



Fremtidig deling av data i kunnskapssektoren

Konseptutredning





Utarbeidet av: Unit
Kontakt: Per Hovde
Forsidebilde: Shutterstock.com

Innholdsfortegnelse

Sammendrag	7
Formålet med konseptutredningen	7
Anbefalt konsept	7
Utvalgte konsepter	8
Veien videre	9
1. Om konseptutredningen	10
1.1. Bakgrunn	10
1.2. Mandat	10
1.3. Vår tilnærming i utredningen	10
2. Dagens situasjon	13
2.1. Produsentsiden av kunnskapsdata	16
2.2. Konsumentensiden	26
3. Problembeskrivelse	31
3.1. Utfordringer på konsumentensiden	31
3.2. Utfordringer på produsentsiden	35
4. Behovsanalysen	41
4.1. Normative behov	41
4.2. Etterspørselsdrevne behov	45
4.3. Interessentbaserte behov	48
4.4. Prosjektutløsende behov	50
5. Mål for den fremtidige infrastrukturen	52
5.1. Om mål	52
5.2. Samfunns mål	53
5.3. Effektmål	53
5.4. Resultatmål	54
6. Krav til infrastrukturen	56
6.1. MÅ-krav	56
6.2. BØR-krav	58
7. Mulighetsstudie	59
7.1. Mulighetsrommet	59
7.2. Konseptene	61
7.3. Utvelgelse av konsepter	67
8. Alternativanalysen	73
8.1. Beskrivelse av konseptene	73

8.2.	Vurdering av kostnadsvirkninger	94
8.3.	Vurdering av nyttevirkninger	102
8.4.	Andre vurderinger	112
8.5.	Samlet vurdering av konseptene	114
9.	Veien videre	117
9.1.	Tiltak med grenseflater til anbefalt konsept	117
9.2.	Avklaringsfasen	119
9.3.	Organisering av avklaringsfasen og forprosjektet	120
9.4.	Suksessfaktorer og fallgruver.....	120
10.	Vedlegg.....	121
10.1.	Vedlegg A: Oversikt over dataprodusenter og datakilder	121
10.2.	Vedlegg B: Sentrale begreper med forklaring	126
10.3.	Vedlegg C: Oversikt over gjennomførte intervjuer	134
10.4.	Vedlegg D: Prosjektnedbrytningsstruktur	136
10.5.	Konseptuell skisse av anbefalt konsept K12 Orden i kunnskapssektoren.....	137
11.	Referanser	138

Figuroversikt

Figur 1: Funksjonalitet i konsept K12 orden i kunnskapssektoren	8
Figur 2: Strukturen i en konseptvalgutredning	11
Figur 3: Illustrasjon over produsent- og konsumentensiden	13
Figur 4: Illustrasjon på verdikjede for dataproduksjon	15
Figur 5: Illustrasjon av produsentsiden innenfor barnehage, grunnutdanning og videregående skole	16
Figur 6: Illustrasjon av produsentsiden innenfor høyere utdanning	19
Figur 7: Illustrasjon av produsentsiden innenfor voksnes læring	21
Figur 8: Illustrasjon av produsentsiden innenfor forskning	22
Figur 9: Illustrasjon av produsentsiden innenfor integrering	24
Figur 10: Illustrasjon av produsentsiden i hele kunnskapssektoren	25
Figur 11: Konsumentgrupper av kunnskapsdata	26
Figur 12: Konsumentgrupper og hovedformål med databruken	30
Figur 13: Konsumentenes behov	50
Figur 14: Oversikt over problembeskrivelsen og samfunnsvirkninger	51
Figur 15: Oversikt over samfunnsmål og effektmål for fremtidig infrastruktur i kunnskapssektoren	52
Figur 16: Prosjekttrekanten	54
Figur 17: MÅ-krav til løsning for dataflyt i kunnskapssektoren	56
Figur 18: Samhandlingsområder i norsk arkitekturrammeverk for samhandling	57
Figur 19: BØR-krav til løsning for dataflyt i kunnskapssektoren	58
Figur 20: Beskrivelse av prosessen med å utvikle konsepter	59
Figur 21: Mulighetsdimensjoner	60
Figur 22: Alternativer til mulighetsdimensjonen funksjonalitet	60
Figur 23: Oversikt over de ulike konseptene delt inn etter endring som kreves	61
Figur 24: Oppsummering av utvelgelsen av konsepter	72
Figur 25: Skisse over organisering i nullalternativet	75
Figur 26: Konsumentenes brukerreise for å få tilgang til og analysere kunnskapsdata	75
Figur 27: Skisse over organisering i konsept K2	78
Figur 28: Funksjonalitet i konsept K2	79
Figur 29: Skisse over organisering i konsept K10	82
Figur 30: Funksjonalitet for konsument i konsept K10	82
Figur 31: Skisse over organisering i konsept K12	86
Figur 32: Funksjonalitet i konsept K12	86
Figur 33: Skisse over datainfrastrukturen i konsept K14	89
Figur 34: Funksjonalitet for konsument i konsept K14	90
Figur 35: Oversikt over kostnader	94
Figur 36: Periodisering av kostnader i analysen	95
Figur 37: Oppsummering av kostnader knyttet til forprosjekt i mnok	96
Figur 38: Totale investeringskostnader per konsept i mnok	98
Figur 39: Årlige drifts- og forvaltningskostnader per konsept i mnok	99
Figur 40: Vurdering av nytten av konseptene	103
Figur 41: Tabell for pluss-minusmetoden	104

Tabelloversikt

Tabell 1: Samlet vurdering av konseptene i alternativanalysen	8
Tabell 2: Beskrivelse av konsept K0	62
Tabell 3: Beskrivelse av konsept K1 – K5	62
Tabell 4: Beskrivelse av konsept K6 – K8	64
Tabell 5: beskrivelse av konsept K9 – K11	65
Tabell 6: Beskrivelse av konsept K12 – K14	66
Tabell 7: Beskrivelse av konsept K15	67
Tabell 8: Konseptene vurdert mot MÅ-kravene	70
Tabell 9: Evalueringskriterier for utvelgelse av konsepter	71
Tabell 10: Sammenligningstabell for konseptene	92
Tabell 11: Basiskostnader i mnok	99
Tabell 12: Basiskostnad for drift og forvaltning i mnok	100
Tabell 13: Optimistisk kostnadsestimat i mnok	100
Tabell 14: Pessimistisk kostnadsestimat i mnok	101
Tabell 15: Neddiskonterte kostnader i mnok	102
Tabell 16: Vurdering av redusert tidsbruk for konsumentene	106
Tabell 17: Vurdering av redusert tidsbruk for dataproducentene	107
Tabell 18: Vurdering av økt bruk av data	108
Tabell 19: Vurdering av riktig bruk av data	110
Tabell 20: Vurdering av bedre personvern	111
Tabell 21: Samlet oversikt over vurderingene av nytten	112
Tabell 22 Vurdering av gjennomføringsrisiko	113
Tabell 23 Vurdering av konseptenes realopsjoner	114
Tabell 24: Samlet vurdering av konseptene	116
Tabell 25: Oversikt over dataproducenter og datakilder	121
Tabell 26: Sentrale begreper med forklaring	126
Tabell 27: Oversikt over gjennomførte intervjuer	134

Sammendrag

Formålet med konseptutredningen

Kunnskapsdepartementet har ambisjon om at arbeidet i departementet, underliggende direktorater og virksomheter innenfor departementets sektoransvar – utdanning, forskning og integrering – skal være kunnskapsbasert. For å lykkes med denne ambisjonen er det avgjørende med et velfungerende system for kunnskap som sikrer at det utvikles kunnskapsgrunnlag med kvalitet, legitimitet og relevans. En viktig del av et slikt system er en infrastruktur som gir rask, rimelig og sikker tilgang til data i sektoren, hvor data kan kobles på tvers av kilder og med datakilder i andre sektorer.

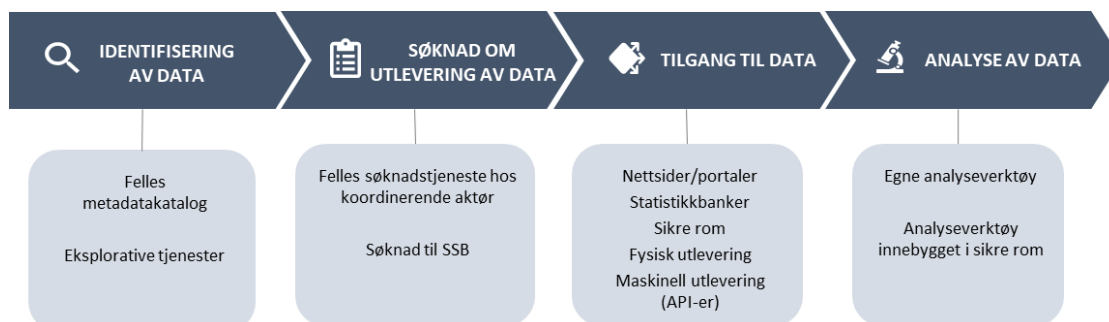
Det er derfor behov for å se nærmere på utfordringer, behov og muligheter for å bedre dataflyten i kunnskapssektoren. Denne konseptutredningen utreder alternative løsninger for fremtidig deling av data i kunnskapssektoren. Formålet med utredningen er å gi et grunnlag for beslutninger om tiltak for å effektivisere infrastruktur for data og statistikk og å gjøre data mer åpne og tilgjengelige. Utredningen tar for seg data for viderebruk, dvs. bruk av data til andre formål enn det dataene ble samlet inn for. I tillegg tar utredningen for seg data fra alle virksomhetsområder Kunnskapsdepartementet er ansvarlig for; barnehage og grunnskole, videregående skoler, høyere utdanning, voksnes læring, forskning og integrering.

Anbefalt konsept

Utredningen anbefaler konsept K12 Orden i kunnskapssektoren. Konseptet innebærer mindre organisatoriske endringer blant aktørene som forvalter dataene i kunnskapssektoren i dag («dataprodusenter»). Dataprodusentene realiserer «orden i eget hus» i henhold til Difis retningslinjer for deling av data innenfor sitt område, noe som gjenspeiles i navnet på konseptet. «Orden i eget hus» er en strategi for å skape god informasjonsforvaltning i offentlig sektor gjennom at den enkelte virksomhet rydder i sine data og beskriver dem med tanke på å dele dem med andre virksomheter. I konseptet realiseres også neste steg i strategien ved at orden skapes, ikke bare isolert i hver enkelt virksomhet, men gjennom en helhetlig tilnærming hvor alle virksomhetene i kunnskapssektoren gjør dette i fellesskap. For å understøtte dette og andre initiativ i sektoren gis en av KDs underliggende etater et overordnet ansvar for samhandling og koordinering på tvers av dataprodusentene for viderebruk av data i sektoren. Aktøren får overordnet ansvar for:

- En helhetlig forvaltning av data i kunnskapssektoren
- Tydeliggjøre roller og ansvar i dataforvaltningen i sektoren
- Følge opp dataprodusentenes arbeid med «orden i eget hus».
- Prosessen knyttet til harmonisering av data og gi retningslinjer for klassifisering av data
- Etablere og forvalte en felles metadatakatalog for all kunnskapsdata for viderebruk
- Felles søknadstjeneste for konsumenter som har behov for tilgang til data med begrenset offentlighet på tvers av dataprodusentene
- Eksplorative tjenester
- Sikre analyserom for analyser av data med begrenset offentlighet
- Etablere et rådgivende forum for juridiske avklaringer knyttet til deling og utlevering av data

Oppsummering av funksjonalitet for konsumentene som tilbys i det anbefalte konseptet er illustrert i Figur 1. Et viktig tiltak er etableringen av en metadatakatalog som vil gi konsumentene oversikt over tilgjengelig data i sektoren. Samtidig vil metadatakatalogen gjøre det enkelt å fange opp inkonsistente begreper og kodeverk knyttet til dataene og iverksette prosesser for å oppnå en begrepharmonisering eller -koordinering på tvers av dataprodusentene.



