysis%20in%20R_2018.pdf

²¹ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8525079/>

Det er verdt å merke seg at behovet for analyseløsninger som kan brukes rett ut av boksen (som Insights og SciVal) er levedyktige nettopp fordi de bygger bro over kløften mellom kompetansen hos brukerne og grensesnittet som løsningene over tilbyr. Tjenestenivå 3 må derfor ledsages av mer enn bare teknisk tilrettelegging. Det vil derfor være et behov for koordinering og opplæring i alle de nevnte løsningene, noe som kan søkes løst og koordineres gjennom f.eks. Bibliometriforum.

Krav til dokumentasjon og dialog med brukerne

I forrige avsnitt betonet vi viktigheten av tilrettelagte datasett med ferdig genererte indikatorer for bruk i analyseverktøy. Hvilke indikatorer som skal beregnes, hvilke avgrensninger som skal foretas i datasett og hvilke verktøy datasettene skal kommunisere med, er avklaringer som må gjøres i dialog med brukerne. Dette må også dokumenteres.

Dialog og dokumentasjon

Sikt og NIB må ha en kontinuerlig dialog med brukerne for å levere tjenester tilpasset deres behov. Denne dialogen må formaliseres, koordineres, kvalitetssikres og dokumenteres. NIB må være behovsdrivet og alle leveranser må kvalitetssikres, primært gjennom NIFU på indikatorer og gjennom f.eks. Bibliometriforum (inkl. NIFU) for andre ting. Dette må være modus operandi for å bygge ut tjenestenivå 3.

Videre må kravet til dokumentasjon gjelde hele NIB og omfatte script, rutiner, databasemodell, databasefelter og endringer når de skjer. Det gjelder også dokumentasjon på alle avgjørelser som har konsekvens for NIB, brukerne, adgangskontroll/passord, teknisk rigg, programvare på server og driftsnivå. Det gjenstår et arbeid før kravet om dokumentasjon er tilfredsstilt.

Sårbarhet- og risiko-analyse

Ved årsskiftet 2020/2021 ble det gjennomført en ROS-analyse av NIB der spesielt to sårbarheter ble løftet frem. NIB er personavhengig, og det er en stor sjanse for lav eller ingen bruk av NIB på grunn av manglende tilrettelegging. Det siste punktet er av en annen karakter enn det som vanligvis påpekes i en ROS-analyse. Risikoen for manglende bruk representerer her en rekke forskjellige utfordringer knyttet til NIB. Mulighetsstudien tar opp mange av disse, i tillegg til at denne rapporten er ment å adressere disse utfordringene i mer detalj.

Sårbarhet og risiko med personavhengighet er en høyest reel utfordring for NIB, der en del av problemet kan bøtes på med god dokumentasjon slik ROS-analysen foreslår. Sårbarhet i forbindelse med personavhengighet kan likevel forstås på to måter: den første er tap av kjernekompetanse som igjen krever redundans, den andre er hvilke muligheter som finnes for utvikling av tjenesten, da det kun er en person som har ansvaret for, og kompetanse på tjenesten som helhet. Dokumentasjon kan bøte på det første, men ikke det andre.

NIB regnes ikke for å være en virksomhetskritisk tjeneste og har et forholdsvis lavt krav til oppetid på 80%. NIB kjører i dag kun på en enkelt fysisk server, uten test-, utvikling- og backup-miljø. Drift er i dag satt ut til UiO, men det bør avgjøres om server skal flyttes til Sikts eget driftsmiljø i Trondheim. Dette kan også gi kostnadsbesparelser. Et høyere driftsnivå bør avklares etter hvert som NIB brukes av flere.

Scenarier for å komme fra NIB 1.0 til NIB 2.0

Vi har valg å tegne opp 5 scenarier for NIB, som i større eller mindre grad vil leve opp til ambisjonene til NIB 2.0. Vi ser for oss tre skalerbare komponenter i scenariene som gir en viss grad av valgfrihet, og som har fordeler og ulemper når det gjelder kvalitet på tjenestene og innfrielse av ambisjonene for NIB. De tre komponentene er kildevalg, tekniske rutiner, samt tjenestenivå.

