

# Behovs- og mulighetsstudie for Nasjonal infrastruktur for bibliometri (NIB)

Desember 2022

# Forord

Prosjektet Nasjonal infrastruktur for bibliometri (NIB) ble etablert i 2017 med finansiering fra Kunnskapsdepartementet. Målet med prosjektet var å anskaffe bibliometriske data for hele verden fra minst en leverandør og etablere en enkel infrastruktur for raskest mulig å gjøre disse dataene tilgjengelige for pilotinstitusjoner med egen kompetanse innenfor bibliometrifeltet og til slutt koble disse dataene med data fra Cristin. Prosjektet ble realisert høsten 2019.

NIB har vist seg nyttig i flere ulike sammenhenger, men vi ser at dagens løsning begrenser bruken da løsningen krever betydelig teknisk kompetanse. Derfor ble det nedsatt en arbeidsgruppe som fikk i oppgave å beskrive hvilke muligheter så lå i å videreutvikle NIB.

Denne behov og mulighetsstudien er utarbeidet av en arbeidsgruppe bestående av deltakere i Bibliometriforum<sup>1</sup>. Arbeidsgruppen er oppnevnt av Tjenesterådet for forskningstjenester. Arbeidsgruppens mandat innebærer blant annet å gi råd om bruk, drift, forvaltning og videreutvikling av NIB<sup>2</sup>.

Arbeidsgruppen hadde oppstart januar 2022 og har avholdt digitale møter, ett arbeidsmøte i Oslo, samt studietur til Stockholm med besøk hos Vetenskapsrådet, Kungliga Tekniska högskolan (KTH) og Karolinska Institutet. Arbeidsgruppen presenterte behov og mulighetsstudien i Bibliometriforum 25.-26. april 2022, og 22.-23. november 2022.

Arbeidsgruppen har i forbindelsen med utarbeidelsen av Behovs- og mulighetsstudien tatt inn et ekstra medlem fra UHR og har bestått av følgende medlemmer:

- Dag Aksnes, NIFU
- Jon Holm, Norges forskningsråd
- Margaret Louise Fotland, UiO
- Heidi Bodal, HiØ
- Caroline Susanne Armitage, UiB
- Tanja Larssen, UiT
- Bart Peeters, NTNU
- Vidar Røeggen, UHR (bidragsyter til studien)
- Hanne Hole Young, Sikt (sekretariat)
- Ivar Ternsell Torgersen, Sikt.

Marta Lorenz, UiB og Ida Eir Lauritzen, Sintef deltok i arbeidsgruppen i begynnelsen.

Arbeidsgruppen presenterer i denne rapporten sin anbefaling, som innebærer en større videreutvikling av NIB. I en videreutviklet NIB vil institusjonene få en fremtidsrettet tjeneste der dataene er ferdig koblet mot andre tjenester og benyttes direkte i andre systemer, og der krav til teknisk kompetanse hos bruker reduseres betydelig. Arbeidsgruppens løsning innebærer en økning fra dagens ressursbruk, fra 1,5 årsverk til 3,5 årsverk.

<sup>1</sup> Bibliometriforum er et nasjonalt nettverk innenfor bibliometrifeltet. Nettverket er åpent for alle som er involvert i bibliometri i norske forskningsinstitusjoner i alle de forskningsutførende sektorene.

<sup>2</sup> Se vedlegg 1 Mandat for arbeidsgruppe for tjenesten Nasjonal infrastruktur for bibliometri (NIB)

## Oppsummering

## Innhold

|       |                                                                                             |    |
|-------|---------------------------------------------------------------------------------------------|----|
| 1     | Innledende presentasjon.....                                                                | 5  |
| 1.1   | Fremtidig potensial.....                                                                    | 5  |
| 2     | Bakgrunn og status.....                                                                     | 6  |
| 2.1   | Forskningspolitisk kontekst.....                                                            | 6  |
| 2.2   | Behovs- og mulighetsstudien.....                                                            | 7  |
| 2.3   | Prosjekthistorie og status.....                                                             | 7  |
| 2.4   | NVI-indikatoren.....                                                                        | 8  |
| 3     | Brukere og behov.....                                                                       | 9  |
| 3.1   | Forskningstrender og samarbeid.....                                                         | 11 |
| 3.2   | Bibliometrisk forskning og kunnskapsoppsummeringer.....                                     | 12 |
| 3.3   | Rapportering og akkreditering.....                                                          | 12 |
| 3.4   | Strategiske evalueringer og styringsinformasjon.....                                        | 12 |
| 3.5   | Vurdering av forskere og prosjekt.....                                                      | 14 |
| 3.6   | Gjenbruk av løsningene til aktører utenfor academia, innovasjon og forskningssektorene..... | 14 |
| 4     | Kapasitet og kompetanse.....                                                                | 15 |
| 4.1   | Bibliometrisk kompetanse.....                                                               | 15 |
| 4.2   | Nasjonalt nettverk – Bibliometriforum.....                                                  | 15 |
| 4.3   | Tilgang til forskningskompetanse på høyt nivå.....                                          | 16 |
| 4.4   | Internasjonalt nettverk.....                                                                | 16 |
| 4.4.1 | Forskning.....                                                                              | 16 |
| 4.4.2 | Forum for egenutviklede bibliometriske databaser.....                                       | 16 |
| 4.5   | Dokumentasjon.....                                                                          | 17 |
| 5     | Policy og etikk.....                                                                        | 17 |
| 5.1   | Etikk.....                                                                                  | 17 |
| 5.2   | Policy for datakvalitet.....                                                                | 17 |
| 5.3   | Deling og gjenbruk av data.....                                                             | 18 |
| 6     | Datakilder og tilrettelegging av data.....                                                  | 18 |
| 6.1   | Datakilder.....                                                                             | 18 |
| 6.2   | Berikende (nye) kilder.....                                                                 | 19 |
| 6.3   | Datasiteringer.....                                                                         | 19 |
| 6.4   | Beregnete verdier – Indikatorsett.....                                                      | 19 |

|      |                                                      |    |
|------|------------------------------------------------------|----|
| 6.5  | Algoritmiske beregninger .....                       | 19 |
| 6.6  | Dugnadsverdier og synergieffekter .....              | 20 |
| 6.7  | Vask av data .....                                   | 20 |
| 6.8  | Kobling av data .....                                | 20 |
| 6.9  | Tilrettelegging og AI .....                          | 21 |
| 6.10 | Nye egne data .....                                  | 21 |
| 7    | Tilgang til dataene .....                            | 21 |
| 7.1  | Data via DUCT – Sentralt bibliometrisk verktøy ..... | 21 |
| 7.2  | CRIS/NVA .....                                       | 22 |
| 7.3  | Datasett .....                                       | 22 |
| 7.4  | API .....                                            | 22 |
| 7.5  | Analyseplattform .....                               | 22 |
| 7.6  | Dashbord .....                                       | 23 |
| 7.7  | Direkte tilgang til detaljerte data .....            | 23 |
| 8    | Bærekraftig drift av NIB .....                       | 24 |
| 9    | Oppsummering og anbefalinger .....                   | 25 |
| 9.1  | Andre prioriteringer og anbefalinger .....           | 27 |
| 10   | Appendiks – Eksempler på bruk av NIB .....           | 28 |
| 10.1 | Tematiske analyser .....                             | 28 |
| 10.2 | Siteringsanalyser .....                              | 29 |
| 10.3 | Analyser av åpen tilgang .....                       | 29 |
| 10.4 | Bruk av NIB ved Forskningsrådet .....                | 30 |
|      | Vedlegg 1 .....                                      | 31 |
|      | Innledning .....                                     | 31 |
|      | Mandat .....                                         | 31 |
|      | Sammensetning .....                                  | 32 |
|      | Funksjonstid .....                                   | 32 |

# 1 Innledende presentasjon

Nasjonal infrastruktur for bibliometri (NIB) ble etablert som prosjekt i 2017, basert på finansiering fra Kunnskapsdepartementet. NIB utgjør en database med bibliometriske verdensdata fra Web of Science (WoS) og har vært operativ siden 2019. Det er Sikt som har vært ansvarlig for opprettelsen og drift av infrastrukturen og for å legge til rette for bruk av dataene for norske institusjoner og institutter.

Siden opprettelsen har databasen vært brukt i flere ulike sammenhenger, blant annet til produksjon av Indikatorrapporten, til evalueringer og i forskningsprosjekter. De store mulighetene databasen gir, kan imidlertid utnyttes enda bedre, både til forskningsformål og til analyser ved forskningsinstitusjoner. Det ligger til rette for store synergieffekter gjennom tettere samarbeid mellom de miljøene som benytter seg av bibliometriske data i Norge, både når det gjelder forskningsrettet og administrativ bruk av bibliometri, i forbindelse med forskningsevaluering og i arbeidet med strategisk utvikling ved institusjonene.

Dette dokumentet beskriver noen av mulighetene som ligger i videreutvikling av tjenesten og tilgang til data. Det beskrives ulike grader av videreutvikling, fra dagens nivå, via en moderat videreutvikling til løsningen som anbefales, en større videreutvikling:

- **Dagens løsning:** Dagens løsning dekker behovet til nasjonal statistikkproduksjon, analyser og forskning. Løsningen fungerer godt til denne bruken. Den stiller imidlertid store krav til teknisk kompetanse og lener seg tungt på bibliometrisk kompetanse, og mangelen på et moderne brukergrensesnitt gjør at NIB har en begrenset nytteverdi for mange av institusjonene. Dagens løsning er også svært sårbar på driftssiden. Løsningen vurderes derfor som lite bærekraftig.
- **Moderat videreutvikling:** Løsningen vil stille noe mindre krav til teknisk og bibliometrisk kompetanse ved institusjonene. Den vil kunne tilby tilrettelagte skript og datasett og enkel kobling mot Cristin-data og bidra til større driftssikkerhet i Sikt (øker fra 1,5 til 2,5 årsverk).

**Større videreutvikling (anbefalt løsning):** Her får institusjonene en fremtidsrettet tjeneste der dataene er ferdig koblet mot andre tjenester og benyttes direkte i andre systemer. Nye tjenester, datakilder og metodeutvikling tas i bruk fortløpende. Krav til teknisk kompetanse hos bruker vil med en slik løsning være lav og gjør at tjenesten lett kan tas i bruk hos flere enn de som allerede har spesialkompetanse. Løsningen innebærer en økning fra dagens 1,5 årsverk til minst 3,5 årsverk.

*En full utvikling av NIB er ikke tatt med i beskrivelsen, da det ansees som urealistisk å få gjennomført på nåværende tidspunkt.*

## 1.1 Framtidig potensial

Som denne rapporten skal vise, har NIB mange anvendelsesområder. Utfordringen er at tilgang til databasen krever kompetanse i SQL som gir en høy brukerterskel. NIBs store potensial ligger i økt bruk, gjennom enklere tilgang til data, med lavere brukerterskel. Dette krever to typer av tilrettelegging, hvor det viktigste er det behov for å utvikle grensesnitt som er tilpasset kompetansenivået til ulike brukere. Derne må dataene tilrettelegges og standardiseres, slik at alle ikke trenger å gjøre den jobben hver for seg for å kunne analysere dataene og eventuelt koble de til andre data. Her under tilrettelegging for bruk av dataene i andre system.

Sikt sin rolle (og innsats) for disse typene av tilrettelegging kan både være å utføre tilrettelegging selv og å organisere bidrag fra aktuelle brukere. For forsknings- og utdanningssektoren som helhet vil vi anta at det kan innkasseres en betydelig stordriftsfordel ved at en del av arbeidet med tilrettelegging utføres sentralt. Bedre tilrettelegging fra Sikt sin side vil sannsynligvis også føre til økt interesse fra ulike brukere til å bidra med sine løsninger og tilpassede datasett som kan deles med brukere som har liknende behov. Samtidig vil resultater, datavask og analyser som ulike brukere gjennomfører, kunne gjenbrukes av andre. På denne måten spares tid og ressurser.

For NIB og Bibliometriforum har det vært viktig med tilgang til forskningskompetanse, og dette vil også være viktig i fremtiden. Det bibliometriske forskningsmiljøet i Norge er lite og sårbart. Gjennom NIB er det investert i data, men det er ikke investert tilsvarende i bibliometrisk forskning. Per i dag finansieres en del bibliometrisk arbeid gjennom oppdrag, men dette gir i liten grad mulighet for faglig utvikling og ny forskning. I Norges forskningsråds grunnlagsnotat fra 2016 om opprettelsen av den nasjonale infrastrukturen, ble det påpekt at man må ha oppmerksomhet på hvordan den grunnleggende bibliometriske forskningen i Norge skal finansieres. Dette behovet gjenstår fremdeles.