FIGUR 1: FUNKSJONALITET I KONSEPT K12 ORDEN I KUNNSKAPSSEKTOREN

Det anbefalte konseptet har en forventet kostnad på 105 MNOK. Dette inkluderer kostnader til forprosjekt og investeringer for å gjennomføre konseptet. I tillegg vil det påløpe årlige drifts- og forvaltningskostnader på 12 MNOK. Investeringsperioden er estimert til å være omkring to år. Det anbefales at forprosjekt gjennomføres i perioden 2020-2021, slik at investeringen av valgt konsept starter i 2022 og fullføres i innen utgangen av 2023.

Utvalgte konsepter

I utredningen har fire konsepter vært valgt ut og gjennomgått en samfunnsøkonomisk analyse. Analysen består av å identifisere og vurdere virkningene av de ulike konseptene, hvor virkningene har vært delt i kostnader, nytte og andre virkninger (gjennomføringsrisiko og realopsjoner). Tabell 1 gir en oversikt over vurderingene i analysen hvor det anbefalte konseptet er uthevet.

TABELL 1: SAMLET VURDERING AV KONSEPTENE I ALTERNATIVANALYSEN

	K2 Orden i eget hus	K10 Koordinerte områdesentraler	K12 Orden i kunnskaps- sektoren	K14 Kunnskaps- sentralen
Forprosjekt	5	10	10	15
Investering	50	100	70	125
Skattefinansieringskostnad	10	20	15	30
Sum (*)	65	130	95	170
Årlig drift- og forvaltning	9	8	12	10
Rangering kostnad	1	3	2	4
Redusert tidsbruk for konsument	+	++	++	+++
Redusert tidsbruk for dataproducenter	0	++	+	++
Økt bruk av data	++	+++	+++	+++
Riktig bruk av data	++	+++	+++	++++
Bedre personvern	++	+++	+++	++++
Rangering nytte	4	2	3	1
Gjennomføringsrisiko	Lav risiko	Medium/ høy risiko	Medium risiko	Høy risiko
Realopsjon	Høy	Medium	Høy	Lav
Samlet vurdering	3	2	1	4

*Den samfunnsøkonomiske kostnaden er ikke identisk med kostnad fra kostnadsanalyse da den er neddiskontert, uten mva og arbeidsgiveravgift og inkluderer skattefinansieringskostnad.

I en helhetsvurdering på bakgrunn av kostnad, nytte, gjennomføringsrisiko og realopsjoner vurderes K12 som det beste konseptet. I K12 gjøres det mindre organisatoriske endringer i kunnskapssektoren sammenlignet med K10 og K14, og konseptet har derfor nest lavest samfunnsøkonomisk kostnad av de vurderte konseptene og er vesentlig rimeligere. Samtidig innebærer tiltakene som gjennomføres i K12 at nytten ikke er vesentlig lavere enn i K10 og K14. Årsaken til dette er at K12, i likhet med K10 og K14, inneholder tiltak som metadatakatalog, eksplorative tjenester, felles søknadstjeneste og sikre analyserom for konsumentene av data. I tillegg er gjennomføringsrisikoen av konseptet vurdert til medium og realopsjonen er høy, noe som gir konseptet høy fleksibilitet.

Veien videre

Det er en rekke planlagte eller pågående tiltak i og utenfor sektoren som vil ha tydelige grenseflater til anbefalt konsept K12. Noen av de viktigste tiltakene er (ansvarlig aktør i parentes):

- Videre utvikling av Felles datakatalog (Brønnøysundregistrene frem til 1. januar 2020 da ansvaret for katalogen overtas av det nye Digitaliseringsdirektoratet underlagt Kommunal- og moderniseringsdepartementet)
- Videreutvikling av microdata.no til versjon 2.0 (samarbeid mellom NSD og SSB)
- Prosjektet Deling av data (Difi med involvering av flere offentlige virksomheter)
- Tjenester for sikre analyserom (bl.a. Tjenester for Sensitive Data (TSD) ved UiO og Sikker Adgang til Forskningsdata og E-infrastruktur (SAFE) ved UiB)

Av tiltakene er særlig microdata.no 2.0 og tjenester for sikre analyserom der det forventes betydelige synergier med anbefalt konsept. Det anbefales derfor en avklaringsfase i forkant av forprosjektet for å kartlegge synergier mellom overnevnte tiltak og anbefalt konsept. Det kan være aktuelt å etablere samarbeidsavtaler for å sikre at synergiene blir realisert.

1. Om konseptutredningen

1.1. Bakgrunn

Kunnskapsdepartementet har ambisjon om at arbeidet i departementet, underliggende direktorater og virksomheter innenfor departementets sektoransvar – utdanning, forskning og integrering – skal være kunnskapsbasert. For å lykkes med denne ambisjonen er det avgjørende med et velfungerende system for kunnskap som sikrer at det utvikles kunnskapsgrunnlag med kvalitet, legitimitet og relevans. En viktig del av et slikt system er en infrastruktur som gir rask, rimelig og sikker tilgang til data i sektoren, hvor data kan kobles på tvers av kilder og med datakilder i andre sektorer.

I forbindelse med organisasjonsgjennomgangen av kunnskapssektoren i 2017, med påfølgende omorganisering fra 2018, ble det nedsatt en arbeidsgruppe som fikk i oppdrag å se nærmere på systemet for kunnskap i sektoren og gi råd om videre utvikling av systemet. Arbeidsgruppen bestod av representanter fra Kunnskapsdepartementet (KD) og underliggende virksomheter, og arbeidet resulterte i rapporten «Et velfungerende kunnskapssystem» (Kunnskapsdepartementet, 2018a). Rapporten peker på utfordringer og presenterer forslag til forbedringer knyttet til datainfrastruktur og publisering av data. Ifølge Kunnskapsdepartementet er det likevel behov for å utfylle problembeskrivelsen og behovsanalysen fra rapporten. I tillegg gir ikke rapporten alene et tilstrekkelig grunnlag for å ta beslutninger om konkrete tiltak eller fremme forslag om spesifikke satsinger på området. Rapporten danner imidlertid et godt utgangspunkt for en utredning av utfordringer, behov og muligheter for å bedre dataflyten i kunnskapssektoren.

I november 2018 ble det besluttet at Unit fikk tildelt oppdraget fra Kunnskapsdepartementet med å gjennomføre en forenklet konseptvalgutredning (KVU) for fremtidig infrastruktur for data, analyse og statistikkpublisering om utdanning, forskning og integrering. Utredningen kan ses på som en videreføring av arbeidet til den nevnte arbeidsgruppen.

1.2. Mandat

Mandatet for konseptfasen ble behandlet av Kunnskapsdepartementet i november 2018. Det er definert følgende mandat for den forenklete konseptvalgutredningen:

«Prosjektet skal utrede konsepter for fremtidig deling av data i kunnskapssektoren. Formålet er å gi et grunnlag for beslutninger om tiltak for å effektivisere infrastruktur for data og statistikk og å gjøre data mer åpne og tilgjengelige.»

Mandatet innebærer at utredningen fokuserer på deling av data som oppstår i kunnskapssektoren. Dette betyr at deling av data fra andre sektorer ikke inngår som en del av mandatet. Utredningen vil likevel ta hensyn til samhandling om data med andre sektorer ettersom flere av aktørene i kunnskapssektoren har behov for data fra andre områder enn kunnskapssektoren, samt at andre sektorer har behov for data fra kunnskapssektoren (for eksempel NAV og SSB).

Deling av data i denne utredningen handler om data for viderebruk. Det vil si bruk til andre formål enn det de ble samlet inn for, såkalt sekundærformål. I denne sammenheng er det data som gir fakta og grunnlag for kunnskap om tilstanden innen alle områder Kunnskapsdepartementet er ansvarlig for.

1.3. Vår tilnærming i utredningen

Unit har hatt ansvaret for gjennomføring av utredningen, og den er utarbeidet i perioden desember 2018 til desember 2019. Oppdraget innebærer å utrede behov, identifisere mål og krav, samt beskrive

og vurdere ulike løsninger på de identifiserte behovene som i størst mulig grad innfrir fastsatte mål og krav. Dette er presentert gjennom en forenklet konseptvalgutredning som til slutt presenterer den anbefalte løsningen. Under dette arbeidet har nærmere 40 interessenter, herunder direktorater, fylkeskommuner, kommuner, universiteter og offentlige virksomheter, vært involvert gjennom intervjuer og innspillmøter.

Utredningen omfatter hele kunnskapsdepartementets sektoransvar, dvs.:

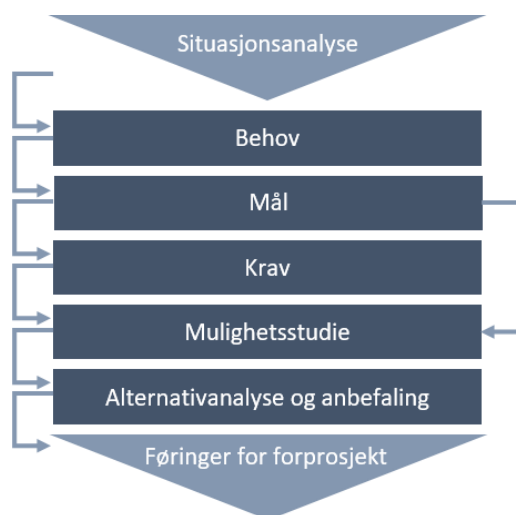
- Barnehage, grunnopplæring og videregående skole
- Høyere utdanning
- Voksnes læring
- Forskning
- Integrering

Utredningen tar ikke for seg presentasjonen av dataene for sluttbrukerne (visualisering), men har fokus på å tilgjengeliggjøre data.

I rapporten benyttes begrepet *kunnskapsdata* om data i kunnskapssektoren. Det innebærer data om utdanning, forskning og integrering, dvs. alle data som produseres innenfor Kunnskapsdepartementets myndighetsområde. Andre sentrale begreper brukt i konseptutredningen er nærmere forklart i begrepslisten i vedlegg B.

Konseptutredningen følger krav, veiledere og god praksis til en konseptvalgutredning i staten så langt det er hensiktsmessig, jf. Finansdepartementets veileder nummer 9 «*Utarbeidelse av KVVU/KL dokumenter*» (Finansdepartementet, 2010).

Figur 2 illustrerer strukturen i arbeidet med konseptvalgutredninger.