Kildevalg

NIB har 4 muligheter. Fortsette med WoS, bytte til Dimensions, bytte til OpenAlex eller supplere med OpenAlex. Vår vurdering vil basere seg på en total-vurdering som inkluderer ressursbruk, kilde-egnethet og fremtidsrettethet. Når det gjelder tidshorisonten, så vil scenariet med WoS som eneste kilde ha en horisont på 4 år som er ut kontrakten.

Tekniske rutiner

NIB må ha rutiner av teknisk art, inkludert import, datarydding, tilrettelegging for analyse, indikatorgenerering, generering av datasett for utlevering og speiling av data fra NIB til Cristin/NVA. Det er mindre handlingsrom på import og datarydding. Dette handler om datakvalitet og å legge best mulig grunnlag for tjenestene.

Tjenestenivåer

Tjenestenivåene kan ikke levere bedre enn kvaliteten i grunnmuren tilsier og som tidligere påpekt, for å nå ambisjonsnivået til NIB 2.0 må NIB levere på tjenestenivå 3. Det er likevel uvisst hvor langt man må strekke tilpassingen av datasett til eksisterende verktøy og hva behov og omfanget faktisk vil bli. Dette må avklares i dialog med brukerne.

Ressurser

Selv om en detaljert ressursvurdering vanskelig kan gjøres på dette nivået, vil vi likevel påpeke med grove estimater der vi vurderer et behov for nødvendige utgifter og/eller der det er et stort potensial for besparelser.

Vi tar utgangspunkt i en vurdering av et null-scenario som beskriver dagens situasjon. Der det er naturlig brukes skalaen under for å visualisere kvalitet og utviklingsgrad av tjenestene. Som utgangspunkt må en tjeneste være på nivå 4 for å være på et akseptabelt nivå, men likevel bør ambisjonene være å komme i «grønn sone».

Skala:	0	1	2	3	4	5	6	7	
--------	---	---	---	---	---	---	---	---	--

0: Null-scenario (ingen endringer)

NIB leverer i dag på grunnmuren og på tjenestenivå 1 og 2, men på rudimentært vis. Som mulighetsstudien og denne rapporten har påpekt, er vurderingen at dette ikke skjer på en teknisk tilfredsstillende måte og at NIB sliter med teknisk gjeld.

I tillegg har ikke løsningen for datautlevering med et generisk verktøy (DUCT/Tableau) som tilbys på tjenestenivå 2 den soliditet som kreves for å være et godt tilbud, blant annet fordi det er usikkerhet knyttet til denne løsningens fremtid. Dette er i alle tilfeller utenfor NIB sitt ansvarsområde.

Sårbarheter som personavhengighet er betydelig, i tillegg til en uakseptabel belastning på tjenesteanstalter.

Tjenester

Kildevalg									
• WoS som fortsatt kilde									
Grunnmur									
• Import									
• Datarydding									
• Tilrettelegging									
Tjenestenivå 1									
• SQL/Generiske datasett									
Tjenestenivå 2									
• Speiling av data fra NIB til Cristin/NVA									
Tjenestenivå 3									
• Tilpassede datasett til analyseløsninger									
Tjenestenivå 4									
• Analyseplattform									

Ressursvurdering

Ressurser-belastning	
• Frigitte personalressurser	Ingen
• Frigitte midler	Ingen
• Tilfang av ressurser	Ingen

1: «Minimums-varianten»_(Reparer import, spar tid, frigi ressurser)

NIB vil levere med styrket grunnmur og med samme kvalitet på tjenestenivå 1 og 2 som i forrige scenario. Scenariet karakteriseres av en full omskriving av importrutiner fra bunn av gjort av kyndig teknisk personell, men ellers foretas ingen forbedringer. Fordelen med scenariet er automatisering, kort importtid som kan måles i uker, og frigjøring av ressurser hos tjenesteansvarlig. Frigjorte ressurser kan i prinsippet omdisponeres til forbedringer andre steder, men det er usikkert hvor mye ressurser som frigis, spesielt i lys av overtidsbruk.

