

Saksdokument

Porteføljestyre for forskning og kunnskapsressurser

Til	Porteføljestyret for forskning og kunnskapsressurser
Dato	28.08.2023
Saksnummer	11/23
Saksarkivnr.	23/00000
Sakstype	V-sak
Saksansvarlig	Vigdis Namtvedt Kvalheim
Saksbehandler	Anita Ellerås

Godkjenning av innkalling og dagsorden

Forslag til vedtak

Porteføljestyret for forskning og kunnskapsressurser godkjenner innkalling og dagsorden.

Bakgrunn

Innkalling i form av kalenderinvitasjon ble sendt medlemmer og observatører 26. juni 2023 (se vedlegg). Dato for møtet ble valgt etter doodle-forespørsel til medlemmer og observatører.

Dagsorden følger vedlagt.

Vedlegg

PS-FK-11-23 Vedlegg: Innkalling

PS-FK-11-23 Vedlegg: Dagsorden for Porteføljestyrets møte 2/2023



Sikt
Kunnskapssektorens
tjenesteleverandør

PS-FK-11-23 Vedlegg: Innkalling

Porteføljestyremøte FK 2/23

Arrangør  Dag Egil Kiberg

Klokkeslett mandag 28. august 2023 09:00-12:00

Plassering

Svar  Godtatt [Endre svar](#)

Hei

Da blir neste porteføljestyremøte 28. august, 09.00 – 12.00 (digitalt).

Agenda og saksdokumenter blir tilsendt ca én uke før møtet.

Jeg vil gjerne benytte anledningen til å ønske dere alle en riktig god sommer!

Med vennlig hilsen

Dag Kiberg

seksjonsleder, Seksjon for forskningsdata

Telefon: (+47) 402 11 301



Sikt

Sikt – Kunnskapssektorens tjenesteleverandør

PS-FK-11-23 Vedlegg: Dagsorden for Porteføljestyrets møte 2/2023

Mandag 28. august 2023 kl. 09.00-12.00. Videomøte.

Tidspunkt	Saks- nr.	Type	Sak	Ansvarlig
09.00 - 09.05	11/23	V	Godkjenning av innkalling og dagsorden	Styreleder
09.05 - 09.10	12/23	V	Godkjenning av protokoll fra møte 1/2023	Styreleder
09.10 - 09.30	13/23	O	Orientering fra Sikt	Vigdis Namtvedt Kvalheim
09.30 - 10.00	14/23	V	Fremtidig kostnadsfordeling for Nasjonalt vitenarkiv	Frode Arntsen
10.00 - 10.30	15/23	O/D	Microdata.no	Ørnulf Risnes
10.30 - 10.40			Pause	
10.40 - 11.10	16/23	O/D	MODANSA- en modernisering av Sikts infrastruktur for forskningsdata	Bodil Agasøster
11.10 - 11.30	17/23	O/D	Kunnskapsressurser og bibliotek tjenester	Asbjørn Risan
11.30 - 11.50	18/23		Eventuelt og møteplan for høsten 2023 Katrine Weisteen Bjerde, HK-dir: Kartlegging av tilstanden i UH-sektoren for digital omstilling	Styreleder
			Oppsummering og evaluering	Styreleder
			Møteslutt	Styreleder

Vedlagte saksdokumenter:

- PS-FK-11-23 Godkjenning av innkalling og dagsorden
- PS-FK-11-23 Vedlegg: Innkalling
- PS-FK-11-23 Vedlegg: Dagsorden for Porteføljestyrets møte 2/2023
- PS-FK-12-23 Godkjenning av protokoll fra møte 1/2023
- PS-FK-12-23 Vedlegg: Protokoll fra møte 1/2023
- PS-FK-14-23 Fremtidig kostnadsfordeling for Nasjonal vitenarkiv
- PS-FK-14-23 Vedlegg: Forslag til kostnadsfordelingsmodell for Nasjonal vitenarkiv
- PS-FK-15-23 Microdata.no
- PS-FK-16-23 MODANSA - en modernisering av Sikts infrastruktur for forskningsdata
- PS-FK-16-23 Vedlegg: Rapport om modernisering av Data om det norske samfunn
- PS-FK-17-23 Kunnskapsressurser og bibliotek tjenester

Saksdokument

Porteføljestyre for forskning og kunnskapsressurser

Til	Porteføljestyret for forskning og kunnskapsressurser
Dato	28.08.2023
Saksnummer	12/23
Saksarkivnr.	23/00000
Sakstype	V-sak
Saksansvarlig	Vigdis Namtvedt Kvalheim
Saksbehandler	Dag Kiberg

Godkjenning av protokoll fra møte 1/2023

Forslag til vedtak

Porteføljestyret for forskning og kunnskapsressurser godkjenner protokoll fra møte 1/2023.

Bakgrunn

Vedlagt følger protokoll fra møtet 2. juni 2023 i Porteføljestyret for forskning og kunnskapsressurser.

Vedlegg

PS-FK-12-23 Vedlegg: Protokoll fra møte 1/2023

PS-FK-12-23 Vedlegg: Protokoll fra møte 1/2023

Sikt – Kunnskapssektorens tjenesteleverandør

Porteføljestyret for
forsknings- og kunnskapsressurser

Protokoll fra møte i Porteføljestyret for forsknings- og kunnskapsressurser 01.06.2023

Møte	Møte nr. 1/2023	
Dato	01.06.2023	
Tid	Kl. 10:00-15:00	
Sted	Vestlandshuset, Bergen	
Medlemmer		
<i>Til stede</i>	Kjersti Melberg (leder)	Ingelin Steinsland
	Maria Carme Torras Calvo	Tord Tjeldnes (digitalt)
	Yngve Foss	Ellen Hermanrud
	Geir Arne Rosvoll	Trude Eikebrokk
Varamedlemmer		
<i>Til stede</i>	Torunn Skofsrud Boger	Anna Løken
Forfall		
	Jan G. Bjålie	Pål Vegar Storeheier
	Lise Amy Hansen	Tone Fløtten
Faste observatører		
HK-dir	Katrine Weisteen Bjerde	
Forskningsrådet	Johannes Waage Løvhaug	
Sikt		
	Vigdis N. Kvalheim (div.dir.)	Dag Kiberg (sekretær)
	Bodil Agasøster	Frode Arntsen
	Nina Karlstrøm	Marianne H. Myhren
	Asbjørn Risan	Julie Storlie (digitalt)

Merknader:

For utdypende informasjon til protokoll, se felles presentasjon og saksunderlag fra møtet:

<https://sikt.no/samstyringsmodell/portefoljestyret-forsknings-og-kunnskapsressurser>

Sak	Tema	Sakstype
1/23	Konstituering av porteføljestyret med kort presentasjon av møtedeltakerne	V-sak
	Kjersti Melberg ønsket velkommen og presenterte seg som leder av porteføljestyret. Deretter presenterte øvrige medlemmer, varamedlemmer, faste observatører og deltakere fra Sikt seg. Vedtak: Porteføljestyret for forskning og kunnskapsressurser ble konstituert med Kjersti Melberg som leder.	
2/23	Godkjenning av innkalling og dagsorden	V-sak
	Vedtak:	

PS-FK-12-23 Vedlegg: Protokoll fra møte 1/2023

Sikt – Kunnskapssektorens tjenesteleverandør

Porteføljestyret for
forsknings- og kunnskapsressurser

	Porteføljestyret for forskning og kunnskapsressurser godkjente innkalling og dagsorden.	
3/23	Porteføljestyrets mandat, roller, oppgaver og arbeidsform	O/D-sak
	Vigdis Namtvedt Kvalheim orienterte om Sikt, samstyringsmodellen og porteføljestyrets og Digitaliseringsstyrets mandat. Se presentasjon for utfyllende informasjon. Kommentarer: Styret drøftet Styrets rolle i forhold til Sikts budsjett, Digitaliseringsstyret og institusjonene. Det ble vist til at Styret ikke disponerer eget budsjett, men har stor påvirkning på utvikling, avvikling og prioritering innen produkt- og tjenesteporteføljen. Hovedoppgaven for Styret er å beslutte og følge opp produktstrategier og årlige prioriteringer. Samtidig vil Styret påvirke gjennom den dialogen og de drøftinger som foregår i Porteføljestyret. Det var enighet mellom Styret og Sikt om at Styrets handlingsrom og påvirkning er avhengig av at det legges tydelige rammer for arbeidet og at det er god dialog mellom sekretariatene for porteføljestyrene og sekretariatet for Digitaliseringsstyret. Det ble informert fra Sikt, om at det allerede er etablert fora for å sikre nødvendig samarbeid og koordinering. Det ble også påpekt fra Sikt at det for å få god porteføljestyling i sektoren er viktig at HK-dir., Forskningsrådet og Sikt samarbeider godt. Vedtak: Styret drøftet styrets roller, oppgaver og arbeidsform og tok for øvrig saken til orientering.	
4/23	Porteføljen for Forsknings- og kunnskapsressurser	O-sak
	Vigdis Namtvedt Kvalheim orienterte om Sikt og om bredden i Divisjon for forsknings- og kunnskapsressursers tjenesteområder. Porteføljen kjennetegnes av mange og ulike typer produkter. Disse dekker tjenester rettet mot ulike deler av kunnskapssektoren, både i og utenfor UH- og forskningssektoren, og er rettet både mot individer og institusjoner. Tjenestene utvikles og drives gjennom ulike finansieringsmodeller og -kilder. Divisjonens samlede inntektsbudsjett er for 2023 på MNOK 242,3. Orienteringen inneholdt også en kort orientering om bakgrunnen for etableringen av Sikt og Sikts mandat og målsettinger. Se presentasjon for utfyllende informasjon. Kommentarer: Det ble fra Sikt vist til at det er viktig for Styret å ha en god og fullstendig oversikt over porteføljen. Senere i dette møtet og i senere møter vil Styret derfor orienteres om og få anledning til å drøfte spesifikke tjenester og produkter mer inngående.	

PS-FK-12-23 Vedlegg: Protokoll fra møte 1/2023

Sikt – Kunnskapssektorens tjenesteleverandør

Porteføljestyret for
forsknings- og kunnskapsressurser

	Vedtak: Porteføljestyret tok saken til orientering.	
5/23	Presentasjon av Nasjonalt vitenarkiv (CRIS/NVA)	O/D-sak
	Fagdirektør Frode Arntsen orienterte om Cris/NVA som produkt/tjeneste. Nasjonalt vitenarkiv (Cris/NVA) skal erstatte dagens Cristin og de institusjonelle vitenarkivene (70+). Dette er ett av virkemidlene fra regjeringens side for å realisere ambisjonen om åpen forskning. Cris/NVA bidrar til dette ved å gjøre informasjon åpent tilgjengelig i en felles nasjonal løsning. Prosjektets opprinnelige plan fra 2018/19 skisserte en ferdigstilling av NVA sommeren 2022. Prosjektplanen er senere justert og ambisjonen er nå å skifte system i november/desember 2023. Forsinkelsen skyldes i hovedsak at prosjektet i store deler av sitt forløp har hatt utfordringer med tilgang på IT-utviklere. Kommentarer: Porteføljestyret drøftet risikoelementer knyttet til utvikling og bruk av den nye tjenesten, om det er satt av nok ressurser i første driftsfase og om ROS-analysen som ble presentert i møtet var for optimistisk. Det ble her særlig lagt vekt på at det i tråd med prosjektplanen, blir gjennomført ende-til-ende-tester i god tid før tjenesten rulles ut i stor skala. Tilbakemeldingen var at forhold knyttet til testing og ressurstilgang er tilstrekkelig ivarettatt. Det vil bli foretatt en kvalitetssikring og eventuelt justering i ROS-analysen. Det ble stilt spørsmål om finansiering av tjenesten når denne er rullet ut. Her ble det vist til at det i tråd med forutsetningen fra Kunnskapsdepartementet, legges opp til brukerfinansiering av vitenarkivdelen av NVA. Forslag til finansieringsmodell vil fremlagt for PS på neste møte. Vedtak: Porteføljestyret tok saken til orientering og ba om at de innspill som kom i gjennomgang av saken inngår i det videre arbeidet knyttet til risikovurdering og veien videre.	
6/23	Strategi for åpen publisering etter 2024	O/D-sak
	Produktområdeleder Nina Karlstrøm orienterte om at det i regjeringens langtidsplan for forskning og høyere utdanning 2023-2032 (Meld. St. 5 (2022-2032)) står at «Regjeringen ber HK-dir, Sikt og Forskningsrådet bistå forskningsinstitusjonene med å utarbeide en strategi for norsk vitenskapelig publisering etter 2024 og en plan for å nå målene i en slik strategi». For å svare på dette oppdraget ble det nedsatt en referansegruppe og en arbeidsgruppe bestående av representanter fra Universitets- og Høgskolerådet (UHR), Forskningsinstituttene Felles Arena (FFA), de regionale	

PS-FK-12-23 Vedlegg: Protokoll fra møte 1/2023

Sikt – Kunnskapssektorens tjenesteleverandør

Porteføljestyret for
forsknings- og kunnskapsressurser

	helseforetakene (RHF), Sikt-Kunnskapssektorens tjenesteleverandør, HK-dir og Forskningsrådet. Utkast til strategi blir lagt ut på openscience.no, og det inviteres til innspill med frist 1.oktober. Endelig rapport skal leveres høsten 2023. Her skal det formuleres et sett av anbefalinger i nær dialog med institusjonene. Det er institusjonene selv som er ansvarlige for å oppnå kravet om åpen tilgang. Kommentarer: Styret viste til betydningen av åpen publisering, men at det samtidig er viktig at man har fokus på kvaliteten. Her ble det bl.a. vist til at Kanalregisteret har en sentral rolle. Det ble vist til at åpen publisering gir mer bruk, ikke bare innen akademia, men også i ulke profesjoner og i forvaltningen. Vedtak: Porteføljestyret tok saken til orientering.	
7/23	Behovs- og mulighetsstudien for Nasjonal infrastruktur for bibliometri (NIB)	O/D-sak
	Produktområdeleder Bodil Agasøster orienterte om formål og drift av NIB, og om anbefalinger fra arbeidsgruppen for NIB. Arbeidsgruppen ble oppnevnt av Tjenesterådet i 2021 og hadde som formål å gi råd om videreutvikling av tjenesten. Dette vil kreve oppbemanning og behov for investeringsmidler og varig økning av driftsmidler. Saken vil komme opp til drøfting i styret i neste møte. Kommentarer: Styret drøftet saken og viste bl.a. til at det er viktig at man gjør nødvendige investeringer og utvikling før man introduserer brukerbetaling. Dette er i tråd med hvordan Sikt eventuelt vil legge opp til. For øvrig viste styret til at man bør se på muligheter for å få til tverrnasjonale løsninger. Det ble også stilt spørsmål om muligheter for å ta i bruk KI-teknologi. Vedtak: Porteføljestyret tok saken til orientering.	
8/23	Møteplan 2023	V-sak
	Til møtet ble følgende møter foreslått: • September (digitalt møte) • Oktober (reservetidspunkt, digitalt møte) • November (digitalt møte) Vedtak: Endelige møtedatoer bestemmes basert på svar på doodle.	
9/23	Eventuelt	
	Ingen saker til eventuelt.	
10/23	Oppsummering og evaluering	

PS-FK-12-23 Vedlegg: Protokoll fra møte 1/2023

Sikt – Kunnskapssektorens tjenesteleverandør

Porteføljestyret for
forsknings- og kunnskapsressurser

	Styreleder takket for møtet og opplyste at de neste møtene vil bli digitale. Kommentarer: Det ble bedt om at det i dagsorden blir gitt tidsrammer på hver sak. Dette vil bli fulgt opp av sekretariatet.	
--	---	--



Saksdokument

Porteføljestyre for forskning og kunnskapsressurser

Til	Porteføljestyret for forskning og kunnskapsressurser
Dato	28.08.2023
Saksnummer	14/23
Saksarkivnr.	23/00000
Sakstype	V-sak
Saksansvarlig	Frode Arntsen
Saksbehandler	Frode Arntsen

Fremtidig kostnadsfordeling for Nasjonalt vitenarkiv

Forslag til vedtak

Årlig brukerbetaling for Nasjonalt vitenarkiv fordeles med en fast årlig avgift, samt en variabel andel basert på den enkelte institusjons antall faglige årsverk. Den variable andelen fratrekkes en trinnvis rabatt ved økende antall årsverk, i tråd med beskrevet kostnadsfordelingsmodell.

Små institusjoner (fem eller færre faglige årsverk) betaler 50% fast avgift. Små institusjoner som samles i klynger, betaler 150% fast avgift for inntil 20 institusjoner, over dette betales 200% fast avgift, samt normal avgift for klyngens totale antall faglige årsverk. Faglige årsverk skal normalt hentes fra autoritative kilder (DBH/NFR/SSB), alternativt benyttes antall faglige årsverk som for institusjoner av tilsvarende størrelse/virksomhet, eller tall oppgitt av institusjonen selv.

For 2024 fastsettes årlig avgift per institusjon til kr 27 500¹, og basispris per faglige årsverk til kr 270. Beløpene justeres normalt årlig i tråd med lønns- og prisstigning, og dette vil spesifiseres i forvarsel fra Sikt.

Bakgrunn

Sikt er i ferd med å slutføre utviklingen av Nasjonalt vitenarkiv, en tjeneste som forener funksjonaliteten fra Cristin og de over 70 institusjonelle vitenarkivene, i hovedsak Brage-arkiver, til en felles produktplattform – Nasjonalt vitenarkiv. Den nye løsningen planlegges satt i drift innen utgangen av 2023.

Dagens Cristin-tjeneste er bevilgningsfinansiert, mens dagens Brage-tjeneste i sin helhet er brukerfinansiert med ca. 4 MNOK/år. Brukerinntektene på Brage-arkivene har i hovedsak vært benyttet til driftsoppgaver, herunder brukerstøtte og jevnlig oppgradering av programvaren. Kunnskapsdepartementet forutsetter at den nye tjenesten skal være delvis brukerfinansiert².

¹ Alle beløp i dokumentet er gitt eks. mva.

² [Stortingsproposisjon 1, 2020-2021](#)



Forventede kostnader for forvaltning og videreutvikling etter produksjonssetting av NVA (se vedlegg) er tidligere detaljert gjort rede for i Tjenesterådet og Digitaliseringsstyret. Totalt forventes kostnader fra 2024 på 42 MNOK/år, med en forventet andel på 32 MNOK av Sikt sin totale bevilgning. Dette gir behov for en brukerfinansiering på 10 MNOK i 2024, og tilsvarende beløp videre fremover. Ved en lavere bevilgningsandel og/eller lavere brukerinntekt vil konsekvensen være at videre utvikling av ønsket funksjonalitet i Nasjonalt vitenarkiv vil ta lengre tid å realisere. Det forventes også at brukerne vil spille inn behov for endret og ny funksjonalitet etter at de får erfaring med tjenesten.

Se vedlegg for en mer detaljert redegjørelse.

Vurdering

Forvaltning og videreutvikling av Nasjonalt vitenarkiv skal finansieres ved en kombinasjon av en andel av Sikt sin bevilgning over Statsbudsjettet og brukerfinansiering fra de institusjonene som benytter tjenesten.

Saken er i flere omganger behandlet av *Tjenesteråd for forskning*, senest som drøftingssak den 14. november 2022 i sak 44/22 om fremtidig finansieringsmodell. Tjenesterådet mente da at det synes fornuftig å benytte antall faglige årsverk som en fordelingsnøkkel i tillegg til et fastledd.

Digitaliseringsstyret ble den 30. nov. 2022, i sak 53/22, orientert om innføringsplan og finansieringsplan for Nasjonalt vitenarkiv, herunder skisse til modell for brukerfinansiering i tråd med retningslinjer for fellestjenester gitt av Digitaliseringsstyret. Digitaliseringsstyret tok saken til orientering uten kommentarer knyttet til finansieringsmodell.

I etterkant av denne behandlingen har Sikt jobbet videre med kostnadsfordelingsmodell for Nasjonalt vitenarkiv, i hovedsak hvordan en skal redusere kostnaden per årsverk med stigende antall. Her er det tatt utgangspunkt i fordelingsmodell benyttet for øvrige fellestjenester i kunnskapssektoren, ref. rammeverksdokumentet [Forvaltning og videreutvikling av fellestjenester i høyere utdanning og forskning](#), samt en intensjon om å holde kostnaden lavest mulig. Det er og vurdert hvordan en skal hensynta de forskjellige brukerinstitusjonenes betydelige spenn i størrelse/antall faglige årsverk mellom de største og de minste institusjonene. Aktørenes behov for en forutsigbar finansieringssituasjon for tjenesten er også en viktig parameter.

I hovedtrekk foreslås en modell der samtlige institusjoner i utgangspunktet betaler samme fastpris per år, med 50% reduksjon for institusjoner med færre enn 5 faglige årsverk. Dette gjelder særlig for små museer og muligens enkelte små private helseinstitusjoner. For de små helseinstitusjonene er det en pågående dialog med HOD/RHF-ene rundt de små institusjonenes fremtidige bruk av tjenesten. De små museene, som i dag fremstår som én institusjon samlet under en «paraply» i Cristin, har en pågående vurdering om hvordan de vil fremstå og bruke Nasjonalt vitenarkiv. Alternativer her er å videreføre dagens modell eller evt. å samle seg under flere mindre og regionale «paraplyer». For en slik samling av institusjoner foreslås det at de betaler en årlig avgift som er 50-100% høyere enn normal fastandel (avhengig av antall institusjoner samlet, 1-20 eller flere enn 20) i tillegg til normal pris for paraplyinstitusjonenes antall faglige årsverk, se eksempel i vedlegg.