FIGUR 2: STRUKTUREN I EN KONSEPTVALGUTREDNING

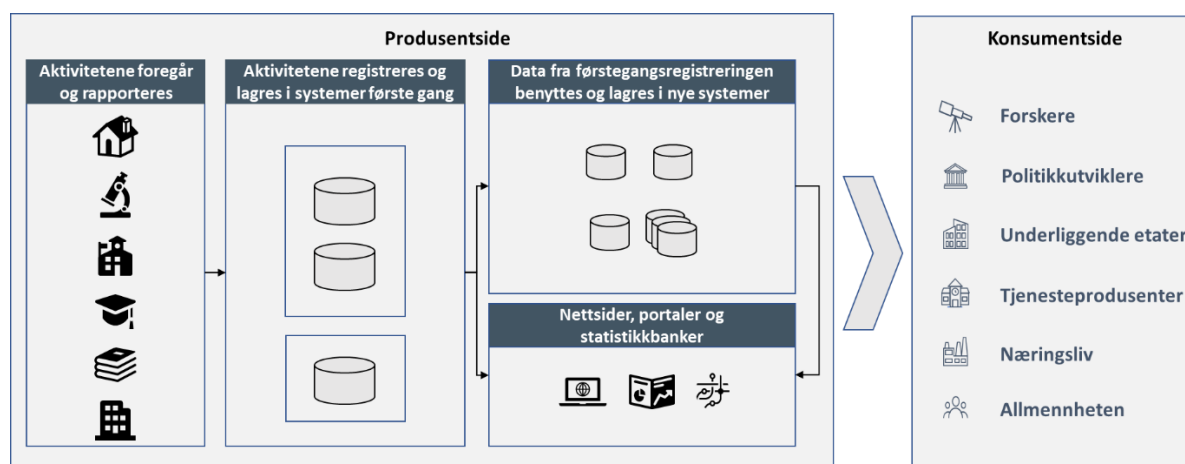
I kapittel 2 beskrives dagens situasjon med deling av data i kunnskapssektoren fra både dataprodusentene og datakonsumentenes side. De identifiserte utfordringene knyttet til dagens situasjon er gjort rede for i kapittel 3 og de reelle behovene for å utvikle en løsning for deling av kunnskapsdata i fremtiden er presentert i kapittel 4.

Med utgangspunkt i problembeskrivelsen og behovene, er det i kapittel 5 og 6 definert hhv. mål for hva som er en ønsket tilstand for den fremtidige løsningen og hvilket krav det skal stilles til løsningen. I kapittel 7 blir selve konseptene for løsningen utarbeidet, og de ulike konseptene gjennomgår en overordnet vurdering («mulighetsstudien»). På bakgrunn av disse vurderingene, blir det valgt ut et begrenset antall konsepter som blir tatt med videre til neste fase i utredningen i kapittel 8 («alternativanalysen»). I alternativanalysen blir de valgte konseptene grundigere analysert og vurdert gjennom en samfunnsøkonomisk analyse, og det anbefales en løsning for den fremtidige delingen av kunnskapsdata. Kapittel 9 tar for seg veien videre og gir føringer for neste steg i prosessen, dvs. forprosjekt, samt hvordan Kunnskapsdepartementet kan realisere det anbefalte konseptet.

2. Dagens situasjon

Dette kapittelet gir en oversikt over dagens situasjon for deling av data i kunnskapssektoren. Beskrivelsen er basert på informasjon fra styrende dokumenter i sektoren, samt intervjuer og innspillseminarer med aktører i og utenfor sektoren. Tabell 27 i vedlegg C gir en oversikt over institusjonene og aktørene som har blitt intervjuet i forbindelse med denne utredningen.

I utredningen skilles det mellom organiseringen av dataene i kunnskapssektoren, omtalt som «produsentsiden», og bruk av dataene, omtalt som «konsumentensiden». Flere av aktørene i kunnskapssektoren er både produsenter og konsumenter. Et eksempel på dette er Direktoratet for internasjonalisering og kvalitetsutvikling i høyere utdanning (DIKU) som bl.a. er produsent av data om studentmobilitet samtidig som de konsumerer data fra bl.a. Unit, NSD/DBH og Statistisk sentralbyrå (SSB). Forholdet mellom produsent- og konsumentensiden er illustrert i Figur 3.



FIGUR 3: ILLUSTRASJON OVER PRODUSENT- OG KONSUMENTSIDEN

Produsentsiden inkluderer alle aktører («produsenter») i kunnskapssektoren som er involvert i produksjonen av data. I kunnskapssektoren skjer dette i all hovedsak der de offentlige tjenestene blir utøvd av både offentlige og private aktører. Dette omfatter blant annet barnehager, skoler, universiteter og asylmottak. Utenfor kunnskapssektoren skjer produksjonen av data hos blant annet private aktører, næringslivet, kommuner og fylkeskommuner. All data som blir produsert blir samlet i et eller flere registre eller databaser.

I Figur 3 er produsentsiden delt mellom førstegangsregistreringen av dataene, og data fra førstegangsregistreringen som benyttes videre og lagres i nye systemer. Flere av produsentene som viderebruker data, kobler disse sammen med data de selv produserer gjennom sin saksbehandling, og skaper nye og unike data som ikke eksisterer i samme form fra før. Et eksempel er Lånekassen som kobler data over studenter med data over vedtak om studielån. Videre er det produsenter som sammenstiller eksisterende data uten tilførsel av nye data. For eksempel kobler SSB data fra kunnskapssektoren med sosioøkonomiske data til analyseformål, uten å tilføre data som ikke allerede er lagret andre steder.

I enkelte tilfeller vil det ikke være klare skiller mellom de ulike stegene i produksjonen av kunnskapsdata og noen av dataprodusentene vil befinne seg i flere av stegene samtidig. Et eksempel er Unit som både lagrer data fra universitetene for første gang i databasen Felles studentsystem (FS) og bruker dataene videre i datavarehuset Star.



De lagrede dataene i de ulike databasene og registrene gjøres tilgjengelig for konsumentene gjennom nettsider og portaler. Spennet på løsningene for dette er stort.

Konsumentsiden omfatter alle brukere som konsumerer kunnskapsdata i og utenfor kunnskapssektoren. I utredningen er konsumentene delt inn i forskere, politikkutviklere, underliggende etater, tjenesteprodusenter, næringsliv og allmennheten. Kapittel 2.2 gir en nærmere beskrivelse av de ulike gruppene og hvilke formål de har med dataene.

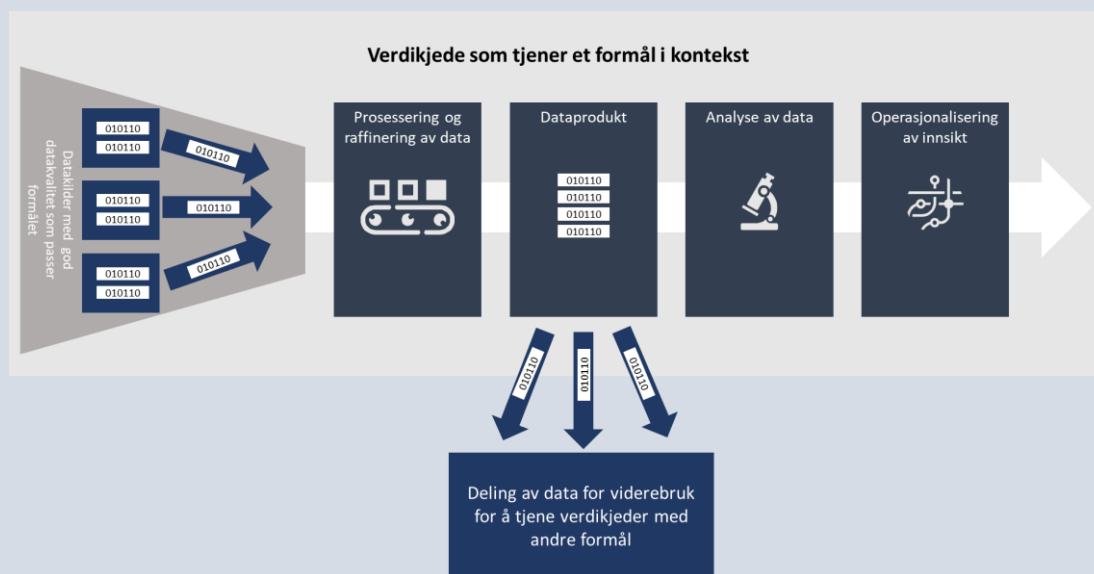
Hele prosessen fra produksjonssiden til konsumentsiden, dvs. fra den unike informasjonen blir skapt til konsumentene har benyttet dataene til sine formål, kan illustreres gjennom en verdikjede for dataproduksjon, jf. boks 1.

BOKS 1 VERDIKJEDE FOR DATAPRODUKSJON OG VIDEREBRUK AV DATA

I Figur 4 er det illustrert en generell verdikjede for dataproduksjon og viderebruk av data. I traktene til venstre i figuren illustreres en innsamling av data fra ulike datakilder som er relevante for aktørens formål. Videre fra traktene så blir disse dataene prosessert og raffinert slik at aktørene sitter igjen med et dataprodukt som konsumenten av dataene kan analysere og hente innsikt fra. Et dataprodukt omfatter både publisert statistikk og data tilrettelagt for analyse gjennom blant annet datavarehus og statistikkportaler.

Verdikjeden synliggjør et sentralt poeng innenfor dataproduksjon: Hvert steg i prosessen påvirker senere steg. For å få til en velfungerende deling av data, må deling derfor være i fokus gjennom hele verdikjeden for både produsent og konsument av data. I tillegg er det nødvendig med god organisering av data, samt tilstrekkelige beskrivelser av dataene og et juridisk rammeverk å forholde seg til. Fokuset på datakvalitet må starte allerede ved innsamlingen av data, blant annet fordi lav kvalitet på de innsamlede data vil få konsekvenser i resten av verdikjeden. Et annet eksempel er at aggregering av data kan fungere i en verdikjede, men ikke dekke behovene for andre verdikjeder som benytter samme data. Database for statistikk om høgre utdanning (DBH) er et eksempel på dette. DBH ble opprinnelig opprettet av Kunnskapsdepartementet for rapportering til departementet om tilstanden i UH-sektoren gjennom en verdikjede dedikert for dette formålet. Utdanningsinstitusjonene har også tilgang til data fra DBH for sin egen virksomhet. Men da dataene er som publiseres er aggregert på et nivå som ikke dekker institusjonenes behov for detaljinformasjon, må de dette innhentes via andre verdikjeder.

Hvis man ser verdikjeden i sammenheng med illustrasjonen over dagens situasjon i kunnskapssektoren, jf. Illustrasjon over produsent- og konsumentens side Figur 3, vil produsentsiden ligge til venstre i Figur 4 frem til og med dataprodukt, mens konsumentens side vil inkludere dataprodukt, analyse av data og operasjonalisering av innsikt. Denne utredningen tar for seg verdikjeden til og med analyse av data, som innebærer at operasjonalisering av innsikt er utenfor omfanget til utredningen.