Minimumsvarianten gjør derimot få grep på datarydding eller tilrettelegging av kilden. Scenariet vil heller ikke hjelpe å gjøre «veien til analyse kortere» for brukere da det trolig ikke vil være mye rom for utvikling på tjenestenivå 3. Scenario 1 er ivarettatt i alle andre scenarier, og står derfor heller ikke godt på egne ben.

Tjenester

Kildevalg	Ressurser
• WoS som fortsatt kilde	-
Grunnmur	
• Import	2 månedsverk
• Datarydding	-
• Tilrettelegging	-
Tjenestenivå 1	
• SQL/Generiske datasett	-
Tjenestenivå 2	
• Speiling av data fra NIB til Cristin/NVA	-
Tjenestenivå 3	
• Tilpassede datasett til analyseløsninger	-
Tjenestenivå 4	
• Analyseplattform	

Ressursvurdering

Ressurser-belastning	
• Frigitte personalressurser	Betydelig
• Frigitte midler	Ingen
• Tilfang av ressurser	Krever ikke tilfang av faste ressurser

2: «Total-renovering»_(Behold WoS og bygg ut tjenestene)

NIB vil levere med solid grunnmur og på alle tjenestenivåer. Scenariet karakteriseres først og fremst av tekniske rutiner får en full overhaling og berikes med ny funksjonalitet. Det betyr at importrutinen omskrives fra bunn av, det prosjekteres rydding og vask av data, samt at data tilrettelegges på best mulig. Kvalitetshevingen på tjenestenivå 1 og 2 (DUCT) er «gratis» da mesteparten av jobben gjøres i grunnmuren. Tilpassing av datasett på nivå 3 gjøres sammen med brukerne og Bibliometriforum for å sikre at ambisjonene i NIB 2.0 tilfredsstilles. Det tekniske arbeidet vil være basert på erfaringsgrunnlaget fra arbeidet som gjøres i dag. Det bemerkes at WoS også kan levere data på patenter og andre dokumenttyper selv om dette ikke er med i dag. Dette vil øke egnetheten til WoS.

Tjenester

Kildevalg	Ressurser
• WoS	-
Grunnmur	
• Import	2 månedsverk
• Datarydding	1 månedsverk
• Tilrettelegging	2 månedsverk
Tjenestenivå 1	
• SQL/Generiske datasett	-
Tjenestenivå 2	
• Speiling av data fra NIB til Cristin/NVA	-
Tjenestenivå 3	
• Tilpassede datasett til analyseløsninger	1 månedsverk
Tjenestenivå 4	
• Analyseplattform	

Ressursvurdering WoS

Ressurser-belastning	
• Frigitte personalressurser	Betydelig
• Frigitte midler	Ingen
• Tilfang av ressurser	Anslag faste tekniske ressurser for utvikling av tjenesten: 4 månedsverk per år

3: «Ny kommersiell NIB»_(Bytte kilde til Dimensions)

NIB vil levere med solid grunnmur og på alle tjenestenivåer. En ny kilde betyr automatisk at alle tekniske rutiner må skrives fra bunn av, men dette vil likevel være basert på felles metoder og regelverk som i forrige scenario.

Scenariet er altså ganske likt som det forrige, men noe mer ressurskrevende da det tekniske arbeidet med WoS vil være basert på et eksisterende erfaringsgrunnlag. Den viktigste vurderingen er om Dimensions ansees som av såpass mye bedre kvalitet at det forsvaret et kildebytte. Vår vurdering av kilden var at dekningsgraden og andre dokumenttyper er attraktiv, men det er også usikkerheter som medfører risiko, vi har derfor under tvil gitt Dimensions ett hakk høyere karakter enn WoS. Grunnet nye kostnader kan Dimensions ikke implementeres før WoS-lisensen utløper i 2024, noe som i praksis betyr at minimumsvarianten (scenario 1) allikevel vil gjelde for 2024.

Et bytte til Dimensions vil medføre transaksjonskostnader med en anbudsprosess og en ulempe med ett kildebytte for brukerne. Det er også usikkert om Dimensions vil kunne innfri eventuelle krav om eierskap til data, en anbudskonkurranse risikerer at Dimensions må diskvalifiseres. Det er trolig at det vil være begrensninger på bruk, likt det WoS har.