For 2024 foreslås en normal fastpris på kr 27 500. for hver institusjon, som justeres med hhv. 50% rabatt for de minste (<5 faglige ÅV) og et tillegg på 50-100% for hver samling av små institusjoner (flere eller færre enn 20). Kostnaden per faglige årsverk foreslås med en basiskostnad på kr 270 per år, fratrasket en trinnvis rabatt ved økende antall årsverk som vist i tabell på neste side.

Premisset har vært at rabatten skal være økende med økende antall faglige ansatte, og gi en kostnad som ikke er vesentlig høyere enn brukerbetaling for dagens Brage-arkiv. Med denne rabattfordelingen og beløpene som foreslått, vil institusjonene som betaler minimumspris blant dagens Brage-brukere få en økning fra kr 18 504 i 2023 til kr 29 000 i 2024. Institusjonen som i dag betaler mest for bruk av Brage, vil få en økning fra kr 234 384 i 2023 til kr 356 000 i 2024. Det foreslås at beløpene normalt justeres for lønns- og prisutvikling årlig, og varsles som del av Sikt sitt årlige kostnadsvarsel til institusjonene.



Rabatt i %	Fra	Til	Pris per ansatt	
0,0 %	0	10	kr	270,00
5,0 %	11	30	kr	256,50
8,0 %	31	50	kr	248,40
12,0 %	51	199	kr	237,60
15,0 %	200	499	kr	229,50
50,0 %	500	1 499	kr	135,00
92,5 %	1 500	99 999	kr	20,25

Merk at det tilkommer en engangsavgift for oppstart, estimert til 25 000 kr. Denne dekker Sikt sine utgifter forbundet med oppsett og oppstart når institusjonen tar NVA i bruk.

Vedlegg

Forslag til kostnadsfordelingsmodell for Nasjonalt vitenarkiv

PS-FK-14-23 Vedlegg: Forslag til kostnadsfordelingsmodell for Nasjonalt vitenarkiv

Detaljert forslag til kostnadsfordelingsmodell for Nasjonalt vitenarkiv (NVA)

Dagens Cristin- og Brage-tjeneste har i hovedsak overlappende brukerinstitusjoner (kunder), men Cristin har totalt sett langt flere brukerinstitusjoner enn Brage. Totalt antall brukerinstitusjoner er avhengig av tellemåte og avhenger av hvordan de mindre institusjonene i både museumssektoren og helsesektoren regnes:

- Om samtlige blir ansett å være frittstående institusjoner, så er sannsynlig kundemasse når tjenesten er i full drift cirka 200 - 220 institusjoner.
- Institusjonene som i dag er samlet under «paraplyer» er
 - Museene (41 stk)
 - Private ideelle Helse Vest (7 stk)
 - Private ideelle Helse SØ (7 stk)

Dersom disse skal videreføres, så er sannsynlig kundemasse cirka 145 - 165 kunder, avhengig av fremtidig strukturering. Sikt har etablert dialog med museums- og helsesektorene (HOD/RHF) for å avklare hensiktsmessige løsninger knyttet til finansiering av de små institusjonene. Det antas å være uforholdsmessig kostbart om de små «paraplyinstitusjonene» pålegges en brukerbetaling eksakt tilsvarende de større og frittstående brukerinstitusjonene.

Sikt mottar jevnlig forespørsler fra institusjoner som i dag ikke bruker hverken Brage eller Cristin, typisk frittstående forskningsinstitusjoner ol. som viser interesse for å ta i bruk NVA, og som vil kunne øke antallet kunder.

Dagens forvaltningskostnader for Brage og Cristin

Brage-arkivene, som er basert på åpen kildekode, benyttes av 72 institusjoner (UiO og UiT har egne institusjonelle vitenarkiv) som samlet betaler ca. 4 MNOK årlig for bruk av tjenesten. Inntektene har finansiert i underkant av 2 årsverk og nødvendig drift/infrastruktur, integrasjoner, vedlikehold og forvaltningsoppgaver.

Cristin-tjenesten er 100% finansiert gjennom en andel av Sikt sin bevilgning over statsbudsjettet. Oppgaver som ivaretas av 9 ÅV er drift/forvaltning av systemet Cristin med relaterte oppgaver som import av publikasjonsdata samt årlig tilrettelegging av tall for rapportering av Norsk vitenskapsindeks (NVI) og tilsvarende tilrettelegging av tall for rapportering av kliniske behandlingsstudier (KBS). I tillegg tilrettelegges data fra Cristin for mer omfattende institusjonelle analyser gjennom DUCT (Data Ut fra Cristin via Tableau).

Samtlige forvaltningsoppgaver som i dag gjennomføres i tilknytning til Cristin må videreføres etter overgangen til Nasjonalt vitenarkiv. Cristin-relaterte forvaltningsoppgaver mottar en bevilgning på ca. 26 MNOK i 2023.



Fremtidige kostnader for NVA

Forventede kostnadssatte årsverksbehov for utviklings- og forvaltningsoppgaver samt direkte driftskostnader for perioden 2024-2026 er vist nedenfor.

Bemanning/ÅV/MNOK	2024	2025	2026
Utviklingsteam (ÅV/MNOK)	9,0/15,9	9,0/15,9	7/12,4
Forvaltningsteam (ÅV/MNOK)	9/15,9	9/15,9	9/15,9
Andre driftskostnader (MNOK)	10,2	10,5	8,3
Sum (Tall i MNOK)	42	42,3	36,6

Produktområdet *Forskningsinformasjon* forventer en fremtidig andel av Sikt sin bevilgning på ca. 26 MNOK for forvaltningsoppgaver knyttet til NVA, noe som er på samme nivå som i 2023.

	2024	2025	2026	Kommentarer
Totale kostnader NVA	-42	-42,3	-36,6	
Bevilgning forvaltning	26	26	26	
Bevilgning utvikling				Uavklart for 2024-2025
Brukerinntekter	10	10	11	
Differanse	-6	-6,3	0	

Med basis i bemanningsbehovene, er det kun størrelsen på utviklingsinnsatsen og størrelsen på brukerbetaling som kan «justeres på». Lavere innsats på utvikling vil medføre lengre tid før etterspurt funksjonalitet er på plass. Brukerinntekter på 10 MNOK/år er basert på at antall betalende kunder øker betydelig og brukerbetaling for NVA per institusjon vil øke sammenlignet med dagens brukerbetaling for Brage. Brukerinntekter ut over det som er foreslått forventes å medføre innsigelser fra brukerne.

Forslag til modell for kostnadsfordeling

For å avklare hvordan den delen av kostnadene knyttet til NVA som skal brukerfinansieres bør fordeles, er det naturlig å ta utgangspunkt i hvilke kostnadsdrivere som foreligger. I tillegg vil de relevante fordelingsmodellene som tidligere er behandlet av Digitaliseringsstyret og vist i rammeverksdokumentet [Forvaltning og videreutvikling av fellestjenester i høyere utdanning og forskning](#), som sist ble revidert i mai 2021, være et relevant grunnlag for hvordan kostnader skal fordeles.

Dokumentets vedlegg 2 – *Prinsipper for enhetlige betalingsmodeller for fellestjenester* – gjør rede for prinsippene for prising av fellestjenester, samt kriterier for enhetlige betalingsmodeller. Generelt diskuteres omkring bruk av en fast kostnad uavhengig av bruk, samt et eller flere element som enkelt kan relateres til faktisk bruk.

For NVA vil de tydeligste kostnadsdriverne være faktisk bruk som reflekteres i prosessortid hos AWS, samt hvor store mengder data den enkelte institusjon lagrer hos AWS. Det vil være vesentlig høyere kostnader knyttet til lagring av lyd- bilde- og videofiler sammenlignet med PDF-filer, som er det vanligste for vitenskapelige artikler, rapporter og studentoppgaver.

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



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

Det vurderes ikke andre parametere ut over de som fremkommer i rammeverksdokumentet som er behandlet og vedtatt av Digitaliseringsstyret.

Med utgangspunkt i disse mulige fordelingsparameterne synes det mest naturlig å benytte antall FoU-ansatte (årsverk), da de øvrige parameterne vanskelig lar seg sammenligne på tvers av de forskjellige forskningsutøvende sektorene. Antall studenter er ikke relevant for denne tjenesten da flertallet av brukerinstitusjoner ikke driver utdanning, antall tilsatte anses ikke som relevant da mange av institusjonene ikke har hovedfokus på forskning og størrelse på bevilgning anses også som lite egnet av samme grunn.

For de fleste brukerinstitusjonene finnes det offentlige, kvalitetssikrede tall for FoU-ansatte i de forskjellige sektorene.

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

Det kan argumenteres med at UH-sektoren skal betale en relativt høyere andel av kostnadene fordi de også lagrer studentoppgaver, men samtidig er situasjonen at et FoU-årsverk ved en UH-institusjon kun benytter ca 50% av tiden til forskning og publisering, mens en ansatt ved et institutt benytter hele årsverket til forskning, utredninger og publiseringer. Den andelen en UH-ansatt benytter til undervisning vil også telles med som FoU-årsverk, og kan slik anses å ivareta studentoppgavens plass i Nasjonalt vitenarkiv.

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

For enkelte fellestjenester benyttes årlige beregningsgrunnlag basert på tre års glidende gjennomsnitt for å unngå store variasjoner hos spesielt de mindre institusjonene. Denne modellen anbefales i utgangspunktet ikke da den gir merarbeid og større risiko for feil.

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

Institusjonenes størrelse er en faktor som må vurderes, spesielt hvorvidt det er hensiktsmessig med en lineær fordeling av årsverksbaserte kostnader eller om det skal være relativt sett dyrere for en liten enn en stor institusjon.

Vitenarkivet Brage benytter en trappetrinnsmodell bestående av seks trinn, basert på et prinsipp der størrelse gir rabatter, hvor man basert på sine FoU-årsverk (og da til dels grove estimater av dette)



plasseres på et trappetrinn med en gitt årlig kostnad. Dette oppleves å ha fungert fint i en årrekke, men det har gitt til dels uheldige utslag for spesielt de nesten største institusjonene som har plassert seg nært inntil grensene for forrige/neste trappetrinn.

Sammenlignet med en lineær modell eller trappetrinnsmodell, vil en modell med en tilnærmet jevnt avtagende kurve gi fordeler og ulemper, men det vil være lettere å se sammenhenger ved å ha flere eller færre årsverk, sammenlignet med en trappetrinnsmodell. En lineær modell indikerer at det er et direkte sammenlignbart forhold mellom kostnader og institusjonens størrelse. Små institusjoner har normalt et mindre støtteapparat, og krever ofte relativt sett mer brukerstøtte og oppfølging. Samtidig vil en del driftsoppgaver være av samme omfang uavhengig av institusjonenes størrelse. Dette taler mot en lineær modell.

For å korrigere for slike «urettferdigheter» benytter andre fellestjenester en modell med en avrundet kurve som gjør at de store institusjonene (f.eks. NTNU og UiO) ikke straffes unødvendig mye for å være nettopp store. Det er også en slik modell som er lagt til grunn for det som nå foreslås for NVA.

Med utgangspunkt i at deler av Sikt sine oppgaver er uavhengig av institusjonenes størrelse anbefales det at modellen har to komponenter: en fast del som i hovedsak er uavhengig av institusjonenes størrelse, og en variabel del som avhenger av antall FoU/faglige årsverk med en avrundet, rabattert kurve.

I hovedtrekk foreslås en modell der samtlige institusjoner i utgangspunktet betaler samme fastpris per år, med en 50% reduksjon for institusjoner med fem eller færre faglige årsverk. Dette gjelder særlig for små museer og enkelte små helseinstitusjoner. For de små helseinstitusjonene er det en pågående dialog med HOD/RHF-ene rundt de små institusjonenes fremtidige bruk av tjenesten.

Mange av de mindre museene, som i dag fremstår som en institusjon samlet under en «paraply» i Cristin, har en pågående vurdering om hvordan de vil fremstå og bruke Nasjonalt vitenarkiv. Alternativer her er å videreføre dagens modell, eller evt. å samle seg under flere mindre og regionale «paraplyer». For en slik samling av institusjoner foreslås det at de betaler en årlig avgift som er 50% høyere enn normal fast andel, i tillegg til en normal pris for paraplyinstitusjonenes samlede antall faglige årsverk.

For 2024 foreslås en normal fastpris på kr 27500 ekskl. mva. for hver institusjon, som justeres med hhv. 50% rabatt for de minste (fem eller færre faglige ÅV) og et tillegg på 50-100% for hver samling av små institusjoner. Konkret foreslås 50% høyere for 1-20 institusjoner samlet og 100% høyere fastavgift for flere enn 20 institusjoner. Eksempel på kalkyle for en samling små institusjoner nedenfor.

Den variable andelen foreslås å starte med en kostnad på kr 270/årsverk opp til 10, avtagende med en gradvis økende rabatt som vist i tabellen nedenfor, slik at institusjoner som har mer enn for eksempel 1500 FoU-/faglige årsverk betaler kr 20,25/årsverk fra og med årsverk nr. 1500.

Rabatt i %	Fra	Til	Pris per årsverk
0,0 %	0	10	kr 270,00
5,0 %	11	30	kr 256,50
8,0 %	31	50	kr 248,40
12,0 %	51	199	kr 237,60
15,0 %	200	499	kr 229,50
50,0 %	500	1499	kr 135,00

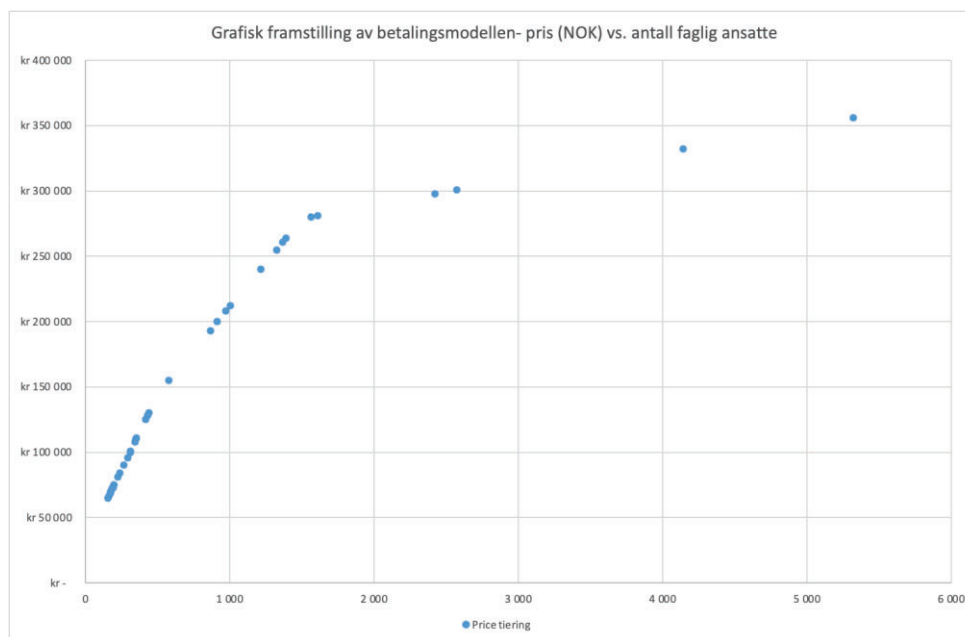


92,5 %	1500	-	kr	20,25
--------	------	---	----	-------

For forskjellige typer institusjoner, eksemplifisert med store, middels og små som vist i tabellen nedenfor, vises hvordan kostnaden for årsverk i de forskjellige intervallene fremkommer. Nederst er det vist 4 små museer med 2 faglige årsverk, først hver for seg med pris basert på 50% av vanlig fastpris og deretter som en paraply bestående av 4 museer med 150% fastprisandel og respektive faglige årsverk som basis for beregning. En «paraply» vil ha en kostnad som er 13750 kroner lavere sammenlignet med 4 individuelle virksomheter, men være 13750 kroner høyere enn en enkeltinstitusjon med tilsvarende totalt antall faglige årsverk.

	TOT ÅV	Fastandel	1-10	11- 30	31- 50	51- 199	200- 499	500- 1499	1500+	Sum	Merknad
UiO	4141	27 500	2 700	5 130	4 968	35 640	68 850	135 000	53 510	333 298	
UiB	2572	27 500	2 700	5 130	4 968	35 640	68 850	135 000	21 721	301 509	
Oslo UniSykehus	1729	27 500	2 700	5 130	4 968	35 640	68 850	135 000	4 653	284 441	
Sintef	1389	27 500	2 700	5 130	4 968	35 640	68 850	120 150		264 938	
NIFU	57	27 500	2 700	5 130	4 968	1 663				41 961	
Museum Hord	2	13 750	540							14 290	Fastandel= 50% av normal
Narviksenteret	2	13 750	540							14 290	Fastandel=. 50% av normal
Nord-Troms museum AS	2	13 750	540							14 290	Fastandel=. 50% av normal
Romsdalsmuseet	2	13 750	540							14 290	Fastandel=. 50% av normal
Museumssamling	8	41 250	2 160							43 410	Fastandel= 150% av normal

Forslaget til kostnadsfordelingsmodell gir en dels avrundet kurve som illustrert under. Modellen har markerte overganger for intervallene 0-500, 501-1499 og 1500 og høyere og kan selvfølgelig granuleres ytterligere for å få en jevnere kurve uten at dette anses å medføre nevneverdige endringer for enkeltinstitusjoner.



FoU-/faglige årsverk

Forslaget er basert på at algoritmen som beregner kostnad for den enkelte institusjon tar utgangspunkt i antall årsverk som benyttes til forskning.

En utfordring er at det ikke foreligger en entydig og omforent definisjon av begrepet FoU-årsverk eller faglige årsverk, og at ulike oversikter legger til grunn ulike definisjoner og ulik beregning av tilsvarende tall. Dette vil gi utslag som vanskeliggjør å kunne si at modellen er 100 % sammenlignbar for de forskjellige sektorene. Dette har alltid vært situasjonen for vitenarkivet Brage, og er derfor kjent for – og akseptert av – en betydelig andel av kundene fra før av.

I dag hentes tallene for årsverk i de forskjellige sektorene slik:

- **Universitets- og høyskolesektoren:** Tilgjengelige tall gjennom [DBH](#) hos HKDir benyttes.

Årsverksbegrepet beskrives på samme nettsted til «Begrepet «undervisnings-, forsknings-, og formidlingsstillinger» er ikke entydig definert, noe som kan føre til at ulike tallstørrelser omtales som det samme. Andre begreper som tilsvarende dette begrepet, er faglige stillinger og vitenskapelige stillinger. Tallet inkluderer faglige-administrative lederstillinger, som i den opprinnelige inndelingen ikke ble regnet med i denne kategorien.» For UH-sektoren vil antall lederstillinger være en svært liten andel av det totale antall FoU-årsverk og anses ikke å påvirke totaltallet.

- **Helsesektoren:** FoU-årsverk tilgjengeliggjøres for oddetallsår i rapporten "[Ressursbruk til forskning i helseforetakene](#)" (se tabell B4), utgitt av Statistisk sentralbyrå. I rapport 2022/42 foreligger tall for helseforetakene fra 2021.
- **Instituttsektoren:** Forskerårsverk for forskningsinstitutter hentes fra [Forskningsrådets nøkkeltall for instituttsektoren](#) (fane for årsverk) og i enkelte tilfeller fra institusjonens årsrapport.
- **Øvrige:** For f.eks. museene finnes i dag ingen samlet oversikt over FoU-årsverk. Enkelte institusjoner rapporterer tall i sine årsrapporter, mens det for andre fremstår mest hensiktsmessig å basere på selvrapportering fra institusjonen selv. Under kategorien øvrige ligger også for eksempler institusjoner som Norges Bank, NVE osv.



Der tallene foreligger som årlig oppdaterte tall, foreslås at disse benyttes. Der det foreligger tall for andre perioder benyttes disse for respektive påfølgende periode.

Sikt vil kontinuerlig dokumentere hvilke tall og beregninger som benyttes i underlaget, for å sikre en så rettferdig og åpen kostnadsfordeling som mulig.

Saksdokument

Porteføljestyre for forskning og kunnskapsressurser

Til	Porteføljestyret for forskning og kunnskapsressurser
Dato	28.08.23
Saksnummer	15/23
Saksarkivnr.	23/00000
Sakstype	O/D
Saksansvarlig	Vigdis Namtvedt Kvalheim
Saksbehandler	Ørnulf Risnes

Microdata.no

Forslag til vedtak

Porteføljestyret tar saken til orientering og ber om å få framlagt forslag til prismodell.

Saksfremstilling

Kort om microdata.no

Microdata.no er datainfrastruktur hvor forskere, analytikere og studenter får umiddelbar og fleksibel tilgang til kobling og analyse av store mengder registerdata som er tilgjengeliggjort der.

Dataprodusenter som Statistisk sentralbyrå, Kreftregisteret og andre offentlige og private dataprodusenter og -eiere kan selvbetjent dele sine registerdata i microdata.no, til bruk for forskning og analyse. Foreløpig er SSB eneste dataeier med data på plattformen, men et tett og langvarig samarbeid med de sentrale helseregistrene har som mål at helseregistrene også skal kunne dele sine registerdata i microdata.no, bl.a. for kobling med demografiske og sosio-økonomiske data fra SSB.

Løsningen kontrollerer all databruk gjennom et automatisert anonymiserende grensesnitt, som hindrer at sluttbrukerne får ser eller kan laste ned/printe ut personopplysninger. Fordi sluttbrukere i **juridisk forstand** ikke selv behandler personopplysninger, kan de få tilgang på dagen og dermed også adgang til å analysere detaljerte individdata uten noen form for søknad eller dokumentasjon av behandlingsgrunnlag. Tilgangen gjelder for personer som er ansatte eller studenter ved en institusjon som har tegnet avtale om bruk av microdata.no.