## 2 Bakgrunn og status

Bibliometri omfatter kvantitative studier av forskning, med utgangspunkt i forskningspublikasjoner. Opprinnelig ble slike studier primært drevet som en del av biblioteks- og informasjonsvitenskapen og til historiske og sosiologiske studier av vitenskap. Etter hvert har forskningspolitikk blitt et viktig anvendelsesområde, hvor to hovedformål er spesielt viktig: å evaluere forskning og dens resultater og et overblikk over utviklingen av forskning og publisering i fag, land og ved institusjoner.

### 2.1 Forskningspolitisk kontekst

Åpen forskning dominerer den forskningspolitiske agendaen internasjonalt. Norge er aktiv deltaker og langt fremme i dette arbeidet. Overgangen til åpen publisering har åpnet døren for en langt bredere agenda som endrer hele kunnskapssektoren. Endringene knytter seg til hvordan forskning finansieres, evalueres, hvordan resultater deles og etterprøves.

I 2022 ble det forhandlet nye utviklingsavtaler mellom Kunnskapsdepartementet og de statlige universitetene og høyskolene. Utviklingsavtalene har som formål å bidra til høy kvalitet og en mangfoldig høyere utdanningssektor gjennom tydelige institusjonsprofiler og god arbeidsdeling mellom utdanningsinstitusjonene. Disse avtalene trenger gode data som kan gi informasjon om forskningsaktiviteten. Sikt fremforhandler avtaler om åpen tilgang med de store forlagene og forhandlingene understøttes av gode data om hvor norske forskere publiserer. Innen 2024 skal all offentlig finansiert forskning være åpen, og vi trenger gode data for å følge utviklingen og evaluere tiltakene som skal bidra til å oppnå dette målet. Norge er i ferd med å reformere hvordan vi vurderer akademiske karriereløp og har utviklet en veileder der mer åpenhet, større bredde i vurderingene og en helhetlig tilnærmingssåte som motsats til ensidig bruk av enkelte indikatorer er et sentralt utgangspunkt. En bredere dokumentasjon av forskningsaktiviteten vil være avgjørende for å realisere dette arbeidet. Sist men ikke minst trenger Norge gode data til å støtte opp om et aktivt forskningsmiljø som kan studere og løpende evaluere de endringsprosessene vi er inne i. NIB har en naturlig plass for å understøtte alle disse endringene.

Flere av disse policy-agendaene er fanget opp i en nylig publisert avtale om reform av forskningsevaluering som er utarbeidet i samarbeid mellom Science Europe, European University Association (EUA) og EU-kommisjonen. Visjonen for reformen er formulert som følger (s.2):<sup>3</sup>

Our vision is that the assessment of research, researchers and research organisations recognises the diverse outputs, practices and activities that maximise the quality and impact of research. This requires basing assessment primarily on qualitative judgement, for which peer review is central, supported by responsible use of quantitative indicators. Among other purposes, this is fundamental for: deciding which researchers to recruit, promote or reward, selecting which research proposals to fund, and identifying which research units and organisations to support.

Som det fremgår gjelder reformen vurdering på mange ulike nivåer. Reformens mål er å styrke kvalitativ evaluering basert på et bredt sett av data som i større grad representerer mangfoldet i forskningsinstitusjonenes aktiviteter: forskning, utdanning, formidling og innovasjon.

Norge er langt fremme i dette arbeidet og en arbeidsgruppe nedsatt av UHR har utviklet en veileder for vurdering av akademiske karriereløp ([NOR-CAM](#)), som ble lansert i 2021. Det er også referert til denne rapporten i den europeiske avtalen, som et av eksemplene på hvordan reformen kan gripes an. UHR har etablert et nasjonalt NOR-CAM-nettverk som skal ha som formål å dele erfaringer og beste praksis om utvikling av lokale institusjonelle versjoner av NOR-CAM-rammeverket og samtidig skape en arena for å diskutere praktiske og strategiske sider ved den europeiske avtalen.

Bibliometri vil fortsatt være relevant som en del av datagrunnlaget for å vurdere alle disse aktivitetene. Hva som skal regnes som ansvarlig bruk av bibliometriske data vil være en viktig diskusjon knyttet til bruk av NIB-data framover. Når det gjelder forskningsinstitusjonenes egen bruk av bibliometriske analyser, har Det nasjonale publiseringsutvalget (HK-dir) utformet en veileder for forskningsinstitusjoner og forskningsledere som alle som bruker NIB-data bør kjenne til.<sup>4</sup> Denne veilederen er også tatt inn i NOR-CAM rapporten.

## 2.2 Behovs- og mulighetsstudien

Behovsanalysen ser på hele det bibliometriske feltet, fra behov for bibliometriske data, vasket og tilrettelagt, til mulighet for koblinger mot andre data og systemer, tekniske løsninger og analyseplattform, forskning, samt kompetanse og nettverk. Hensikten er å beskrive hvordan vi kan ta ut større deler av potensialet som ligger i NIB og få gevinstrealisering ute i institusjonene.

## 2.3 Prosjekthistorie og status

I 2016 utarbeidet Norges forskningsråd et notat til Kunnskapsdepartementet om etablering av nasjonal infrastruktur for bibliometri. Notatet beskrev de grunnleggende forhold vedrørende datainfrastruktur, behov og forslag til løsninger. Notatet tok utgangspunkt i arbeidet til en arbeidsgruppe om opprettelse av en slik infrastruktur, opprinnelig foreslått av NIFU i 2014.

NIB består i dag av en teknisk infrastruktur, et datasett og flere ulike nettverk.

Prosjektet ble etablert 1. april 2017 med følgende mandat:

<sup>3</sup> [Agreement on reforming research assessment](#), European commission 20 July 2022

<sup>4</sup> [Om bruk av bibliometri på individnivå. En veiledning for forskningsinstitusjoner og forskningsledere](#), UHR-Publisering, mai 2020

- Anskaffe bibliometriske data for hele verden fra minst en leverandør
- Etablere en enkel infrastruktur for raskest mulig å gjøre dataene tilgjengelige for pilotinstitusjonene, som har egen bibliometrikompetanse
- Koble bibliometridata for hele verden med data fra Cristin.

Dette ble realisert høsten 2019, med 10 årganger med data fra Web of Science (WoS) levert av Clarivate. 97,5 % av alle poster i Cristin, hvor tidsskriftene indekseres i WoS, er matchet opp mot WoS.

Serverkapasiteten ble økt betydelig i 2021 for å kunne ta inn alle årganger med data (1980-2021). Det er i dag 13 aktive brukere av NIB. Dataene brukes hovedsakelig av NIFU til indikatorrapporten og til forskning. I tillegg brukes den til en doktorgrad (UiT) og UiB har brukt basen til ulike analyser. Vi er også i gang med å fase inn tre nye brukere (to nye institusjoner). Ut over dette har vi et ukjent, men dog betydelig, antall brukere av dataene gjennom [DUCT](#)<sup>5</sup>, som oppdateres med siteringstall og indikatorer fra NIB.

Det ble tidlig klart at NIB ikke bare kunne være en teknisk infrastruktur, men også måtte involveres i bygging av kompetansenettverk. Prosjektet startet allerede i 2016 med å bygge et nordisk nettverk med de som har tilsvarende egenutviklede bibliometriske databaser, for å sikre NIB tilgang til kompetanse. I 2019 ble Bibliometriforum etablert for å samle norske bibliometrikere, slik at vi sammen kan bygge analyse- og bibliometrikompetanse i Norge, utnytte NIB-dataene best mulig og lage gode analyser.

*Status i dag er at NIB har levert på det som var mandatet. Samtidig ser vi at mulighetene dataene gir ikke utnyttes godt nok. Gevinstrealiseringen kommer den dagen dataene er godt tilrettelagt, og alle kan ha en enkel tilgang til og nyttiggjøre seg av dataene. Til det må det investeres i mulighetene til å tilrettelegge data, hente ut data og analysere dataene.*

## 2.4 NVI-indikatoren

Kunnskapsdepartementet nedsatte et ekspertutvalg som fikk i oppdrag å gjennomgå dagens finansieringssystem for universitets- og høgskolesektoren (UH). Utvalget la frem sine anbefalinger 18. mars 2022, og disse har senere vært ute på høring i forskningssektorene. Utvalget anbefaler blant annet å fjerne forskningsindikatorene i dagens budsjettmodell, herunder publiseringsindikatoren. Instituttsektoren og helseforetakene, som også benytter publiseringsindikatoren i finansieringssammenheng, har valgt å opprettholde den. Flere har uttrykt bekymring for at datakvaliteten kan svekkes om indikatoren fjernes fra budsjettmodellen, eller at institusjonene vil kunne miste et insentiv for å opprettholde et høyt nivå på det som rapporteres. Metadataene i Cristin vil som resultat kunne svekkes. Det er imidlertid viktig å understreke at utvalgets anbefalinger knytter seg til indikatoren og dens kobling til finansieringssystemet. Utvalget er tydelige på at rapporteringen i UH-sektoren skal fortsette som før. Hvis metadataene svekkes som effekt vil NIB kunne bli et viktig supplement og muligens få en mer sentral rolle i å fremskaffe oversikt over institusjonens egne publikasjoner og hvem man samarbeider med ved publisering. Gjerne i et godt og flytende samarbeid med det nye forskningsinformasjons- og arkivsystemet NVA.

<sup>5</sup>Via DUCT kan en hente ut data fra Cristin, for bearbeiding. Tableau er et visnings- og analyseverktøy.



## 3 Brukere og behov

I Norge benyttes bibliometriske analyser i mange ulike kontekster. Dette omfatter blant annet beregning av statistikk og indikatorer på nasjonalt nivå, til utarbeiding av forskningsprosjekter og bruk i resultatbasert finansiering. Bibliometri er også brukt for å støtte opp om forskningsledelse, som datagrunnlag ved forskningsevaluering og til strategisk utvikling – både på nasjonalt og institusjonelt nivå. Behovene varierer med tanke på hvem som driver med analysene og hva som er formålet. Eksempler av brukere og behov er gitt i dette kapitlet.

Interessentene i Norge befinner seg på ulike nivåer i forsknings- og innovasjonssystemet. I utgangspunktet kan alle norske ikke-kommersielle forskningsinstitusjoner få tilgang til NIB. Følgende institusjoner er plukket ut som viktige interessenter:

- De opplagte interessentene er institusjonene i forskningssektorer: både små og store institusjoner innenfor UH-sektoren, instituttsektoren og helsesektoren. Institusjonene har behov for en egen analyseberedskap med dedikert personale, som kjøpt kan håndtere aktuelle problemstillinger og fremskaffe analyser som grunnlag for innsikt og viktige beslutninger. Institusjonene er med andre ord avhengig av gode analyser som belyser institusjonens forskningsaktivitet, som blant annet kan dokumentere institusjonens særegne profil og oppdrag. Bibliometri kan gi institusjoner på alle nivåer en oversikt over hvem man samarbeider med, hvem som bidrar til samarbeidet (medforfattere, prosjektledere), hva man forsker på, hvor åpen man er og hvilke fagmiljøer som er aktive innenfor sine felt. Videre kan forskningsmiljøenes økende behov for kunnskapsoppsummeringer med avanserte analyser og spørringer eller tekstutvinningsmetoder mot NIB betjenes. Bibliotekene sitter på en spesialkompetanse her som med fordel kan kombineres med bibliometrikompetansen. Noen institusjoner har allerede begynt å ta NIB i bruk i analyser, mens noen trenger andre innganger til NIB før den kan tas i bruk (kap. 7).
- Forskningsinstituttet NIFU er allerede en viktig bruker av NIB. NIFU har vært det sentrale fagmiljøet for utviklingen av fagfeltet bibliometri i Norge. Instituttet benytter NIB-data i mange ulike prosjekter, blant annet til statistikk og ved indikatorproduksjon, evalueringer, analyser av fag og fagfelt, samt forskningsprosjekter.
- Forskningsrådet har ifølge sine [vedtekter](#) ansvar for 'å sørge for evaluering av norsk forskning' (§2j). Dette ansvaret følges særlig opp gjennom fagevalueringer som gjennomføres med cirka 10-års intervaller for alle forskningsinstitusjoner som mottar offentlig støtte. Bibliometri inngår som en sentral del av datagrunnlaget for slike evalueringer, men også i Forskningsrådets forvaltning av egne bevilgninger (porteføljestyling). For fagevalueringene som gjennomføres innenfor MNT-feltet, helse og medisin, produseres de bibliometriske analysene av NIFU basert på NIB-data.
- Det nasjonale publiseringsutvalget har sitt mandat fra Kunnskapsdepartementet og har i oppgave å gi myndighetene bredt faglig forankrede råd i saker som berører vitenskapelig publisering, å fremme faglige diskusjoner og ansvarlig bruk av publiseringsstatistikk. Til å støtte opp om denne oppgaven benytter utvalget seg av forskerkompetanse i bibliometri (3 månedersverk fra et forskningsinstitutt), blant annet til å gjennomføre analyser av endringer i publiseringslandskapet, mønstre for samarbeid og for å følge overgang til åpen publisering. Utvalget har også ansvar for forvaltningen av publiseringsindikatoren, og i arbeidet med indikatoren behøves verdensdata som i dag hentes fra Scopus, som kan erstattes av NIB. Det

ligger et stort potensial for utvalget i å samarbeide tettere med NIB-miljøet rundt levering av slike data i fremtiden.