Tjenester

Kildevalg	Ressurser
• Dimensions	-
Grunnmur	
• Import	2 månedsverk
• Datarydding	1 månedsverk
• Tilrettelegging	2 månedsverk
Tjenestenivå 1	
• SQL/Generiske datasett	-
Tjenestenivå 2	
• Speiling av data fra NIB til Cristin/NVA	-
Tjenestenivå 3	
• Tilpassede datasett til analyseløsninger	1 månedsverk
Tjenestenivå 4	
• Analyseplattform	

Ressursvurdering Dimensions

Ressurser-belastning	
• Frigitte personalressurser	Betydelig
• Frigitte midler	Kostnad på Dimensions er ukjent.
• Tilfang av ressurser	Transaksjonskostnader kildebytte: 2 månedsverk Anslag faste tekniske ressurser for utvikling av tjenesten: 4 månedsverk per år.

4: «Gå åpent»_ (Bytt kilde til OpenAlex)

NIB vil levere med solid grunnmur og på alle tjenestenivåer. Dette scenariet skiller seg heller ikke nevneverdig fra de to foregående rent teknisk, men et kildebytte til en åpen kilde vil gi umiddelbare besparelser fra 2025. OpenAlex er i likhet med Dimensions en attraktiv kilde som er gratis, med enda sterkere dekning, men OpenAlex som også har noen usikkerheter som gjør at vi ikke umiddelbart vurderer denne klart bedre egnet enn en totalrenovering av WoS. Et kildebytte kan i prinsippet iverksettes umiddelbart, hvis man velger å anse 2024-kostnader til WoS som tapt. Et bytte vil ikke kreve anbud.

Tjenester

Kildevalg	Ressurser
• OpenAlex	-
Grunnmur	
• Import	1 månedsverk
• Datarydding	1 månedsverk
• Tilrettelegging	2 månedsverk
Tjenestenivå 1	
• SQL/Generiske datasett	-
Tjenestenivå 2	
• Speiling av data fra NIB til Cristin/NVA	-
Tjenestenivå 3	
• Tilpassede datasett til analyseløsninger	1 månedsverk
Tjenestenivå 4	
• Analyseplattform	

Ressursvurdering OpenAlex

Ressurser-belastning	
• Frigitte personalressurser	Betydelig
• Frigitte midler	Betydelig
• Tilfang av ressurser	Anslag faste tekniske ressurser for utvikling av tjenesten: 4 månedsverk per år.

5. «Kjør dobbelt» (Ny og forbedret NIB- Supplere WoS med OpenAlex).

NIB vil levere med solid grunnmur og på alle tjenestenivåer. Scenariet er i all hovedsak en kombinasjon av scenario 2 med WoS som kilde og scenario 4 med OpenAlex. Denne kombinasjonen er mulig da NIB allerede er bundet av en kontrakt og utgifter. Scenariet gir også en mulighet til å starte med scenario 2 for senere bevege seg over i scenario 4 og skaffe seg et vurderingsgrunnlag for å bestemme hvorvidt NIB skal fortsette med WoS som kilde også i en ny treårsperiode.

Vår vurdering er at scenariet totalt sett gir en mer effektiv innføring av begge kilder, opp mot kildene én og én. Dette kommer av gode muligheter for overføring av både erfaringer og teknisk kompetanse, samt overlappende infrastruktur. Effektivisering vil, som i scenario 2, komme NIB umiddelbart til gode, men frigitte midler vil naturlig først komme ved en avgjørelse om eventuelt å fase ut WoS. Å ha to kilder kan likevel gi frigitte midler: å ha på plass OpenAlex vil være et godt argument ved en evt. reforhandling med Clarivate om lavere pris i neste treårs-periode av kontrakten.