Microdata.no eies, utvikles og driftes av Sikt og SSB i fellesskap. Utviklingen har i hovedsak vært finansiert gjennom Forskningsrådets infrastruktursatsning i to omganger. Tjenesten ble lansert våren 2018, inneholder i dag ca. 440 variabler¹ med data om hele befolkningen på områder som demografi, inntekt, gjeld, utdanning, migrasjon, sykefravær og trygdeytelser tilbake til 1960-tallet. Så langt har data fra microdata.no vært benyttet i over 170 publikasjoner², fra nivå 2-publikasjoner til masteroppgaver, rapporter og NOU-er.



Brukerne (p.t. ca. 200 aktive brukere pr. mnd, ca. 500 aktive pr. år) kommer fra 105 institusjoner i mange ulike sektorer og segmenter; universiteter og høyskoler, forskningsinstitutter, departementer, direktorater, andre organer i statsforvaltningen, kommuner/fylkeskommuner samt konsultantselskaper med oppdrag for offentlig sektor³. Det forventes økt interesse fra helseforskningssiden når helsedata blir tilgjengelige.

Hvorfor microdata.no?

Norge har som ett av få land store mengder koblingsbare registerdata tilgjengelig. Slike data har høy verdi for forskning og analyse. Begrepet «registerdata» handler typisk om data som har blitt registrert om individer som ledd i en administrativ prosess forbindelse med en hendelse eller faste rapporteringer (f.eks. fødsel, flytting, migrasjon, selvangivelse, gjennomført utdanning, innleggelse på sykehus, trygdeytelser, etc.)

Dette er normalt data som er registrert på individnivå. Begrepet «mikrodata» brukes om slike data for å skille dem fra aggregerte data og statistikk. Koblingsbare mikrodata kan belyse langt flere problemstillinger enn det som er mulig med aggregerte data.

SSB har hovedansvaret for produksjon av offisiell statistikk. Statistikkloven gir SSB anledning til å avkreve data fra ulike organer som NAV, Skatteetaten, Lånekassen, kommunene, m.fl. for å kunne produsere offisiell statistikk. Statistikkloven sier samtidig at SSB skal stille disse dataene til disposisjon for forskning og analyseformål når det er mulig.

SSB har betjent norske forskningsmiljøer med registerdata i en årrekke, bl.a. i samarbeid med NSD som nå er en del av Sikt. SSBs registerdatatilbud startet med data knyttet til velferd/trygd, og har senere blitt utvidet med andre typer demografiske og sosio-økonomiske data. SSBs tilbud samt det faktum at Norge som ett av få land globalt har fødselsnummer tilordnet hver person, har gjort det mulig for forskningsmiljøer å få utlevert koblete mikrodata i relativt stor skala, og gjennom det studere samfunnsutviklingen over tid, mellom regioner, m.m. Slike muligheter er fraværende i de fleste land utenom Norden, grunnet mangel på fødselsnummer og/eller mangel på sentraliserte registre.

Situasjonen i Norge er likevel langt fra problemfri, og i praksis har registerdatatilfanget vært dramatisk underutnyttet også i Norge. Både EU-kommisjonen, OECD og ulike norske regjeringer har over lang tid pekt på viktigheten av å utnytte denne type data bedre i forskning og analyse.

Det er mange årsaker til at registerdata har vært underutnyttet i Norge. , Hovedproblemet sett fra et forskerståsted har vært at søknads- og utleveringsprosessen har vært så tid- og ressurskrevende at den i praksis har ekskludert store brukergrupper fra å få tilgang.

Det er hensynet til både hensynet til personvernet og behovet for tilrettelegging av data for hver enkelt bestilling, som gjør at den tradisjonelle søknads- og utleveringsprosessen er tid- og ressurskrevende. Hvert forskningsprosjekt som ønsker å bruke registerdata må spesifisere nøyaktig hvilke data de trenger ut fra tekstlige beskrivelser av data, for hvilken delpopulasjon og for hvilket tidsrom. De må legitimere behandlingsgrunnlag og forskningsmessige behov for data i forhandling med SSB. Når partene er enige, produserer SSB datafiler ihht bestillingen og leverer ut til forskningsprosjektet. SSB tar ikke betalt for data, men de tar betalt for arbeidet med skreddersydd datatilrettelegging og saksbehandling ifm utlevering.

I praksis kan slike søknads- og utleveringsprosesser ta måneder og i verste fall år, og medføre store kostnader for forskningsprosjektet. Mens enkelte store og ressurssterke forskningsmiljøer hadde anledning til å gjennomføre slike søknadsprosesser mer eller mindre rutinemessig, og har ressurser til å bygge datainfrastruktur og sterke registerforskningsmiljøer lokalt, har praksisen virket ekskluderende for mange potensielle brukermiljøer.

Tjenesten microdata.no ble utviklet for å øke bruken av registerdata i norsk forskning. Virkemiddelet var å utvikle en løsning og etablere en tjeneste som demokratiser tilgangen til registerdata og gjøre slike data tilgjengelig for langt flere på en kostnads- og tidsmessig effektiv måte.



Samarbeidet mellom SSB og Sikt (tidligere NSD)

Sikt og SSB eier, videreutvikler og drifter altså microdata.no i fellesskap. Samarbeidet er regulert i en eierskapsavtale og tilhørende samarbeidsavtale. Hva er historien bak dette samarbeidet?

Langvarig samarbeid

Samarbeidet mellom SSB og NSD startet tidlig på 1970-tallet. NSD fikk ansvaret for formidling av SSB-data til forskningsmiljøene. I første omgang handlet samarbeidet mest om data fra utvalgsdataene (spørreundersøkelser), men omfanget ble senere utvidet til kommunedata og registerdata. NSDs nasjonale rolle som forskningsdataarkiv og tette bånd til Norges forskningsråd og forskningsmiljøene og kunnskap om forskningens behov var bakgrunnen for arbeidsdelingen.

Etter et innledende samarbeid mellom NSD, SSB, Rikstrygdeverket m.fl. om formidling av registerdata fra trygdefeltet (KIRUT) på 90-tallet, etablerte SSB selv et tilbud for formidling av registerdata til forskning med utgangspunkt i tryggedata (FD-Trygd) som senere altså ble utvidet med data fra andre sektorer/aktiviteter. Etter hvert fikk NSD oppdraget med å formidle SSB-registerdata til studenter og forskere fra et forenklet og nedskalert registerdatatilfang (FD-Trygd 20%-utvalget).

NSD som teknologiutvikler og standardiseringspådriver

Fra 80-tallet og fremover etablerte NSD seg som global programvareleverandør av statistikk- og analyseverktøy (NSDstat og Nesstar). Parallelt med dette var NSD aktive pådrivere i internasjonalt arbeid rundt standardisering av metadata, og etter hvert maskinlesbare metadata som kunne drive selvbetjente datatilgangsløsninger for forskere. Statistikkprogramvareløsninger fra NSD var i perioder relativt utbredt i Europa og Nord-Amerika (primært hos universiteter og forskningsdataarkiver).

Første felles søknad til Forskningsrådet infrastrukturetsatsing i 2011

I 2009 etablerte Forskningsrådet ordningen «Nasjonal satsing på forskningsinfrastruktur». I forbindelse med utlysningen i 2011 ble NSD og SSB oppfordret til å skrive en felles søknad om midler til utvikling av en forbedret infrastruktur for registerdatatilgang. Det ble pekt på de respektive institusjonenes komparative fortrinn, herunder NSDs kompetanse om forskningens behov samt teknologiutviklingserfaring rundt formidling og bruk av data. Oppfordringen ble tatt til følge, og NSD og SSB leverte en felles søknad til 2011-utlysningen. Søknaden nådde opp, og 35 MNOK ble bevilget til et 5-årig utviklingsprosjekt under navnet RAIRD (Remote Access Infrastructure for Research Data)⁴. Prosjektet munnet ut i tjenesten microdata.no som altså ble lansert våren 2018.

Konseptet i microdata.no

Visjonen for RAIRD-prosjektet var å kunne tilby forskere registerdatatilgang på dagen og uten noen form for søknad. Visjonen la noen begrensninger, men åpnet samtidig mange muligheter for infrastrukturen som skulle utvikles.

Begrensninger

For å eliminere søknadsprosess og dokumentasjon av prosjekt- og behandlingsgrunnlag var det nødvendig å utvikle et anonymiserende grensesnitt hvor brukerne ikke kunne se eller laste ned data, men likevel kunne koble og analysere data. Nøkkelen var å unngå at sluttbruker behandler personopplysninger i juridisk forstand.

Det er viktig å påpeke at det ikke er *dataene* som anonymiseres i microdata.no; isteden er det analyseresultatene (statistikk, grafikk og annen output skapt av bruker gjennom systemet) som anonymiseres. Data-anonymisering er kostbart og reduserer informasjonsverdien av data, og benyttes ikke. Registerdataene i løsningen ligger inne i full detaljrikdom. Men det anonymiserende grensesnittet fungerer altså slik at sluttbrukerne ikke ser eller får ut personopplysninger av systemet.



Det australske statistikkbyrået (ABS) hadde på dette tidspunktet (rundt 2012) allerede pilotert statistikk- og analyseløsninger med et relativt simpelt anonymiserende grensesnitt-konsept. Tidlig pilotering og testing i RAIRD-prosjektet demonstrerte at et slikt konsept var mulig. Piloteringen avdekket også at brukere som var vant til å kunne se direkte på data ville trenge kompenserende tiltak for å kunne orientere seg godt i dataene gjennom et slikt anonymiserende grensesnitt. Det ble også klart at anonymiserende tiltak ville måtte designes og utvikles løpende for de statistiske metodene infrastrukturen skulle støtte, og at løsningen derfor bare ville kunne dekke en delmengde av alle analysebehov – og bare en relativt liten del ved lansering.

Muligheter

Konseptet åpnet samtidig helt nye muligheter for selvbetjent bruk av data, hvor forskerne interaktivt ville kunne koble variabler fra hele bredden av tilgjengelig datagrunnlag samt løpende definere egne analysepopulasjoner. Det var en kjent sak at forskere ikke sjelden ønsket seg flere variabler eller andre populasjoner etter at data var mottatt/utlevert i tradisjonell utlevering på slutten av en lang og smertefull søknadsprosess. Fleksibiliteten i microdata.no-konseptet og selvbetjeningsløsningene der helt eliminerte denne type problemer.

Muligheter for rik og integrert datadokumentasjon (definisjoner, koding, avvik, etc) i verktøyet åpnet seg også. Dette var viktig av flere grunner, ikke minst for å kompensere for bortfallet av muligheten til å inspisere selve dataene som konseptet innebar. Men også for å gjøre det enklere for brukerne å orientere seg best mulig om definisjoner, endringer, forbehold o.l. i datagrunnlaget. Registerdata er ikke designet for forskning, og det kan være krevende å forstå definisjoner og endringer uten rike og tilgjengelige metadata.

Styrker og svakheter

Konseptet har noen sentrale avveininger. På den ene siden ble det mulig med rask, fleksibel og bred tilgang til registerdata. På den andre siden manglet mulighet til å se data direkte. Funksjonalitet og metode var samtidig begrenset til analyseverktøyene som til enhver tid kunne gjøres tilgjengelige i det anonymiserende grensesnittet.

Fordi hovedmålet med satsingen nettopp var å senke terskelen for forskertilgang til registerdata ble det besluttet å satse på konseptet. Samtidig ble det besluttet å jobbe kontinuerlig og målrettet med implementasjon av mest mulig nødvendig funksjonalitet i samarbeid med brukermiljøene – samt å jobbe for at microdata.no kunne komplementere og støtte best mulig opp rundt tradisjonell søknadsbasert utlevering av registerdata.

Samspill med andre tilgangsformer

Microdata.no dekker mange registerdatabehov, men ikke alle. Det er fortsatt behov for tilbudet om tradisjonell søknadsbasert utlevering av data. Miljøer som driver metodeutvikling, skal koble med data som ikke ligger i microdata.no, som trenger spesialprogramvare eller metoder som ennå ikke er på plass i løsningen er avhengige av å få data utlevert på tradisjonell måte. Det betyr søknad med dokumentasjon av behandlingsgrunnlag, saksbehandling/forhandling om dataminimering, SSB-tilrettelegging av datasett, etc.

Microdata.no har blitt videreutviklet med en eksportfunksjon⁵ slik at løsningen kan strømlinjeforme og forenkle deler av den tradisjonelle søknadsprosessen.

Eksportfunksjonen gjør microdata.no til en avansert variabelutforsker der brukerne kan skape og teste seg fram til datasettene de trenger, og gjennom det får oversikt over det reelle databehovet, slik at det blir lettere å begrunne det i søknaden.

Dette fjerner en kilde til misforståelser og gir mindre arbeid for SSB fordi brukerne selv har laget/tilrettelagt datasettet. SSB henter ut de ferdige filene som brukerne selv har laget. Det er bra for både leveringstid og pris.



Videreutvikling i samarbeid med helseregistre

SSBs registerdatatilfang inneholder hovedsakelig demografiske og sosio-økonomiske data, mens helseregisterdata ligger i egne helseregistre.

De sentrale helseregistrene er også etablert for å ivareta landsomfattende oppgaver. De brukes først og fremst til helseanalyser, statistikk, kvalitetsforbedring av helsetjenester, forskning, administrasjon, styring og beredskap. Helseregistrene er opprettet med hjemmel i helseregisterloven og forskrifter. Register med direkte personidentifiserbare opplysninger som ikke er basert på samtykke, er opprettet etter behandling i Stortinget. Fra 1.1.2024 samles alle de sentrale helseregistrene i FHI.

Søknad til Forskningsrådet infrastruktursatsing i 2018 med Kreftregisteret som partner

NSD og SSB inviterte Kreftregisteret til å være partner i en ny søknad til Forskningsrådets infrastruktursatsing i 2018. Kreftregisterets rolle skulle være å delta i utviklingen av et teknisk og juridisk opplegg som kunne gjøre det mulig for andre enn SSB å dele koblingsbare registerdata via microdata.no.

Prosjektet fikk navnet Microdata 2⁶ og ble tildelt 42.5 MNOK av Forskningsrådet i 2019. Prosjektoppstart var 2020, og prosjektet løper ut 2023.

Prosjektets hovedmål var altså å legge til rette for at flere dataeiere kunne dele data via plattformen. Dette arbeidet har krevd mye teknisk utvikling og en rekke juridiske avklaringer, og målsetningen i prosjektet oppnås i den forstand at infrastrukturen høsten 2023 blir klar for at andre dataeiere kan dele sine data der. Andre delmål handlet om å øke mengden SSB-data, videreutvikle analysefunksjonalitet, pilotere mulighet for å koble forskerinnsamlede surveydata med registerdata.

Helseanalyseplattformen og gjenbrukssporet

Parallelt med utviklingen av microdata.no kjørte Direktoratet for e-helse Helsedataprogrammet. En sentral komponent her skulle være Helseanalyseplattformen (HAP). Tanken med HAP var å bygge en teknisk løsning for mer effektiv deling av helseregisterdata, samt en organisasjon (Helsedataservice/HDS) for å sikre «én vei inn» og sentralisert søknadsbehandling for forskere med behov for helseregisterdata.

Organisasjonen HDS ble etablert i 2021. Men den tekniske løsningen HAP ble skrotet⁷ i 2023. Dette gjorde at Direktoratet for e-helse måtte se seg om etter eksisterende nasjonale løsninger som kunne bidra til å oppfylle flest mulig av målene med HAP. Det såkalte «gjenbrukssporet» ble åpnet, og direktoratet etablerte dialog med etablerte tekniske løsninger (TSD/UiO, SAFE/UiB, HUNT Cloud/NTNU) samt microdata.no (Sikt og SSB). Intensjonen bak gjenbrukssporet å både å gjenbruke delresultater fra Helsedataprogrammet, men ikke minst å utnytte etablerte og modne løsninger for tilgang til og bruk av sensitive data på norsk jord. Gjenbruksaktørene (herunder Sikt/SSB) har pekt på at det er et stort potensial i de tjenestene som eksisterer, men at det krever ressurser å utnytte potensialet og få tjenestene til å spille sammen og gå opp i en høyere enhet sammen med HDS og SSB som hver seg forvalter registergrunnlaget.

Gjenbrukssporet og ulike aktiviteter igangsatt av Direktoratet for e-helse har gitt microdata.no økt synlighet i helseregister- og helseforskningsfeltet. Dette har skapt arenaer for dialog og grobunn for videre samarbeid mellom microdata.no og sentrale helseregistre samt andre nøkkelaktører i sektoren.

Våren 2023 ble det besluttet i revidert nasjonalbudsjett at Direktoratet for e-helse legges ned i sin nåværende form og slås sammen med Helsedirektoratet fra 1.1.2024. Samtidig ble det som nevnt også besluttet at de sentrale helseregistrene Kreftregisteret, Norsk pasientregister/Kommunalt pasient- og brukerregister (NPR/KPR) samt FHIs registre samles under FHI fra 1.1.2024.

Helseforvaltningen og ikke minst helseregisterlandskapet forandres altså i 2024. I praksis betyr dette at FHI med alle sine helseregistre sammen med SSB blir de to største aktørene på registerdata fremover.



Søknad til Forskningsrådet infrastruktursatsing i 2023 med FHI som partner

Søknadsfristen for årets utlysning er 15.11.2023. For å kunne sende søknad stilte Forskningsrådet krav om obligatorisk innsending skisse som grunnlag for søknad om forskningsinfrastrukturmidler innen 21.06.2023. Sikt, SSB og FHI sendte inn en felles skisse med arbeidstittel «Microdata.no Cross-FAIR. FAIR access to microdata across sectors and disciplines». Hovedmålsetningen for søknaden er å etablere mest mulig koblingsbare helseregisterdata på infrastrukturen, samt å fortsette å styrke løsningens funksjonalitet og relevans for både samfunns- og helseforskning.

Veien videre

Inneværende prosjekt Microdata 2 løper ut 2023. Innspurten i prosjektet handler om å slutføre arbeidet med etableringen av microdata.no som plattform for deling av data for Kreftregisteret og andre relevante dataeiere. Planen er å tilgjengeliggjøre data fra Kreftregisteret på infrastrukturen før året er omme. Det gjenstående arbeidet vil fokusere på sikkerhetsarbeid, slutføring av ROS-analyse samt endelig tilrettelegging av data fra Kreftregisterets side.

Utforming og forankring av søknad til Forskningsrådets infrastruktursatsing 15.11.2023 er høyt prioritert. Sikt er vertsorganisasjon for søknaden.

Finansiering etter 2023

Kunnskapsdepartementet (KD) satte høsten 2021 ned et utvalg som skulle vurdere og foreslå løsninger til videreutviklingen av datainfrastruktur i Norge. Dette infrastrukturutvalget overleverte en rapport til KD 10. mai 2022. Finansieringsmodellen for microdata.no fremover utformes i tråd med utvalgets anbefaling om en tredeling:

1. **Langsiktig finansiering** – stabil og forutsigbar finansiering.

Finansieringen kan for eksempel komme i form av grunnfinansiering fra ett relevant departement eller fra flere departementer etter en fordelingsnøkkel basert på hvilke forvaltningsområder infrastrukturen leverer forskning og forvaltning til. Slik finansiering kan også komme i form av spesifikke forvaltningsoppdrag eller gjennom forpliktende medlemskap eller partnerskap fra institusjoner som bruker infrastrukturen.

Langsiktig finansiering skal dekke kontinuerlig videreutvikling i dialog med brukermiljøene for å sikre relevans (adaptivt vedlikehold), herunder nødvendig videreutvikling av funksjonalitet i tråd med teknologiske endringer, og endrete brukerbehov. Utnytte ny teknologi og metoder for kontinuerlig å styrke analysekapabilitet, teknisk kapasitet og informasjonssikkerhet kombinert.

Sikt og SSB har bedt Kunnskapsdepartementet og Finansdepartementet, respektive, om langsiktig finansiering fra 2024. Departementenes beslutning kommer i statsbudsjettet for 2024.

2. **Konkurransesatt finansiering** – finansiering som et resultat av konkurranse mellom flere aktører som vurderes etter gitte kriterier, for eksempel Forskningsrådets infrastrukturutlysning eller EUs rammeprogrammer.

Konkurransesatt finansiering (ofte omtalt som prosjektf finansiering) gjelder typisk større løft/innovasjoner som påvirkes av endringer i teknologi-, data- eller brukerlandskapet. Eksempler kan være nye typer data, fødererte analyser med internasjonale datapartnere, støtte for sanntidsdata, maskinlæring etc.

Her deltar Sikt/SSB/FHI altså med en felles søknad knyttet til microdata.no til årets infrastrukturutlysning fra Forskningsrådet.

1. **Brukerfinansiering eller brukerbetaling** – finansiering som kommer fra de som bruker tjenestene. Dette kan enten være i form av abonnement, stykkpris eller



forbruksfinansiering, eller en kombinasjon av disse.

Brukerbetalingen skal dekke kostnadene bruken påfører infrastruktureierne. Noen kostnader påføres av enkeltbrukere, andre av samlet bruk. Det er hensiktsmessig å dele bildet i 3:

Korrektivt vedlikehold

Feilretting, håndtering av feilsituasjoner skapt gjennom bruk, gjenoppretting.

Ytelsesrelatert vedlikehold

For å håndtere økt bruk, økte datamengder, tyngre analysemetoder, effektiv og rettferdig fordeling av beregningsressurser, oppskalering av hardwareressurser etc.

Rene brukerrelaterte driftskostnader (support, opplæring og rapportering)

Kostnader knyttet til direkte brukersupport, dokumentasjon, on-boarding, opplæring, tilbakerapportering av bruken av data til dataeier og til brukerorganisasjoner.

Sikt og SSB jobber med å ferdigstille en pris- og betalingsmodell for bruk av tjenesten.