- SIKT har ansvaret for å organisere forhandlinger med forlagene om tilgang til vitenskapelige tidsskrifter. Disse avtalene omhandler de senere årene også adgang til å publisere åpent. I forberedelser til disse forhandlingene benyttes data om norsk publisering, for å se hvilke tidsskrifter og forlag som er viktige for norsk forskning. NIB vil kunne være et uvurderlig redskap for å understøtte forhandlingsoppdraget til SIKT.
- Kunnskapsdepartementet (KD) og andre departementer – NIB er finansiert av KD for å håndtere en komplett bibliometrisk database og legge til rette for bruk av denne i Norge. KD<sup>6</sup> uttrykte her en tydelig forventning til at bibliometri skulle brukes mer aktivt som en av kildene i rapportering og styring av forskningssektorene. I Norge har alle departement ansvar for forskning innenfor sitt fagområde. NIB kan levere datagrunnlag og analyser til bruk i departementene, ut fra departementenes fagområder eller til mer overordnet politikktutforming og koordinering, slik som på bærekraftsmålene til FN. Med bruk av knytningen mot PubMedCentral kan for eksempel Helse- og omsorgsdepartementet få mer spesialiserte sammenligninger mot utlandet innen spesifikke medisinske felt.
- Andre statsorgan og offentlige institusjoner – Innovasjon Norge kan for eksempel få nettverksanalyser til sine næringslivsbesøk i utlandet, slik at de kan oppdage og finne frem til forskning og forskere som det er relevant å besøke eller ta med på slike besøk. KS har fokus på innovasjon og kunnskap. De vektlegger at forskning skal skje i, for og med kommunene. NIB kan gi tallene på finansiering og publikasjoner knyttet til denne forskningen, og ikke minst hvilke nettverk forskerne har i dette arbeidet.
- Studenter og forskere. Bibliometri er også en vitenskapelig disiplin, og det drives mye forskning internasjonalt. I motsetning til i mange andre land, finnes det imidlertid ikke et utdanningstilbud innen området, men OsloMet tilbyr et kurs i vitenskapelig kommunikasjon som del av utdanningen i bibliotek- og informasjonsvitenskap, og her inngår noen tema om bibliometri. En kartlegging gjennomført av NIFU viser også at omfanget av forskningen i Norge er mer begrenset enn i flere andre nordiske land. I Norge er det primært NIFU som har stått for forskningen innenfor feltet, og bidraget fra UH-institusjonene er begrenset, noe som trolig skyldes mangel på kompetanse og fagmiljøer. NIB representerer en unik datakilde, og NIB kan tilby relevante data for studenter i bibliometri og tilknyttete felt på alle nivåer, inkludert doktorgradsstipendiater. Dette er en viktig for å sikre rekruttering til faget. NIB gir gode analysemuligheter fordi basen kan aksesseres med SQL. Å stimulere til økt bruk av NIB i forsknings- og utdanningssammenheng er noe det bør arbeides videre med.
- Allmenheten kan få tilgang til analyser fra dataene. Lisensen gir anledning til å opprette en Web-portal på norsk, hvor vi kan vise frem aggregerte data. Dette kan for eksempel være til nytte for en rådmann som ønsker å få oversikt over hvilken forskning som utføres i kommunen eller regionen, og hvilke nettverk disse forskerne har, for å kunne planlegge bedre. Det kan også være nyttig for frivillige organisasjoner som vil orientere seg innenfor eget interessefelt.

Arbeidsgruppa erfarer at mange UH-institusjoner og offentlige organer som eksempelvis Forskningsrådet står klare til å utnytte mulighetene i NIB etter som de kommer på plass. Lisensen for bruk av NIB-data omfatter alle ikke-kommersielle forskningsinstitusjoner som mottar offentlig støtte, samt alle offentlige organer. Det er rimelig å anta at flere av disse ville kunne ha nytte av bibliometriske analyser hvis det utvikles et forenklet brukergrensesnitt.

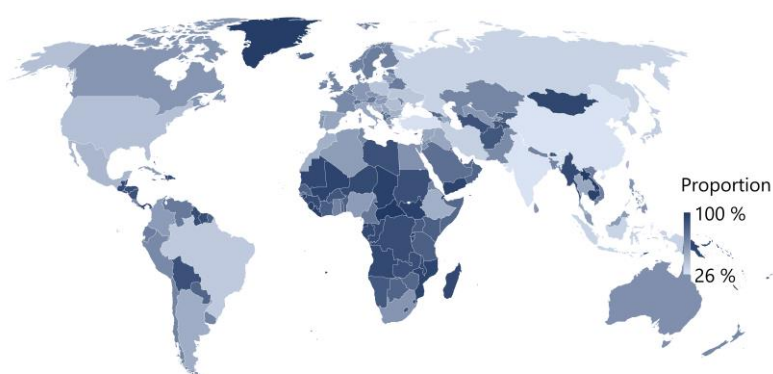
<sup>6</sup> Møte med KD fredag 21. april 2017.

Bibliometrimiljøet i Norge kan også ses som en interessant. NIB er et praktisk og formelt samlingspunkt for utvikling og utveksling av bibliometrisk kompetanse i Norge. Særlig for personale som driver analyse, men ikke er i ordinære forskningsstillinger, er kompetansenettverket rundt NIB viktig for fagutvikling og kvalitetssikring. NIB fungerer allerede i dag som en arena for kontakt mellom profesjonelle bibliometrikere og praktikere som arbeider med ledelsesstøtte og kunnskapsutvikling for strategi og politikk ved forskningsinstitusjoner og offentlige organer som Forskningsrådet.

### 3.1 Forskningstrender og samarbeid

Ved forskningsinstitusjoner, er det vanlig å få spørsmål som ikke kan besvares med «standard» analyser, som må besvares med tilpasset analyser av bibliometriske data. Slike forespørsel kan komme fra ledelse, forskningsadministrasjon, prosjektledere, og forskningsgrupper.

Et typisk oppdrag er analyse på samarbeid, internasjonalt eller med andre norske institusjoner. Slike analyser brukes i opprettelse av nytt samarbeid, strategiarbeid, eller besøk til og fra utlandet. Bibliometri kan gi en fagprofil som kan si noe om de institusjonene en besøker eller får besøk fra. Via nettverksanalyser kan institusjonene se hvilke fagområder en har til felles, eller hvilke forskere ved institusjonene som samarbeider. NIB sine samarbeidsdata er mer fullstendig enn de i Cristin, særlig for publikasjoner med mange forfattere, og tillater sammenligning med andre land (Figur 1).



Figur 1 Andel av publiseringen med internasjonalt samarbeid per land, 2020-2021. Kilde: NIFU. Indikatorrapporten bruker også NIB data til [internasjonale sammenligninger](#).

Tilsvarende finner vi igjen utenfor academia. Når Innovasjon Norge arrangerer næringslivsbesøk til utlandet, kan forskningssamarbeid være en del av dette. Bibliometri vil kunne gi en oversikt over hvilke nettverk som allerede eksisterer eller hvor begge land gjør det godt. Ved offisielle besøk eller statsbesøk kan regjering eller statsoverhode også fungere som døråpner innenfor forskning, så sant de har det rette kunnskapsgrunnlaget.<sup>7</sup>

Et annet eksempel på denne type aktivitet er forespørsel fra aktører som har en interesse i et bestemt forskningstema. For eksempel har «Plastnettverket» i Bergen oppsøkte hjelp fra bibliometrikere for å finne forskere som jobber med om plast, uansett fag, for å hjelpe med å få folk inn i nettverket. NIB gir tilgang til abstracts for publikasjoner i maskinlesbar form, som gir bedre muligheter til å identifisere

<sup>7</sup> I lisensen med Clarivate har vi mulighet til å dele denne type analyser på en åpen web-løsning (på norsk).

publikasjoner knyttet til et tema. Et viktig spørsmål i denne sammenheng er hvordan Sikt kan legge til rette for maskinlæring og bruk av algoritmiske metoder for Natural Language Processing (NLP).

### 3.2 Bibliometrisk forskning og kunnskapsoppsummeringer

Forskningen trenger full tilgang til detaljerte data og kan ellers trekke på den tilrettelegging og vasking som utføres med tanke på andre oppgaver og behov. NIB databasen omfatter data om mer enn 70 millioner publikasjoner tilbake til 1980. Det er dermed en verdifull datakilde også til forskningsformål. Basert på publiseringsdataene i NIB er det mulig å gjennomføre mange ulike typer vitenskapelige studier, fra rent bibliometriske, til studier som også omhandler andre tema, for eksempel vitenskapelig kommunikasjon, akademisk karrierer, kjønn i forskning og vitenskapshistorie.

Kunnskapsoppsummeringer omfatter diverse type litteraturoversikter, inkludert «systematic reviews» og «scoping reviews». Denne type analyser er det økende interesse for i instituttsektoren, helsesektoren og UH-sektoren. Behov for kunnskapsoppsummeringer, data og analyse er også utpekt i Langtidsplanen for forskning og høyere utdanning<sup>8</sup>. I slike analyser kan NIB benyttes som en fullstendig oversikt over Web of Science innhold. For eksempel, har NIFU gjennomført en analyse av diskriminerings- og likestillingsforskning i Norden<sup>9</sup>.

### 3.3 Rapportering og akkreditering

I flere sammenhenger produseres det ulike publikasjonslister, som med fordel kan tas direkte ut fra Cristin eller NIB. Institusjoner som ikke har selvakkrediteringsrett til alle studietilbud, må søke NOKUT om akkreditering av nye mastergrad- og ph.d-studium. Til disse søknadene skal både forsknings- og utviklingsarbeid kartlegges for fagmiljøene. Her er det ønskelig med rask og enkel tilgang til publikasjonsoversikter med nivå, og siteringer for forskernes CV'er. Siteringer er data som per i dag ikke finnes i Cristin men finnes i NIB.

### 3.4 Strategiske evalueringer og styringsinformasjon

Det er etterspørsel etter strategisk styringsinformasjon både ved institusjonene og i nasjonale forskningspolitiske organer. Forskningsrådet bruker systematisk data fra Cristin i evalueringer av norsk forskning. NIFU benytter data på tilsvarende måte i evalueringsprosjekter, et eksempel her er evalueringen av Strukturreformen.

Et eksempel på slik bruk er oppdateringen av forskningsstrategien for norske maritime næringer, Maritim21. Som grunnlag for denne oppdateringen fikk Forskningsrådet laget en analyse av norsk maritim forskning. Avgrensningen av dette forskningsfeltet finnes ikke ferdig kodet i noen kommersiell database. Forskningsrådet brukte derfor algoritmiske metoder for å identifisere hva som kan regnes som maritim forskning. Slike metoder kan både være basert på tekstanalyse (topic modelling) og siteringskjeder (bibliographic coupling eller co-citation)<sup>10</sup>. Fordi NIB-data ligger på en SQL-server med god beregningskapasitet, er den godt tilrettelagt for bruk av algoritmiske metoder.