Tjenester

Kildevalg	Ressurser
• WoS og OpenAlex	-
Grunnmur	
• Import	4 månedsverk
• Datarydding	1 månedsverk
• Tilrettelegging	2 månedsverk
Tjenestenivå 1	
• SQL/Generiske datasett	-
Tjenestenivå 2	
• Speiling av data fra NIB til Cristin/NVA	-
Tjenestenivå 3	
• Tilpassede datasett til analyseløsninger	1 månedsverk
Tjenestenivå 4	
• Analyseplattform	

Ressursvurdering OpenAlex

Ressurser-belastning	
• Frigitte personalressurser	Betydelig
• Frigitte midler	Avgjørelse utsettes
• Tilfang av ressurser	Anslag faste tekniske ressurser for utvikling av tjenesten: 4 månedsverk per år.

Dette er de fem scenarioene arbeidsgruppen har sett for seg som de mest realistiske og relevante. Det er også mulig å stokke om og bruke ulike komponenter på tvers av de opptegnede scenarioene slik de står her. Scenario 1 er som nevnt ivarettatt i alle andre scenarier. Scenario 5 kan i prinsippet være en kombinasjon av 4 og 1, hvis ressursituasjonen skulle tilsi det. Det bør da i tilfelle noteres at, etter arbeidsgruppens mening, vil det føre til at NIB ikke vil levere på tjenestenivå 3.

Anbefaling og oppsummering

Anbefaling

Vårt valg: Scenario 5

Det er ønsket om et fremoverlent NIB ligger til grunn for vår anbefaling av scenario 5. Her foreslår vi en opprustning av dagens WoS-kilde, men at denne suppleres med en ny hovedkilde i OpenAlex. Det betyr å kjøre dobbelt, som minimum ut inneværende kontrakts-periode for 2024.

Forslaget forutsetter flere premisser: WoS er allerede betalt for i minst et år til og leverer som intendert. Å kaste ut kilden før tiden til fordel for en annen, innebærer en stor risiko som vi ikke mener kan forsvares. Arbeidsgruppen mener snarere at muligheten for å kjøre dobbelt, med både WoS og OpenAlex som kilder, har tydelige fordeler. WoS representerer det etablerte og trygge, mens OpenAlex representerer fremtiden innen bibliometri.

At drivet ligger i åpne kilder, understrekes av at helt sentrale miljøer som Leiden satser på OpenAlex, og at Nordic Bibliometric Infrastructure har samarbeid om OpenAlex som overordnet tema i perioden 2024-2029. Vår anbefaling gjør NIB til en del av en internasjonal utvikling og best rustet til være den ressursen som møter behovet til norske institusjoner fremover samtidig som vi sikrer oss med en solid og veletablert kilde. NIB er i en unik situasjon som har mulighet til sikre kontinuerlig drift med WoS og samtidig kjøpe tid til å vurdere OpenAlex i ro og mak gjennom 2024. Fra et rent faglig ståsted, vil det i tillegg være ønskelig at WoS også beholdes ut hele kontraktsperioden, altså også neste tre-års periode, men vi er innforstått med at dette er en avgjørelse som er avhengig ressurs situasjonen. WoS kan også levere data på patenter, noe som også kan være gjenstand for forhandlinger. Hvorvidt NIB skal fortsette med WoS eller ikke, er en avgjørelse som uansett bør forankres som hos NIFU og NFR.

Fra et ressursmessig ståsted på kort sikt er det derimot betimelig å spørre om hvorfor NIB skal ta inn en ekstra ny hovedkilde i tillegg til WoS, all den tid NIB sliter med store tekniske utfordringer som også vil kreve ressurser.

Ressursvurdering av scenario 5

Dette ser vi som et overkommelig hinder, arbeidsgruppen har derfor pilotert en import av OpenAlex for å vise hva som er mulig med små ressurser. I vårt «proof-of-concept» har vi lastet ned en komplett datadump, pakket ut filer, importert data inn i en database, med en funksjonell prototype av en OpenAlex-basert NIB som resultat. Dette ble gjennomført på under en uke. Selv om OpenAlex databasen har lavere kompleksitet enn WoS, inneholder den også nesten tre ganger så mange dokumenter. Dette forteller oss flere ting: å komme i gang med OpenAlex er ikke ressurskrevende. Dette er en konsekvens av spesial-tilpassede import script, en enkel databasestruktur og et moderne datadesign som OpenAlex leverer i en samlet pakke for nedlasting. Det vil i alle tilfeller kreve mer ressurser å gjøre OpenAlex produksjonsklar. Dette er ikke virkeligheten med WoS som har et mer konvensjonelt design og leverer rådata i XML-format, noe som krever en større innsats.