Prismodellen for brukerbetaling av microdata.no skal dekke brukerrelaterte kostnader, og er utformet med utgangspunkt i følgende kriterier:

Prismodellen skal:

- Være enkel å forstå og forklare
- Være transparent og mest mulig forutsigbar
- Ikke påføre institusjonene unødig høye kostnader
- Ikke påføre institusjonene unødig merarbeid og administrasjon
- Dekke brukerrelaterte kostnader for infrastruktureierne
- Bidra til god kapasitetsplanlegging for infrastruktureierne
- Gi infrastruktureierne incitament til å øke bruk og nytteverdi av infrastrukturen

Pris- og betalingsmodell sirkuleres til de 105 eksisterende brukerinstitusjonene så snart den er klar. Betaling innføres fra 2024.

Saksdokument

Porteføljestyre for forskning og kunnskapsressurser

Til Porteføljestyret for forskning og kunnskapsressurser
Dato 28.08.2023
Saksnummer 16/23
Saksarkivnr. 23/00000
Sakstype

Saksansvarlig Vigdis Namtvedt Kvalheim
Saksbehandler Bodil Agasøster

MODANSA- en modernisering av Sikts infrastruktur for forskningsdata

Forslag til vedtak

Porteføljestyret tok saken til orientering.

Saksfremstilling

MODANSA står for Modernisering av Data om det Norske samfunn. Datasamlingene er et resultat av en stor og bevisst satsing fra Norges forskningsråd (NFR) gjennom mange tiår og et forpliktende samarbeid mellom forskningsmiljøene representert ved NSD og sentrale dataprodusenter, særlig Statistisk sentralbyrå. Det kanskje viktigste resultatet av denne satsingen er Kommunedatabasen, som var NSDs mest brukte datatilbud gjennom tidene. Kommunedatabasen er også et godt eksempel på hovedintensjonen med datatilbudene fra Sikt:

1. Mest mulig data samlet under samme inngang, definert som data over tid
2. Godt dokumenterte data
3. Åpent tilgjengelig for forskning og undervisning
4. Med kort vei til analyse.

I 1993 passerte Kommunedatabasen sine første 100.000 variabler, allerede i 2001 ble 200.000 variabler passert. Da NSD ble en del av Sikt ved inngangen til 2022, hadde databasen nesten 400.000 variabler for hver enkelt kommune. Enkelte strekker tilbake til Norges første folketelling i 1769.

Kommunedatabasen og de øvrige datasamlingene om det om det norske samfunn har en kompleks struktur og består av både kvantitative og kvalitative data. De er driftet i eldre



løsninger som i liten grad oppfyller krav til universell utforming og FAIR-prinsippene¹. Behovet for å modernisere datatilbudet og formidlingsplattformen for disse datasamlingene

er stort. Satsingen som beskrives i dette notatet vil gi brukervennlig tilgang til datatyper som ikke faller inn under surveydata og microdatasatsingen til Sikt, som har vært prioritert de siste 5-10 årene. Tekstdata og data organisert i databaser (registerdata, forvaltningsdata, andre typer stordata, etc.) blir stadig viktigere for enkel tilgang til å studere det norske samfunnet langsiktig og komparativt. Kunnskapsutviklingen er avhengig av at norsk forskning har god tilgang også til denne typen data.

Målet er å øke bruken av denne typen data i norske forskningsprosjekter. Virkemiddelet er å utvikle en helhetlig plattform som tilbyr nye koplingsmuligheter og avanserte spørringer ved hjelp av kunstig intelligens. Det vil gi stor verdiøkning av eksisterende data og legge til rette for nye dataområder og datatyper. Satsningen vil også gi bedre tilgang og kortere vei til analyse for forskere, studenter og andre brukere.

Bakgrunn

Data om det norske samfunn omfatter blant annet Kommunedatabasen, Forvaltningsdatabasen, Kirkedatabasen, og Data om det politiske system. Disse datasamlingene har en unik nasjonalhistorisk verdi. De strekker seg fra så tidlig som 1796 frem til i dag, noe som gir mulighet til å analysere samfunnsutviklingen på politisk, geografisk og individuelt nivå. Data er levende i den forstand at de brukes til forskning, undervisning og av studenter, samt at de oppdateres kontinuerlig.

Denne type data er særlig egnet til sammenstillinger med individdata fra intervjuundersøkelser og registre. I et nasjonalt perspektiv er disse datasamlingene helt unike og de eneste i sitt slag. Det innebærer at et tap av datasamlingene ville fått store konsekvenser på et nasjonalt nivå, både for forskningsmiljøene, og offentlig forvaltning. Særlig gjelder dette data som finnes i Kommunedatabasen² og Forvaltningsdatabasen³. Datasamlingene er mye brukt og har siden 2019 resultert i 184 publikasjoner*:

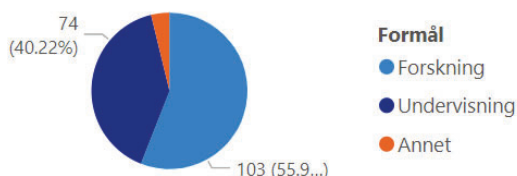
¹ FAIR står for Findable, Accessible, Interoperable og Re-usable og handler i all hovedsak om å håndtere og beskrive data på en måte som setter andre i stand til å forstå og bruke dem i fremtiden. På norsk kan man si at dataene skal være *gjenfinnbare, tilgjengelige, samhandlende og gjenbrukbare* ([Open Science | FAIR-prinsippene](#))

² Kommunedatabasen består av omfattende registeropplysninger om norske kommuner fra 1796 til 2018. Totalt inneholder Kommunedatabasen nærmere 400 000 variabler om hver kommune, om tema som demografi, økonomi, arbeidsliv, offentlige finanser, politikk, helse, utdanning, kultur, samferdsel og naturressurser

³ Forvaltningsdatabasen er en kartlegging av departement og ytre etater i Norge etter 1947. Den gir oversikt over den hierarkiske organiseringen og hvordan denne har endret seg over tid



Fordeling mellom data brukt til undervisnings og forskningsformål



Antall publikasjoner	Type
2	Bachelorgrad
6	Bok
12	Bokkapittel
9	Doktorgrad
11	Forskningsrapport
72	Mastergrad
7	Offentlig utredning
65	Tidsskriftartikkel
184	

*T.o.m mars 2023. Tallene er basert på manuell telling av publikasjoner. Bacheloroppgaver blir vanligvis ikke publisert og det reelle tallet på bruk er sannsynlig mye høyere.

I 2022 ble «forprosjektet» etablert for å lage et konseptuelt grunnlag for hvordan «Data om det norske samfunn» kan moderniseres. Dette var på bakgrunn av vedtak i styringsrådet for etablering av Sikt ([Referat styringsråd 2021-11-30, sak 6](#)): «Forprosjektet bør identifisere etterspørsel og brukerbehov, forprosjektet skal også lage en plan for finansiering av et gjennomføringsprosjekt». Vedtaket var en oppfølging og delvis stadfesting av vedtak i NSDs styre høsten 2021 om å prioritere MODANSA som en satsing finansiert gjennom NSDs egenkapital.

Forprosjektet kartla og vurderte datasamlingene, gjennomførte intervjuer og en spørreundersøkelse med brukermiljøene, og arrangerte en intern workshop. Prosjektet leverte «[Rapport om modernisering av data om det norske samfunn](#)⁴» som inneholder en rekke anbefalinger for hvordan datatilbudet og drift av «Data om det norske samfunn» bør moderniseres. Rapporten la grunnlaget for «Satsingsforslag for MODANSA» som ble vedtatt i ledergruppen til Sikt den 27. juni 2023.

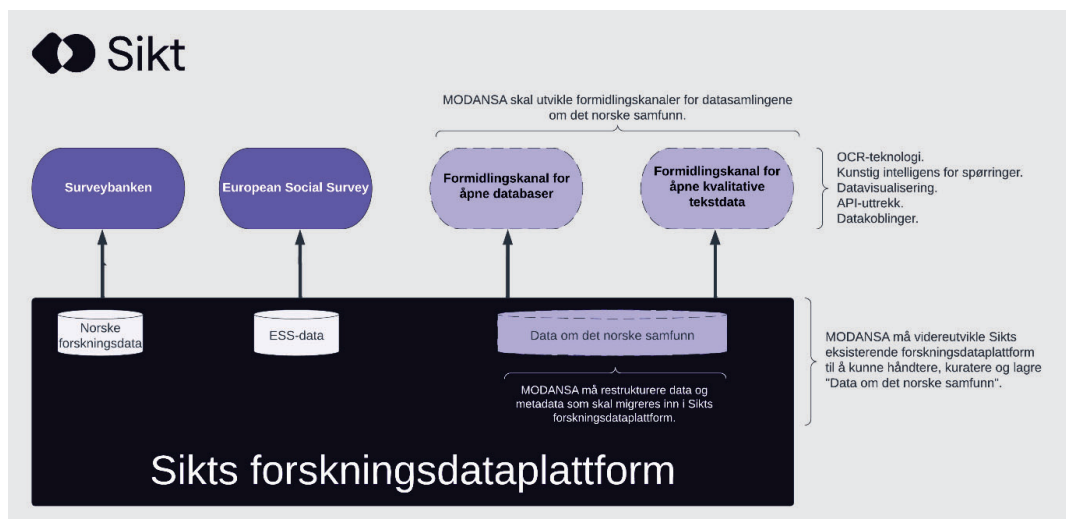
Satsingsforslag

MODANSA skal **utvikle Sikts forskningsdataplattform**, som i dag er innrettet mot kvantitative surveydata til å håndtere, kuratere og lagre ikke-surveybaserte data, dvs. kvalitative tekstdata og registerdata, offentlige data, stordata, etc.

Målet er å styrke Sikt sin datainfrastruktur generelt, slik at flere verdifulle typer data vil løftes opp. Det betyr at investeringen i MODANSA vil ha nytte utover datasamlingene under «Data om det norske samfunn». Sikt vil være i stand til å ta imot, gi merverdi til og tilgjengeliggjøre flere ulike sammensetninger av forskningsdata.

Satsningen vil utvide Sikt sin forskningsdataplattform med to nye datalagrings- og formidlingskanaler. Den ene vil formidle åpne kvantitative data organisert i databaser, mens den andre vil formidle åpne kvalitative tekstdata.

⁴ Se vedlegg



Formidlingskanalen for databaser vil åpne opp for koblinger på tvers av datasamlinger. For hver ny datasamling som legges inn, vil det oppstå et mangfold av nye relasjoner, som lar forskere konstruere datasett på tvers av datasamlingene. Dette er koblinger som tidligere ikke har vært tilgjengelig i Sikts forskningsdataplattform. Formidlingskanalen for tekstdata vil benytte seg av en språkmodell som lar brukere stille avanserte spørringer på tvers av store mengder tekstdata, som for eksempel: "Gi meg alle partiprogram fra Høyre, Venstre og Arbeiderpartiet i perioden 1980 til 2000 som omhandler "klima".

Gevinster

MODANSA åpner opp for forskning og analyse på samfunnsmessig viktige data om det norske samfunn av høy kvalitet, et datagrunnlag som vil få økt relevans som kunnskapsgrunnlag for samfunnet i tiden som kommer. Moderniseringen av Sikts datainfrastruktur vil gi følgende gevinster:

- Tilgjengelighet og brukervennlige løsninger
- Universell utforming av tjenesten
- Sikrer FAIR-forskningsdata

Leveranser

Basert på anbefalingene har forprosjektet foreslått en liste over leveranser som kan inngå i satsingen. Leveransene definerer satsingens krav til kvalitet og omfang. Dette må sees i kontekst av tids- og kostnadsramme og justeres etter det som blir vedtatt som MODANSAs rammeverk.

Forslag til leveranser:

- Formidlingskanal for åpne databaser bygd på eksisterende forskningsdataplattform.
- Formidlingskanal for åpne kvalitative tekstdata bygd på eksisterende forskningsdataplattform.
- Omstrukturering og migrering av data fra "Data om det norske samfunn" til Sikts forskningsdataplattform.
- Videreutvikling av Sikts forskningsdataplattform til å håndtere kuratering og langsiktig lagring av nye type data.
- Standardisering av maskinlesbar metadata som beskriver "Data om det norske samfunn" og migrering til Sikts forskningsdataplattform.



- Verktøy for innhenting av data via API.
- Etablering av ny arbeidsflyt som er tilrettelagt for eksisterende løsninger.

MODANSA må levere løsninger som tilrettelegger for at "Data om det norske samfunn" og nye datasamlinger, får samme gjenbruksverdi og arbeidsflyt som Sikts øvrige datatilbud.

Finansiering

Under er en foreslått investeringsplan basert på det estimerte budsjettet i «Sikts satsningsområder innen forskning 2023-2026». Kostnadssommene er oppført i 1000 kr. Total budsjettert ramme for prosjektet er MNOK 35. Drift og utvikling skal etter prosjektslutt finansieres via bevilgning, men det vil være aktuelt å ta betalt fra bl.a. private dataprodusenter for tilrettelegging, dokumentasjon, kvalitetssikring og tilgjengeliggjøring av data. Det skal ikke tas betalt for *bruk* av data. På sluttbrukersiden skal datavære åpent tilgjengelig for alle.

Leveranser	2023*	2024	2025	2026	2027	Totalt
Planlegging og kompetansebygging	500					500
Videreutvikling, migrering og standardisering		5250	4750	2 500	500	13000
Utvikle verktøy for innhenting av data via API		1 000	1 000			2 000
Utvikle formidlingskanal for åpne databaser			2 500	4 000	2 250	8 750
Utvikle formidlingskanal for åpne kvalitative tekstdata			2 500	4 000	2 250	8 750
Administrasjon, organisering og drift		250	250	500	1000	1000
Totalt	500	6 500	11 000	11 000	6 000	35 000

*Kvartal 4.

Veien videre og medfinansieringspartnere

Det legges opp til at Sikt dekker deler av prosjektkostnadene over sitt budsjett/risikokapital, mens resten av prosjektmidlene søkes dekket gjennom medfinansiering. Hvordan og hvem som skal dekke den resterende finansieringen er ikke avklart. Det er sendt et satsningsforslag til KD om finansiering fra 2024, men dette vil ikke bli avklart før i oktober 2023. Dersom satsningsforslaget til KD ikke går igjennom, må alternative finansieringsmuligheter utredes. Sikt vil gå i dialog med relevante aktører/brukere høsten 2023 om mulig medfinansiering.

Vedlegg

Rapport om modernisering av "Data om det norske samfunn"

Rapport om modernisering av "Data om det norske samfunn"

Sammendrag	2
1.0 Forprosjektet	4
1.1 Bakgrunn	4
1.2 Prosjektmandat.....	5
1.3 Internt samarbeid ble et sentralt suksesskriteria.....	5
2.0 Utredning av datatilbudene.....	6
2.1 Bruk av datatilbudene.....	7
2.2 Kommunedatabasen.....	7
2.3 Forvaltningsdatabasen.....	8
2.4 PolSys	9
2.4.1 Andre arkiv.....	10
2.5 Kirkedatabasen	10
2.6 Data fra Innovasjon Norge.....	11
3.0 Kartlegging av brukerbehov.....	11
3.1 Kartlegging av forskningsmiljøene og brukere	12
3.2 Intervjuguide og spørreundersøkelse	12
3.3 Omfang og gjennomføring av brukerintervjuer og spørreundersøkelse	14
3.4 Resultater	14
4.0 Anbefalinger for modernisering av "Data om det norske samfunn"	16
4.1 Modernisering av datatilbudet.....	16
4.1.1 Innholdsmessig oppdateringer	16
4.1.2 Nye formidlingskanaler for databaser og kvalitative tekstdata	17
4.1.2.1 Formidling av data fra databaser.....	18
4.1.2.2 Formidling av åpne kvalitative tekstdata	19
4.1.2.3 Universell utforming	19
4.2 Modernisering av drift og vedlikehold	19
4.2.1 Organisere i team	20
4.2.2 Tett kontakt med forskermiljøene.....	20
4.2.3 Verktøy for automatisering av datainnhenting	21
5.0 Veien videre for MODANSA-prosjektet	22

5.1 Finansiering av gjennomføringsprosjektet	22
5.2 Hovedleveranser	23
5.3 Rammebetingelser	24
5.4 Organisering og forankring	24
5.5 Forventet gevinster.....	25
5.6 Risikoer.....	27
5.7 Interessenter.....	28

Sammendrag

Data og infrastruktur har et stort potensial for utvikling av ny kunnskap og innovative tjenester som skaper betydelig verdi for samfunnet og sektoren.

En sentral del av Sikt sitt oppdrag er å arkivere, tilrettelegge og formidle data til forskning og analyse. Dette samfunnsoppdraget er blant annet en overføring fra NSD. Tidligere NSD har i tråd med oppdraget akkumulert data om norske kommuner, statsforvaltningen og politiske systemer langt bakover i tid. Disse samlingene har i dag unik nasjonalhistorisk verdi, og de er fortsatt levende i den forstand at de brukes og oppdateres kontinuerlig. Dataene knytter sammen historie, nåtid og framtid. Disse lange tidsseriene med data har stor interesse for forskning, blant annet ved sammenstillinger med individdata fra intervjuundersøkelser og registre. Noe som gir mulighet til å analysere samfunnsutviklingen på politisk, geografisk og individuelt nivå. Disse datasamlingene inngår i det som kalles "Data om det norske samfunn".

Datasamlinger om det norske samfunn har det til felles at lagring og formidling (deling) er bygd på løsninger som er utdatert og ikke kan tilfredsstille nye krav som for eksempel FAIR-prinsippene.

I Forprosjektet har tre oppgaver blitt gjennomført for å bedre datatilbudet relatert til "Data om det norske samfunn". Den første oppgaven var å ivareta dagens datatilbud, mens de to andre gikk ut på å lage et konseptuelt grunnlag for hovedprosjektet MODANSA som skal modernisere "Data om det norske samfunn" fra 2023 til 2026.

Under er et kort sammendrag av hva forprosjektet har levert på de tre oppgavene:

1) Utredet hvordan datatilbudene kan opprettholdes etter nedstenging av on-prem servere, og implementere dette.

Forprosjektet gjorde en kartlegging av alle databeholdningene under "Data om det norske samfunn". Etter kartleggingen ble det besluttet hvordan datatilbudene skulle driftes etter nedstengingen av on-prem servere. Forprosjektet var ansvarlig for å implementere beslutningene.

Dette førte til at noen databeholdninger – som i stor grad var konstruert som surveydata - ble overført til eksisterende forskningsdataplattform og formidlet gjennom Sikts portal for forskningsdata.

Andre databeholdninger, slik som Kommunedatabasen, Forvaltningsdatabasen og "PolSys" ble flyttet til AWS. Det var planlagt at disse datatilbudene skulle stenges i en periode, men i samarbeid med Sikts DI-divisjon fikk forprosjektet beholdt drift av datatilbudet. Dette var viktig for å unngå

nedetid på tjenester som har en stor brukergruppe. Kommunedatabasen fikk et forbedret brukergrensesnitt og ble åpnet opp til alle brukere etter samtale med SSB.

Enkelte datatilbud under "Data om det norske samfunn" ble vurdert som verdiløse og ble avsluttet.

[Kp. 2 gir en detaljert beskrivelse av databeholdningene og beslutningene som ble implementert.](#)

2) Hentet innspill om innholdsmessig og funksjonell oppgradering av datasamlingene fra forskningsmiljøer og andre.

Forprosjektet gjennomførte seks brukerintervju og en brukerundersøkelse for Kommunedatabasen.

I forkant av dette ble brukermiljøene kartlagte og det ble rekruttert informanter som representerte ulike type brukere.

Resultatene fra intervjuene og undersøkelsene ga innblikk i brukerbehovene som må ivaretas i moderniseringen av "Data om det norske samfunn".

[Kp. 3 gir en beskrivelse av kartleggingen av brukermiljøene, hvordan data ble samlet inn og en presentasjon av funnene om brukerbehov.](#)

3) Utarbeidet anbefalinger for modernisering av "Data om det norske samfunn"

Basert på kartleggingen av databeholdningene, brukerbehovene, og et internt arbeidsseminar har forprosjektet utarbeidet en rekke anbefalinger for modernisering av "Data om det norske samfunn".

Anbefalingene som foreslås vil endre hvordan datatilbudet til de ulike datasamlingene formidles til brukerne. Det vil også endre hvordan produktteamet arbeider med drift og vedlikehold. Det har blitt identifisert behov for både teknisk og organisatorisk endring.

Forprosjektet anbefaler at det bygges videre på eksisterende forskningsdataplattform og at det utvikles to nye kanaler for formidling av databaser og kvalitativ tekstdata. Dette er data som Sikt ikke leverer et godt tilbud på i dag med hensyn til universell utforming og FAIR-prinsippene.

Dette skyldes at tidligere modernisering av Sikts forskningsdataplattform har vært fokusert rundt surveydata. Nå ser vi en trend hvor surveydata er truet av lav deltagelse som går utover datakvaliteten. Det er å forvente at andre type data kommer til å bli mer sentral i forskning fremover. Det er nødvendig at Sikt videreutvikler forskningsdataplattformen slik at kunnskapssektoren har et godt datatilbud for data ikke generert fra surveyforskning.

Videreutviklingen av forskningsdataplattformen innebærer en sammenslåing av flere datatilbud og datasamlinger. Formidlingskanalene skal ta utgangspunkt i Forvaltningsdatabasen, Kirkedatabasen, Kommunedatabasen og "PolSys", men løsningene skal være tilrettelagt for å ta imot flere nye datakilder. Dette gir Sikt en skalerbar infrastruktur.

Et hovedfokus i formidlingskanalen for databaser vil være å åpne opp for koblinger mellom datasamlinger. Det betyr at brukerne enkelt kan koble og konstruere datasett på tvers av samlinger av databaser i den nye kanalen. For å realisere dette må metadata standardiseres på tvers av datasamlingene.

For hver ny datasamling som legges inn i den nye formidlingskanalen vil det oppstå et mangfold av nye relasjoner. Kanalen vil åpne opp for koblinger som tidligere ikke har vært tilgjengelig i Sikts forskningsdataplattform.