FIGUR 4: ILLUSTRASJON PÅ VERDIKJEDE FOR DATAPRODUKSJON

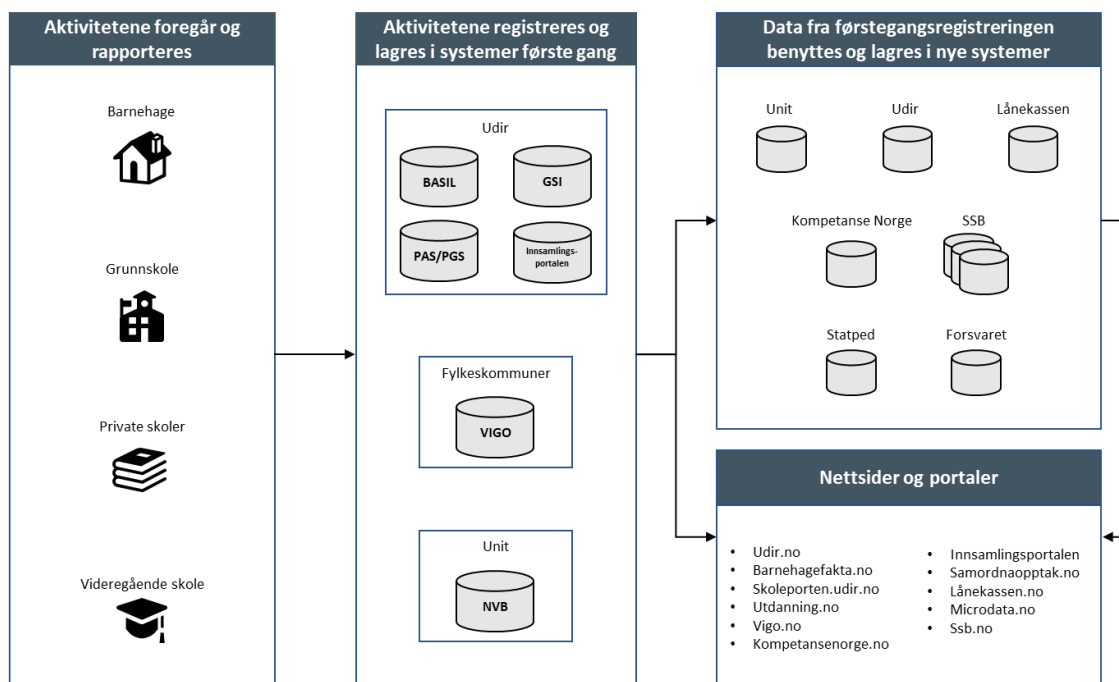
2.1. Produsentsiden av kunnskapsdata

Beskrivelsen av produsentsiden er delt inn i områdene som omfattes av Kunnskapsdepartementets sektoransvar:

- Barnehage, grunnskole og videregående skole
- Høyere utdanning
- Voksnes læring
- Forskning
- Integrering

2.1.1. Barnehage, grunnskole og videregående skole

Innenfor dette området driftes barnehager i dag av både kommunale og private aktører. Opplæring på grunnskolenivå (barne- og ungdomskoler) er et kommunalt ansvar, mens fylkeskommunene har ansvaret for videregående opplæring. Private skoler med godkjenning tilbyr også grunnopplæring og videregående opplæring. Dataflyten for området er illustrert i Figur 5.



FIGUR 5: ILLUSTRASJON AV PRODUSENTSIDEN INNENFOR BARNEHAGE, GRUNNUTDANNING OG VIDERE GÅENDE SKOLE

I dag rapporterer barnehage, grunnskole og private skoler inn til Utdanningsdirektoratet (Udir) sine ulike registre. Videregående skoler rapporterer i dag via det fylkeskommunale systemet VIGO. UBAS er Udir sitt brukeradministrasjonssystem for pålogging der eksterne konsumenter logger seg inn for å levere data eller se aggregerte data Udir har tilgjengeliggjort.

Barnehager rapporterer via rapporteringsløsningen BASIL til Udir. Barnehagen fyller ut skjema manuelt og sender inn et elektronisk resultatregnskap til sin kommune. Kommunen er ansvarlig for å gi supplerende opplysninger, før all informasjon sendes videre til Udir. Data rapporteres ikke på individnivå, og det laveste nivået på data er derfor hver enkelt barnehage. Innsamlingen foregår årlig. Data fra BASIL gjøres tilgjengelig på nettsiden barnehagefakta.no og på udir.no/statistikk. Hver enkelt



barnehage eller kommune kan få tilgang til BASIL sin rapportportal via UBAS, der er det mulig for konsumenten å sammenstille informasjon som blir registrert i BASIL.

Grunnskolen informasjonssystem, kjent som GSI, er systemet for registrering av opplysninger om grunnskolen i Norge. Alle barne- og ungdomsskoler må rapportere til systemet som driftes og forvaltes av Udir. I dag rapporteres det inn omkring 1000 opplysninger om hver eneste grunnskole. Data inkluderer elevtall, årstimer, ressurser, spesialundervisning, språklige minoriteter, målform, fremmedspråk, fysisk aktivitet, leksehjelp, SFO, valgfag og PPT (Utdanningsdirektoratet, u.d.). Dataene rapporteres på aggregert nivå. Data fra GSI rapporteres videre til blant annet SSB. Statped benytter GSI som kunnskapskilde (manuelt) for sin virksomhet.

Data om elever i videregående opplæring rapporteres inn til det fylkeskommunale systemet VIGO gjennom ulike skoleadministrative systemer (iSkole, SATS, Extens m.m.). VIGO forvaltes av det interkommunale selskapet VIGO IKS, som er ansvarlig for utvikling og forvaltning av fylkeskommunenes felles IT-systemer innen videregående opplæring. VIGO er delt inn i ulike databaser som VIGO Voksen, VIGO OT og VIGO Koderverksdatabase. Databasene inneholder data om videregående skoleopplæring inkl. fagopplæring, samt grunnskolen 10. trinn. Dataene i VIGO er på individnivå, og omfatter bl.a. karakterer, fravær, fag og programområder. Data fra VIGO rapporteres videre til Udir, SSB, Lånekassen og Forsvaret. VIGO er den største eksterne datakilden Udir benytter.

Det arbeides i dag med et nytt felles skoleadministrativt system (Visma in School) for alle offentlige videregående skoler. Systemet er innført i Akershus, men er allerede rapportert å være forsinket (Braathen, 2019). Systemet vil være en modernisering av dagens ulike system, og gi en mer automatisert og effektiv arbeidsflyt. Systemet vil også gjøre det enklere for VIGO å motta data, da det bare er et felles system å forholde seg til.

I systemene PAS (PrøveAdministrasjonsSystem) og PGS (PrøveGjennomføringsSystem) samles data fra ulike type prøver og eksamener på ungdomsskole og videregående nivå. Systemet administreres av Udir. Udir samler i tillegg data fra brukerundersøkelser som elevundersøkelsen i grunnskolen og foreldreundersøkelsen i barnehagene.

Frittstående videregående skoler rapporterer blant annet elevtall, programområder, fag og karakterer via Innsamlingsportalen. Portalen er en web-portal for innrapportering til Udir. Dataene sendes videre til VIGOs sentralbase, som dermed samler data om all ungdom med rett til videregående opplæring.

De fleste videregående skoler overfører elektroniske vitnemål for sine avgangselever til Nasjonal vitnemålsdatabase (NVB) som forvaltes av Unit. Vitnemålene brukes hovedsakelig til kontroll og poengberegning i opptaket til høyere utdanning gjennom Samordna opptak.

Data fra de ulike systemene tilgjengeliggjøres for konsumentene på en rekke ulike nettsider og portaler. Hver dataprodusent har i mange tilfeller sin egen nettside eller portal der de tilbyr konsumentene data fra sine registre. Eksempelvis publiserer både Udir, Lånekassen og Kompetanse Norge data fra sine registre på egne nettsider. Når det gjelder utlevering av data på forespørsel, utleveres disse fra dataprodusentene i hovedsak som flatfiler. Før utlevering gjør produsentene som behandlingsansvarlige, en vurdering av om det foreligger et rettslig grunnlag for utleveringen er innenfor gjeldende regelverk. SSB sammenstiller også data fra de ulike registrene og publiserer mye av dette i sin statistikkbank (ssb.no). En del data tilgjengeliggjøres også til forskere gjennom løsningen microdata.no, jf. nærmere omtale i boks 2.

BOKS 2 MICRODATA.NO

Microdata.no er en tjeneste som samler og sammenstiller data fra kunnskapssektoren og andre sektorer. Microdata.no er et samarbeid mellom NSD og SSB, og gir konsumenter tilgang på registerdata fra SSB uten tid- og kostnadskrevende søknadsprosesser og lang behandlingstid for utlevering av data. I dag har forskere og studenter ved godkjente forskningsinstitusjoner tilgang til å bruke tjenesten.

For å løse utfordringer knyttet til personvern, tilbyr tjenesten bruk av ikke-anonymiserte data gjennom en konfidensialitetssikrende plattform. Dette gjør at brukerne kun er i indirekte kontakt med personopplysninger, og at data ikke forlater SSB, i tillegg til at plattformen sørger for at all output er anonym.

Begrensninger ved løsningen er at konsumentene ikke får se eller laste ned data direkte. Konsumentene kan heller ikke koble dataene mot egne data, og de må benytte microdata.no sin innebygde statistikkprogramvare for analyser.

Fra 2020 får NSD og SSB bevilget 42 millioner kroner fra Norges forskningsråd for å videreutvikle microdata.no-løsningen (microdata.no 2.0) i perioden 2020-2023. I tillegg bidrar SSB med seks millioner kroner i prosjektet. Det planlegges flere utvidelser av tjenesten:

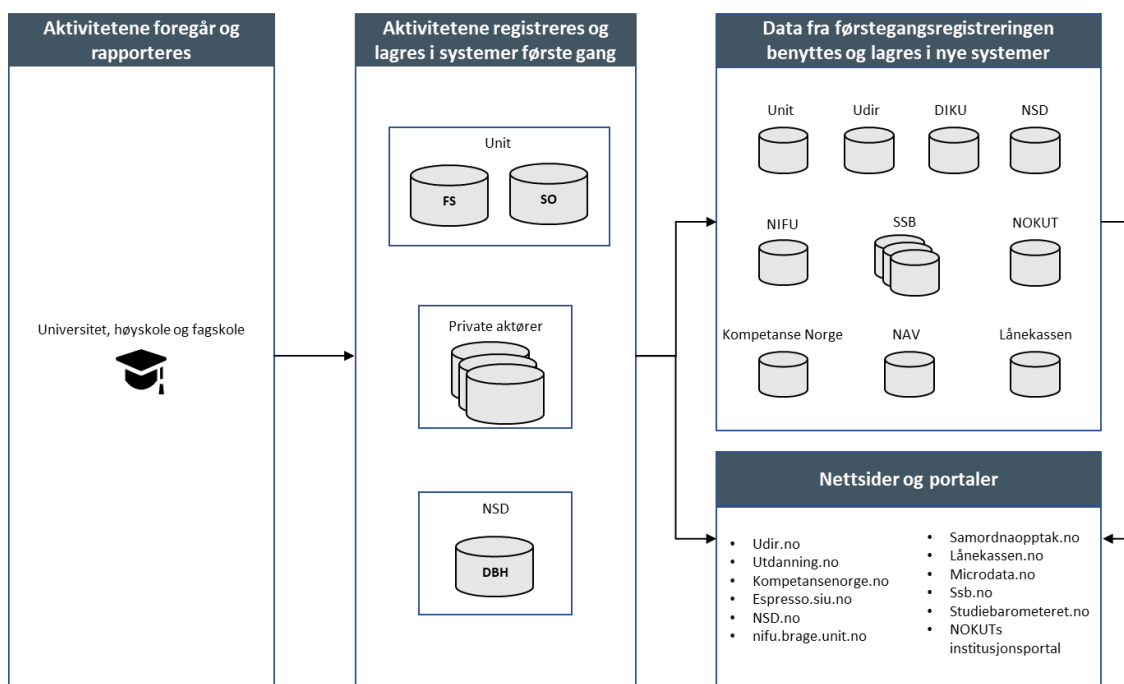
- Økt tilfang av registerdata inn i løsningen (fra SSB, NSD og Kreftregisteret)
- Data fra andre registreiere skal kunne gjøres tilgjengelig for analyse i tjenesten
- Internasjonalisering (oversatte metadata og internasjonal innlogging). Internasjonal tilpasning legger til rette for at metadata kan oversettes til engelsk, men prosjektet omfatter ikke oversettingsarbeid
- Digitalisert samspill mellom microdata.no og tradisjonell/fysisk utlevering av data
- Utvikling av metoder og visualiseringer
- Løsning for at forskere kan laste opp egne data for kobling med registerdata

I begynnelsen av 2020 starter arbeidet med Microdata.no 2.0. Prosjektet har en varighet på fire år, dvs. innenfor samme tidsperioden som anbefalt konsept planlegges for. Prosjektet planlegger løsninger som forenkler og automatiserer tilrettelegging og innlasting av nye data. Omfanget av data vil gradvis økes, med første utvidelse ved utløpet av 1. kvartal 2020 hvor det også vil bli lagt inn noen flere variabler fra utdanningssektoren.