Budskapet er at scenario 5 er godt innen rekkevidde og utgjør de første skrittene mot en fremtidsrettet tjeneste.

Vi fraråder derfor et rent bytte fra konservative WoS til fremtidsrettede OpenAlex selv med store kostnadsbesparelser i sikte. Dette fordi vi mister en viktig overgangsperiode der vi kan vurdere OpenAlex i ro og mak og NIB får et kontinuitetsbrudd som må forankres i miljøet. For en forholdsvis liten investering (dvs. også ruste opp WoS), får vi et langt høyere nivå på tjenesten.

Uansett hvilke valg som treffes fremover, så må det allokeres ressurser til videreutvikling av teknisk art. Import, rydding og tilrettelegging av indikatorer og datasett er gjentakende oppgaver, men også noe som kommer som nye bestillinger fra miljøet, og som vil kreve nyutvikling. En kvalifisert gjetning er at opp mot et halvt årsverk langt på vei vil dekke behovet. Vi tror også det er gjenbruksmuligheter i intern organisering med andre tjenester (primært NVA) som gjør tilsvarende oppgaver.

NIB i Sikt

I tråd med mandatet er rapporten et dokument med søkelyset på det tekniske, på kilder, på ressurser og på muligheter og begrensninger som ligger i en database med bibliografiske data. Men et annet element er minst like viktig; NIB sin rolle i Sikt. NIB eksisterer i dag i et slags vakuum med en forholdsvis svak kobling til helheten i Sikt og liten koordinasjon med andre av Sikts tjenester. Vi tror det er mye å hente ved en sterkere forankring av NIB. Det innebærer å identifisere og utnytte grenseflater for å maksimere nytteverdien av NIB og andre Sikt-tjenester. Hvilke tjenester i Sikt kan NIB bidra til? Og motsatt, hvilke tjenester i Sikt kan bidra til NIB?

NIB har allerede grenseflater mot to produktområder - NVA og LÅT. Det er stort overlapp mellom data i NVA og NIB, men det er svært mye data i NIB som ikke hører hjemme i NVA. NIB har dermed et potensial til å berike NVA, både vertikalt og horisontalt dvs. både som import-kilde for flere poster i NVA og som berikende kilde til mer data på de postene som allerede finnes. NIB har også en direkte avhengighet til NVA, som gjennom DUCT, tilbyr brukerne data etter at de er speilet i NVA-basen. Det ligger også et potensial i gjenbruk av rutiner for rydding, vask og sammenkobling av samme type data. LÅT er i dag en «nyter» av NIB-data. LÅT henter informasjon om norske publikasjoner i NIB som igjen beriker datasettene LÅT bruker som grunnlag for forhandlingsforberedelser og statistikk på åpen tilgang. Et eksempel på nytteverdien av denne koblingen er å finne i en rapport fra JISC, Sikt sin storebror i UK. I rapporten (som publiseres i januar 2024) er Norge ett av kun to land blant over 20 inviterte, som har levert gode nok datasett til en global undersøkelse og kritisk analyse av transformative forlagsavtaler. Datasettet som LÅT har utlevert er beriket med data fra NIB. I fremtiden kan LÅT berike NIB med mer informasjon om åpen tilgang enn det WoS har i dag, så koblingen som går en vei i dag, kan gå begge veier i morgen.

Dette er eksempler på eksisterende grenseflater med Sikt-tjenester og er etter vårt syn synergier i praksis. Dette er poeng som har en organisatorisk side: NIB må kobles på hver gang relevante prosjekter er et tema.