Formidlingskanalen for kvalitative tekstdata vil benytte seg av optisk tegngjenkjenning (OCR1) teknologi til å la brukerne søke både i og på tvers av dokumenter og innhold. Den vil tilgjengeliggjøre tekstdata på en helt annen måte enn i dag. Dette åpner opp døren for maskinlæring som kan tiltrekke nye brukergrupper. Søket vil kunne generere statistikk fra dokumentene som kan brukes til kvantitativ innholdsanalyse. Det må etableres en ny arbeidsflyt som ligger tett på hvordan produktteamet "Lagring, kuratering og tilgjengeliggjøring" arbeider i Sikts eksisterende forskningsdataplattform.

Drift og vedlikehold av "Data om det norske samfunn" må bli organisert i team, ikke av enkeltpersoner slik som i dag. Dette er viktig for å tilrettelegge for den planlagte sammenslåing av produktteamene under produktområdet Dataforvaltning.

For å effektivisere drift anbefales det at innhenting av data i større grad benytter seg av APIer. Det bør lages et verktøy som åpner opp for å automatisere prosesser som i dag tar mye tid.

Det anbefales en økt satsing på å komme tettere på forskningsmiljøene. Dette er nødvendig for å treffe på hvilke nye datasamlinger som skal gå inn i de nye formidlingskanalene. Det er gjenbruk av data som har verdi, ikke arkivering i seg selv.

Kp. 4 redegjør for anbefalingene til modernisering i dypere detalj.

Styringsgruppen ønsket at forprosjektet skulle komme med et forslag til veien videre for hovedprosjektet som skal gjennomføre moderniseringsarbeidet basert på forprosjektets konseptgrunnlag. Dette hovedprosjektet har fått navnet MODANSA (**MOD**ernisering av **DA**ta om det **Norske SAM**funn).

Prosjektgruppen vil anbefale at MODANSA får midler fra enten bevilgningsfinansiering eller egenfinansiering. Brukerfinansiering er det ikke tradisjon for og det kan hemme Sikt med å treffe på virksomhetsmålene.

I kp. 5 legges det frem drøfting om finansiering av prosjektet, samt innspill til leveranser, organisering, rammeverk og forankring. Forprosjektet har også innspill til forventete gevinster, og identifisering av risikoer og interessenter.

MODANSA-prosjektet vil skape en verdikjening for data som bare vil få en økt relevans i forskning fremover. Det vil bidra til å treffe på Sikts virksomhetsmål om å bedre datainfrastruktur og få nye kunder. Prosjektet vil bidra til mer åpen tilgang til data, og det oppfyller Sikts plan om å utvikle på eksisterende tjenester og infrastruktur – slik at sektorens behov for analyse, forskning og innsikt dekkes på en bedre måte. MODANSA vil bidra til at produktteam kan arbeide smidig, som også er et av Sikts virksomhetsmål.

1.0 Forprosjektet

1.1 Bakgrunn

Databeholdningene som omfatter det vi kaller "Data om det norske samfunn" var ikke en del av moderniseringen som foregikk i tidligere infrastrukturprosjekter i tidligere NSD. Dette har ført til at dataene ikke er harmonisert med andre prosesser og systemer.

¹ [OCR – Store norske leksikon \(snl.no\)](https://snl.no/OCR)

30. november 2021 ble det vedtatt i Styringsrådet at konseptutredning om modernisering av "Data om det norske samfunn" var strategisk viktig for Sikt. Det ble vedtatt at et forprosjekt skulle gjennomføres:

"Forprosjektet bør identifisere etterspørsel og brukerbehov, forprosjektet skal også lage en plan for finansiering av et gjennomføringsprosjekt." - Referat styringsråd 2021-11-30.

Forprosjektet startet opp i 2022 for å legge grunnlaget for et gjennomføringsprosjekt som skal modernisere "Data om det norske samfunn". Dette prosjektet omtales nå som MODANSA-prosjektet (Modernisering av Data om det norske samfunn).

1.2 Prosjektmandat

Formålet med forprosjektet var å produsere et praktisk og konseptuelt grunnlag for å utvikle

1. Mer brukervennlige datatjenester, både med hensyn til funksjonalitet, datatilbud og kvalitet (FAIR).
2. Mer effektiv drift og vedlikehold gjennom tettest mulig integrering i Sikts forskningsdataplattform.

For å oppfylle formålet ble det satt tre hovedmål:

- 1) Utrede hvordan datatilbudene kan opprettholdes etter nedstenging av on-prem servere, og implementere dette.
- 2) Hente innspill om innholdsmessig og funksjonell oppgradering av datasamlingene fra forskningsmiljøer og andre.
- 3) Utarbeide rapport med anbefalinger av modernisering.

Arbeidet i forprosjektet – som presenteres i denne rapporten - danner grunnlaget for det som blir utgangspunktet for MODANSA-prosjektet.

Prosjektmandatet kan leses i sin helhet her: <https://isikt.sharepoint.com/:w:/s/SeksjonforforskningsdataNSD/EeeT-Ry3hI9JiReHXx80cm0BBFs8yczuQnvdFysWHSFOsQ?e=EpcWsj>.

1.3 Internt samarbeid ble et sentralt suksesskriteria

For å lykkes med å levere på forprosjektets tre hovedmål ble samarbeid med Divisjon for Datainfrastruktur (DI), Servicesenteret, Seksjon for Kommunikasjon og Seksjon for Forskningsløsninger avgjørende.

Parallelt med forprosjektet ble det gjennomført et "Skyprosjekt", som hadde i oppgave å stenge ned on-prem servere lokalisert på Marineholmen (som inneholdt både data og datainfrastruktur for "Data om det norske samfunn") og migrere data til andre servere.

Ettersom forprosjektet skulle beslutte hva som skulle skje med data etter nedstengingen av on-prem servere, var det en konstant dialog mellom prosjektene. Dette hjalp oss med å raskt identifisere hvilke data som kunne berges i Sikts forskningsdataplattform, og hvilke data som stod i fare for å ha nedetid.

Sammen med "Skyprosjektet" ble det satt i gang en dialog med DI. Det viste seg at DI kunne overta den tekniske driften til Kommunedatabasen, Forvaltningsdatabasen og "PolSys" ved å opprette virtuelle maskiner på deres områder.

Før Forprosjektet var opprettet var det tenkt at disse databasene skulle enten ha fullstendig eller delvis nedetid – som minimum ville vært i 18 måneder, men trolig en del lenger. Dette anså Forprosjektet som en uakseptabel løsning da dette ville 1) ha en negativ innvirkning for et stort antall brukere, 2) føre til brudd i avtaler med kunder og samarbeidspartnere og 3) føre til at datasamlinger ikke ble oppdatert i en lengre periode. Uroen av dette ville vært et forstyrrende element i det fremtidige moderniseringen av "Data om det norske samfunn". Heldigvis førte godt samarbeid mellom DI, Skyprosjektet og Forprosjektet til at det meste av datatilbudet kunne fortsette uendret, samt med enkelte forbedringer.

Forprosjekt initierte en dialog med Servicesenteret som førte til at de tok på seg oppgavene tilknyttet overvåkning av tjenestene som Skyprosjektet migrerte til DIs områder. I dialogen med Servicesenteret kom vi frem til at de overtok førstelinjeansvaret med brukerne og saksbehandling/brukerstøtte knyttet til tjenestene gikk fra epost til RT-systemet.

Beslutningene relatert til hvilke datatilbud som ble flyttet til Colectica, DI eller stengt ned ble i tett samarbeid med Kommunikasjon formidlet til brukerne via ulike kanaler.

Sikts Kommunikasjonsseksjon var også en viktig samarbeidspartner i arbeidet med kartleggingen av brukerbehov. De hjalp oss med rekruttering av deltakere og gjennomføring av seks brukerintervju. I tillegg bistod de med oppretting og formidling av en spørreundersøkelse rettet mot Kommunedatabasens brukere.

Forprosjektet ville ikke kunne gi anbefalinger til modernisering av "Data om det norske samfunn" uten samarbeid med PO Forskningsløsninger, samt rådgivning fra Ørnulf Risnes fra PO Microdata. 30. august ble det gjennomført et arbeidsseminar med forprosjektets prosjektmedlemmer og utviklere identifisert som sentrale for MODANSA. I seminaret ble brukerbehov, datastrukturering i andre prosjekter, og problemstillinger og løsninger i fremtidig modernisering diskutert.

I Forprosjektets bevilgning lå det en forventning fra styringsrådet om samarbeid med DI og Kommunikasjon. Det viste seg å være et viktig innspill. I tillegg utnyttet forprosjektet andre samarbeidsflater til å levere på hovedmålene.

2.0 Utredning av datatilbudene

I første fase av forprosjektet var det nødvendig å evaluere alle databeholdningene under "Data om det norske samfunn". Prosjektet hadde i oppgave å beslutte hvilke datatilbud som skulle avsluttes eller videreføres ved stengingen av on-prem servere. For data som skulle videreføres måtte prosjektet beslutte i hvilken løsning det ville inngå, og hvorvidt det var en permanent eller midlertidig løsning.

Alle vurderinger er dokumentert i gitlab: <https://gitlab.sikt.no/nsd-no/dapla/data-migration-from-marineholmen-to-DAPLA/-/issues>.

Selv om det var Skyprosjektet som var ansvarlig for å koordinere migrering av data, var det Forprosjektet som besluttet og implementerte løsningene etter stenging av on-prem servere.

I dette kapitlet redegjøres det for utredningen av datatilbudene til "Data om det norske samfunn". Dette innebærer en grundig beskrivelse av hvilke data som inngår i datasamlingen, Sikts forpliktelser/avtaler tilknyttet dataene og hvordan datatilbudene i dag driftes etter stengingen av on-prem servere. Utredningen danner et viktig kunnskapsgrunnlag for videre modernisering i MODANSA-prosjektet.

2.1 Bruk av datatilbudene

Å redegjøre for bruk av alle datatilbudene under "Data om det norske samfunn" er komplekst, siden de brukes ulikt. Det gjør at sammenligning med andre datatilbud blir vanskelig.

Det er vrient å telle bruk av data som ikke lastes ned, men som presenteres visuelt. Fordi man vet ikke om brukeren benyttet seg av data som ble presentert. I tillegg er det en utfordring å definere hva som er bruk av data; skal man telle hvor mange som blir eksponert for data, hvor mange som laster ned data, eller hvor mange publikasjoner som refererer til data? Skal man telle brukere som enkelt personer eller forskningsprosjekt? Hvordan teller man et forskningsprosjekt som får tilgang til data som brukes til flere ulike formål? Skal man telle all bruk eller kun bruk fra den tenkte målgruppen?

Til tross for hindrene som gjør det vanskelig å måle bruk av data er det likevel nyttig å ramme inn bruken av "Data om det norske samfunn". Det hjelper oss med å se verdien til datasamlingene.

Datatilbudene Forvaltningsdatabasen og "PolSys" på i samme database før migreringen fra on-prem servere. I 2021 akkumulerte de totalt 349 075 sidevisninger. Til sammenligning hadde search.nsd.no – som omfatter alle søk etter surveydata fra NSDs nettside – 15 237 sidevisninger i 2020 (google analytics hadde ikke komplette tall fra 2021).

Fra 2016 til 2021 har antall brukere i Kommunedatabasen variert fra 98 til 136 per år. En registrert bruker har tilgang i ett til databasen i ett år. En brukerundersøkelse tyder på at personer som har fått tilgang til databasen bruker data til flere formål. Sammenlignet med andre tidsserier i Sikt er dette høye brukertall. Under "Norske forskningsdata" har Sikt publisert over 2500 datasett som er tilgjengelig via Sikts datasøk (search.nsd.no). Her finner man populære tidsserier som Arbeidskraftsundersøkelsen, Ungdata, livsløpstudiet NorLAG, Norsk Medborgerpanel, Levekårsundersøkelsen, og mye mer. Dette inkluderer også datasett som blir arkivert av forskningsprosjekter via Sikts deponeringstjeneste. I perioden 2016 til 2020 varierte antall databestillinger av "Norske forskningsdata" fra 300 til 342 databestillinger (merk at en databestilling kan ha flere datamottakere). Det at Kommunedatabasen alene har 1/3 av bruken viser at det er en populær datasamling.

For de andre databeholdningene under "Data om det norske samfunn" finnes det ikke god statistikk på bruk. Enkelte data blir trolig lite brukt. For "Kirkedatabasen" finnes det ikke statistikk på bruk, men henvendelser tyder på at den benyttes.

2.2 Kommunedatabasen

Databeholdning	Data	Beslutning
Kommunedatabasen	Kommunedatabasen	Midlertidig DI-løsning.

I Kommunedatabasen finnes statistikk om sentrale tema for norske kommuner, som demografi, økonomi, arbeidsliv, politikk, helse, kultur, samferdsel og naturressurser. Unikt for basen er det at data kan hentes ut for gjeldende kommunegrenser for et hvilket som helst år. Databasen inneholder følgende data:

- **Naturressurser og miljø (1865-2017).** Omfatter data om areal, avfall og avløp
- **Befolkning (1769-2019).** Omfatter alle folketellinger og demografiske data

som fødte, døde, flytting, ekteskap, skilsmisser, fødested.

- **Arbeidsliv (1835-2017).** Yrkes- og næringsstatistikk, arbeidstakerstatistikk, ansatte i stat og kommune, arbeidsledige.
- **Næringsvirksomhet (1875-2012).** Statistikk om industri, varehandel og byggeareal, og bedrifts-, jordbruks- og håndverkstallinger.
- **Valg og folkeavstemninger (1829-2017).** Valgmennsvalg, valg til Storting, fylkesting og kommunestyre. Stemmerettstillinger og folkeavstemninger.
- **Offentlige finanser (1945-2016).** Statistikk over skatt, inntekt, formue og arbeidsgiveravgift. Inneholder en omfattende oversikt over kommuneregnskap.
- **Utdanning (1950-2017).** Statistikk som gjelder barnehager, grunnskoler,

videregående skoler, høyere utdanning og voksenopplæring. Inneholder også data om utdanningsnivå.

- **Helse og sosiale forhold (1867-2017).** Helsedata, personell, tiltak og institusjoner. Inneholder også Kommunal monitor og noe kriminalstatistikk.
- **Bygg og boliger (1865-2017).** Boligstatistikk og byggearealstatistikk.
- **Samferdsel (1977-2008).** Veitrafikkulykker og motorkjøretøy.
- **Fritid, kultur, språk, religion (1845-2016).** LO-medlemmer, kino, målvedtak, kirkestatistikk, trossamfunn og etniske grupper.
- **Kommunetypologier og annet (1918-2018).** Kommunetypologier, regionale inndelinger, tipping, TV-aksjoner, personminutter, åpningstider, kommuneundersøkelser osv.

Kommunedatabasen har også en omregningsmekanisme som gjør det mulig å sammenligne data over tid til tross for kommuneendringer. Databasen har også kartkoordinator som i seg selv kan ha verdi i forskning.

Tidligere var det kun godkjente norske forskningsinstitusjoner som kunne søke tilgang til databasen. Dette ble endret etter et møte med SSB 21.03.2022, hvor det ble avklart at databasen skulle åpnes for alle, i tråd med endringene i statistikkloven.

I april 2022 ble Kommunedatabasen migrert over til og driftet av Divisjon for Datainfrastruktur. Dette er en midlertidig driftsløsning for å ivareta datatilbudet under moderniseringsprosjektet.

Tidligere hadde det blitt utviklet en stor forbedring av brukergrensesnittet som ikke hadde blitt implementert. Forprosjektet valgte å implementere dette, slik at den midlertidige driftsløsningen ga Kommunedatabasen et betydelig løft. Kostnaden av å ta i bruk et forbedret brukergrensesnitt var lav.

2.3 Forvaltningsdatabasen

Databeholdning	Data	Beslutning
Forvaltningsdata	Utvalgsarkivet Data om statsansatte Enhetene/hierarkiet i statsforvaltningen Spørreundersøkelser Forvaltningsdata Forvaltningslitteratur	Midlertidig DI-løsning. Midlertidig DI-løsning. Midlertidig DI-løsning. Sikts forskningsdataplattform. Midlertidig DI-løsning.

Forvaltningsdatabasen er en kartlegging av departement og ytre etater i Norge etter 1947. Den gir oversikt over den hierarkiske organiseringen og hvordan denne har endret seg over tid. Utover det inneholder Forvaltningsdatabasen koblinger til:

- Forvaltningslitteratur som omhandler litteratur relevant for de ulike virksomhetene i Forvaltningsdatabasen, også årsmeldinger og tildelingsbrev.
- Oversikt over statsansatte som gjør det mulig å følge utviklingen i tallet på statlige arbeidstakere, 1980-2014.
- Utvalgsarkivet med tilrettelagt informasjon om utvalg, styrer og råd, samt om medlemmene i disse, for perioden 1980-1997 samt 2002-2022.
- Nøkkeldata fra Enhetsregisteret er lagt inn i databasen og koblet til for årene etter 2015.
- Det er lagt inn kobling til Arkivportalen for samsvarende enheter samt kobling til Lovdata som viser lovene relatert til den aktuelle enhet.

Forvaltningsdatabasen inneholder både kvantitative og kvalitative data. På mange måter kan det anses som et komplekst oppslagsverk med relasjoner mellom flere enheter.

Noen relevante undersøkelser var også knyttet til Forvaltningsdatabasen, bl.a.

«sentraladministrasjonsundersøkelsene», studien av «statlig autonomi og styring i staten» samt «forvaltningens EU-tilpassing», men de har blitt (eller skal bli) kuratert og publisert i Sikts forskningsdataplattform.

Eierskapet til Forvaltningsdatabasen og ansvaret for drift ble overført fra Rokkansenteret til NSD i 2006. Databasen har aldri vært fullfinansiert, men har i tillegg til midler fra NSDs grunnbevilgning, alltid fått offentlig tilskudd til drift fra hhv departement og seinere direktorat. Tilskudd er de siste årene fram til 2022 gitt fra Direktoratet for forvaltning og økonomistyring (DFØ) til NSD. I Prop. 1 S Kunnskapsdepartementet (2021–2022) foreslår departementet å flytte dette driftstilskuddet på 0,7 mill. kroner fra Finansdepartementets kap. 1605, post 01 til kap. 273 Kunnskapssektorens tjenesteleverandør, post 50.

Likt som Kommunedatabasen besluttet forprosjektet at Forvaltningsdatabasen skal driftes av DI i en midlertidig løsning for å ivareta datatilbudet under moderniseringsprosjektet.

2.4 PolSys

Databeholdning	Data	Beslutning
Data om det politiske system	Stortingets dokument 8-arkiv	Sikts forskningsdataplattform.
	Stortingets fraksjonsdata fra komiteinnstillinger	Sikts forskningsdataplattform.
	Stortingets biografidatabase	Midlertidig DI-løsning.
	Stortingets voteringsarkiv	Midlertidig DI-løsning.
	Regjeringens statssekretærarkiv	Midlertidig DI-løsning.
	Regjeringens statsrådsarkiv	Midlertidig DI-løsning.
	Partidokumentarkivet	Midlertidig i Sikts
	Valglistearkivet	forskningsdataplattform. Midlertidig DI-løsning.

«PolSys» er en samlebetegnelse for data om det politiske system, og det innebærer blant annet data om Stortinget, regjering og politiske parti.

Dette er informasjon om institusjoner, personer og ressurser innenfor Norges politiske og sentraladministrative system som stilles tilgjengelig gjennom disse arkivene. Dette er ulike type data, både matrisedata og tekstdata.

Enkelte av dataområdene ble vurdert som egnet for Sikts forskningsdataplattform, og det ble besluttet at de skulle publiseres der permanent. Det var fordi de kunne deles som flate analyseklare filer og hadde ikke behov for interaksjonsmuligheter.

Partidokumentarkivet hadde en egen søkefunksjon som det ikke var mulig å få på plass hos DI. Derfor valgte vi å samle alle dokumentene i en zip.fil og publiserte det som en midlertidig løsning i eksisterende forskningsdataplattform. Beslutningen sikret at man unngikk fullstendig nedetid på datatilbudet, men eksisterende forskningsdataplattform er ikke egnet for disse dataene uten videreutvikling, da det er veldig vanskelig å navigere i dem uten et søk. Dette er en lite brukervennlig løsning. Det har kommet flere henvendelser fra brukere som savner det tidligere søket.

For de resterende dataområdene ble det besluttet at de gikk inn i midlertidig DI-løsning.

2.4.1 Andre arkiv

“Andre arkiv” var en samlebetegnelse for data som ikke passet under stortinget, regjering eller parti i PolSys.

Databeholdning	Data	Beslutning
Andre	Akademikerarkivet Bedriftsstyrearkivet Organisasjonsarkivet Embetsmannsarkivet Manifestoarkivet	Sikts forskningsdataplattform. Sikts forskningsdataplattform. Sikts forskningsdataplattform. Sikts forskningsdataplattform. Avsluttet.

Dataområdene under «Andre arkiver» oppdateres ikke lenger og har lav verdi. Det ble i prosjektet bestemt å ta ned dagens datatilbud og heller formidle data via Colectica som flate filer. Brukere vil nå finne data via Sikts søkeportal.

Unntaket er Manifestoarkivet. Det ble besluttet at dette dataområdet ikke skulle videreføres i Colectica eller en fremtidig ny løsning. Dette skyldes av at «[The Manifesto Project](#)» allerede formidler disse dataene til forskning, og det er ikke i Sikts interesse å bruke ressurser på noe som allerede er løst. Særlig siden The Manifesto Project har oppdaterte data over en større tidsperiode og på tvers av flere land.

2.5 Kirkedatabasen

Databeholdning	Data	Beslutning
Kirkedata	Kirkedatabasen	Sikts forskningsdataplattform.