Det er ingen hindre i lovverk for å åpne for nye konsumentgrupper så lenge personvern hensynet fortsatt ivaretas i løsningen. Prioriteringen av nye konsumentgrupper er mer et praktisk spørsmål, hvor samarbeid om nye data og nye konsumenter vil drives av behov og interesse, men fortsatt i kontrollerte former. Det er et mål i prosjektet at microdata.no også skal være en løsning for at andre register-/dataeiere skal kunne gjøre dataene tilgjengelig for kobling og analyse i tjenesten. Kreftregisteret inngår som samarbeidspartner for å utvikle konsepter og løsninger for dette formålet. Per i dag er det vanskelig for prosjektet å forskutere når en slik tjeneste vil være moden nok for utrulling..

2.1.2. Høyere utdanning

Høyere utdanning omfatter universiteter, høyskoler og fagskoler i Norge (UH-sektoren). Data knyttet til forskning på universiteter og høyskoler blir omtalt i kapittel 2.1.3, og dataflyten innen høyere utdanning er illustrert i Figur 6.



FIGUR 6: ILLUSTRASJON AV PRODUSENTSIDEN INNENFOR HØYERE UTDANNING

De fleste universiteter og høyskoler og noen fagskoler i Norge benytter det studieadministrative systemet Felles studentsystem (FS) som forvaltes av Unit. I dag tilsvares det 34 utdanningsinstitusjoner, mens enkelte andre skoler bruker studieadministrative system fra private aktører, f.eks. bruker Handelshøyskolen BI systemet Banner. FS inneholder informasjon fordelt på 14 ulike moduler som inkluderer blant annet informasjon om opptak, undervisning, vurderinger, betaling og kvalifikasjoner. I tilknytning til FS ligger tjenesten Samordna opptak som håndterer opptak til høyere utdanning. Samordna opptak ble også omtalt under området «Barnehage, grunnutdanning og videregående skole».

De fleste universiteter, høyskoler og fagskoler rapporterer data til DBH, som driftes og forvaltes av NSD.¹ Dataene om regnskap og personal kommer fra institusjonenes personal- og økonomisystemer og overføres til DBH via «DBH datalevering», mens studiedata overføres til DBH gjennom FS. Fagskoler rapporterer direkte til DBH. Totalt innrapporteres omtrent 490 variabler fra UH-institusjonene og 90 variabler fra fagskolene, samt administrative variabler som NSD legger til for å strukturere dataene.

I dag benytter en rekke ulike aktører data fra UH-sektoren videre i sine systemer og i sitt arbeid:

¹ Det er kun to høyskoler som ikke rapporterer til DBH per i dag (Noroff og Bjørknes høyskole). Private institusjoner rapporterer også til DBH, og i 2020 vil de resterende private fagskolene rapportere til DBH.

- *STAR*: Datavarehus som samler studiedata (som eksamensdata og gjennomstrømming av studenter) fra FS for analyseformål og rapportering. I tillegg kobler STAR data fra studiebarometeret fra NOKUT og Samordna opptak. STAR driftes av Unit.
- *Lånekassen*: Mottar data fra blant andre Folkeregisteret, FS, DBH/DBH-F, VIGO, UDI og NAV (AA-registeret) om studentstatus og eksamensresultat på individnivå. Disse dataene kobles sammen med data over vedtak om studielån. Lånekassen leverer data om lån og stipend til studenter direkte til SSB. Lånekassen leverer også data brukt i saksbehandling til NAV, Utlendingsdirektoratet (UDI) og Helseøkonomiforvaltningen (Helfo).
- *DIKU*: Mottar bakgrunnsinformasjon om studenter og utvekslingsdata på individnivå for statistikkformål. Kilder for dette er blant andre FS, DBH, Lånekassen, Studiebarometeret (NOKUT) og SSB. Espresso er DIKUs søknads- og rapporteringsverktøy hvor det blant annet er mulig for universiteter og høyskoler å søke om støtte til enkelte prosjekter.
- *Nasjonalt organ for kvalitet i utdanningen (NOKUT)*: Mottar bakgrunnsinformasjon om studenter og kunnskapsdata som danner listegrunnlag og populasjon for den nasjonale, årlige studentundersøkelsen som henter inn data til Studiebarometeret. NOKUT mottar data om indikatorer fra DBH og fra Unit ifm. gjennomføring av Studiebarometeret og Underviserundersøkelsen. NOKUT leverer grunnlagsdataene fra undersøkelsene til arkivering i NSD og til UH-institusjonene. NOKUT leverer også data om godkjenning av utenlandsk utdanning hos enkeltpersoner til SSB.
- *SSB*: Mottar data fra DBH og Samordna opptak, og kobler disse sammen med andre sosioøkonomiske data for statistikkproduksjon og analyser.
- *Nordisk institutt for studier av innovasjon, forskning og utdanning (NIFU)*: Mottar data fra DBH bl.a. til gjennomføring av kandidatundersøkelsene og til statistikkproduksjon.

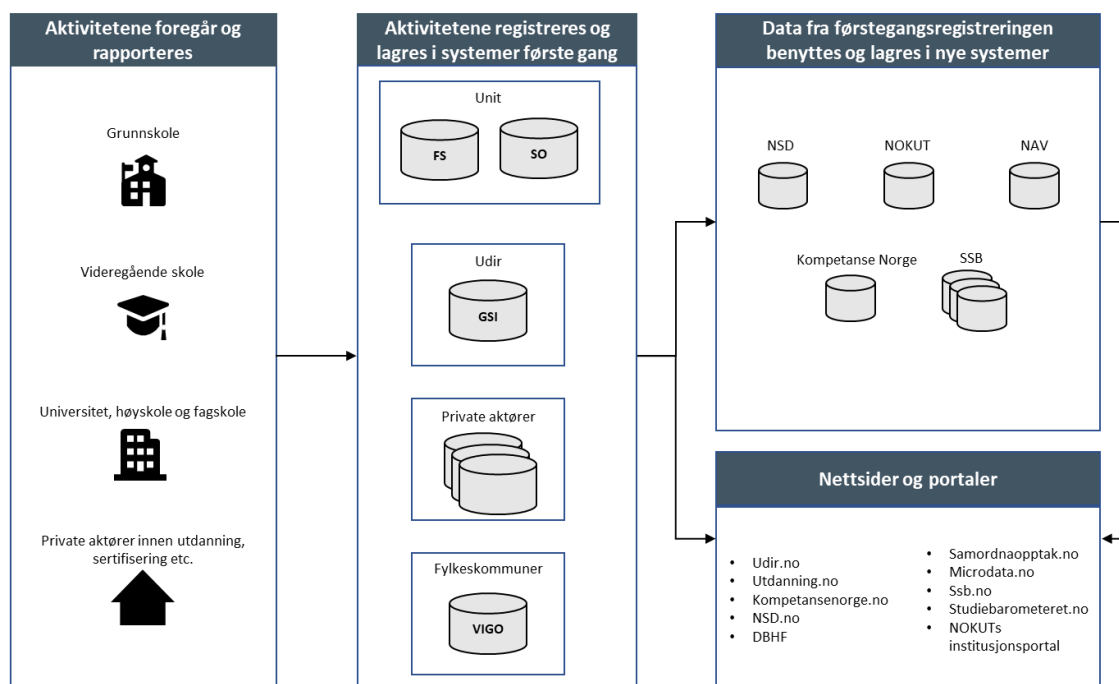
De senere årene er det blitt gjort enkelte grep for å forenkle dataflyten og unngå dobbeltrapportering innenfor området høyere utdanning. Tidligere gikk blant annet informasjon direkte fra institusjonene til NIFU og SSB også, slik at det var flere registre for «førstegangsregistrering» av de samme type data. I tillegg har aktørene arbeidet med å samordne begreper og innføre felles standarder. Blant annet satte KD i 2017 ned en arbeidsgruppe som fikk i oppgave å se på begrepsbruk, definisjoner og tellemåter i statistikk for høyere utdanning og fagskoleutdanning med sikte på mer enhetlig statistikk.

Som i området barnehage, grunnutdanning og videregående utdanning, tilgjengeliggjøres dataene i de ulike systemene for konsumentene på en rekke ulike nettsider og portaler, jf. eksempler i Figur 6.

2.1.3. Voksnes læring

Voksnes læring er et område som omfatter alle former for læring etter at man er ferdig med skolegangen, og eventuelt har tilegnet seg en grad fra en høyskole, fagskole eller universitet. Dette innebærer at organisert læring utenfor det formelle utdanningssystemet som kurs, organisert utdanning i organisasjonslivet og frivillig arbeid «gjennom livet» er inkludert her.

Figur 7 viser en illustrasjon av dataflyten innenfor området.



FIGUR 7: ILLUSTRASJON AV PRODUSENTSIDEN INNENFOR VOKSNES LÆRING

Dataflyten innenfor området er i dag preget av at informasjon er spredt hos en rekke ulike aktører. I dag eksisterer det ikke et sentralt register for førstegangsregistrering, men Kompetanse Norge samler inn mye data om voksnes læring fra andre aktører (Kompetanse Norge, 2019). Data som samles inn hos Kompetanse Norge omfatter blant annet data fra FS, DBH, GSI, VIGO og SSB. Kompetanse Norge har ansvar for Kompetansepluss², norskopplæring, norsk- og samfunnskunnskapsprøver for voksne innvandrere, og statsborgerprøven. Kompetanse Norge produserer statistikk over norsk- og samfunnskunnskapsprøver og statsborgerprøven. Disse benyttes hovedsakelig av Integrerings- og mangfoldsdirektoratet (IMDi) og SSB.