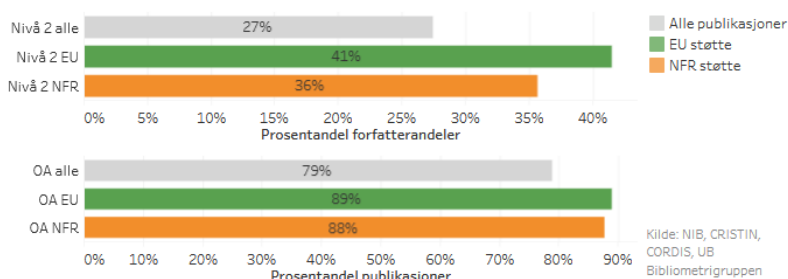
<sup>8</sup> [Langtidsplan for forskning og høyere utdanning 2023–2032](#), 3.5.1.8 «Kunnskap i kriser»

<sup>9</sup> NIFU - [Forskning om diskriminering, trakassering og likestilling](#)

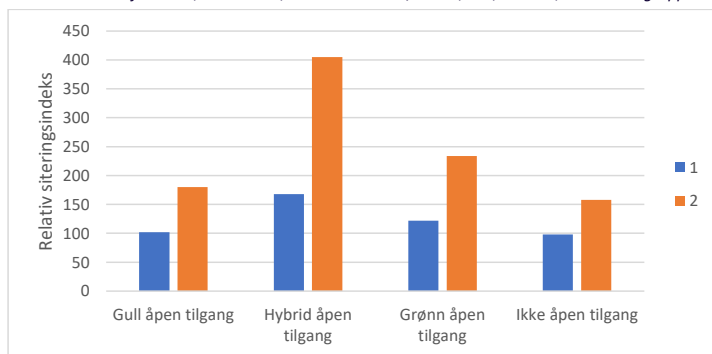
<sup>10</sup> <https://www.forskningsradet.no/siteassets/publikasjoner/1254021973383.pdf>  
<https://www.forskningsradet.no/siteassets/publikasjoner/2021/maritim-forskning-i-norge-2011-2020-bibliometrisk-analyse.pdf>

Behovet for informasjon til strategiske evalueringer omfatter også informasjon som institusjonsledelsen vil bruke strategisk i styring. Gjennom ulike analyser kan ledelsen følge med på utviklingen over tid samt benytte dette på et mer strategisk plan. Behovene er ulike alt etter hvilket nivå lederen opererer på.

Til dette formålet, er analyser basert på den årlige NVI-rapporteringen en typisk oppgave for bibliometrikere ved Cristin-institusjoner<sup>11</sup>. Cristin dekker som kjent en større del av norsk forskningsproduksjon enn de kommersielle databasene. NIB kan gi siteringstall og mer data om utenlandske samarbeidspartnere. Således kan Cristin og NIB berike hverandre. Siterings-, samforfatterskaps- og fagklassifiseringsdata fra NIB kan tas i bruk som del av årlig rapportering for å få mer innsikt. Den kan også være viktig for, eller muliggjøre andre strategiske analyser og evalueringer: for eksempel i interne evalueringer av tematiske satsningsområder og undersøkelser av effekt av ekstern finansiering (Figur 2). Siteringsanalyser, som ikke er mulig med kun Cristin data, er mulig med NIB data (Figur 3).



Figur 2 Prosentandel forfatterandeler på nivå 2 og prosentandel tidsskriftsartikler med åpen tilgang som krediterer Norges Forskningsråd eller EU som medfinansjør, 2019-2021, UiB. Kilde: DUCT/Cristin, NIB, CORDIS; Bibliometrigruppen ved UiB.



Figur 3 Relativ siteringsindeks for publisering med ulike typer åpen tilgang etter tidsskriftsnivå, 2016-2018. Kilde: NIFU, fra Indikatorrapporten 2021.

Utenom den årlige rapportering, kan bibliometriske data kobles mot andre ledelsesdata for spesifikke strategiske spørsmål. For eksempel kan bibliometrisk data kobles til stillingstype og data fra studieadministrasjonen for å undersøke publiseringsmønstre blant stipendiater, eller etter stillingskategori.

<sup>11</sup> Se eksempler her: <https://bibliometri.w.uib.no/>

NIB som datakilde er viktig i disse aktivitetene ikke kun fordi den gir ekstra data. NIB er også viktig fordi egne analyser kan tilpasses, og dermed gir den informasjonen som er strategisk viktig for institusjonen. Uten NIB og egne analyser er man mer avhengig av grovkornete internasjonale rangeringer, for eksempel Times Higher Education, eller egne institusjonelle abonnement til andre datakilder. Rangeringer er lite egnet for særskilte strategiske spørsmål (t.d. evaluering av et fagområde), siden de gir kun begrenset, ferdige tall. I tillegg er bruk av rangeringer er omstridt fordi de kan slå ut uheldig mot bestemte fagområder (for eksempel som har mye publisering på norsk), som likevel kan være fagområder av stor betydning for institusjon og samfunn. En kombinasjon av NIB og Cristin unngår disse problemene.

### 3.5 Vurdering av forskere og prosjekt

Vurdering og evaluering av forskere og forskning er et felt med stort potensial som er under utvikling. Tanken er at man skal vurdere bredere enn dagens evalueringer og vurderinger. EUA og Science Europe har forfattet en europeisk avtale på vegne av EU «Agreement on reforming research assessment» CoARA som beskriver en bredere måte å vurdere forskere og evaluere forskning. Avtalen er nå til vurdering hos institusjonene, og Forskningsrådet har allerede signert avtalen. I Norge har UHR i 2021 publisert rapporten «Veileder for vurdering i akademiske karriereløp» som er en veileder for vurdering i akademiske karriereløp. Mange institusjoner holder nå på å kartlegge egen praksis på feltet. UiO har allerede kartlagt egen institusjon i rapporten «Rapport fra arbeidsgruppe for vurdering av forskere» 2021. Rapporten støtter langt på vei UHRs rapport, samt at UiO har institusjonalisert matrisen NOR-CAM og kalt den UiO-KVM. Hovedbudskapet i rapportene er et sterkt ønske å fjerne bruk av journal impact factor og H-indeksen for søknader til stilling, opprykk og søknad til finansierer. Forslaget som kommer fram i rapportene er at det skal balanseres mellom kvantitativ og kvalitative mål. Bibliometriske indikatorer bør brukes med varsomhet og suppleres med annen informasjon. Alle resultater, aktiviteter og kompetanser vurderes i lys av mål om åpen forskning. De som skal vurderes må vite hvilke kriterier som vil benyttes og få innsyn i hvordan de er brukt.

Forskningsrådet innhenter per i dag ikke bibliometrisk informasjon i vurderingen av forskningsprosjekter. Det har vært drøftet om slik informasjon skulle innhentes på en standardisert måte for alle prosjekter for å unngå at fageksperter som deltar i søknadsbehandling søker opp enkle og mulig misvisende bibliometriske indikatorer fra Internett (H-index etc). Forskningsrådet vurderer for tiden om det skal gjøres endringer i CV-mal som går i retning av mer kvalitative beskrivelser av forskningskarrierer (narrativ CV).

### 3.6 Gjenbruk av løsningene til aktører utenfor academia, innovasjon og forskningssektorene

I utgangspunktet kan alle norske ikke-kommersielle forskningsinstitusjoner få tilgang NIB. I tillegg åpner lisensen som NIB har med Clarivate for en norsk web-portal, hvor vi viser aggregerte data. En åpen web-portal kan utvikles ved at vi gjenbruker løsningene som utvikles for NIB-institusjonene. Det avgjørende for Clarivate er at dataene som gjøres tilgjengelig er aggregert på et nivå hvor data fra enkeltpublikasjoner ikke kan gjenbrukes.

Det er ikke foretatt noen systematisk undersøkelse av behovet, siden dette ikke er noen primærbruker, men vi har uansett fanget opp noen eksempler. Kommunene vil kunne bruke NIB-data til å se på hvilken forskning som foregår i deres nærområde. Enkelte kommuner finansiere også forskning, og KS har vært opptatt av å følge slik forskning.

Innovasjon Norge arrangerer næringslivsbesøk til utlandet, og en bredere samarbeidsoversikt på forskningen kan bidra til at de kan lage et mer treffsikkert program.

## 4 Kapasitet og kompetanse

For å møte behovene trengs kapasitet og kompetanse på flere nivåer. Den utførende kompetanse og kapasitet internt i Sikt er vesentlig for å kunne møte behovene som er beskrevet i dette dokumentet. Arbeidsgruppen går ikke spesifikt inn på hvordan Sikt skal løse den delen, men fokuserer på sektorens behov for å bli møtt av kompetanse og kapasitet.

### 4.1 Bibliometrisk kompetanse

Bibliometri er et relativt ungt fagfelt, som er i stor utvikling, og kompetansenivået i forskningssektorene er varierende. Tilgang på kompetanse er sentralt i utviklingen av NIB som tjeneste og bibliometri som verktøy ute ved institusjonene, samtidig som den er viktig for å kunne ta i bruk tjenesten effektivt. Per i dag er det organisert noe opplæring som en del av Bibliometriforum (se nedenfor), men det vil være et behov for å styrke denne opplæringen. Den gjensidige avhengigheten av kompetanse i tilknytning til NIB og ute i institusjonene, tilsier at Sikt bør ta et tungt ansvar for å organisere opplæring/kurs for deltagere i Norge.

### 4.2 Nasjonalt nettverk – Bibliometriforum

Bibliometriforum er en felles arena hvor norske bibliometrikere møtes for å ta opp ulike temaer som berører bibliometri. På møtene har det deltatt opptil 65 personer og over 40 institusjoner har vært representert på disse møtene.

Gjennom Bibliometriforum har vi erfart at bibliometrisk kompetanse varierer veldig mellom institusjoner. Noen er relativt ny i feltet, eller jobber lite med bibliometri som en av mange andre oppgaver, mens andre har høyere kompetanse med personer som for eksempel har bibliometri som en av primæroppgavene, har skrevet bibliometrisk forskning, eller har fått formell opplæring (kurs fra CWTS Leiden eller European summer school for scientometrics).

I forhold til teknisk kompetanse viste en rask undersøkelse ved Bibliometriforum i april 2022 at en rekke verktøy er i bruk ute ved institusjonene. Et flertall bruker DUCT og Excel, men mange ønsker også bedre kompetanse i disse verktøy. Det er noen få miljøer hvor programmering er i bruk i dag (R, Python, SQL og APIs), men en større andel som har ønske/planer å få bedre kompetansen i disse (særlig for R og APIs).

Teknisk og bibliometrisk kompetansenivå er viktig å ta stilling til under vurdering av ulike måter å levere ut data fra NIB (se *7 Tilgang til dataene*). For flere av løsningene kan det være hensiktsmessig å organisere opplæring via Sikt eller i regi av Bibliometriforum for å få korrekt og mest bruk av dataene. Ferdiglagde analyser i dashboards, eller tilpassede analyser på en analyseplattform, er et alternativ for miljø som ikke har kapasitet til teknisk kompetansebygging, men bibliometrisk kompetanse er likevel nødvendig for å kunne tolke dataene riktig (uavhengig av de tekniske kravene). Derfor anbefales det at opplæring også bygges ut med grunnleggende bibliometrisk kompetanse for de som trenger det (for eksempel passende bruk av indikatorer fra dashboards).

### 4.3 Tilgang til forskningskompetanse på høyt nivå

Gjennom Nordisk institutt for forskning og utdanning (NIFU) har nettverket rundt NIB tilgang til bibliometrisk kompetanse på høyt internasjonalt nivå. Dette er viktig for å utvikle bruken av bibliometri ved institusjonene og evnen til å nyttiggjøre seg de mulighetene som NIB gir.

For driften av NIB er det viktig med høyt kompetente brukere, som har et aktivt forhold til dataene. Dette sikrer at feil oppdages lettere. For utvikling av indikatorer og avansert tilrettelegging av data (kategorisering på ulike egenskaper av dataene), trengs også en aktiv rettleiding og kvalitetssikring.