Grunnlaget for Kirkedatabasen er årlige KOSTRA-data, som Den norske Kirke rapporterer til SSB på soknenivå. Dataene blir supplert med tall fra kirkens medlemsregister og demografiske data fra SSB.

En del utvalgte variabler fra årsstatistikken blir samlet i såkalte nøkkeltall hvor soknestrukturen er regnet om til dagens for enklere å kunne lage tidsserier og hvor manglende verdier er estimert. Databasen skal være lagt til rette for forskere og for kirkens administrasjon.

Årsstatistikken faller pent inn i modellen med nedlastbare datasett. For tiden pågår arbeid med publisering til Sikts forskningsdataarkiv. Når denne prosessen er fullført, kjente feil i arkivet er rettet og direkte tilgang til datasettene er på plass, kan vi regne dette som en løsning som oppfyller kontraktsforpliktelsene. Ferdigstilt statistikk er fortsatt tilgjengelig for brukere, men med redusert brukervennlighet. Kirkedatabasen er finansiert av Kirkerådet. Siste faktura lød på kr 700.000 pluss moms.

2.6 Data fra Innovasjon Norge

Data om Innovasjon Norge	Kundeeffektundersøkelser Engasjementsregister Regnskapsregister SYSFIN	Sikts forskningsdataplattform. Sikts forskningsdataplattform. Avsluttet, men tar vare på data. Avsluttet, men tar vare på data.

Data fra Innovasjon Norge dreier seg om støtten fra Innovasjon Norge, samt forgjengerne Statens nærings- og distriktsutviklingsfond, og før det igjen, Distriktenes utviklingsfond, til foretak og enkeltpersoner. Bedrifts- og engasjementsregisteret er årlige tabeller med data på prosjektnivå over støtten som blir gitt. Kundeeffektundersøkelsene er en todelt spørreundersøkelse til mottakerne om effekten av støtten. Første del, Førundersøkelsen, kommer kort tid i etterkant. Andre del, Etterundersøkelsen, er en tilsvarende undersøkelse som kommer fire år etter første del. Respondentene er anonyme overfor Innovasjon Norge, og oppgaven er derfor satt bort til et eksternt konsulentfirma, for tiden Oxford Research.

Sikts database for Innovasjon Norge inneholder årlige data for Bedrifts- og engasjementsregisteret, Førundersøkelsen og Etterundersøkelsen. Tidligere fikk vi også regnskapsdata med finansierings- og sysselsettingsdata (SYSFIN) i prosjektene som hadde fått støtte. Dette ble aldri Nesstar-publisert, men var nevnt på nettsidene til NSD som tilgjengelige data.

Både Bedrifts- og engasjementsregisteret og Kundeeffektundersøkelsene kan publiseres som nedlastbare datasett, men er på grunn av ressursmangel fortsatt ikke publisert etter at Nesstar ble stengt ned. For Kundeeffektundersøkelsene kan vi også legge til at Oxford Research ikke har levert data på mange år.

Prosjektet er finansiert av Innovasjon Norge. Siste faktura lød på kr 160.000 pluss moms.

3.0 Kartlegging av brukerbehov

I andre fase av forprosjektet ble brukermiljøene kartlagt og innspill fra sentrale brukere samlet inn. I den sammenheng gjennomførte prosjektet seks dybdeintervju med brukere og en brukerundersøkelse for Kommunedatabasen.

Brukerinvolvering er avgjørende for å danne et konseptuelt grunnlag til modernisering som treffer behovet i sektoren.

3.1 Kartlegging av forskningsmiljøene og brukere

Medarbeidere i forprosjektet hadde dyp kjennskap til forskningsmiljøene som benyttet seg av datatilbudene. Deres erfaringer ble brukt til å kartlegge hvilke typer brukere det var aktuelt å kalle inn til brukerinvolvering.

I tillegg ble Google Scholar brukt for å plukke opp publikasjoner som hadde referert til data fra Datasamlingen om det norske samfunn.

Det ble identifisert fire type brukergrupper som det var særlig interessant å forholde seg til:

1) Spesialister i forskning

Dette er superbrukere fra forskningssektoren som har brukt tjenesten(e) i flere år på mange prosjekter. De har ofte hatt tett samarbeid med rådgivere hos Sikt/tidligere NSD. De kjenner data, datainfrastruktur og organisasjonen vår svært godt.

2) Generalister i forskning

Generalistene er forskere og studenter som gjerne har brukt tjenestene til ett prosjekt. De har ikke hatt behov for særlig mye rådgivning og støtte. De har brukt tjenesten i stillhet, fullført forskningsprosjektet sitt, og gått videre til andre datasamlinger. Denne gruppen representerer de fleste brukerne.

3) Ny brukere i forskning

Dette er gruppen kjente ukjente. Det betyr at dette er brukere som ikke bruker datatilbudene våre i dag, men som kunne ha hatt nytte av det. Grunnen til at de ikke har brukt datatilbudet kan skyldes manglende kunnskap, juridiske restriksjoner eller hull i datatilbud. Dette er potensielle nye brukere.

4) Brukere utenfor forskning

"Data om det norske samfunn" skiller seg ut fra andre datasamlinger, siden det inneholder data som også er ettertraktet av brukere utenfor forskningssektoren. Forvaltningsbrukere vil falle innenfor denne gruppen.

Ut fra disse kategoriene og de ulike dataområdene lagde forprosjektet en matrise i Miro for å identifisere de ulike brukerne.

Arbeidsdokumentet er tilgjengelig her: <https://miro.com/app/board/uXiVOErRkaA=/>.

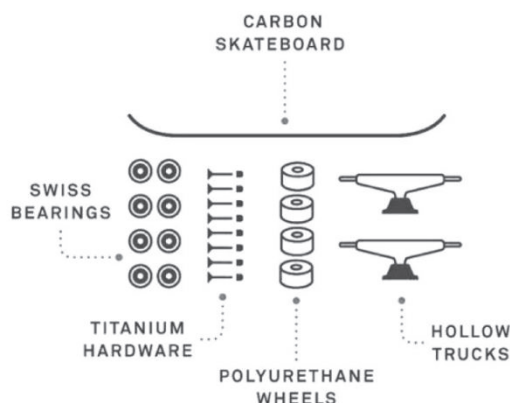
Forprosjektet hadde ikke ressurser eller tid til å intervju brukere fra alle kategoriene eller datatilbudene. Derfor valgte prosjektgruppen å snevre inn kartleggingen til de tre største datatilbudene: Forvaltningsdatabasen, Polsys og Kommunedatabasen.

3.2 Intervjuguide og spørreundersøkelse

For prosjektet var det viktig å utforske brukerens behov og problemer. Hvilke oppgaver skal de løse? Hvilke utfordringer, kostnader, risikoer og stresselementer møter de på? Hva ønsker de å oppnå, hva gir de muligheter, hvilke avhengigheter har de og hva forventer de?

Det er brukerens **utfordringer i forskningshverdagen** vi må lære, ikke hvilke spesifikke funksjoner brukerne tror de ønsker fra en ny formidlingskanal.

Even though customers buy this...



...they really want this.



Intervjuguiden ble revidert etter første brukerintervju. Det ble tydelig for teamet at den var for omfattende. Det førte til lite diskusjon og samtale rundt spørsmålene, men heller korte svar. Derfor ble intervjuguiden forenklet til disse spørsmålene:

1. Fortelle om dine erfaringer med bruk av våre data. Gjerne presenter en artikkel eller annet som var resultatet fra dataene du benyttet.
2. Hvordan pleier du å finne frem til data?
3. Hvordan vil du motta data? Hvilke krav har du til dokumentasjon, datakvalitet, koblingsmuligheter, filformat, etc.?
4. Har du innspill til nye forskningstemaer og/eller utvidelse av datatilbudet? Hva vil du data skal inneholde av informasjon? Har du innspill til hvordan dataene kan bli mer attraktive?
5. Hvordan kan datatilbudet fjerne stresserelementer fra din forskningshverdag? Hvordan hjelper vi deg med å løse problemer? Hvordan kan vi redusere risiko og kostnader?"

Denne intervjuguiden ble brukt i alle intervjuene etter det første. Den førte til at brukeren kom med dype innspill og gode diskusjoner. Dette hjalp oss med å få nye innspill og bekrefte/avkrefte tanker prosjektgruppen hadde om brukerbehovet.

Spørreundersøkelsen som ble sendt til brukere av Kommunedatabasen bestod av 15-17 spørsmål (avhengig av brukerens svar). Spørsmålene var en kombinasjon av fritekst og avkryssing, og tematisk lå de tett til intervjuguiden. Det var en fordel at flere av intervjuene var gjennomført før spørsmålene ble formulert. Dette hjalp med å spisse spørsmålenes relevans og tilleggsverdi til intervjuene.

Spørreundersøkelsen er tilgjengelig her:

<https://forms.office.com/Pages/DesignPageV2.aspx?subpage=design&FormId=RiE0y2V5XkKDq4j47LmeDTPxOHzkrqBMhpBB3A-EaM1URUFKNVRHWFBDdJUEJXUFdJTEcySIBERC4u&Token=9e9df68838a14182910dbe51a9d48e1c>.

3.3 Omfang og gjennomføring av brukerintervjuer og spørreundersøkelse

I forprosjektet ble det gjennomført seks dybdeintervju med brukere og det var 19 brukere som svarte på spørreundersøkelsen. Alle som ble kontaktet takket ja til intervju og flere uttrykte stort engasjement for Sikts datatilbud.

For intervjuene ble det rekruttert to brukere fra hver av Kommunedatabasen, Forvaltningsdatabasen og "PolSys". Prosjektgruppen konkluderte med at de største datatilbudene skulle prioriteres. Dog hadde flere av informantene brukt mer enn kun det ene datatilbudet. I tillegg bestod intervjuguiden av spørsmål som var relevant på tvers av alle datatilbudene, selv om samtalen med informantene var spisset inn på datatilbudet de hadde brukt mest.

Det var ønskelig med spredning i ulike type informanter. To av informantene kjente Sikts oppdrag svært godt fra å ha vært medlemmer i NSDs styre, mens de fire andre kjente oss kun fra å ha brukt tjenestene. Enkelte av informantene hadde kun brukt Sikt en gang, mens andre var gjentakende brukere som hadde brukt flere av Sikts tjenester. Alle informantene kom fra fagfelt innenfor Sikts hoveddomene; økonomi, historie, adm.org og statsvitenskap. Fire av informantene var svært erfarne forskere, en var en relativt ung forsker, og den siste var en doktorgradsstipendiat.

Intervjuene varte i overkant av en time hver og ble gjort virtuelt over Microsoft Teams. Det ble tatt opptak av samtalen som slettes innen 31.12.22. Det ble gjort et forsøk å bevare anonymiserte transkripsjoner, men amberscript var ikke egnet for norsk tale. Det ville vært for ressurskrevende å transkribere intervjuene manuelt.

Spørreundersøkelsen ble sendt ut til brukere av Kommunedatabasen i perioden 2017-2022. Det var kun 19 brukere som fullførte undersøkelsen, men alle hadde svart utfyllende.

Ettersom flere av spørsmålene skulle svares med fritekst var 19 deltakere absolutt tilstrekkelig til å danne seg et bilde som samsvarte med det som ble identifisert i intervjuene. 15 hadde brukt databasen til forskning, 5 til doktorgrad, 8 til mastergrad, 4 til undervisning og 1 til utredning (merk at enkelte brukere hadde benyttet seg av Kommunedatabasen til flere formål). De aller fleste kom fra enten statsvitenskap eller samfunnsøkonomi, men det var også deltakere med fagbakgrunn fra sosiologi, historie og siviløkonomi.

3.4 Resultater

Noen tema og innspill gikk igjen i de råd vi fikk i intervjuene og som også synes å være gyldig for flere av datasamlingene våre.

Mer brukervennlig

Flere trakk fram at de ønsker at data gjøres lettere tilgjengelig på en mer brukervennlig måte, for eksempel gjøre det enklere å ta ut og laste ned data. Med tanke på dataformat kom det fram at Excel er tilstrekkelig for de fleste. Noen ønsket seg også å kunne ha "min side" med favoritter, lagrede funn, lister etc, slik som hos for eksempel Nasjonalbiblioteket. Et annet innspill var å lage gode og oversiktlige måter å orientere seg i det omfattende datamaterialet. Navigasjonsspalten til venstre på de gamle websidene til "PolSys" ble trukket fram som svært nyttig for å finne fram i datasamlingene. Betydningen av å ivareta god dokumentasjon på nettsidene, gjerne overtydelig, ble også vektlagt. For eksempel ble det nevnt at i Kommunedatabasen er det vanskelig å få greie på hva omregningsår er og hvordan det fungerer. Tilbakemeldingene med tanke på DOI var litt mer sprikende; noen ønsker deg dette på datauttrekk mens andre mener det er tilstrekkelig å ha en DOI for datasamlingen (kilden).

Fullstendighet

En annen tydelig tilbakemelding var behovet for fullstendighet. Sikt må komplettere tidsserier og tette datahull samt gi data merverdi. Tidsseriedata er viktigere enn tverrsnitt. Gode tidsseriedata er stort sett en mangel, spesielt internasjonalt. Hull i tidsserier bør dokumenteres. Også viktig å vise hva vi ikke har, gjerne med en pekepinn om andre som kan tenkes å ha det. Videre, viktig å utvide og legge til andre / nye data / nye forskningstema. Spesielt viktig å systematisere og legge til eldre data, særlig før år 2000, nyere data er som regel lett å finne.

Gjøre datatilbudet bedre kjent

Det kom videre fram fra flere at datatilbudet er lite kjent, det er et stort uutnyttet potensial. Samlingene er underbrukt av andre enn forskerne selv. Kjennskap til samlingene spres gjennom nettverk, eller gjennom å google. Data er også underbrukt innenfor humaniora, som hos historikere. For å ivareta gode relasjoner med forskningsmiljøene ble understreket veldig tydelig at medarbeidere tilknyttet Sikt sine datasamlinger må være i kontakt og dialog med forskerne og delta på ulike relevante arenaer. Sikt bør være til stede med informasjon om sine datasamlinger på tilsvarende måte som representanter for forlag er til stede med tilbud om bøker og tidsskrift. Et annet ønske som kom fram, var at Sikt burde kunne bidra med sin kompetanse på kvantitativ metode og analyse inn mot ulike fagmiljøer. Mange har mest kompetanse på kvalitativ metode (historikere, for eksempel).

Kontakt og rådgivning

Viktigheten av at man får rask respons på henvendelser ble understreket. Det ble også trukket fram som svært nyttig å ha en kontaktperson og rådgiver som kjenner datasamlingene godt, for eksempel til å bistå med å rydde i materialet eller å legge til rette for å kople sammen ulike datasett, bl.a. å kople strukturdata sammen med individdata. Dette bidrar til å redusere risiko, kostnader og stressелеmenter. Det ble videre trukket fram at NSD var kjent som en trygg og etterrettelig kilde. Også i fortsettelsen bør data være korrekte. NSD var også kjent i miljøet for alltid å levere gode tjenester og service overfor brukerne, noe man håper vil fortsette i Sikt sine moderniserte tjenester.

Viktig datatilbud for sentrale fagmiljø

En viktig tilbakemelding var at datasamlingene utgjør et svært sentralt datatilbud innenfor de statsvitenskapelige fagene. I motsetning til for eksempel økonomene, som har tilgang til store relevante datamengder fra SSB, så mangler statsvitere og organisasjonsforskere systematiske kilder til gode og relevante data hos andre dataeiere og register. Betydningen av at data også er tilgjengelige for internasjonale brukere ble samtidig trukket fram. Norske data er kjent for å være svært gode.

Dokumentsamlingene

Det kom også fram tydelige ønsker om å gjøre de omfattende dokumentsamlingene våre søkbare, samt å tilrettelegge for dokumentanalyse. Noen forskere hadde også egne dokumentsamlinger som kunne avleveres og inngå et slikt datatilbud. Det kom også fram ønsker om å legge til rette innfallsporter til kvalitative studier samt pekere videre til annen relevant informasjon, noe som vil redusere tiden man bruker på å grave fram data.

Koblingsmuligheter

I både intervju og spørreundersøkelsen kom det fram at kobling av data var interessant. Flere gjorde det, og de som ikke hadde gjort det sa det hadde vært spennende med muligheter for det. I spørreundersøkelsen viste deg seg at 19 av 19 hadde koblet Kommunedatabasen med andre

datakilder. Mange hadde benyttet seg av blant annet annen data fra SSB, annen registerdata, uttrekk fra kartverket og egne data.

4.0 Anbefalinger for modernisering av "Data om det norske samfunn"

Forprosjektets mandat er formålet å utforme et konseptuelt grunnlag for modernisering med tanke på:

1. Mer brukervennlige datatjenester, både med hensyn til funksjonalitet, datatilbud og kvalitet (FAIR).
2. Mer effektiv drift og vedlikehold gjennom tettest mulig integrering i Sikts forskningsdataplattform.

I dette kapittelet vil anbefalinger knyttet til de nevnte formålene beskrives. Anbefalingene er basert på kartlegging av alle dataområdene, intervju- og spørreundersøkelse blant brukerne og et internt arbeidsseminar 30. august 2022.

4.1 Modernisering av datatilbudet

Det første formålet i mandat omhandler modernisering av datatjenestene knyttet til "Data om det norske samfunn". Data og infrastruktur har et stort potensial for utvikling av ny kunnskap og innovative tjenester som skaper betydelig verdi for samfunnet og sektoren. Disse anbefalingene er rettet mot å bedre Sikts datainfrastruktur.

Prosjektgruppen har anbefalinger om

- Hvordan datatilbudet kan forbedres innholdsmessig.
- Videreutvikling av Sikts forskningsdataplattform for formidling av databaser og kvalitativ tekstdata.
- FAIR-prinsippene og universell utforming.

4.1.1 Innholdsmessig oppdateringer

Resultatene fra brukerintervju og spørreundersøkelser viser at det er et gjennomgående ønske om at Sikt først og fremst bør oppdatere data og tette hull i tidsserier. Deretter kommer respondentene med konkrete forslag til utvidelser av datatilbudet, som da gjerne passer deres egen forskning.

Prosjektgruppen har dermed noen konkrete tips om ulike data som kan gå inn i forvaltningsdatabasen, Kommunedatabasen og "PolSys". Likevel er ikke disse tilstrekkelig til å beskrives i et konseptuelt grunnlag om innholdsmessige oppdateringer. Anbefalingene er gjerne tett knyttet til informantens egne forskningsprosjekt – som alene ikke er tilstrekkelig til å utforme en strategi for datafangst.

Det største funnet er derimot at forskeren selv anerkjenner dette. De uttrykker at Sikt må involveres i forskningsmiljøene for å selv kunne ta vurderinger på hvilke innholdsmessige oppdateringer av data som vil ha gjenbruksverdi hos forskningsmiljøene.

Prosjektgruppen anbefaler at Sikt styrker involveringen med forskningsmiljøene. Det er data som brukes som har verdi. Derfor må Sikt lykkes i å velge ut hvilke nye data som skal inn i datatilbudet. Derfor må Sikt være tett på brukerne.

Prosjektgruppen utgreier om hvordan tettere kontakt med forskningsmiljøene bør gå inn som en del av drift og vedlikehold av datatilbudet i kp.4.2.2.

4.1.2 Nye formidlingskanaler for databaser og kvalitative tekstdata

I 2016 til 2021 ble kvaliteten på infrastrukturen til kvantitativ surveydata løftet. Dette forgikk i flere infrastrukturprosjekter, deriblant NFR-finansierte NORDi (Norwegian Open Reserach Data Infrastructure)-prosjektet, EU-finansierte SSHOC (Social Sciences & Humanities Open Cloud)-prosjektet og OsloMet-finansierte SurveyBanken-prosjektet. Dette omfatter over 2500 datasett fra spørreundersøkelser som ble dokumentert etter DDI-standard for metadata. I oppstarten av NORDi-prosjektet, som bygde grunnmuren de andre infrastrukturprosjektene videreutviklet, var det plan om å samle all NSDs forskningsdata i denne infrastrukturen, men ambisjonene ble justert utover i prosjektet.

Det har medført at Sikt har veldig gode tjenester for arkivering og formidling av kvantitativ surveydata, men andre type data driftes på eldre løsninger som i liten grad oppfyller FAIR-prinsippene. Dette er dataene som ligger under "Data om det norske samfunn".

Det har dermed oppstått et betydelig kvalitetsskille på datatilbud mellom kvantitative datasett, som stort sett er surveydata, og 1) data som finnes i databasestruktur, som stort sett er registerdata og forvaltningsdata og 2) kvalitative tekstdata.

MODANSA vil ta opp tråden som ble etterlatt av tidligere infrastrukturprosjekter. Prosjektet vil bygge videre på eksisterende forskningsdataplattform for å lage:

- en formidlingskanal som kan formidle data fra åpne databaser.
- en formidlingskanal som kan formidle åpne kvalitative tekstdata.

For at disse kanalene skal kunne berikes med data må eksisterende systemer videreutvikles og ny arbeidsflyt for drift av "Data om det norske samfunn" må etableres. Det er viktig at dette baserer seg på løsningene som ligger under produktteamet "Lagring, kuratering og tilgjengeliggjøring. MODANSA kan ikke bygge på eksisterende datastruktur uten å bygge videre på tilhørende prosedyrer, rutiner og arbeidsflyt.

Forslaget om to formidlingskanaler medfører en sammenslåing av systemene og datatilbudene under "Data om det norske samfunn". Istedenfor å ha flere ulike systemer for å bearbeide disse, vil alt samles i en forskningsdataplattform og formidles gjennom to nye kanaler.