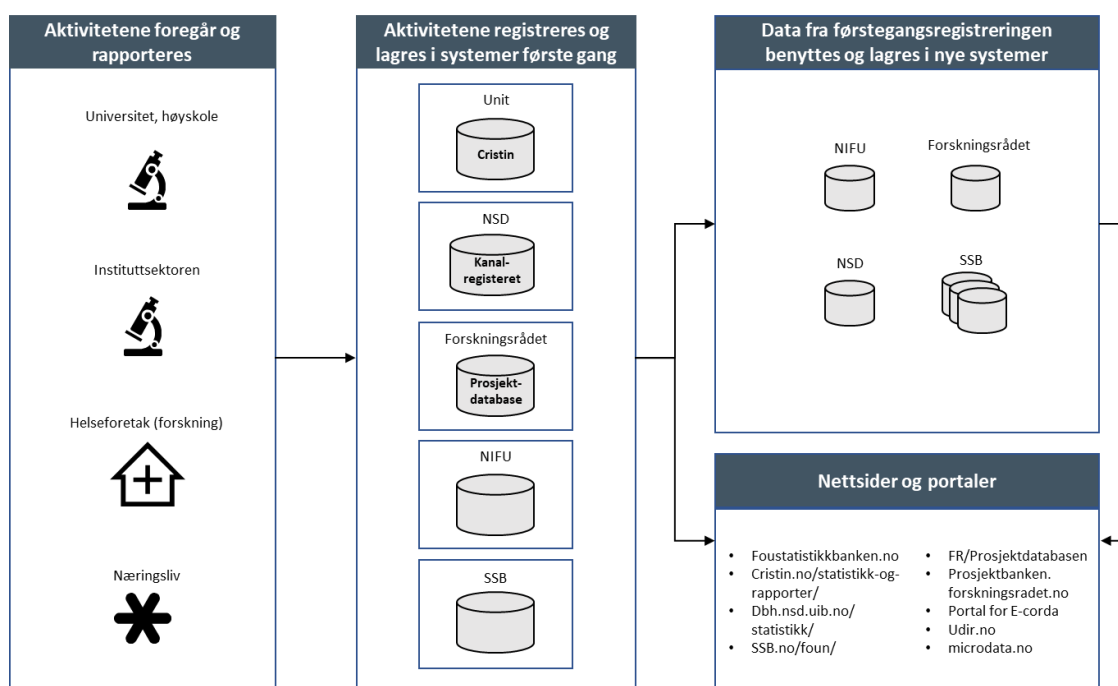
I Norge finnes det ikke registerdata om deltakelse i ikke-formell opplæring som for eksempel private aktører med kursvirksomhet. Dataene er derfor begrenset til data fra undersøkelser, blant annet fra Lærevilkårsmonitoren som gjennomføres en gang i året som en del av SSBs arbeidskraftsundersøkelse. I tillegg deltar Norge i internasjonale undersøkelser som Adult Education Survey (AES), Continuing Vocational Training Survey (CVTS) og PIAAC som kartlegger voksnes deltagelse i formell og ikke-formell opplæring. AES og CVTS gjennomføres hvert femte år.

Flere større aktører som tilbyr kurs og etterutdanning rettet mot bedriftsmarkedet har register på gjennomførte kurs og sertifiseringer blant sine kursdeltakere, men dette rapporteres ikke til et sentralt register.

² Kompetansepluss er en tilskuddsordning som skal bidra til at flere voksne, gjennom opplæring i grunnleggende ferdigheter, kan ta del i videre opplæring og utdanning og styrke sin deltagelse i arbeids- og samfunnsliv.

2.1.4. Forskning

Område forskning omfatter forskningsaktiviteter som utføres av universiteter, høyskoler, helseforetak, instituttsektoren og næringslivet. Utredningen tar for seg datainfrastrukturen knyttet til data om forskningsaktiviteter og ikke data som benyttes i selve forskning.³ De sentrale aktørene for data om forskningsaktiviteter er Unit, NSD, Forskningsrådet, NIFU og SSB. Dataflyten er illustrert i Figur 8.



FIGUR 8: ILLUSTRASJON AV PRODUSENTSIDEN INNENFOR FORSKNING

Data som skapes gjennom forskningsaktivitetene gir grunnlag for kunnskap om forskning gjennom:

- Administrative datasystemer ved de utførende institusjonene (opplysninger om prosjekter, personell og utgifter).
- Administrative datasystemer i underliggende etater som gir tilskudd til finansiering av forskning m.m. (som Forskningsrådet, m.fl.).
- Informasjon innhentet for statistiske formål både om forskningen og aktørene som er involvert (personer, institusjoner og foretak).
- Informasjon som produseres som ledd i forskningen og relaterte aktiviteter (vitenskapelige publikasjoner, forskningsdata, samarbeidsplattformer, patenter, m.m.).

Data knyttet til vitenskapelige publikasjoner og andre relevante resultater fra forskning samles i det nasjonale forskningsinformasjonssystemet Cristin (Current Research Information System in Norway). Systemet inneholder blant annet forskningspublikasjoner, prosjekter, forskningsenheter og årsrapportering. Cristin samler i dag informasjon om forskningsaktiviteter ved rundt 140 norske medlemsinstitusjoner som inkluderer både universiteter, høyskoler, helseforetak og ulike forskningsinstitutt. Cristin driftes og forvaltes av Unit.

³ Forskningsdata som benyttes i selve forskningen, dekkes gjennom andre initiativ knyttet til åpen forskning, som blant annet Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata (Kunnskapsdepartementet, 2017d).

De ulike forskningsinstitusjonene innenfor UH-, institutt- og helsesektoren kontrollerer og godkjenner sin rapportering av vitenskapelige publikasjoner i Cristin. Etter at institusjonene er ferdig med sin rapportering overfører Cristin dataene videre til DBH, som årlig publiserer den offisielle statistikken for UH-sektoren som inngår i den resultatbaserte finansieringen av UH-institusjonene. DBH driftes og forvaltes av NSD. Cristin utveksler informasjon med systemene til Forskningsrådet, Regionale komitéer for medisinsk og helsefaglig forskningsetikk (REK) og NSD.

Cristin samler data og metadata (informasjon om dataene) om norsk forskning til bruk i rapportering av vitenskapelig produksjon, kjent som NVI-rapporteringen. Norsk vitenskapsindeks (NVI) er en sentral del av finansieringen for institusjoner som mottar deler av sin finansiering gjennom resultatbasert omfordeling (RBO). Informasjon om publikasjoner som gir uttelling i RBO, blir rapportert til Kunnskapsdepartementet og Helse- og omsorgsdepartementet årlig for å bestemme fordeling av finansieringen. Forskningsrådet benytter nøkkeltall om instituttsektoren som er samlet inn av NIFU, som grunnlag for resultatbasert omfordeling av basisbevilgningen til forskningsinstitutter.

Kanalregisteret hos NSD gir oversikt over vitenskapelige publiseringskanaler, samt publikasjonskanaler som ikke er godkjent i rapporteringssystemet. NSD har ansvar for drift og vedlikehold av registeret, mens Det nasjonale publiseringsutvalget (NPU) har ansvar for godkjenningen av selve publiseringskanalene. Vedtak om publiseringskanalene blir registrert i kanalregisteret og gjøres offentlig tilgjengelig, i tillegg til å bli sendt til Cristin. NSD tilbyr et API for tilgjengeliggjøring av opplysninger om publiseringskanaler som Cristin benytter for å oppdatere sine data.

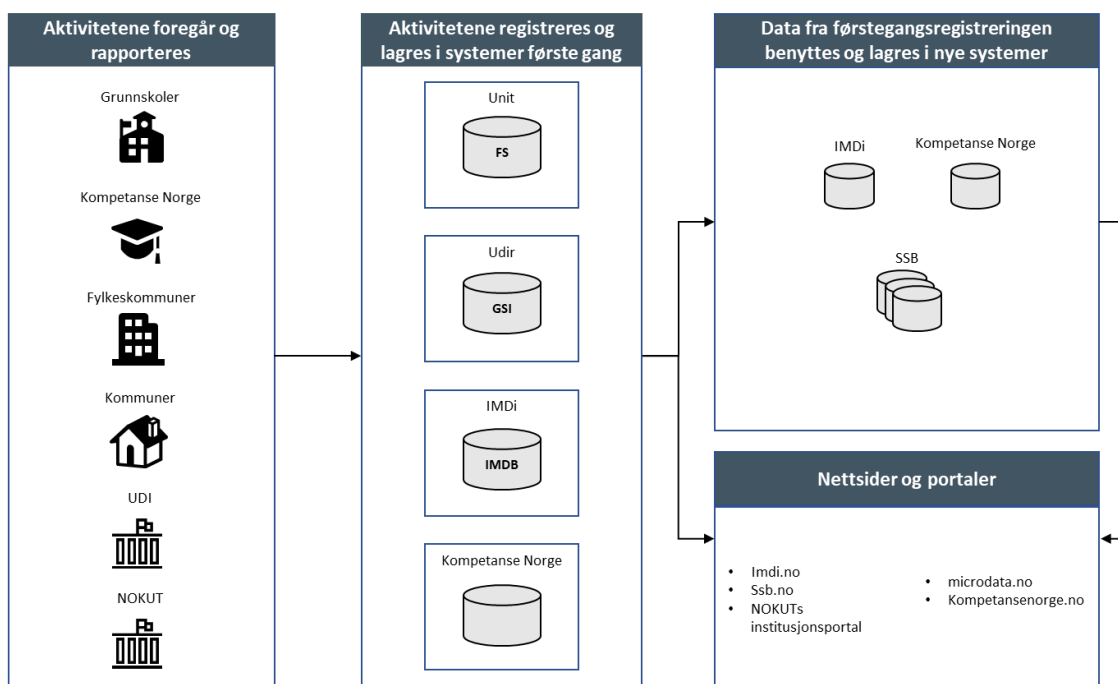
Forskningsrådets prosjektdatabase inneholder data om søknader om og prosjekter med bevilget støtte fra Forskningsrådet. For prosjekter som får støtte lagres det på et senere tidspunkt ytterligere data som prosjektdeltagere, evaluering, tema for forskningen og resultater av forskningen. Prosjektdatabasen gjenbruker data fra eksterne kilder som Cristin-databasen og NIFUs institusjonsregister, samt data fra Brønnøysundregistret.

I dag produserer SSB FoU-statistikk for næringslivet og NIFU for de øvrige sektorene. FoU-statistikk for næringslivet er i stor grad basert på spørreundersøkelser. De siste årene er undersøkelsene i UH-sektoren betydelig forenklet ved at NIFU får økonomi- og personelldata via DBH, slik at instituttene bare behøver å oppgi den informasjon som ikke allerede er tilgjengelig. Data om forskning tilgjengeliggjøres i de ulike systemene for konsumentene på en rekke ulike nettsider og portaler, jf. eksempler i Figur 8. Viktige kilder for data om forskning er DUCT (Cristin) og NIFUs FOU-statistikk som begge tilbyr analyseverktøyet Tableau for å gjøre analyser og hente ut rapporter.

2.1.5. Integrering

Området inkluderer data om integrering som utføres av blant annet Kompetanse Norge, grunnskoler, kommuner, fylkeskommuner, NOKUT og UDI. Den sentrale aktøren innenfor dette området er Integrerings- og mangfoldsdirektoratet som er ansvarlig for databasen IMDB og plattformen IMDinett. IMDinett er IMDIs digitale plattform for applikasjonene Nasjonalt introduksjonsregister (NIR), modul for bosetting og kompetansekartlegging samt tilskudd til norskopplæring. NIR blir benyttet av kommunene for å registrere informasjon om personer omfattet av introduksjonsloven. Dette omfatter opplæring i norsk og samfunnskunnskap for nyankomne innvandrere, introduksjonsprogram for flyktninger og opplæring i norsk og norsk kultur og norske verdier for asylsøkere.

Figur 9 viser en skisse over dataflyten innenfor integrering.