Det nasjonale publiseringsutvalget benytter seg aktivt av bibliometrisk kompetanse ved å finansiere 3 månedersverk av en forskerstilling for å understøtte sekretariatet i det løpende arbeidet i utvalget. Bibliometrisk kompetanse bidrar til sentrale analyser som står helt sentralt når utvalget fatter vedtak. Slik kompetanse brukes i forbindelse med godkjenning av nye publiseringskanaler, ved vurdering av bekymringsmeldinger fra forskningsmiljøer, for å beskrive endring i publiseringsmønstre og for å kvalitetssikre data som ligger til grunn for å beregne verdensproduksjon og nasjonal produksjon på fagfeltnivå.

For NIB og Bibliometriforum har det vært viktig med tilgang til uavhengig forskningskompetanse, og dette vil også være viktig i fremtiden. Det bibliometriske forskningsmiljøet i Norge er lite og sårbart. Gjennom NIB er det investert i data, men det er ikke investert tilsvarende i bibliometrisk forskning.

Det er allerede etablert god kontakt mellom NIB og miljøer i Norden som forvalter tilsvarende bibliometriske databaser. I utveksling av kompetanse og erfaringer med slike miljøer har det vært viktig at NIFU har deltatt med tung forskningskompetanse fra norsk side. Arbeidsgruppen påpeker at potensialet for bruk av NIB i strategisk og forskningspolitisk arbeid kun kan realiseres om det sikres tilstrekkelig tilgang på forskningskompetanse. På sikt vil tilgang til NIB også kunne stimulere til mer bibliometrisk forskning i Norge slik at det blir en positiv dynamikk. I den videre utvikling av NIB vil det fortsatt være nyttig å samarbeide med utenlandske miljøer, som driver tilsvarende bibliometrisk infrastruktur. Arbeidsgruppen mener at slikt internasjonalt samarbeid vil fungere best om vi fra norsk side deltar med forskningskompetanse på tilsvarende nivå som de utenlandske institusjonene.

### 4.4 Internasjonalt nettverk

Bibliometri er et lite miljø i Norge, slik at både forskning, administrativ bruk av NIB-data og selve driften av NIB er avhengig av internasjonale miljøer.

#### 4.4.1 Forskning

Bibliometri er et internasjonalt forskningsfelt hvor nesten all forskning skjer utenfor Norge. Feltet har egne konferanser og vitenskapelige tidsskrifter slik som *Scientometrics* og *Quantitative science studies*. Det er viktig med nasjonal forskningskompetanse som er i kontakt med forskningsfronten på feltet. Etablering av forskningsmessige nettverk muliggjøres blant annet ved deltakelse på internasjonale konferanser. Gjennom NIB har forskere i Norge tilgang til store datasett som kan gjøre en attraktiv som samarbeidspartnere for andre lands forskere.

#### 4.4.2 Forum for egenutviklede bibliometriske databaser

NIB er primus motor for et skandinavisk/nordisk nettverk for institusjoner med egenutviklede bibliometriske databaser, med fokus på administrativ bruk av dataene. Det er her vi har hentet den kritiske kompetansen som vi har vært avhengig for å etablere, drifte og utvikle tjenesten. Vårt viktigste bidrag ligger i det å bygge nettverket.



Foruten kompetanseutvikling, har vi konkrete planer på samarbeide om blant annet vasking av institusjonsadresser og utveksling av ferdig vaskede poster fra eget land. Det er planer om et «nordisk symposium» om dette temaet i 2023, hvor en formalisering av nettverket vil bli drøftet på et høyere nivå.

**Kommentert [BA1]:** Vi kuttet budsjettet til en konferanse i Norge, dette må bli i 2024

## 4.5 Dokumentasjon

For å kunne ta i bruk de detaljerte dataene, direkte på serveren, trengs god dokumentasjon. Det samme gjelder datasett, slik at det fremgår tydelig hvilke data som er med og hvilke som er utelatt, samt hva dataene i feltene betyr. I utgangspunktet vil den dokumentasjonen Web of Science har gjelde for NIB, med noen tillegg. NIB bør imidlertid tilstrebe å tilby egen dokumentasjon og veiledning for hele basen samlet og tilgjengelig for brukerne av NIB-data.

# 5 Policy og etikk

## 5.1 Etikk

Det er rettet kritikk mot uinformert bruk av kvantitative indikatorer (slik som H-indeks) i forskningsanalyse.<sup>12</sup> Diskusjoner om hvordan vi kan bruke bibliometri på en ansvarlig måte er blitt en stadig større del av profesjonell bibliometri. UHR-Publisering har utgitt et notat med råd om lokal bruk av publiseringsindikatoren.<sup>13</sup> Notatet var revidert mai 2020, og utvidet versjon omtaler i tillegg til publiseringsindikatoren også Journal Impact Factor (JIF) og H-indeks. De tre indikatorene ses i lys av DORA-erklæringen og Leiden-manifestet, som begge er internasjonalt kjent kilder til råd og anbefalinger om bruk av bibliometri på individnivå, og rådene fra UHR-Publisering er i tråd med disse. Som vist i kapittel 3.5 er vurdering av forskere og forskning et felt under utvikling, med nasjonale og internasjonale initiativ.

Nasjonalt har Bibliometriforum en viktig funksjon som arena for å drøfte og løfte etiske problemstillinger knyttet til bruk av bibliometri. Bibliometriske analyser kan anvendes på ulike aggregeringsnivå. Det kan sies å fungere best på institutt- og institusjonsnivå og høyere, mens å bruke bibliometri på individnivå er mer problematisk. Det er viktig å ha etablert gode etiske standarder for ansvarlig bruk av bibliometri.

## 5.2 Policy for datakvalitet

Clarivate indekserer de dataene som fremkommer i publikasjonene rimelig riktig. Når vi finner feil i NIB-dataene, er det gjerne på publikasjonen feilen har oppstått. En typisk feil er at en av forfatterne får feil affiliasjon (institusjonsaffiliasjon som tilhører en av de andre forfatterne) eller at landkoden settes lik det en annen forfatter har. Det vi (hittil) ikke har funnet feil på er ORCID. Hvis en forfatter oppgir ORCID, ser vi at den er korrekt. De «norske» dataene kan med andre ord bli veldig mye bedre hvis alle norske forfattere oppgir ORCID.

<sup>12</sup> <https://sfdora.org/>, Feil! Bokmerke er ikke definert.<https://www.nature.com/articles/520429a>, Feil! Bokmerke er ikke definert.NORCAM

<sup>13</sup> [Om bruk av bibliometri på individnivå - lang.pdf \(hkdir.no\)](#)

Cristin har lenge vært regnet som en meget god datakilde, som kan tilskrives godt instruksverk og opplæring. Gjennom krav på bruk av ORCID, ROR og DOI, samt oppfølging av dette, vil vi få gode data etter hvert som Clarivate indekserer publikasjonene. Det å kreve ORCID og ROR ved registrering i Cris/NVA, og for søknader til Forskningsrådet, kan være et viktig grep for å øke bruken av disse identifikatorene.

NIB kan også brukes som et verktøy for å følge opp eventuelle pålegg, og da ikke bare relatert til ORCID og ROR, men også finansieringsinformasjon med mer. Det å utarbeide systematiske oversikter, vil kunne vise hvem som er gode til å følge disse påleggene.

En annen policy-tilnærming er å hjelpe de store institusjonene og store norske private selskap, som driver med mye forskning, til å få sin adresse «unified» av Clarivate<sup>14</sup> og å ta i bruk ORCID. Dette vil gi forbedret datakvalitet for analyse av hele den norske forskningsporteføljen.

### 5.3 Deling og gjenbruk av data

Dataene i NIB er kommersielle data, kjøpt inn til et spesifikt formål, og det er ikke tillatt å dele disse utenfor NIB-nettverket uten direkte avtale med Clarivate. Lisensen med Clarivate tillater derimot at forskere innenfor NIB-nettverket deler data med forskere fra andre institusjoner til konkrete forskningsprosjekter som de samarbeider om.

## 6 Datakilder og tilrettelegging av data

Den viktigste gevinstrealiseringen får vi når dataene vaskes og tilrettelegges sentralt. Da gjøres dette ett sted en gang, helt likt for alle institusjoner. Dette er også kritisk når det gjelder å kunne bruke data i andre systemer og koble til andre data. En viktig gevinst er at alle bruker akkurat de samme dataene, slik at ulike analyser kan repeteres på ulike institusjoner og være fullt ut sammenlignbare.

Arbeidsgruppen anser koblingen mellom NIB og Cristin (CRIS/NVA - MIME) som den viktigste tilretteleggingen av dataene, på nåværende tidspunkt. Det bes om at dette prioriteres fremfor all annen tilrettelegging, uavhengig av nivå på videreutvikling.

### 6.1 Datakilder

NIB bruker Web of Science-data (WoS) fra Clarivate som primærkilde. Det finnes alternative globale kilder, slik som Scopus, Dimension og OpenAlex. Dekningsgraden mellom WoS og Scopus er relativt lik og metadataene er relativt like. Dimensjon har en litt annen innretning på sin indeksering, med kun 5 dokumenttyper og ingen ISBN, men med høy andel records med doi. OpenALEX er basert på Microsoft Academic Graf og har en vesentlig flere records fra ulike kilder. De jobber med å tilrettelegge egne data gjennom bruk av AI.

---

<sup>14</sup> Clarivate bruker en algoritme til å sette et standard navn på en institusjon, ut fra en liste med ulike varianter av institusjonsnavnet, som er satt opp i samarbeid med institusjonen. Det gjør at for eksempel USN er USN for alle kjente og sannsynlige varianter av navnet, samt navnene til de sammenslåtte institusjonene.

En nærmere vurdering og beskrivelse av de ulike datakildene vil være krevende og lite hensiktsmessig for denne studien. Videre er en anbudsprosess svært krevende og det å skifte data leverandør vil også kreve en del teknisk tilrettelegging. Fordelene vi kan få ved å endre dataleverandør, gjennom endret kvalitet og dekningsgrad, må vurderes opp mot disse kostnadene. Arbeidsgruppen 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.

NIB-databasen dekker primært publisering i internasjonale tidsskrifter, men også noen bøker og konferanseserier, samt en del nasjonale tidsskrifter. Det vil si at ikke all vitenskapelig publisering er dekket av databasen, og dekningsgraden varierer både mellom institusjoner og fag. Dette er tidligere analysert i en studie som NIFU har gjennomført.<sup>15</sup>

## 6.2 Berikende (nye) kilder

Flere datakilder kan gi merverdi til NIB-dataene. For å kunne analysere OpenAccess-publisering, er allerede Unpaywall implementert i databasen, for å gi oss OA-statusen til publikasjonene. Dette er gjort i samarbeid med NIFU, og det er lagt til rette for oppdateringer hver gang Unpaywall gir ut oppdaterte lister.

Det finnes flere slike berikende kilder, hvor de som baserer seg på DOI er enklest å implementere. Det kan være snakk om patenter og innovasjoner, som siterer akademisk forskning. Siteringer finnes også i offentlige dokumenter, utredninger og lovarbeid, eller for den saks skyld sosiale medier både innenfor og utenfor academia.

Arbeidsgruppen vil prioritere vedlikehold av eksisterende data, fremfor nye datakilder.

## 6.3 Datasiteringer

De siste årene er det gjort mye rundt gjenbruk av data (FAIR-prinsippene) og flere publikasjoner kommer nå med datasitering. Etter hvert som datasiteringene indekseres, kan vi få studier som forteller noe om gjenbruken av data. Ved bruk av andre registre kan vi også se både hvem som har finansiert dataene og sammenligne med hvem som har finansiert studiene hvor dataene er brukt.

Det er forventet et økt fokus på datasettsiteringer i fremtiden og blant annet har Sikt vist en egeninteresse for datasiteringer på data fra forskningsdataplattformen, som automatisk utstyrer sine datasett med egne doi.

## 6.4 Beregnede verdier – Indikatorsett

Per i dag beregnes siteringstall og siteringsindekser, kvalitetssikret av NIFU. Dette kan med fordel utvides til flere indikatorer, basert på NIB-data. Hvilke indikatorer som skal implementeres, vil bli utredet separat av Arbeidsgruppen, i tett samarbeid med NIFU. Dette er krevende arbeid, med tanke på at indikatoren skal stå seg i møte med ulik bruk og kritikk, og være godt dokumentert og kvalitetssikret.