Vi tror denne tenkningen går rett inn i datastrategien til Sikt, der 'standardisering', 'datadeling', 'et datadrevet Sikt', 'ny teknologi', og 'felles føringer' er alle relevante størrelser for NIB. NIB er tuftet på ideen om deling av sentralt standardiserte datasett, bruk av ny teknologi (i dette tilfelle til virksomhetsstyring) og vi har allerede argumentert for NIB sin rolle i «et datadrevet Sikt». Det er derfor et relevant spørsmål hva Sikt vil med NIB.

Utenfor Sikt er Bibliometriforum viktig. Som et brukerforum bør de ha en sterk stemme i styringsstrukturen etter mønster fra andre produktområder og som bestiller inn til NIB. Det er også Bibliometriforum som kan ta NIB over i en eventuell fremtidig utvikling av nye analyseløsninger, altså opp på det tjenestenivået som Sikt ikke skal betjene, men som mulighetsstudien ønsker. I forlengelsen av NIB og Bibliometriforum foregår det viktig internasjonalt samarbeid, spesielt i Nordic Bibliometric Infrastructure (NBI), der det norske bibliometrimiljøet kan hente store gevinster i samarbeidet med nordiske partnere.

Vi ser følgelig for oss en NIB som er forankret både i og utenfor Sikt og som evner å snu seg rundt og møte behovene i fremtiden på en smidig måte. NIB må ha kapasitet og beredskap for omstilling, f.eks. ved å inkorporere berikende kilder ved behov og kunne levere ut tilpassede datasett til brukerne på bestilling. Dette krever ressurser slik at man ikke blir hengende bakpå fordi brannslukking og historiske utfordringer sluker eksisterende kapasitet.

Oppsummering

Som indikert i scenario-diskusjonen, så er det noen elementer som vi tror må gjøres i sin helhet, men også flere elementer som er «tøyelige» og som kan strekkes ut i tid. Tilpassing av datasett vil ikke komme som en komplett ferdig bestilling, men strekke seg over tid i porsjoner, datarydding kan også skje i etapper med stor overføringsverdi mellom kildene.

Listen under er derfor tiltak som samlet vil berede grunnen for en NIB som leverer data til bruk i virksomhetsstyring og andre formål, spesielt for den gruppen av brukere som i dag er avskåret fra tjenesten. Da tror vi at Sikt har «gjort veien til analyse så kort som mulig».

På kort sikt (2024)

- Integrer NIB ytterligere i Sikt sin tjenesteportefølje og sørg for at NIB bidrar til, og kan utnytte ressursene i andre Sikt-tjenester.
- Forankre arbeidet med NIB 2.0 ytterligere i Bibliometriforum. Sørg for forventningsavklaring spesielt med tanke på analyseplattform.
- Avdekke behov for kompetanse og analysetjenester sammen med Bibliometriforum.
- Prosjektering: Omskriv import av WoS fra bunnen av. Krever ca. 3-4 månedsverk av utvikler.
- Prosjektering: Utarbeid rutiner for datarydding og tilrettelegging av WoS i ny teknisk drakt, inkluderer tilrettelegging av en første pulje med datasett tilpasset eksisterende bibliometriske verktøy. Krever 2-3 månedsverk av teknikere.
- Sikre samkjøring og synkronisering med NVA når Cristin avløses.
- Sikre dagens DUCT-struktur, så dagens tilbud ikke faller mellom stoler men fortsettes.
- Sette opp en instans av OpenAlex i basisversjon. Krever ett månedsverk, kan håndteres internt i samarbeid med LÅT for å få status som beta-tjeneste.
- Utarbeide og sikre dokumentasjon av alle ledd i tjenesten.

På litt lengre sikt (2024 ->)

- Sette OpenAlex i full produksjon.
- Vurdere å beholde WoS som kilde, eventuelt reforhandle kontrakten for neste tre-årsperiode med forventninger om langt lavere pris og/eller flere datatyper som ikke leveres i dag.
- Når drift og dokumentasjon er på plass, berede grunnen for berikende kilder.
- Prioritere berikende kilder med dokumenttyper som ikke er dekket (patenter, offentlige dokumenter etc.). Vår første kandidat er OpenAIRE, men bibliometri-forum og brukerbehov bør være styrende.
- Koordinere evt. utvikling av analyseplattform i NIB med ekstern finansiering.