FIGUR 9: ILLUSTRASJON AV PRODUSENTSIDEN INNENFOR INTEGRERING

IMDi er et fagdirektorat, forvaltningsorgan og nasjonalt kompetansesenter med ansvar for integreringsfeltet. Direktoratet skal iverksette regjeringens integreringspolitikk. IMDi skal styrke kommunenes, sektormyndighetenes og andre samarbeidspartneres kompetanse på integrering. I mai 2018 ble ansvaret for styring av dette direktoratet overført fra Justis- og beredskapsdepartementet til Kunnskapsdepartementet. Dette var den fjerde flyttingen av direktoratet mellom ulike departement siden det ble opprettet i 2006 gjennom utskilling fra UDI. IMDi samarbeider med kommuner om bosetting og norskopplæring, og samarbeider med UDI om grunnleggende kvalifisering, gjennom den kompetansekartlegging av flyktninger som skjer i mottak.

IMDis database IMDB håndterer blant annet data knyttet til bosetting, tilskudd til norskopplæring og kompetanseregistrering i mottak for innvandrere, i tillegg til data fra NIR om rettigheter og plikter i henhold til introduksjonsloven. Informasjonen blir innhentet fra UDI, kommuner og fylkeskommuner, Det Sentrale Folkeregisteret og Kompetanse Norge. Kommunene er lovpålagt å registrere deltakelse i ordningene under introduksjonsloven i NIR. Dataene i IMDinett brukes blant annet i kartleggingsfasen av den enkelte flyktning, og benyttes videre som datagrunnlag for SSB's publisering av statistikk og analyse om deltakelse i ordningene under introduksjonsloven. Dataene IMDi samler inn brukes blant annet i kartleggingsfasen av den enkelte flyktning og for å finne en egnet bosettingskommune, og benyttes videre av SSB som datagrunnlag for deltakelse i introduksjonsordningen.

Data om barn med innvandrerbakgrunn i grunnskolen blir registrert i Udirs system GSI, på lik linje som informasjon om andre barn i grunnskoleopplæring. Fra GSI blir data overført videre til IMDis registre.

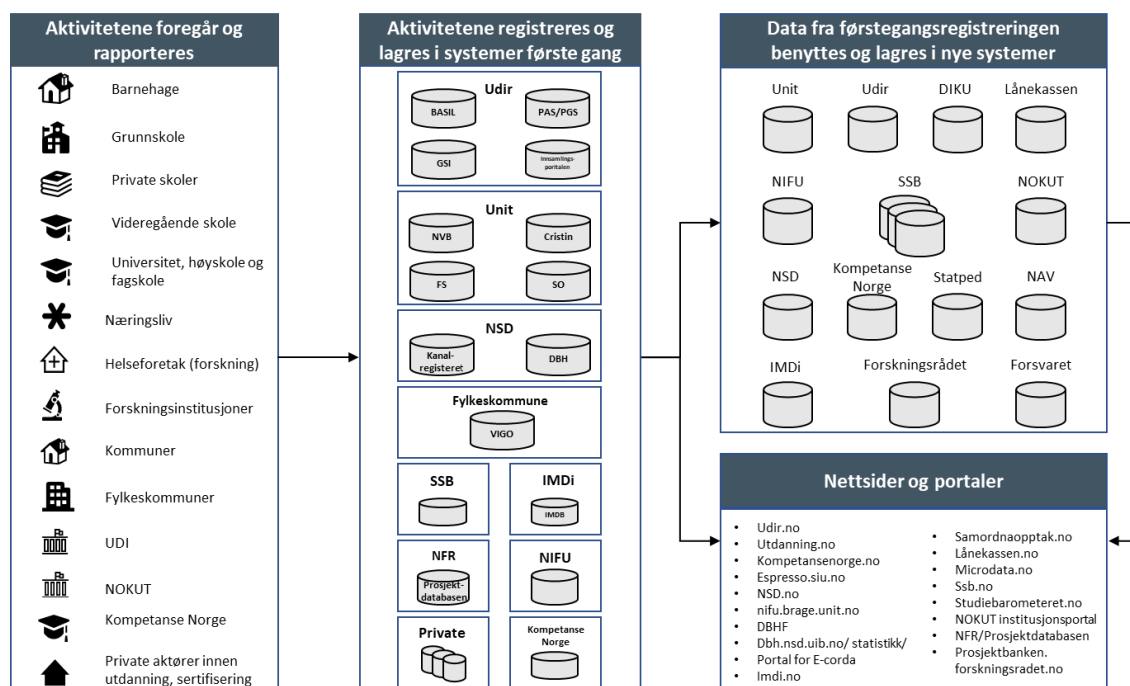
Kompetanse Norge gjennomfører prøver i norsk og samfunnskunnskap som er et vilkår for å kunne få permanent oppholdstillatelse eller statsborgerskap i Norge. IMDi får resultatene av disse prøvene på individnivå. Mange innvandrere tar også del i voksenopplæring, som registreres av kommunene.

Prøvedata fra Kompetanse Norge inngår i Prøveadministrativt system (PAD)⁴, som forvaltes av Kompetanse Norge. Opplysninger om prøver i norsk og samfunnskunnskap fra PAD overføres til IMDB. Kommunene har egen innlogging til PAD. Data om integrering tilgjengeliggjøres i de ulike systemene for konsumentene på en rekke ulike nettsider og portaler, jf. eksempler i Figur 9.

2.1.6. Oppsummering

Det er en stor produksjon av data i kunnskapssektoren og produksjonen er fordelt på mange aktører, jf. Figur 10. Bildet er preget av at det er mange databaser og registre hvor data registreres og det er i mange tilfeller manuelle prosesser knyttet til å utveksle data mellom kildene, grunnet mangel på API-er for automatisert utveksling av data. Verdikjedene for data er til dels uoversiktlige ved at data flyter litt på «kryss og tvers» mellom aktørene/kildene, og at samme aktør i noen tilfeller er involvert ulike steder i verdikjeden. Noen av dataprodusentene fungerer også som tjenesteprodusenter for enkelte tjenester. Flere aktører har kilder for førstegangsregistrering for ett formål, i tillegg til at de benytter data fra andre aktører videre til andre formål, dvs. at de befinner seg i flere av boksene i figuren.

Måten data fra dataprodusentene tilgjengeliggjøres for konsumentene på, er lite oversiktlig ved at den enkelte dataprodusent gjerne har egne statistikkportaler for å synliggjøre sine data. Det finnes ingen samlet, overordnet oversikt over data og statistikk som finnes i sektoren. Portalene er heller ikke bygget etter en felles, omforent standard måte å bygge slike portaler på, og blir i stor grad heterogene.



FIGUR 10: ILLUSTRASJON AV PRODUSENTSIDEN I HELE KUNNSKAPSEKTOREN

⁴ PAD må ikke forveksles med Udirs prøveadministrative system PAS som ble omtalt innenfor området «barnehage, grunnutdanning og videregående skole».

2.2. Konsument siden

Det er mange konsumenter av kunnskapsdata i samfunnet. I denne utredningen er konsumentene gruppert i enkelte kategorier, jf. Figur 11. Enkelte institusjoner vil være representert i flere av konsumentgruppene. Et eksempel er universiteter som både har vitenskapelige ansatte som faller innenfor gruppen «forskere» og administrative ansatte som faller inn under gruppen «tjenesteprodusenter». Kommuner og fylkeskommuner er representert innenfor både politikktutviklere og tjenesteprodusenter gjennom administrativt ansatte i kommunen og fylkeskommune med ansvar for politikktutvikling, eierstyring og virksomhetsstyring.

	Forskere	Vitenskapelig ansatte på universiteter og høyskoler, samt forskningsinstitusjoner og –institutter
	Politikkutviklere	Kunnskapsdepartementet, ansatte i kommune og fylkeskommune
	Underliggende etater	Direktorater og etater under Kunnskapsdepartementet, og andre relevante etater
	Tjenesteprodusenter	Administrativt ansatte på universiteter, høyskoler, fagskoler, videregående skoler, grunnskoler og barnehager, samt eierstyring i kommuner og fylkeskommuner
	Næringsliv	Næringslivsaktører (inkl. konsulentvirksomheter) som benytter utdannings- og integreringsdata
	Allmennheten	Foreldre, barn, media, interesseorganisasjoner osv.

FIGUR 11: KONSUMENTGRUPPER AV KUNNSKAPSDATA

2.2.1. Forskere

Forskere er ansatt på universiteter, høyskoler og i forsknings- og utredningsinstitusjoner. De benytter kunnskapsdata primært til forskning og andre former for analyse, f.eks. i forbindelse med utredninger. Dataene benyttes både i kvalitative og kvantitative analyser, hvor kvantitative analyser benytter mest data. Forskere fra flere ulike fagdisipliner benytter kunnskapsdata, hvor det er en overvekt av forskere fra samfunnsvitenskapelige disipliner. Data som benyttes er både individdata og aggregerte data fra ulike kilder (administrative registre, eksperimenter, spørreundersøkelser, intervjuer m.m.).

De siste tiårene har forskning på individdata økt som følge av forbedret teknologi og utvikling i statistiske metoder som gir økte muligheter til å koble sammen data og analysere store datasett. Dette gjør at forskerne ofte har behov for data fra flere dataprodusenter i kunnskapssektoren, samt koble kunnskapsdata med andre typer data, f.eks. demografiske og sosioøkonomiske data, helsedata eller egeninnsamlede data. Foruten å henvende seg til den enkelte dataprodusent med forespørsel om utlevering av data til forskning, er SSB en viktig instans for forskere. Siden SSB besitter store mengder data fra flere sektorer og låner ut sammenkoblede data til forskningsprosjekter, har SSB blitt et viktig knutepunkt for utlevering av flere typer data til forskning.

Innenfor forskning har blant annet regjeringen og Forskningsrådet oppfordret forskere til å følge retningslinjer med at forskningsdata må være tilgjengelige, gjenfinnbare og gjenbrukbare (tilsvarende FAIR-prinsippene).

2.2.2. Politikktutviklere

Ansatte i Kunnskapsdepartementet benytter kunnskapsdata til flere formål; politikktutforming, etatsstyring og andre forvaltningsformål, samt til å utføre analyser og gjennomføre utredninger. Ansatte i kommune og fylkeskommune er også inkludert i denne konsumentgruppen.

De ansatte i Kunnskapsdepartementet er særskilte brukere av kunnskapsdata ettersom departementet har sektoransvaret for opplæring, utdanning og integrering. På kunnskapsfeltet benytter departementet data til flere formål, blant annet som grunnlagsinformasjon i finansieringssystemet for høyere utdanning, kvalitetsindikatorer for utviklingen i grunnskolen, statistikk over frafall på videregående skole og endringer i utdanningstilbudet for voksne. I tillegg er det etterspørsel etter kunnskapsdata fra andre departement til ulike formål, blant annet til å analysere sammenhengen mellom utdanning og sosioøkonomiske utfall, f.eks. overgangen fra studier til arbeid og karrieremuligheter. På integreringsfeltet er det flere departement som bruker kunnskapsdata, blant annet Kunnskapsdepartementet, Arbeid- og sosialdepartementet, Justis- og beredskapsdepartementet og Kommunal- og moderniseringsdepartementet. Formålet med bruken av dataene varierer mellom departementene, men fellesnevner er statistikk og analyse som grunnlag for å følge med på området, regelverksutvikling og videreutvikling av virkemidlene i integreringsarbeidet.

Ansatte i kommune og fylkeskommune er interessert i kunnskapsdata innenfor sine geografiske områder, men også fra andre kommuner eller fylker for å kunne sammenligne indikatorer på tvers av landet. I tillegg benytter de data til å gjennomføre analyser om historisk og fremtidig utvikling, for å kunne gjennomføre analyser som danner grunnlag for fremtidig politikk og hvilke tiltak som bør settes i gang. I tillegg har kommunene og fylkeskommunene behov for kunnskapsdata når de utarbeider tilstandsrapporter.