## 6.5 Algoritmiske beregninger

I NIB er det lagt til rette for algoritmebasert fagklassifisering på artikkelnivå. Dette blir viktigere med fremveksten av multidisiplinære «megajournaler», slik at det ikke gir mening å følge journalens

<sup>15</sup> <https://www.sciendo.com/article/10.2478/jdis-2019-0001>

(manglende) fagklassifisering. NIB har en algoritme, utviklet på Karolinska Institutet, som kan kjøres, men det krever både økt teknisk kompetanse og kapasitet til å tilrettelegge for bruken. Blant annet må ulike parameter justeres for å få ønsket granularitet og navngivning av klassene.

## 6.6 Dugnadsv verdier og synergieffekter

Det foregår mye tilrettelegging av data ute ved institusjonene. Etter hvert som NIB blir mer tilgjengelig, vil vi kunne risikere at flere institusjoner tilrettelegger akkurat de samme data, men på ulik måte. Det er flere måter vi kan å unngå denne typen dobbeltarbeid og få ut synergier. Sentral tilrettelegging av de viktigste og største datamengdene er den viktigste. Videre må det komme på plass systemer som kan ta vare på den tilrettelegging som gjøres av enkeltinstitusjoner og som kan nyttiggjøres ved andre institusjoner. NIB bør kunne overta korrigerte data og eventuelt ta på seg forpliktelsen til å vedlikeholde disse dataene.

Vi kan også se for oss en modell hvor for eksempel UiB kategoriserer SDGene og vedlikeholder disse på vegne av alle NIB-brukere.

## 6.7 Vask av data

Vasking av data er viktig for kvalitet på analyser. Dataene i Web of Science er rimelig korrekte i forhold til det som faktisk står oppført på publikasjonene. Det betyr ikke at alle data er korrekte. Blant annet kan institusjonsnavnene være korrekte, men landtilhørigheten være feil, fordi det er feil oppført på publikasjonen. Alle institusjoner er heller ikke omfattet av Organization Enhanced, som er Clarivates egen adressevask. Dette gjelder blant annet Akershus universitetssykehus og de aller fleste store firmaer med en vesentlig forskningsproduksjon i Norge (Aker biomaritim, Equinor, ABB, osv.).

## 6.8 Kobling av data

For å kunne koble data trenger vi entydige nøkler eller en persistent identifikator (PI eller PID). PID er et vedvarende referansenummer til en enhet eller et objekt, som identifiserer det gjennom hele livssyklusen. PIDs er ikke bare unike og vedvarende, men også handlingsbare: du kan skrive det i en nettleser og direkte spore opp den identifiserte kilden.

Digital object identifier (DOI) er en global og anerkjent PID for pålitelig identifisering av elektroniske dokumenter, og den er allerede godt integrert i Cristin og institusjonelle vitenarkiv. DOIs gir økt kvalitetssikring av metadata til elektroniske dokumenter som artikler, men også kildekode, og gjør det enklere å koble og supplere ulike bibliometriske data mellom databaser (Cristin, WoS, Dimensions etc.).

Tilsvarende bruk av PIDs til identifisering av personer og organisasjoner vil øke kvalitetssikring av data i Cristin/NVA/NIB og øke muligheten for automatisering og interoperabilitet. [ORCID](#) (Open Researcher and Contributor ID) og [ROR](#) (Research Organization Registry) er godt etablerte identifikatorer til henholdsvis personer og organisasjoner. Med bruk av disse vil vi lettere kunne koble riktig person og organisasjon, samt utfyllende informasjon, mot publikasjonene med bruk av ORCID og ROR. Likevel må ORCID og ROR først bli opprettet av brukere og kreves det noe vedlikehold over tid. ORCID kan for eksempel åpne opp muligheten for å analysere bibliometriske data (publikasjoner, siteringer) i lys av forfatterens utdannings- og ansettelseshistorie, samt tildelte finansiering fra forskningsråd og andre støtteordninger, forutsatt at forfatterne aktivt bruker og oppdaterer sin ORCID profil. Det legges opp til bruk av ORCID i CRIS/NVA.

For helsesektoren vil PubMed-id (PMID) og MeSH-termer være viktig. NIB har allerede alle PMIDer for WoS-publikasjonene, slik at NIB-data kan kobles direkte mot PubMed-data uten videre tilrettelegging. Det kreves en lisens, men ingen betaling, for å legge PubMed data inn i NIB. Det å ha dataene direkte tilgjengelig i NIB (og videre inn i DUCT) vil senke brukerterskelen betydelig og vil være en «lavthengende frukt» for en videreutvikling av NIB.

## 6.9 Tilrettelegging og AI

De vanligste metodene som brukes er enten tekstbaserte (Natural Language Processing) eller basert på siteringer (bibliographic coupling, co-citations). Et viktig spørsmål i denne sammenheng er hvordan Sikt kan legge til rette for maskinlæring og bruk av algoritmiske metoder. Slik tilrettelegging kan for eksempel handle om å gi tilgang til å utføre beregninger på NIB-serveren, bistand til å utforme SQL-skript som henter relevante data fra NIB, deling av algoritmer for analyse etc.

## 6.10 Nye egne data

Spennet her er stort. UiB har testet å hente inn tag-er fra fulltekst, som peker til forskningsdata. Med NVA vil muligheter til fulltekstøk på norske publikasjoner øke, og vi kan hente inn den type data. Ved å gjøre det sentralt, lagres det sentralt. Tekstutvinningsmetoder kan brukes på større skala sentralt.

# 7 Tilgang til dataene

For å realisere gevinsten i NIB og arbeid som har allerede gått i å sikre god datakvalitet, er det viktig at brukerne får tilgang til dataene. Det er flere måter dette kan gjøres; disse har ulike brukspotensial, og stiller ulike krav til brukernes bibliometriske og tekniske kompetanse.

## 7.1 Data via DUCT – Sentralt bibliometrisk verktøy

Å tilgjengeliggjøre og forbedre NIB data i DUCT foreslås som et første trinn for å gjøre NIB-data mer tilgjengelig. Mange norske institusjoner bruker i dag DUCT som verktøy for innhenting/analyser av bibliometriske data. I en uformell undersøkelse ved Bibliometriforum (april 2022), svarte over to tredjedeler av respondentene at de bruker DUCT i dag (27/38 personer). Å sørge for at noe av den mest relevante NIB data ligger i DUCT vil gi flere institusjoner relativ lav terskel for tilgang til NIB data, uten behov for å lære nye verktøy. Eksempler av data som kunne inkluderes i DUCT fra NIB er forskningsfinansierer, abstracts, og fagkategorisering. Et flertall av respondenter pekte også på behov for videre opplæring i DUCT/Tableau (41/55 personer). Hvis denne løsningen benyttes, kan det være hensiktsmessig å organisere opplæring i regi av Bibliometriforum eller via Sikt.

Det må også presiseres at DUCT løsningen er mest aktuell for å supplere Cristin data med NIB data om norske publikasjoner, og at DUCT løsningen ikke kan brukes for hele NIB datasettet. Da må denne løsningen brukes sammen med en eller flere av de andre mulighetene beskrevet. Det er kartlegging- og analysebehov ved brukerinstitusjoner som fordrer at tilgang på data via DUCT både opprettholdes og videreutvikles.

Arbeidsgruppen vil prioritere tilrettelegging av data for bruk i DUCT, uavhengig av den videre utviklingen av NIB.

## 7.2 CRIS/NVA<sup>16</sup>

Cristin skal legges ned og dataene skal inngå i nytt CRIS/NVA-system. I overgang og videreføring av nytt system er det viktig å ivareta funksjonaliteten Cristin har knyttet til NIB og motsatt. For å opprettholde datakvalitet slik den er i dag, og enda bedre i framtiden, må forskningsresultat, person, sted og prosjektdata ivaretas i nytt system. Dataene er viktig for institusjonene for enkel tilgang til å kunne koble Crisdata opp mot NIB-data. DUCT er koblingsnøkkelen for å berike dataene som allerede ligger i NVA/Cris med NIB-dataene, for eksempel siteringer.

## 7.3 Datasett

NIB har hele tiden vært planlagt med fokus på bruk av datasett, siden det ikke ble satt av ressurser til utvikling av andre egnede løsninger med moderne brukergrensesnitt. Det vil si at institusjonene skulle laste ned datasett og bruke datasettene i egne analyseverktøy (BI, Tableau, R og til nød Excel). Erfaringene er at utlevering av datasett ikke er en løsning som vil fungere spesielt godt, men dette er eneste alternativ ved en moderat utvikling av NIB. I den anbefalte løsningen vil datasett-server for utlevering av datasett avvikles og erstattes med en mer brukervennlig analyseplattform, som vil utlevere oppdaterte data på en mer fleksibel måte.

## 7.4 API

Bruk av API vil for noen kunne erstatte bruken av datasett, men også åpne opp for bruk av data inn i andre datasystemer. Et API forutsetter at data blir vasket og tilrettelagt, slik at vi vet at vi får korrekte data overført. Hvis data skal hentes ut pr DOI, må alle aktuelle poster ha DOI. Hvis det skal hentes ut per organisasjon, må alle organisasjoner ha riktig organisasjonsnavn. Utvikling av et API må derfor sees i sammenheng med investering i vask og tilrettelegging av data, og kompetansen til å bruke et API.

Det er allerede ytret ønsker/planer fra flere deltakere på Bibliometriforum om å få bedre kompetanse i APIs (40% respondenter i en rask undersøkelse ved Bibliometriforum april 2022), og vi ser at dette kan effektivisere arbeidet med analyser.

Ut fra kompleksiteten med vask av data, utviklingen av selve APIet og behov for opplæring, ønsker ikke Arbeidsgruppen å prioritere API i denne omgang, da andre løsninger vil gi flere personer muligheten til å ta i bruk data.

## 7.5 Analyseplattform

En analyseplattform vil øke mulighetene for de minste institusjonene til å bruke bibliometri mer aktivt i sin virksomhet, ved at det blir en lavere terskel til å ta i bruk dataene. Ønsket er å legge seg på samme linje som våre nordiske kollegaer, med bruk av fri programvare og komponentbasert utvikling.

I praksis betyr dette at vi tar i bruk R og en Shinyserver<sup>17</sup>, som allerede har ferdige biblioteker for bibliometriske analyser. De som har kompetanse innen R kan utvikle analyser (komponenter) som deles med de andre på Shinyserveren. Dette kan vi gjøre fordi alle bruker akkurat de samme dataene, som vil

<sup>16</sup> Midlertidig navn på prosjektet [ny fellestjeneste for forskningsinformasjon](#).

<sup>17</sup> R er et kommandobasert (programmerbart) statistikkverktøy og minner om en avansert excel. Shinyserver er en programvare som lar oss legge ut data og ulike R-skript på en sentral server.

se helt likt ut uansett hvilken institusjon eller nivå de analyserer på. Dette fordrer noe sentral kompetanse på R, slik at vi kan kvalitetssikre, standardisere og samkjøre behov og utviklingen.

I de miljøene vi sammenligner oss med, har de med relativt enkle ressurser utviklet egne analyseplattformer for bibliometri, basert på de dataene de selv rår over.

Karolinska Institutet har en intern web-løsning<sup>18</sup> tilpasset medisinsk forskning, utviklet i både R, Python og med ulike åpne visualiseringsløsninger (GGplott, m.m.). De kan blant annet drille inn på spesifikke MeSH-termer og se hvem som samarbeider innen forskningen på de feltene. De kan også sammenligne seg selv mot et gjennomsnitt av andre forskere i forhold til ulike siteringsindekser.

Ved KTH er det utviklet en analyseplattform som inneholder statistikk om institusjonens publikasjoner, citation impact og sampubliseringer. En lignende analyseplattform for norske institusjoner vil gi enklere tilgang til statistiske data om institusjonenes egen forskning, informasjon som kan benyttes som grunnlag inn i institusjonenes tilstandsrapporter og utviklingsavtalen med KD.

## 7.6 Dashbord

Dashbord bør her sees i sammenheng med en analyseplattform. Tanken er at det kan settes opp flere standard statistiske indikatorer som vises visuelt/grafisk, til bruk for ledelsen. Dashbordene utstyres med enkle justerbare parametere, slik at det er mulig å analysere ulike tidsspenn og gjøre sammenligninger.

Dashboards gir mulighet for institusjoner til å hente ut visualiseringer og aggregerte data uten høye krav til teknisk kompetanse, og er velegnet til standard analyser. Det er likevel nødvendig at brukerne har en viss bibliometrisk kompetanse (se seksjon 4.1 Bibliometrisk kompetanse). Dashboards har mer begrensede muligheter for å tilpasse analyser og indikatorer til egne formål og behov, men gir tilgang til flere enn det en analyseplattform gir.