Forvaltningsdatabasen, Kommunedatabasen, Kirkedatabasen og "PolSys" er gode cases for slike formidlingskanaler. Databeholdningene inneholder kvantitative og kvalitative data. Det er enkelte koblingsmuligheter, men det er også klare utfordringer i koblingene. Metadata er svært variert – slik som man vil oppleve i denne type data som ikke kan lene seg på en surveybasert DDI-standard for metadata.

Ved å bygge en infrastruktur som får ut verdien av disse databeholdningene vil Sikt kunne koble på flere datakilder, som vil skape enda flere koblingsrelasjoner, og gi merverdi til all data. Det er helt sentralt at infrastrukturen som videreutvikles i MODANSA er skalerbar. Infrastrukturen vil modernisere "Data om det norske samfunn" og andre fremtidige datasamlinger som ikke er basert på surveydata.

I Sikt har ikke FAIR-prinsippene fått preget databaser og kvalitative tekstdata på likt nivå som med kvantitative surveydata. Det blir helt sentralt for MODANSA at FAIR-nivået for databaser og kvalitative tekstdata heves til samme nivå som surveydata.

I den sammenheng bør prosjektet gå tilbake til grunnprinsipper i FAIR og se hvordan prinsippene bør anvendes disse type dataene. Det vil være feil å presse databaser og tekstdata inn i rammene som ble anvendt for å få surveydata til å oppfylle FAIR-prinsippene.

Det er unødvendig kostbart at "Data om det norske samfunn" driftes i separerte systemer og tjenester. Dersom Sikt legger til funksjonalitet på en tjeneste vil ikke den investeringen heve kvaliteten på de andre tjenestene. En samordning av datatilbudene vil effektivisere tjenestetilbud og være bærekraftig. Det er en fordel at utviklere og fagpersoner forholder seg til én infrastruktur. Da vil man bygge tverrfaglig kompetansemiljø og hindre oppbygging av siloer.

4.1.2.1 Formidling av data fra databaser

Trenden viser at færre og færre respondenter ønsker å delta i surveyundersøkelser. Trolig vil andre type data få en stadig viktigere betydning i forskning. Det er derfor viktig at Sikt har en infrastruktur bygd for å fremme gjenbruksverdien for registerdata, offentlige data, store data, etc. Data som organiseres i databaser med relasjoner.

Prosjektet må utvikle en formidlingskanal som kan formidle store mengder kvantitative data med rike koblingsmuligheter. Formidlingskanalen må støtte både datavisualisering og datauttrekk. Dette innebærer at generalisten fortsatt vil kunne navigere i datasamlinger og få data visuelt presentert, mens spesialisten kan lage koblinger og utføre datauttrekk.

Kanalen vil la brukerne konstruere datasett ut ifra flere datakilder med koblinger på tvers av flere enhetstyper og laste de ned. Brukeren må ha mulighet til å finne data, visualisere data og koblinger, konstruere datasett og trekke ut data. Prosjektet bør også utforske muligheten til datauttrekk via API-spøringer. Utrekk av data bør generere en DOI som lar brukeren referere til det spesifikke datauttrekket.

Ved å legge samle flere databaser i en kanal, kan det enkelt og raskt skapes koblinger mellom ulike datakilder. Dog krever dette at metadata blir standardisert og tilrettelagt for maskinlesing – som vil være en stor oppgave. Det er en særlig utfordring av disse type data kan ha begrenset med metadata. Koblingsmuligheter mellom datasamlinger kan eksponere forskere for datakilder de ikke tidligere kjente. Det kan belyse koblingsmuligheter som ikke har blitt utforsket. Dette kan gi forskningssektoren nye og spennende muligheter.

Selv om ambisjonen er å skape relasjoner og koblinger, vil det finnes begrensninger på hvilke data som kan knyttes sammen. Enten grunnet manglende datagrunnlag for å bygge koblinger mellom enhetstyper eller manglende ressurser til å konstruere slike koblinger. Sånn er virkeligheten og det vil alltid finnes begrensninger – likevel vil infrastrukturen ha verdi for datasamlinger uten koblinger til andre data grunnet god datavisualisering og uttrekksmuligheter.

Det er i første omgang koblinger mellom Kommunedatabasen, Kirkedatabasen, Forvaltningsdatabasen og "PoISys" som vil være aktuelt, men kanalen som utvikles må være i stand til å koble på nye datakilder. Det er her den store verdiskapningen fra MODANSA ligger. Et datasett som går fra 100 variabler til 101 variabler, øker ikke bare med 1 variabel. Den øker med 100 nye relasjoner.

Det vil være nødvendig å omstrukturere og migrere data fra "Data om det norske samfunn" til forskningsdataplattformen det skal bygges videre på.

Prosjektet bør se hvordan FD-trygd (og andre dataregistre) ble strukturert inn i microdata.no. Selv om MODANSA ikke skal levere en analyseplattform, slik som Microdata, er det klare likheter i hvordan man kan bygge datasett på tvers av datakilder.

4.1.2.2 Formidling av åpne kvalitative tekstdata

Sikt har store mengder kvalitative tekstdata som i stor grad kun er et vedlegg som kan lastes ned eller bestilles via de ulike tjenestene. Unntaket var partidokumentarkivet som lot brukere søke på tvers av partidokumenter. Disse dokumentene hadde metadata som tillot komplekse søk.

Infrastrukturen for formidling av kvalitative tekstdata trenger modernisering for økt brukervennlighet og FAIR. I disse dataene har Sikt en mulighet til å være innovative på formidling av kvalitative tekstdata. Sikt kan bruke teknologiske fremskritt innenfor optisk tegngjenkjenning (OCR)² til å øke tilgjengeligheten. Data og kunstig intelligens kan skape nye og innovative tjenester som bidrar til sektorens og samfunnets utvikling. Dette kan åpne Siks relevans for flere fagfelt og forskningsmiljøer i kunnskapssektoren.

I "Data om det norske samfunn" finnes det koblinger til mengder av politiske og organisatoriske dokumenter – relevant både i dag og som historiske data – som har potensial til å få en betydelig høyere gjenbruksverdi.

Tidligere har Sikt hatt en peker til dokumentene. For eksempel vil forvaltningsdatabasen peke til alle årsmeldinger til Equinor. "PolSys" vil peke til alle partiprogram fra Venstre. Dette er en veldig enkel form for tilgjengeliggjøring.

MODANSA kan utvikle en formidlingskanal som leser gjennom alle dokumentene.

Formidlingskanalen vil ha et søk som ikke bare finner dokumenter, men den finner relevant innhold i. En slik formidlingskanal kan tiltrekke seg aktører som vil koble på egne kilder til kvalitative tekstdata. Nasjonalbiblioteket har blant annet tilgjengeliggjort kvalitative tekstdata, slik som "Stortingsforhandlinger 1814-2000", hos [Felles datakatalog](#). Sikt kan spille en sentral rolle i å tilrettelegge og formidle data fra aktører som Nasjonalbiblioteket til forskning.

En slik formidlingskanal kan være egnet for å dele kvalitativ intervjudata, men da må Sikt være sikker på at datamaterialet er anonymisert og kan deles åpent.

4.1.2.3 Universell utforming

Tjenestene tilknyttet "Data om det norske samfunn" oppfyller muligens ikke kravene til universell utforming. Det har ikke blitt gjennomført en kartlegging av dette, men basert på levetiden til disse tjenestene er det å forvente.

Dette er en god grunn til å prioritere modernisering av disse datatilbudene – slik at alle i kunnskapssektoren kan få nytte av disse rike dataene på like betingelser.

4.2 Modernisering av drift og vedlikehold

Det andre formålet beskrevet i prosjektmandat omhandler effektivisering av drift og vedlikehold, samt integrering med allerede eksisterende forskningsdataplattform og arbeidsflyt.

² [OCR – Store norske leksikon \(snl.no\)](#)

Prosjektgruppen har anbefalinger om:

- Organisering av drift og vedlikehold i team.
- Tettere kontakt mellom Sikt og forskningsmiljøene.
- Verktøy for automatisering av datainnhenting.
- Etablering av felles prosedyrer for arkivering, databearbeiding og formidling av data.
- Sammenslåing av produktteamene "Lagring, kuratering og tilgjengeliggjøring" og "Data om det norske samfunn".

4.2.1 Organisere i team

En konsekvens av at det ble utviklet ulike systemer til å bearbeide og formidle datasamlingene under "Data om det norske samfunn" er fragmentert kompetanse og personavhengighet. Fordelene med at enkelt personer drifter datatilbud er:

- 1) At brukerne får tilgang til en ekspert med spisskompetanse innenfor den enkelte datasamlingen.
- 2) At medarbeideren som har ansvar for et datatilbud jobber veldig effektivt og kunnskapsrikt med det enkelte datatilbudet.

Problemet er at en slik organisering er sårbart av flere grunner:

- 1) Dersom en medarbeider blir utilgjengelig eller forlater Sikt oppstår det et stort kunnskap- og kompetansehull.
- 2) Sikt kan ikke skalere opp dekningen på et datatilbud ved behov.
- 3) Det må bygges opp kompetanse til ulike system og arbeidsflyt. Både for fagpersoner og utviklere.
- 4) Det er begrenset hvordan produktområdet kan flytte på ressurser. Det er særlig relevant dersom belastningen blir for stor på en enkeltperson som alene sitter på ansvaret for drift av et datatilbud.
- 5) Det er problematisk dersom finansiering av et datatilbud reduseres eller stoppes.
- 6) Det er få muligheter for synergieffekter.

En ny team-organisering vil skape bredere dekning og fleksibilitet – som vil gjøre det enklere å få mest ut av ressursene satt av til drift. I tillegg vil et team, med felles systemer og arbeidsflyt, være kostbesparende i seg selv.

Den største utfordringen med teamorganisering blir balansen mellom bred dekning og spisskompetanse. Det vil være naturlig at enkelte medarbeidere spesialisere seg mer i et datatilbud enn et annet. Det kan være nyttig med tanke på effektiv drift og rådgivning til brukere. Produktteamet må finne en balanse som legger til rette for både dybde og bredde.

De siste årene har det vært få personer som har driftet "Data om det norske samfunn". For at teamorganiseringen skal være vellykket er produktteamet avhengig av at det er satt av tilstrekkelig midler til drift og vedlikehold.

4.2.2 Tett kontakt med forskermiljøene

Sikt er avhengig av å ha tett kontaktene med forskningsmiljøene for å ha suksess med datafangst til "Data om det norske samfunn". Sikt skal jobbe etter innsikt og ikke antagelser om hva kunden faktisk vil ha. Derfor er det så viktig å få samstyring og samarbeid med sektoren til å fungere godt.

Selektiv datafangst er sentralt i drift av datatilbudet. Det vil være for ressurskrevende å oppdatere alt av data innenfor datatilbudet. Datafangst kan brukes for å legge til nye dataområder, dekke hull i manglende tidsserier og oppdatere etablerte dataområder. Verdien til "Data om det norske samfunn" vil styrkes av å treffe på innhenting av data med høy gjenbruksverdi. Enten gjennom å ha verdi isolert eller ved relasjonene innhentingene skaper mellom data.

Teamet som skal utvide, drifte og vedlikeholde datatilbudet skal identifisere satsingsområdene for datafangst. Dette krever tett godt samspill med forskningsmiljøene. Dialogen sikres gjennom deltakelse på forskningskonferanser som er relevant for denne type data. Funnene i intervjuene tyder på at forskningsmiljøene vil ha Sikt "med på laget". De ser at det er i deres interesse å involvere Sikt i prosjekter og konferanser. Det blir vanskelig for den enkelte forskeren å fortelle Sikt hva som bør satses på, men i dialog med flere av dem vil Sikt kunne se trendene og behovet.

4.2.3 Verktøy for automatisering av datainnhenting

Det er et stort potensial for å effektivisere innhenting av data fra ulike datakilder. En rekke register legger ut åpne data via APIer. Dette gjør at Sikts datatilbud kan utvides gjennom å koble slike datasett til våre data.

Det anbefales at MODANSA utforsker muligheten til å benytte seg av APIer for å automatisere innhenting av data. Dette kan redusere timene som går inn i å regelmessig oppdatere de ulike datasamlingene. Det kan frigjøre tid til å hente inn nye datakilder eller tette hull i tidsserier.

Registerdata fra SSB er sterkt representert i "Data om det norske samfunn". Det bør undersøkes om det er mulig å benytte seg av SSBs APIer for å hente data automatisk.

4.2.4 Etablere fellesprosedyrer i produktområdet

Ettersom "Data om det norske samfunn" har vært personavhengig og bygd på ulike systemer, så er det begrenset med dokumenterte prosedyrer på inntak og bearbeiding av data.

Ved modernisering av systemene knyttet til "Data om det norske samfunn" må det etableres nye felles prosedyrer for drift og vedlikehold. Disse prosedyrene bør ligge tett på allerede etablerte prosedyrer tilknyttet infrastrukturen MODANSA skal bygge videre på. Dette medfører blant at dataprosesseringsmetoder bør tilrettelegges for å gjennomføres i Python.

Det samme gjelder å forene hvordan RT og Servicesenteret brukes i produktområdet "dataforvaltning". Produktteamene har allerede gått over til å benytte seg av RT, men det løses på ulike måter. Ideelt sett bør man gjennom prosjektperioden ha en fellesprosedyre for henvendelser. Det er særlig to fordeler med å samle alt tilknyttet datahenvendelser i en RT-kanal: 1) medarbeiderne lærer fra hverandres svar og 2) terskelen for å kunne svare på henvendelser knyttet til det som i dag oppleves som noen andres ansvar reduseres.

4.2.4 Sammenslåing av produktteamene "Lagring, kuratering og tilgjengeliggjøring" og "Data om det norske samfunn"

I Sikts planer skal produktteamet "Data om det norske samfunn" gå inn i produktteamet "Lagring, kuratering og tilgjengeliggjøring". Det er et ønske om å bygge produktteams med harmonisert arbeidsflyt og bruk av systemer.

MODANSA-prosjektet handler ikke kun om teknisk utvikling, men også organisatorisk utvikling. For å styrke datatilbudet må Sikt organisere drift og vedlikehold rundt et produktteam som arbeider sammen for å få det beste ut av Sikts datatilbud.

Dette legger føringer for den tekniske løsninger som utvikles i MODANSA, da det må legges tett opp til arbeidsflyt som blir brukt i forskningsdataplattformen som det bygges videre på.

Et av prosjektets hovedmål bør være at produktteamene har blitt sammenslått ved prosjektslutt.

5.0 Veien videre for MODANSA-prosjektet

[OBS! Kp.5 er ikke lenger gyldig. Se satsingsforslaget til MODANSA for oppdatert informasjon om veien videre: [Satsingsforslag MODANSA.docx](#)]

5.1 Finansiering av gjennomføringsprosjektet

I dokumentet "Sikts satsingsområder innen forskning 2023-2026" ligger MODANSA-prosjektet inne med et estimert budsjett på totalt 35 millioner kroner. Budsjetteringen er fordelt over fire år: 7 millioner i 2023, 11 millioner i 2024, 11 millioner i 2025 og 6 millioner i 2026. Det er foreslått at disse midlene skal hentes fra bevilgning.

Under ligger en foreslått investeringsplan basert på det estimerte budsjettet. Kostnadssummene er oppført i 1000kr. Prosjektleveransene er beskrevet i kp.5.2.

Prosjektleveranser	2023	2024	2025	2026	
Videreutvikle Sikts forskningsdataplattform	3 000	4 000			
Standardisere maskinlesbar metadata	1 500	1 000			
Omstrukturere og migrere data	1 500	2 000			
Utvikle verktøy for innhenting av data via API	1 000	1 000			
Etablere ny arbeidsflyt			1 000		
Utvikle formidlingskanal for åpne databaser		1 500	5 000	2 000	
Utvikle formidlingskanal for åpne kvalitative tekstdata		1 500	5 000	2 000	
Re-organisere drift til ny teamstruktur				500	
Sammenslå produktteamene				500	
Totalt	7 000	11 000	11 000	6 000 000	

Kostnadene er i all hovedsak knyttet til interne arbeidstimer. Trolig vil det ikke være behov for ekstern hjelp fra konsulenter. For de tekniske løsningene vil det i tillegg være utgifter knyttet til utvidelse av teknisk infrastruktur.

Utgiftene til teknisk infrastruktur vil gå inn i fremtidig drift av datatilbudene.

For eksempel: tidligere med bruk av on-prem servere kjøpte man 100GB lagringsplass. Dette var et engangskjøp og når lagringsplassen var fylt måtte man kjøpe mer lagringsplass. I dag driftes Sikts forskningsdatainfrastruktur derimot på Microsoft Azure. Dersom man velger å kjøpe 100GB lagringsplass i denne skybaserte infrastrukturen betaler man for hver måned. Dette medfører at utgifter knyttet til teknisk infrastruktur er løpende i drift av tjenestene. Når alle datasamlingene migreres inn i Sikts forskningsdatainfrastruktur vil utgiftene økes da det vil være behov for mer lagringsplass.

Moderniseringen av "Data om det norske samfunn" har blitt enda viktigere etter Sikt ble opprettet. Samfunnsoppdraget har blitt bredere i Sikt enn det det var i tidligere NSD. Nå skal organisasjonen levere datatjenester til hele kunnskapssektoren. Da er det ikke tilstrekkelig med datainfrastruktur optimalisert for surveydata. Det er et klart behov for å forbedre datatilbudet for andre typer data gjennom å videreutvikle eksisterende forskningsdataplattform og arbeidsflyt.

Dette er i tråd med Sikts virksomhetsmål om å utvide bruken av våre produkter og øke verdien av data og infrastruktur. Sikt er avhengig av å investere i egen infrastruktur for å støtte morgendagens behov i forskningssektoren som vil se til andre typer data enn surveydata.

Prosjektgruppen stiller seg bak forslaget at MODANSA-prosjektet finansieres gjennom bevilgning. Eventuelt kan det benyttes egenfinansiering. NSD tok med seg en egenkapital inn i Sikt som var tenkt til å brukes på videreutvikling av datainfrastruktur.

Et alternativ til bevilgningsfinansiering og egenfinansiering er at MODANSA-prosjektet blir finansiert gjennom brukerbetaling. Dette fraråder prosjektgruppen da det vil heve terskelen til å be om tilgang til data. Det kan argumenteres for at brukerfinansiering strider med prinsippet om åpen datadeling. Å sette en pris på tilgang til data og kunnskap kan bidra til å ekskludere forskningsmiljøer fra å delta i forskning.

Det er ikke tradisjon for brukerfinansiering og en innføring vil trolig blir negativt mottatt av forskningsmiljøene. Sikt kan ikke forsvare kostnadene knyttet til å bygge opp en datainfrastruktur eller ressursene som går til kuratering og kvalitetssikring av forskningsdata – dersom det ikke blir brukt til sitt høyeste potensial. Det er gjenbruk av data som skaper verdier, ikke arkivering av data.

Dersom det likevel er et ønske om brukerfinansiering må det noteres at det er lite hensiktsmessig å kreve betaling fra aktørene som står bak dataene i "Data om det norske samfunn". Det er Sikt som innhenter data, i motsetning til surveydata hvor det er forskerne som leverer data til Sikt. Disse dataene hadde ikke vært tilrettelagt for forskning dersom Sikt ikke hadde samlet dem inn.

Det betyr at det er brukerne som ber om tilgang til data som bør betale for det. I Sikts tilfelle måtte man argumentert for at det ikke ble tatt betalt for tilgang til selve dataene, men for arbeidet med å tilrettelegge data for forskningsformål. Dersom Sikt skal anvende brukerfinansiering, bør kostnaden rettes mot institusjonene, ikke sluttbrukerne.

5.2 Hovedleveranser

Basert på anbefalingene har forprosjektet lagd en liste over leveranser som kan inngå i MODANSA-prosjektet. Merk at leveransene definerer prosjektets krav til kvalitet og omfang. Dette må alltid sees i kontekst av tidsramme og kostnadsramme. Derfor må leveransene justeres etter det som blir vedtatt å være MODANSAs rammeverk.

Forslag til leveranser:

- Formidlingskanal for åpne databaser bygd på eksisterende forskningsdataplattform.
- Formidlingskanal for åpne kvalitative tekstdata bygd på eksisterende forskningsdataplattform.
- Omstrukturering og migrering av data fra "Data om det norske samfunn" til Sikts dataplattform.

- Videreutvikle Sikts forskningsdataplattform til å håndtere kuratering og langsiktig lagring av nye type data.
- Standardisering av maskinlesbar metadata som beskriver "Data om det norske samfunn".
- Verktøy for innhenting av data via API.
- Etablering av ny arbeidsflyt som er tilrettelagt eksisterende løsninger.
- Re-organisere drift av "Data om det norske samfunn" til ny teamstruktur.
- Sammenslå produktteamene "Lagring, kuratering og tilgjengeliggjøring" og "Data om det norske samfunn".

Leveransene er sammenvevd, og det er derfor ikke mulig å si at en er mer prioritert enn en annen. MODANSA må levere en videreutvikling av Sikts forskningsdataplattform som tilrettelegger for at "Data om det norske samfunn", samt nye datasamlinger, får samme gjenbruksverdi og arbeidsflyt som resten av Sikts datatilbud.

Prosjektet må vurdere hvordan anbefalingene i denne rapporten best kan realiseres med hensyn til de foreslåtte leveransene og budsjettet.

5.3 Rammebetingelser

I dokumentet "Sikts satsingsområder innen forskning 2023-2026" er MODANSA satt opp som et fire-årig prosjekt med et estimert budsjett. Sikts ledelse må ta stilling til hva som veier tyngst av tidsrammen, kostnadsrammen og krav til kvalitet og omfang. Det er fordi det må være klart for MODANSA-prosjektet hvilke elementer som kan justeres dersom det oppstår problemer underveis.

Forprosjektet foreslår at det er tidsrammen er justerbar. Basert på tidligere erfaringer ser man at selv om det er satt av nok ressurser til et prosjekt, så er det ikke tilstrekkelig kapasitet blant utviklere og fagpersoner. Det hjelper lite å ha et stort budsjett om prosjektmedarbeidere ikke har kapasitet til å bruke timene.

