

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 deltakelse 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 (**MO**dernisering av **DA**ta om det **Norske SA**mfunn).

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 verdøkning 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 \(snl.no\)](#)

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	Sikts forskningsdataplattform.
	Bedriftsstyrearkivet	Sikts forskningsdataplattform.
	Organisasjonsarkivet	Sikts forskningsdataplattform.
	Embetsmannsarkivet	Sikts forskningsdataplattform.
	Manifestoarkivet	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/uXjVOErRkaA=.

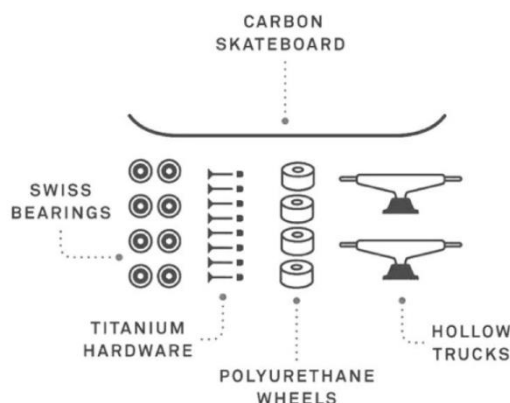
Forprosjektet hadde ikke ressurser eller tid til å intervju brukere fra alle kategoriene eller datatilbudene. Derfor valgte prosjektgruppen å snevre inn kartleggingen til de tre største datatilbudene: Forvaltningsdatabasen, Polsys og Kommunedatabasen.

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 stresselementer 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 stresserelementer. 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 "PoSys". 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.

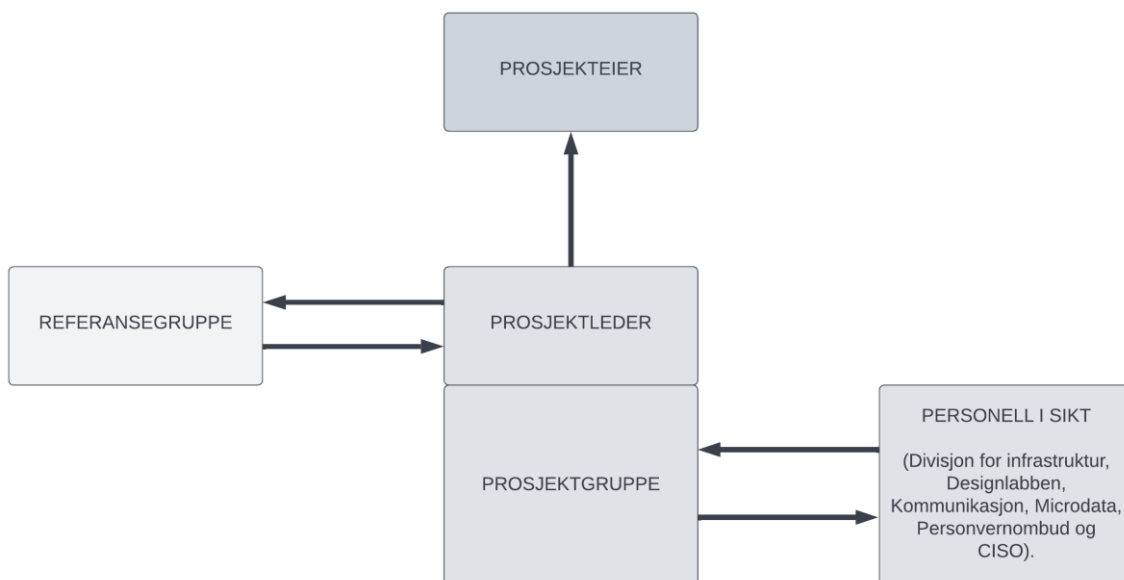
5.4 Organisering og forankring

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 gevinst

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 APler 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 porteføljestyring. 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.