UiT har utviklet dashbord med Cristin-data. Flere andre institusjoner melder at de også vurderer å utvikle et slikt dashbord.

## 7.7 Direkte tilgang til detaljerte data

Lisensen åpner for at forskning skal ha tilgang til de detaljerte dataene. Serverkapasiteten er også dimensjonert ut fra dette. Dette er allerede implementert, og løsningen fungerer etter hensikten, men har en høy brukerterskel.

Det er også en del tjenester som fungerer best om de implementeres direkte på serveren. Forskningsrådet har en algoritme for tematisk klassifisering av forskningspublikasjoner. Den krever direkte tilgang til SQL-serveren (NIB), men også teknisk kompetanse på serversiden, for at den skal være reelt tilgjengelig.

Det er sentralt for bruken av NIB at Sikt innehar nok kapasitet og kompetanse til å sette opp tilganger og implementere algoritmer direkte på serveren.

En analyseplattform, eller et dashboard, kan ikke erstatte direkte tilgangen til detaljerte data.

<sup>18</sup> Eksempel: <https://abm.sys.kth.se/public/>

## 8 Bærekraftig drift av NIB

NIB som tjeneste bygger på en egenutviklet bibliometrisk database, slik den også gjør hos de nordiske miljøene vi sammenligner oss med, og blant annet hos CWTS i Leiden og Bibliometrisentrum i Berlin.

Egenutviklede bibliometriske databaser er en spesialisert tjeneste, hvor det trengs dyp domenekunnskap, god bibliometrisk forståelse, god teknisk innsikt i IT-drift og utviklerkompetanse. I tillegg har NIB en viktig oppgave med å drifte Bibliometriforum og Arbeidsgruppen, hvor det må forventes gode samarbeidsevner og god rolleforståelse.

Det kan ikke forventes at en og samme person har nødvendig kompetanse og kapasitet til å drifte dette alene. Det er uansett for sårbart å basere seg på ett menneske til å fylle disse rollene. Det må med andre ord flere personer til for å få en mer robust drift, samtidig som det også gir nødvendig kapasitet til å videreutvikle tjenesten.

Den tekniske driften av server og operativsystem blir i dag ivarettatt av UiO/USIT og noen timer fra Sikt (programvare-drift). All annen drift ivaretas av ett årsverk ved Sikt. I tillegg kommer et kvart årsverk, som i hovedsak går til å drifte Arbeidsgruppen til NIB og å gjennomføre Bibliometriforum, men også til å være den eneste støtten tjenesteansvarlig har internt. Utenom dette kjøpes en del støtte fra NIFU.

Utfordringene til NIB kan være lettere å forstå hvis vi ser til andre miljøer, som leverer gode tjenester til sine brukere. På det tekniske planet er NIB ikke unik, men ressursmessig er NIB underdimensjonert, med én stilling som skal ta seg av alt fra policy til teknisk drift.<sup>19</sup>

Selv om de tilsvarende bibliometriske databasene i Sverige og Danmark har samme struktur som NIB, vil de aldri bli helt like fordi vi kjøper inn ulike datasett, har ulikt bruksområde og ulike brukere.

Karolinska Institutet har vært viktig for oppbyggingen av NIB.<sup>20</sup> De har 6 personer (5,5 årsverk) som jobber med bibliometri og drift og utvikling av basen og tjeneste rundt. De har en egen webløsning hvor ansatte selv kan se på samarbeid og noen statistiske indikatorer.

KTH har 6 personer som jobber med bibliometribasen, mens Vetenskapsrådet har 4 personer som jobber med den bibliometriske databasen. For alle disse miljøene er de fleste stillingene også knyttet opp mot direkte bruk av dataene, slik at de har både domenekunnskap og bred bibliometrisk kompetanse, som er viktig med tanke på det å lage gode tjenester.

Danmarks Tekniske Universitet (DTU) jobber nasjonalt, på vegne av Undervisnings og forskningsministeriet (UFM), med data på danske publikasjoner. De tilbyr analysemuligheter gjennom en egen analyseplattform og bruker, i likhet med mange andre, fri programvare og åpen kildekode i utviklingen. De er 6 personer<sup>21</sup> som jobber med konseptet og har i tillegg innleid tre utviklere på heltid for å utvikle systemene.

<sup>19</sup> I tillegg får NIB uvurderlig støtte fra Hanne Hole Young til å organisere aktiviteter og miljøet rundt NIB.

<sup>20</sup> NIB har fått databaseskjema og importskript fra KI, slik at vi har helt lik konfigurasjon.

<sup>21</sup> Ivar Ternsell Torgersen, har jobbet et år i 70 % stilling på dette prosjektet, som et ledd i kompetanseutvikling og for å øke samarbeidsmulighetene.



## 9 Oppsummering og anbefalinger

Tabellen nedenfor beskriver situasjonen slik den er i dag, og to ulike nivå for videreutvikling. Arbeidsgruppen anbefaler at en større videreutvikling i første omgang igangsettes. Den vil svare på mye av brukernes behov og vil gi en mer bærekraftig og mer langsiktig drift av NIB. Det anbefales videre økte ressurser til forskning innen bibliometri. Dette kan skje gjennom egenprioriteringer ved institusjonene og spesifikke utlysninger av forskningsprosjekter innen området.

|                 |                                         | Dagens nivå                                   | Moderat videreutvikling                                  | Større videreutvikling (anbefalt løsning)                    |
|-----------------|-----------------------------------------|-----------------------------------------------|----------------------------------------------------------|--------------------------------------------------------------|
| Funksjonalitet  |                                         |                                               |                                                          |                                                              |
|                 | Datatilgang                             | Kun egne datauttrekk                          | Fikserte datasett                                        | Uttrekk etter behov                                          |
|                 | Analyseverktøy                          | SQL-databasetilgang                           | Tilrettelagte skript (R) og Power BI, SAS, Tableau       | NIB-analyseplattform                                         |
|                 | Kobling til andre system                | -                                             | DUCT og utvalgte systemer                                | «Analyseplattformen» <sup>22</sup> og andre ledelsessystemer |
|                 |                                         |                                               |                                                          |                                                              |
| Brukerterskel   |                                         |                                               |                                                          |                                                              |
|                 | Krav til teknisk kompetanse hos brukere | Ekspert med avansert programmeringskompetanse | Moderat                                                  | Lav                                                          |
|                 | Krav til bibliometrisk kompetanse       | Ekspert                                       | Basal                                                    | Noe                                                          |
|                 | Antall brukere                          | Begrenset (13 pr i dag)                       | Ca 80 (Bibliometriforum)                                 | Mange                                                        |
|                 | Brukermiljø                             | De største institusjonene med egne fagmiljø   | Forskningsinstitusjoner med utdannet personell           | Alle forskningsinstitusjoner samt andre offentlige organ     |
|                 |                                         |                                               |                                                          |                                                              |
| Datakilder      |                                         |                                               |                                                          |                                                              |
|                 | Beriking av database                    | Kun OA-status (NIFU)                          | Geografiske regioner og lignende standard informasjon    | Data fra Medline/Pubmed og lignende kilder/tjenester         |
|                 | Kobling til andre datakilder            | Cristin-data                                  | Enkelte tilrettelegginger av nøkler mot andre datakilder | Mulighet til integrasjon mot andre systemer                  |
|                 |                                         |                                               |                                                          |                                                              |
| Tilrettelegging |                                         |                                               |                                                          |                                                              |
|                 | Beregnete verdier                       | Siteringstall, enkle siteringsindekser        | + Flere siteringsindekser                                | + Fleksibelt siteringsvindu                                  |

<sup>22</sup> Analyseplattformen er et eget utviklingsprosjekt ved Sikt

|                           |                                       |                                                             |                                                       |                                                                                                                      |
|---------------------------|---------------------------------------|-------------------------------------------------------------|-------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|
|                           | Fagfelt og tematiske inndelinger      |                                                             | Identifisering av SDG-er                              | Algoritmebasert fagfeltinndeling, nasjonale satsninger og andre tematiske inndelinger                                |
| <b>Bruk av data</b>       |                                       |                                                             |                                                       |                                                                                                                      |
|                           | Bibliometrisk forskning               | Store muligheter                                            | Meget store muligheter                                | Krav om mer forskning, forskningskapasitet og kompetanse for å møte behov for aktiv veiledning i bibliometrimiljøene |
|                           | Bibliometrisk analyse                 |                                                             | Mulig avhengig på teknisk og bibliometrisk kompetanse | Utvidet avansert bruk                                                                                                |
|                           | Indikatorbruk                         |                                                             | Begrenset                                             | Store muligheter/dashboards                                                                                          |
|                           | Rapportering og ledelse               |                                                             | Kan legge til rette for enkelte rutinemessige oppdrag | Tilgang til dashboard og data inn i andre system                                                                     |
|                           | Forhandlinger (LÅT)                   |                                                             | Mulig, men krevende                                   | Enkelt tilgjengelig                                                                                                  |
| <b>Gevinster</b>          |                                       |                                                             |                                                       |                                                                                                                      |
|                           | Sparte ressurser ved felles datavask  | Ingen                                                       | Noe                                                   | Store fordeler med likt datagrunnlag som er klar til å analyseres på                                                 |
|                           | Sparte ressurser med sentralt innkjøp | Kun til forskningen                                         |                                                       | Stor nytte og høy bruk på investeringen i data                                                                       |
|                           | Nasjonalt nettverk (Bibliometriforum) | Usikker gevinst og Bibliometriforum vil bli mindre relevant | Aktivt miljø som kan ta tak i felles utfordringer     | Krevende miljø som stiller store krav til NIB og Sikt                                                                |
| <b>Bemanning og drift</b> |                                       |                                                             |                                                       |                                                                                                                      |
|                           | Driftssikkerhet                       | Veldig sårbart                                              |                                                       | Robust                                                                                                               |
|                           | Bemanningsbehov                       | 1,5 årsverk                                                 | 2,5 årsverk                                           | Minst 3,5 årsverk                                                                                                    |
|                           | Budsjett-konsekvenser ut over årsverk | Kjøp av tjenester fra NIFU eller annet kompetent miljø      |                                                       | Økt drift av Bibliometriforum, tilrettelegging for kursing, Behov for noe mer IT/server-kapasitet                    |
|                           | Kompetansebehov                       | Domenekunnskap                                              | Domene- og utviklerkompetanse                         | Bred domenekunnskap, SQL/R og annen verktøykompetanse, avansert                                                      |

|                   |             |                                                                                                                   |                                                                                                                                |                                                                                                         |
|-------------------|-------------|-------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
|                   |             |                                                                                                                   |                                                                                                                                | programmeringskunnskap                                                                                  |
|                   | Driftsmiljø | Fortsatt hyppige tilgangsproblemer                                                                                | Mer formalisert driftskompetanse må på plass                                                                                   | Server drift må styrkes og tydelige driftskrav må avtales                                               |
|                   | Server      | Importhastighet er den største begrensningen                                                                      |                                                                                                                                | Det må investeres i fullskala servermiljø (utvikling, tests- og testmiljø, og større lagringskapasitet. |
|                   |             |                                                                                                                   |                                                                                                                                |                                                                                                         |
| <b>Anbefaling</b> |             |                                                                                                                   |                                                                                                                                |                                                                                                         |
|                   |             | Løsningen er ikke bærekraftig i dag og det anbefales å vurdere annen organisering av NIB, eventuelt utenfor Sikt. | Dette sikrer noe fremdrift i utviklingen av NIB som tjeneste, i en periode, men vil ikke sikre en langsiktig og god utvikling. | Løsningen muliggjør stor gevinstrealisering, med høy bruk og høy kvalitet.                              |
|                   |             |                                                                                                                   |                                                                                                                                |                                                                                                         |

Ut fra de signalene vi får fra Sikt, er det ikke realistisk å forvente et høyere nivå enn det som er anbefalt her. Det anbefalt nivået er ikke en full utvikling og en full utvikling er heller ikke beskrevet. De bibliometriske miljøene i Norge i rivende utvikling, og det er grunn til å tro at en større utvikling av NIB vil aksellerere denne prosessen. Økt bruk gir mer kompetente brukere, som igjen vil stille nye krav til tjenesten. Hva en eventuell full utvikling vil innebære må derfor utredes nærmere når det igjen blir aktuelt.