En myk tidsramme vil gjøre det enklere å justere prioriteringer i porteføljestyringen.

I tillegg har vi sett at myke tidsrammer åpner opp for synergieffekter. Det er gjerne slik at det som utvikles i ett prosjekt vil ha nytteverdi for tre andre prosjekter. Ved harde tidsrammer kan ikke prosjektet styres smidig til å benytte seg av slike muligheter.

Forprosjektet foreslår at MODANSA-prosjektet og Sikts ledelse blir enig om et kostnadsestimat basert på det endelige vedtatte prosjektmandatet.

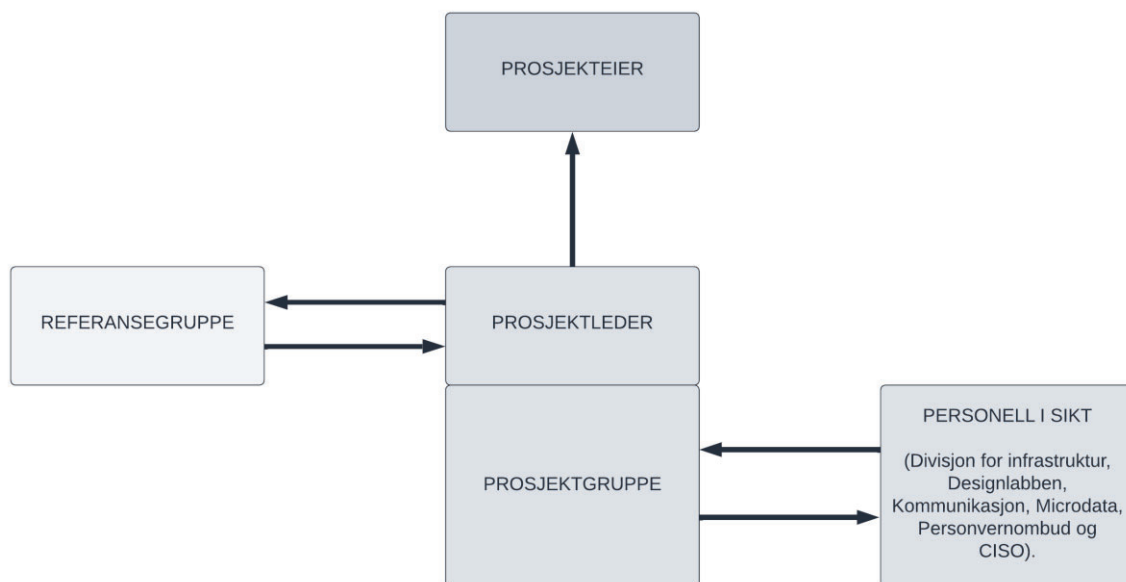
5.4 Organisering og forankring

MODANSA-prosjektet bør forankres under produktområdet Dataforvaltning. I all hovedsak bør prosjektgruppen bestå av prosjektmedarbeidere hentet fra Seksjon for Forskningsdata og Seksjon for Forskningsløsninger.

Prosjektgruppen må benytte seg av kompetanse på tvers av Sikt, deriblant kommunikasjon, Divisjon for data og infrastruktur, Designlaben og produktområdet Microdata. Det vil også være behov for rådgivning fra CISO og Personvernombud.

Prosjektgruppen vil fraråde at det opprettes en styringsgruppe. En styringsgruppe vil ha stor verdi i et samarbeidsprosjekt på tvers av organisasjoner, men for MODANSA - et internt utviklingsprosjekt - vil det være mer hensiktsmessig med en prosjekteier. For å styrke forankring bør prosjekteier være produktområdeleder for Dataforvaltning.

En referansegruppe bestående av interessenter bør vurderes. Det kan være nyttig å få tilbakemeldinger underveis i prosjektperioden som kan sikre at prosjektet er på riktig vei og treffer på det kunnskapssektoren trenger. Referansegruppen kan bestå av personer som representerer de ulike forskningsinstitusjonene.



5.5 Forventet gevinster

I oppstarten av MODANSA bør det gjennomføres en gevinstrealiseringsplan som identifiserer hvilke forventete gevinster prosjektet vil oppnå, avklarer roller og ansvar for å realisere gevinstene, konkretiserer tiltak og aktiviteter som er nødvendig for å hente gevinstene, gjennomfører basemålinger og bestemmer tidspunkt for fremtidige målinger av gevinster, og estimerer hvilke ressurser som kreves for å kunne vurdere gevinstene.

Det er for tidlig å utforme en gevinstrealiseringsplan, men det er likevel mulig å se potensielle gevinster Sikt kan forvente fra MODANSA-prosjektet:

- **Økning i antall brukere.**
Ved å bygge videre på den eksisterende forskningsdataplattformen for å gi et godt datatilbud for databaser og kvalitativ tekstdata blir Sikt relevant for en større brukergruppe enn tidligere. Jo flere kunder som bruker et produkt, jo mer robust blir det.
- **Økning i bruk blant eksisterende brukere.**
Ved å samle Sikts databaser i et datatilbud med muligheter for nye koblinger på tvers av datakilder, får datasamlingene en ny verdi. Det kan forventes at tidligere brukere av "PolSys", Forvaltningsdatabasen og Kommunedatabasen vil utforske forskningsspørsmål som blir mulig gjennom nye koblinger mellom data.
- **Ny datainfrastruktur unngår fremtidig reduksjon i bruk.**
Surveydata vil fortsette å prege forskning i tiden fremover, men i et langtidsperspektiv må Sikt være forberedt på å håndtere andre type forskningsdata for å beholde relevans for brukere. Med en stadig lavere svarprosent i spørreundersøkelser er det å forvente at

forskningsmiljøene vil se til andre type data.

- **Sikt tiltrekker seg nye samarbeidspartnere som ønsker å tilgjengeliggjøre data.**
Utvidelsen av forskningsdataplattformen gjør Sikt til en relevant samarbeidspartner for aktører som vil tilgjengeliggjøre forskningsdata som ikke passer inn i surveydata-infrastrukturen. Det er et særlig stort potensial for fagmiljøer innenfor historie og lingvistikk som kan se anse formidlingskanalen for kvalitative tekstdata som ideell for formidling av deres data.
- **Ny forskningsdatainfrastruktur lar Sikt utvide datatilbudet.**
Videreutvikling av datainfrastrukturen vil i første omgang modernisere datatilbudet til "Data om det norske samfunn", men den største verdien vil være i videreutvikling av Sikts forskningsdataplattform og formidlingskanalene. De nye kanalene lar Sikt formidle data som tidligere ikke passet i Sikts forskningsdataplattform. Sikt vil kunne ta imot nye type data og gi dem høy merverdi.

Gevinsten av MODANSA stopper ikke med moderniseringen av "Data om det norske samfunn". Prosjektet bygger en grunnmur for ikke-surveybasert data som legger til rette for å ta inn nye datasamlinger.
- **Effektivisering av innhenting av forskningsdata**
MODANSA kan utvikle verktøy som lar Sikt benytte seg av eksisterende APIer hos andre datakilder for å automatisere innhenting av data. Dette vil være kostnadsbesparende.
- **Sikt bidrar til mer åpen tilgang til data, slik at mer data blir tilgjengelig på like betingelser og til lavest mulig kostnad.**
Kanalene vil formidle data som er åpent tilgjengelig. Ved å tilby forskningssektoren infrastruktur som lar forskere enkelt dele data åpent for gjenbruk bidrar Sikt til verdiskapning. Sammenslåingen av systemene og arbeidsflyten som tilgjengeliggjør "Data om det norske samfunn" vil bidra til at kostnaden for tilgjengeliggjøring reduseres.
- **Positiv oppmerksomhet for Sikt.**
MODANSA er et moderniseringsprosjekt som skal heve kvaliteten på forskningsdatainfrastruktur som forskningssektoren har nytte av. Prosjektet vil dermed vekke positiv oppmerksomhet fra kundene våre. Sikt bør delta på konferanser for å prate om moderniseringen som foregår og legge opp til bred brukerinvolvering. Det bør settes opp en nettside som forteller om hva prosjektet skal levere og hvilke muligheter det vil gi til forskning i Norge.
- **Tverrfaglig smidig produktteam.**
En ny teamorganisering i arbeidet med "Data om det norske samfunn" vil gi produktteamet muligheten til å ledes smidig. Det vil bli enklere å flytte ressurser til ulike oppgaver. Slik at støtte til et datatilbud kan skaleres opp og ned ved behov. Dette vil la produktteamet justere kursen underveis.

- **Nedbygging av silo.**
Produktområdet "Dataforvaltning" består i dag av fem produktteam. Ved oppretting var det tenkt at "Data om det norske samfunn" etter hvert skulle gå inn i "Lagring, kuratering og tilgjengeliggjøring". I forprosjektet har en del forskningsdata fra "Data om det norske samfunn" allerede gått inn i eksisterende forskningsdataplattform (og ny arbeidsflyt), men det er arbeidet i MODANSA som vil tilrettelegge for sammenslåingen av produktteamene.
- **FAIR.**
Ikke-surveybasert data vil heves til et høyere FAIR-nivå med modernisering. Data vil bli lettere å finne, siterbart, interoperabelt, og ha høyere gjenbruksverdi.
- **Universell utforming.**
Å oppgradere fra gamle formidlingskanaler vil gjøre at den nye datatjenesten kan være tilgjengelig for alle under like betingelser med tanke på universell utforming.

5.6 Risikoer

Alle prosjekter er preget av usikkerhet, det vil si at de støter på uventete situasjoner og hendelser. I oppstart av MODANSA-prosjektet må det gjennomføres en risikoanalyse. I forkant av dette har forprosjektet identifisert risikoer som prosjektet bør følge opp:

- **Vage leveranser.**
Det endelige prosjektmandatet må inneholde konkrete leveranser for prosjektgruppen. Utviklingsprosjekter kan fort preges av vage og store ambisjoner. Det er viktig å konkretisere leveransene til realistiske oppgaver med klare rammer. Det er avgjørende at en dyktig systemarkitekt er med på å definere de endelige leveransene i MODANSA.
- **Tilgjengelighet til nøkkelpersoner.**
Siden drift av "Data om det norske samfunn" har vært så personavhengig er fagkunnskapen og kompetansen til datasamlingene fordelt på få personer. Det medfører at prosjektet er særlig sårbart for at nøkkelpersoner blir utilgjengelig.
- **Drift av datatilbud under utvikling.**
MODANSA vil kreve mye tid og oppmerksomhet fra medarbeidere som drifter dagens datatilbud. Derfor bør disse personene skjermes fra andre oppgaver i Sikt for å unngå overbelastning over tid.
- **At formidlingskanalene som lages blir for spesialisert datasamlingene Sikt allerede har.**
Ettersom de nye kanalene først og fremst skal formidle data fra Kommunedatabasen, Forvaltningsdatabasen og "PoISys" risikerer man at kanalene blir for spesialisert disse dataene. Formidlingskanalene må være bygd for å håndtere nye datasamlinger.
- **At formidlingskanalene som lages blir for generelle til å formidle verdiene til datasamlingene**
På den andre siden er det viktig at kanalene er spisset nok til at gjenbruksverdien i datasamlingene ikke reduseres. Det blir en viktig balansegang å fremme innholdet i det som

kan være ganske ulike type data, men også å ha formidlingskanaler som er egnet for nye datasamlinger.

- **Tilgang på utviklerressurser**

Tidligere har prosjekter i tidligere NSD blitt stanset som konsekvens av manglende porteføljestyling. Dette har skapt forsinkelser i prosjekter da det ikke har vært nok utviklere tilgjengelig. Et tiltak for å unngå dette kan være en myk tidsfrist på MODANSA.

5.7 Interessenter

En interessent er en person eller organisasjon som er aktivt involvert i prosjektet, eller som har interesser som blir berørt av prosjektets gjennomføring eller sluttresultat. MODANSA er et prosjekt som vil ha både interne og eksterne interessenter. I oppstarten av prosjektet må det gjennomføres en interessentanalyse som vil identifisere interessenter, analysere hvordan interessentene berører eller berøres av prosjektet og velge en strategi for hvordan interessentene skal håndteres.

Det er tidlig i prosessen, men under er det skissert en liste over mulige interessenter som prosjektet må håndtere:

- **Sluttbrukere fra forskningssektoren.**

En av de største interessentene i prosjektet vil være sluttbrukerne som opplever at datatilbud de har brukt tidligere endres drastisk. Det må forventes misnøye av at prosjektet vil kunne gå utover ressursene til drift av dagens løsninger under utviklingsperioden. Det blir viktig å formidle at nedsatt kapasitet til drift vil lede til en modernisering av datatilbudet som er i deres langsiktige interesse.

- **Forskningsinstitusjonene.**

Innenfor kunnskapssektoren – som Sikt skal levere tjenester til - berører dette prosjektet først og fremst forskningsinstitusjonene. Sikt vil tilby nye formidlingskanaler som vil øke deres muligheter til å få tilgang til data og tilgjengeliggjøre egen data for gjenbruk. Det kan være lurt å involvere forskningsinstitusjonene i en referansegruppe. På den måten kan prosjektet få tungtveiende innspill, promotere det nye datatilbudet til sektoren og styrke Sikts omdømme.

- **Direktoratet for forvaltning og økonomistyring (DFØ).**

Det er Sikt som driver forvaltningsdatabasen, men den mottar et tilskudd fra DFØ. De bør involveres slik at de opplever at moderniseringen av datatilbudet ivaretar deres interesser.

- **Stortingets datatjeneste.**

Videreutviklingen av infrastrukturen kan potensielt gjenåpne et tett samarbeid med stortingets datatjeneste. De bør informeres om prosjektet og følges opp.

- **Medarbeidere i produktteamet "Data om det norske samfunn".**

Anbefalingene i dette dokumentet foreslår endringer som vil ha betydning på medarbeideres arbeidshverdag. Det er viktig at det deres innsikt og innspill blir ivare tatt under prosjektet. De må være en aktiv part i prosjektet og være involvert i videreutvikling. Prosjektet må unngå krysspress mellom drift og utvikling som potensielt kan overbelaste

kapasiteten.

- **Produktteamet “Lagring, kuratering og tilgjengeliggjøring”.**

Produktteamet “lagring, kuratering og tilgjengeliggjøring” må være involvert i prosjektet slik at arbeidsflyt og systemer blir riktig harmonisert. Dette er sentralt for at fremtidig sammenslåing av produktteamene vil gå knirkefritt.



Saksdokument

Porteføljestyre for forskning og kunnskapsressurser

Til	Porteføljestyre for forskning og kunnskapsressurser
Dato	28.08.23
Saksnummer	17/23
Saksarkivnr.	23/00000
Sakstype	O/D-sak
Saksansvarlig	Vigdis Namtvedt Kvalheim
Saksbehandler	Asbjørn Risan

Kunnskapsressurser og bibliotekttjenester

Forslag til vedtak

Porteføljestyret tar saken til orientering.

Saksfremstilling

Kort historikk

BIBSYS ble etablert som et prosjekt i 1972 for å utvikle og levere bibliotekttjenester til UH-sektoren. BIBSYS Bibliotekssystem ble tatt i bruk i 1980 og har blitt utviklet, vedlikeholdt, driftet og videreutviklet løpende frem til 2016. Parallelt med etableringen av bibliotekssystemet ble BIBSYS-konsortiet etablert bestående av de institusjonene som benyttet bibliotekssystemet. Gjennom konsortieorganiseringen er de deltakende institusjonene sikret medvirkning og styring av de tjenestene som leveres.

Ett sentralt element i utviklingen av systemet var at det skulle være et felles system med en felles katalog. Alle jobbet inn mot de samme dataene og en endring på en post kom alle til gode. Det skulle også være et integrert system som ivaretok alle de sentrale biblioteksfunksjonene i et og samme system.

Tjenestetilbudet ble etter hvert utvidet til å også omfatte andre funksjoner og tjenester som søkegrensesnitt, institusjonelle arkiv, forskningsdokumentasjon m.fl.

I 2008 ble det besluttet at man ikke skulle fortsette å videreutvikle et egenutviklet system, men gå over til et av de kommersielt tilgjengelige bibliotekssystemene på markedet. Høsten 2015 tok konsortiet i bruk bibliotekssystemet Alma levert fra Ex Libris.

Dagens tjenestetilbud



Det overordnede målet for produktområdet er:

Sikt skal sikre gode og kostnadseffektive bibliotekstjenester gjennom de systemene og støttetjenestene vi leverer, slik at bibliotekene i konsortiet kan tilby sine brukere tjenester av høy kvalitet og tilgang til de ressursene de trenger for læring og forskning, også etter utløp av dagens kontraktperiode.

Dagens tjenestetilbud består hovedsakelig av 3 hovedleveranser:

- Alma biblioteksystem – et moderne biblioteksystem som inneholder alle funksjoner for å drifte et bibliotek
- Oria søkesystem – en felles søkeportal til det samlede materialet som finnes ved de fleste norske fag- og forskningsbibliotek. Oria er basert på Primo fra Ex Libris.
- Leganto pensumlistesystem – et system for opprettelse, vedlikehold og tilgang til pensumlister for faglærere og studenter

I tillegg vurderes det å tilby lenketjenerefunksjonalitet for enklere administrasjon og tilgang til e-ressurser (e-bøker/e-artikler) til institusjoner som ikke har behov for et komplett biblioteksystem.

Utover systemleveransen og forvaltningsoppgaver knyttet til denne, leverer produktområdet veiledning, dokumentasjon og brukerstøtte på systemene, samt leverandøroppfølging og administrasjon av konsortiet.

Det etableres og leveres også verdikjennende integrasjoner og tilleggfunksjonalitet knyttet til systemene tilpasset konsortiets ønsker og behov. Eksempler på dette er felles brukerimport, autoritetsregister, fjernlånsfunksjonalitet m.m.

Organisering

Tjenesteleveransene er organisert som to konsortier; BIBSYS-konsortiet (brukere av Alma og Oria) og Pensumlistekonsortiet (brukere av Leganto). BIBSYS-konsortiet har 78 deltakende institusjoner, mens Pensumlistekonsortiet har 26 institusjoner.

Hvert konsortie har et konsortiestyre med representanter fra deltakerinstitusjonene. Det er også opprettet ulike faggrupper som bistår Sikt med råd og anbefalinger basert på brukerbehov i tjenesteleveranse/-utviklingen. BIBSYS-konsortiet har 4 faggrupper; Fulfillment (utlån/innlån/tilgang), Resource management (katalog), Akkvisisjon (innkjøp) og Oria, mens Pensumlistekonsortiet har 1 faggruppe.

Noen hovedpunkter fra mandatet til faggruppene for Alma:

- Gi råd til Sikt angående oppsett og konfigurasjon, samt vurdere muligheter for standardisering på tvers i konsortiet.
- Behandle saker og utredninger fra Sikt, styret og konsortiet.
- Ha en overordnet tilnærming til ansvarsområdet og ikke kun være fokusert på system eller modul.
- Ta initiativ til og bidra med dokumentasjon, veiledninger og annet informasjonsmateriell.
- Identifisere behovet for kompetanseutvikling hos deltakerinstitusjoner i konsortiet og bidra til gjennomføring av forskjellige arrangement.

Hovedpunkt fra mandatet til faggruppen for Oria:

- Faggruppen skal gi råd til Sikt om oppsett, konfigurasjon og utvikling av ekstra funksjonalitet og tilleggstjenester tilknyttet produktet, slik at konsortiet får et best mulig og hensiktsmessig produkt.

Faggruppene er særskilt viktige for å ivareta brukerperspektivet og sikre at tjenestene er så gode og hensiktsmessige som mulig for de som bruker systemene.



Dagens konsortiestyre for BIBSYS-konsortiet ble oppnevnt av Tjenesteråd for forskning etter innspill fra UHR. Det er styret som oppnevner medlemmer til faggruppene. Det søkes å ha en sammensetning av styret som er representativ for konsortiet.

For Pensumlistekonsortiet velges medlemmer til styre og faggrupper på det årlige konsortiemøtet i oktober.

Anskaffelse

Kontrakten med Ex Libris om leveranse av biblioteksystemet Alma og søkesystemet Oria (Primo) er tidsavgrenset og har en varighet ut 2025. Sikt har sammen med styret for BIBSYS-konsortiet startet en prosess for å danne grunnlag for en ny anskaffelse.

Strategisk har Sikt i samarbeid med styret og konsortiet tilnærmet seg anskaffelsesprosessen med følgende spørsmål:

- Har brukerbehovene endret/utviklet seg siden forrige innkjøpsrunde?
- Har produktene/tjenestene i markedet endret seg siden forrige innkjøpsrunde
- Har markedet endret seg siden siste innkjøpsrunde?
- Hva er trendene og fremtidige brukerbehov?

For å svare ut disse spørsmålene har det de siste årene gjennomført følgende aktiviteter:

- Brukertilfredshetsundersøkelse
- Mikro-strategi
- Scenario-analyse
- Markedsanalyse/Request for information (RFI)

I forlengelsen av innsiktsarbeidet er det etablert et anskaffelsesprosjekt som skal ta arbeidet videre inn mot en anbudsrunde (RFP). Styringsgruppen for prosjektet består av medlemmer i styret for BIBSYS-konsortiet.

Innsynsarbeidet er planlagt avsluttet med en spørreundersøkelse med biblioteklederne og mulig noen flere møter med potensielle leverandører, før fokuset flyttes til utarbeidelse av kravspesifikasjon. Faggruppene vil bli involvert i arbeidet med denne. Målsetningen er å kunngjøre anbudet rundt årsskiftet og gjennomføre anbudsprosess og forhandlinger med leverandørene i H1 2024 og inngå avtale i H2 2024. Tidsplanen er tentativ og kan bli justert.