## 9.1 Andre prioriteringer og anbefalinger

Studiens anbefaling går på en fremtidig videreutvikling av NIB og spesielt med tanke på et godt brukergrensesnitt. Arbeidsgruppen noen anbefalinger til hvordan Sikt skal prioritere frem til dette kan komme på plass. Prioritering for 2023:

- Mer og bedre data fra NIB inn i DUCT, med dokumentasjon på felter, innhold og tidspunkt for oppdatering
- Utrede konsept for dashboard/analyseplattform, lage prosjektplan og jobbe med finansiering av en større utvikling av NIB
- Arrangere Bibliometriforum og øke kursaktiviteten.

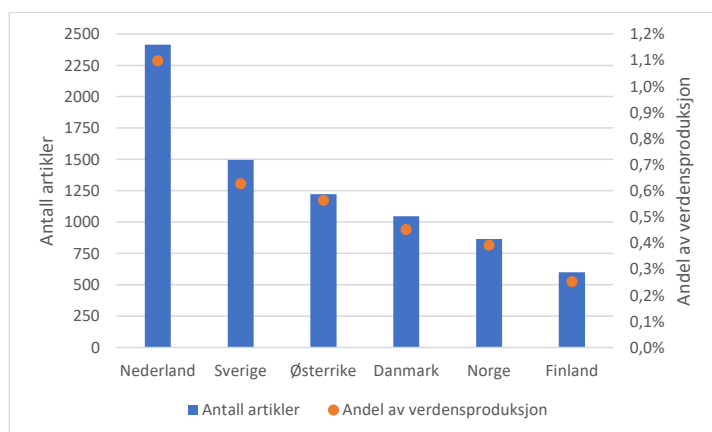
# 10 Appendiks – Eksempler på bruk av NIB

For å demonstrere noen av de mulighetene som ligger i NIB-data har arbeidsgruppen samlet eksempler på hvordan NIB data og sammenstilling av NIB-data og andre datakilder brukes ved institusjoner i dag for en mengde ulike analyser.

Noen eksempler er gitt i hovedteksten; andre er samlet nedenfor. De viser analyser til indikatorrapporten, tematiske analyser blant annet av forskning knyttet til Covid-19 pandemien og klimaendringer, og siterings- og åpen publiseringsanalyser.

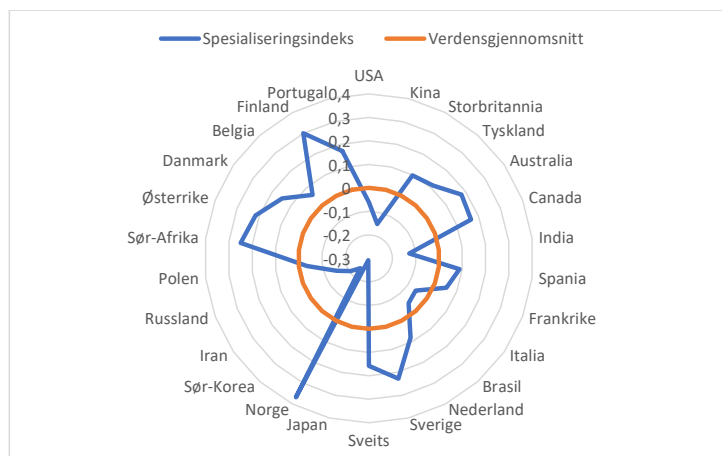
## 10.1 Tematiske analyser

### 1. [Forskning knyttet til Covid-19 pandemien](#)



Figur 4 Antall artikler med COVID-19-tema for utvalgte land (barometerlandene) og andel av verdensproduksjonen. 2020–2021.  
Kilde: NIFU.

### 2. [Forskning om klimaendringer](#)

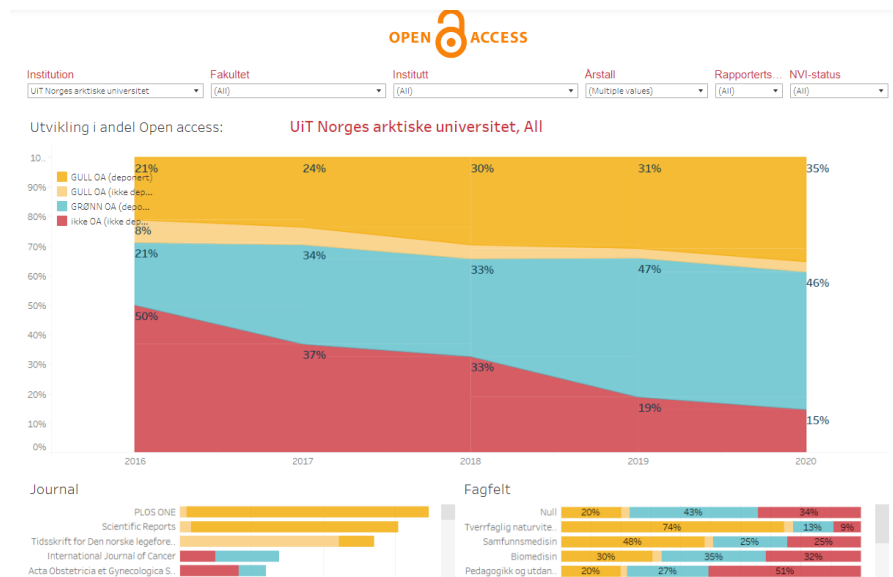


Figur 5 Relativ spesialiseringsindeks, forskning på klimaendringer. Kilde: NIFU.

## 10.2 Siteringsanalyser

- [Per fag i Indikatorrapporten](#)
- [Per institusjon i Indikatorrapporten](#)

## 10.3 Analyser av åpen tilgang



Figur 6 Dashboard i Statistikkportalen UiT. Utvikling i andel Open Access. Kilde: UiT; DUCT/Cristin.

## 10.4 Bruk av NIB ved Forskningsrådet

For Forskningsrådet er det interessant å bruke NIB til å analysere vitenskapelig publisering som et resultat av ulike former for finansiering og ulike bevilgningsmottakere. Ved å knytte NIB-data til data om Forskningsrådets prosjekter, er det mulig å undersøke sammenhengen mellom egenskaper ved prosjektene (f.eks. søknadstyper, finansieringskilde og bevilgningsmottaker) og prosjektenes resultater (f.eks. grad av tverrfaglighet, internasjonalt samarbeid, siteringer og patenter).

Forskningsrådet benytter bibliometriske analyser i evaluering av norsk forskning og som grunnlag for forskningspolitisk rådgivning. Noen analyser gjøres in-house, mens større prosjekter gjennomføres som anskaffelser. For den 3. runden med fagevalueringer innen MNT-fag og helsefag utfører NIFU en bibliometrisk analyse hvor både Cristin-data og NIB-data inngår.

Forskningsrådet utfører selv systematiske analyser som underlag for styring av egen prosjektportefølje (porteføljestyling). Data om prosjektenes publikasjoner hentes fra Cristin. DUCT (Data Ut av Cristin via Tableau) brukes til å klassifisere publiseringsdata etter fag og kvalitetsnivå. Forskningsrådet har også tatt i bruk bibliometriverktøyet InCites for analyser av publisering, finansieringskilder og siteringer. Det kan være aktuelt å benytte NIB for slike analyser framover. Særlig er det interessant å benytte algoritmiske metoder på NIB-data for strategiske analyser knyttet til forskningspolitiske mål. Dette gjelder f.eks. analyser av tematiske porteføljer som ikke kan avgrenses gjennom de etablerte fagklassifikasjonene i Web of Science eller den norske publiseringsindikatoren (NPI).

Forskningsrådet utførte i 2021 en [bibliometrisk analyse](#) som underlag for [Maritim21-strategien](#). I denne analysen ble det benyttet algoritmer for identifikasjon av relevante forskningspublikasjoner basert på siteringskjeder (bibliographic coupling, co-citation) og tekstanalyse (Latent Dirichlet Allocation). Det aktuelle prosjektet ble gjennomført i samarbeid med Clarivate, men det vil være mulig å bruke tilsvarende metoder for analyser av NIB-data. Sikt har lagt til rette for å kjøre disse beregningene direkte mot NIB-databasen ved hjelp av SQL-script, men dette vil avhenge av stabil tilgang på kompetanse for å kunne tilby dette til alle brukere.

# Vedlegg 1

## Mandat for arbeidsgruppe for tjenesten Nasjonal infrastruktur for bibliometri (NIB)

*Mandatet er godkjent av Tjenesteråd for forskning den 19 april 2021 i sak 13/21*

### Innledning

Brukermedvirkningen for tjenesten Nasjonal infrastruktur for bibliometri (NIB) organiseres med utgangspunkt i det etablerte Bibliometriforum som har representanter fra de institusjoner som benytter NIB, gjennom en arbeidsgruppe med oppgaver som beskrevet nedenfor. Bibliometriforum er et nasjonalt nettverk innenfor bibliometrifeltet. Nettverket er åpent for alle som er involvert i bibliometri i norske forskningsinstitusjoner i alle de forskningsutførende sektorene.

### Mandat

Arbeidsgruppen gir råd om bruk, drift, forvaltning og videreutvikling av NIB.

Arbeidsgruppen gir råd til tjenesteansvarlig, som forelegger saker til Tjenesterådet for beslutning der dette er nødvendig.

Arbeidsgruppen skal bidra til samarbeid og koordinering mellom institusjonene som bruker NIB.

Arbeidsgruppen skal bidra til at institusjonene skal kunne opptre samlet innenfor bibliometri med utvikling av felles policy og faglige føringer.

Arbeidsgruppen skal avklare og foreslå prinsipper for blant annet:

- utregning og bruk av felles indikatorer tilrettelagt i NIB
- tilrettelegging av data i NIB, herunder vasking av data
- oppbygging og utlevering av datasett

Arbeidsgruppen skal foreslå prioriteringer av oppgaver i utviklingen av NIB.

Arbeidsgruppen skal gi informasjon og hente innspill til sitt arbeid i Bibliometriforum og i andre relevante kanaler og sammenhenger.

Arbeidsgruppen kan ta initiativ i form av utredninger og anbefalinger i aktuelle saker.

Arbeidsgruppen kan opprette egne ad-hoc-grupper for utredning og løsning av aktuelle problemstillinger.

## Sammensetning

Arbeidsgruppen skal bestå av 8 deltagere og ledes av representant fra en av forskningsinstitusjonene som benytter NIB.

Arbeidsgruppen settes sammen av deltagere fra Bibliometriforum. NIFU – som Norges fremste miljø for bibliometri – har fast plass. Medlemmene i Arbeidsgruppen deltar i kraft av sin funksjon og kompetanse på vegne av alle deltagerinstitusjonene, og skal benytte erfaringer fra egen virksomhet for å ivareta et felles perspektiv. Forslag til deltagere i Arbeidsgruppen fremmes av Bibliometriforum til Unit som saksbehandler og fremmer saken for behandling i tjenesterådet. Arbeidsgruppen oppnevnes av Tjenesterådet for forskningstjenester. Arbeidsgruppen foreslår selv sin leder.

Sammensetningen skal hensynta føringer gitt i rammeverksdokumentet for samstyringsmodellen. Ut over dette skal sammensetningen så godt som mulig dekke bredden av de ulike institusjonene og behovene som NIB skal løse. Det skal være representasjon fra alle de tre forskningsutøvende sektorene. Arbeidsgruppen skal representere bredden av teknisk, forskningsfaglig og bibliometrifaglig kompetanse samt policy-kompetanse innen temaet. Institusjoner med mindre grad av bibliometrikompetanse og tilgang på bibliometriressurser, skal være representert for å sikre at tjenesten blir tilrettelagt slik at flest mulig kan ta den i bruk. Evne til å tilrettelegge for dialog og kunnskapsutveksling mellom alle deltagerinstitusjonene skal også vektlegges.

## Funksjonstid

Deltagerne i Arbeidsgruppen velges for en periode på 2 år med mulighet for reoppnevning. Det skal søkes kontinuitet over funksjonsperiodene, slik at hele Arbeidsgruppen ikke byttes ut samtidig.

### Sekretariat og møter

Unit er sekretariat for Arbeidsgruppen.

Arbeidsgruppen skal ha møte minimum 3 ganger i året. Utgifter i forbindelse med deltagelse i Arbeidsgruppen forutsettes dekket av den enkeltes institusjon.