I hovedsak er både departementene, kommunene og fylkeskommunene interessert i data på aggregert nivå, og ikke individnivå. De innhenter data fra ulike registre og databaser i kunnskapssektoren, både fra offisielle publiseringskanaler og gjennom spesialbestillinger, og sammenstiller selv aggregerte data fra flere aktører ved behov.

2.2.3. Underliggende etater

Ansatte i underliggende etater til Kunnskapsdepartementet og andre departement benytter kunnskapsdata til mange av de samme formål som konsumentgruppen politikktutviklere. Dette innebærer at direktoratene og etatene også har behov for data til politikktutforming, styring av underliggende virksomheter og andre forvaltningsformål. Samtidig har de ansatte i de underliggende etatene behov for data til å utforme statistikk, utføre analyser og gjennomføre utredninger. Underliggende etater under Kunnskapsdepartementet inkluderer blant annet Utdanningsdirektoratet, IMDi, NOKUT, DIKU, Unit, Lånekassen og Statped. Flere av disse konsumentene er også dataprodusenter i kunnskapssektoren. De viktigste konsumentene av kunnskapsdata utenfor sektoren er SSB, NAV og Forsvaret.

Direktoratene og etatene skiller seg imidlertid fra politikktutviklere og departementene ved at de er mer spesialiserte innenfor enkelte fagområder. De har en større faglig rolle hvor de utvikler, forvalter og formidler kunnskap innenfor sine fagområder som kan danne grunnlag for politikktutforming. Dette medfører at konsumentgruppen ofte har et mer omfattende behov for data enn departementene med et større fokus på statistikk og analyser.

De ansatte i forvaltningsorganene er interessert i kunnskapsdata både på aggregert nivå og individnivå. Data på aggregert nivå innhentes både fra egen virksomhet (dersom de selv er dataprodusenter innenfor sitt fagområde), dataprodusenter de har forvaltningsansvar for eller andre dataprodusenter i kunnskapssektoren. De innhentede dataene sammenstilles ofte på tvers av

dataprodusenter i og utenfor sektoren. I motsetning til forskere får ikke de underliggende etatene utlevert individdata fra SSB.

2.2.4. Tjenesteprodusenter

Tjenesteprodusentene i kunnskapssektoren står for og tilbyr tjenester til befolkningen. Dette omfatter asylmottak, barnehager og utdanningsinstitusjoner som grunnskoler, videregående skoler og institusjoner innen høyere utdanning og etter- og videreutdanning. Tjenesteprodusentene tilhører alle tre forvaltningsnivåer i Norge; stat (f.eks. UH-institusjoner), fylkeskommune/region (f.eks. videregående skoler) og kommunene (f.eks. asylmottak, barnehager og grunnskoler).

Innen høyere utdanning benytter universitetene og høyskolene kunnskapsdata primært til virksomhetsstyring og analyser. Dette betyr at data benyttes blant annet til ledelsesstyring, mål- og resultatstyring, økonomistyring, HR/personaladministrasjon, tjenesteutvikling og ulike kvalitetsfremmende tiltak. I tillegg benytter institusjonene kunnskapsdata til å følge med på utviklingen i egen institusjon, f.eks. gjennomstrømming av studenter og studieresultater. UH-institusjonene rapporterer om utviklingen av egen virksomhet til Kunnskapsdepartementet, blant annet gjennom formidling av statistikk og analyser til tilstandsrapporten for forskning og høyere utdanning.

Skoleeiere (kommuner og fylkeskommuner) og skolene benytter kunnskapsdata i hovedsak til samme formål med virksomhetsstyring og analyser, men i et mindre omfang enn innen høyere utdanning. Dette skyldes blant annet at skolene er mindre enheter enn UH-institusjoner, at tildelingen av midler ikke er resultatbasert og at det er færre målbare indikatorer for utviklingen av skolene. Barne- og ungdomsskoler har rapporteringsplikt til kommunene, mens videregående skoler rapporterer til fylkeskommunen. Kommunene og fylkeskommunene driver virksomhetsstyring for sine ansvarsområder og har oppfølging av enkeltvirksomhetene basert på de innrapporterte dataene fra underliggende virksomheter.

Tjenesteprodusentene er interessert i både individdata fra egen virksomhet og aggregerte data. Individdataene benyttes for å følge opp enkeltindivider i egen virksomhet. De aggregerte dataene benyttes for å sammenligne på tvers av virksomheter, for både enkeltvirksomhetens ledelse og virksomhetsstyringen på tvers av underliggende virksomheter på kommunalt- og fylkeskommunalt nivå.

2.2.5. Næringsliv

Store deler av næringslivet bruker kunnskapsdata primært som innsatsfaktor i produksjon av varer og tjenester, dvs. til forretningsmessig virksomhet. I tillegg eksisterer det private forsknings- og analyseinstitusjoner som benytter kunnskapsdata til analyser og utredninger på vegne av offentlige institusjoner.

I produksjon av varer og tjenester er variasjonen i bruken av kunnskapsdata stor, og behovet for data varierer deretter. Enkelte selskaper bruker aggregerte data som bakgrunnsmateriale, andre utfører mer inngående analyser av disaggregerte data (f.eks. for å avdekke behov for nye tjenester rettet mot skoler) osv. Informasjonen som brukes, hentes i stor grad inn via offisielle publiseringskanaler eller gjennom spesialbestillinger hos dataprodusenter.

Private forsknings- og analyseinstitusjoner har på sin side ofte et større behov for disaggregerte data, ofte på individnivå, for å kunne utføre sine aktiviteter. I tillegg har de behov for å koble kunnskapsdata med andre typer data på linje med forskere ansatt på universiteter, høyskoler og offentlige forskningsinstitusjoner, jf. kap. 2.2.1. De private institusjonene er derfor i stor grad avhengig av å

bestille data fra dataprodusentene eller SSB. En utfordring er imidlertid at utlevering av kunnskapsdata på individnivå til næringslivsformål er begrenset etter lovverket.

2.2.6. Allmennheten

Allmennheten er en variert konsumentgruppe og inkluderer innbyggere, ansatte i medier og ansatte i ulike interesseorganisasjoner som for eksempel fagforeninger. Aktørene benytter kunnskapsdata innenfor en rekke ulike områder.

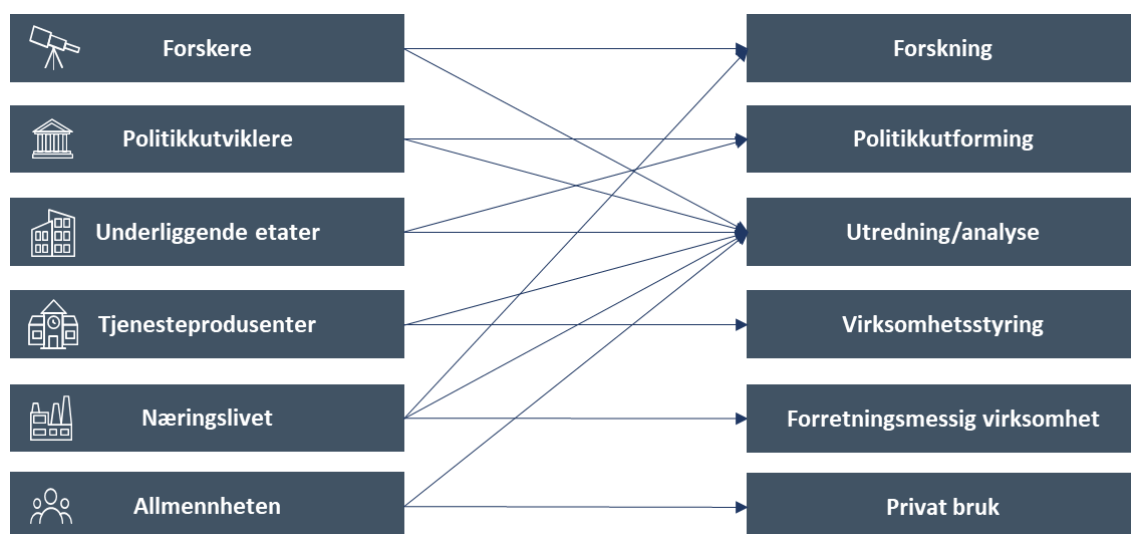
Innbyggere bruker kunnskapsdata i hovedsak til privat bruk. Bruken er variert og inkluderer data fra alle de ulike områdene i kunnskapssektoren. Eksempelvis ønsker foreldre informasjon om kvaliteten på barnehager og skoler, ungdommer ønsker informasjon om utdanningsmuligheter på videregående skole og innen høyere utdanning, mens voksne ønsker informasjon om muligheter innen høyere utdanning og etter- og videreutdanningstilbud. I tillegg har personer som har gjennomført utdanning behov for opplysninger om seg selv, blant annet karakterutskrift, vitnemål, attester, samt at de etter lovverket har rett til innsyn i opplysninger som er registrert om dem. I tillegg til individuell bruk, er det viktig for innbyggernes deltagelse i den offentlige debatten at data er tilgjengelig dersom de ønsker å fremme sine meninger eller gjøre egne undersøkelser innenfor et tema.

Media bruker i hovedsak kunnskapsdata til utredning, analyser og dokumentering av utviklingen i samfunnet. Dette innebærer å innhente informasjon og data som skal formidles til befolkningen, gjerne informasjon om dagsaktuelle temaer eller fagområder som bygger opp under budskapet de ønsker å formidle. Etterspørselen etter kunnskapsdata avhenger derfor av hva som er aktuelt til enhver tid, og hvilke medier som etterspør dataene. Til utredninger og analyser kreves det større mengder data enn mengden data som benyttes for å dokumentere en enkelt sak.

Interesseorganisasjoner for både arbeidstakere og arbeidsgivere, og organisasjoner som Private Barnehagers Landsforbund, studentorganisasjoner og liknende, benytter kunnskapsdata primært til å utføre analyser. Dette kan være analyser i forbindelse med utredninger, profesjonsutvikling og lønnsforhandlinger. De har ved flere anledninger behov for data fra flere datakilder i kunnskapssektoren, og ønsker å sammenstille dataene med andre typer data, både fra offisielle publiseringskanaler og gjennom spesialbestillinger. Interesseorganisasjoner har som hovedmål å fremme interessene til sine medlemmer, og kanalisere disse til politikere og administrativ ledelse. I sitt interessepolitiske arbeid kartlegger de områder som blant annet fullføring av høyere utdanning, søkertall og gjennomstrømning av studenter.

2.2.7. Oppsummering

Konsumenter benytter kunnskapsdata i ulik grad og til mange forskjellige formål. Bruken omfatter blant annet forskning og analyse, utarbeiding av statistikk, virksomhetsstyring og gjennomføring av utredninger. Samtidig benyttes det data til forretningsmessig virksomhet og privat bruk. Figur 12 oppsummerer konsumentgruppens hovedformål med bruken av kunnskapsdata.



FIGUR 12: KONSUMENTGRUPPER OG HOVEDFORMÅL MED DATABRUKEN

3. Problembeskrivelse

Kapittelet gir en oversikt over utfordringer knyttet til deling av kunnskapsdata i og utenfor kunnskapssektoren. Beskrivelsen bygger i hovedsak på dokumentanalyser og intervjuer med produsenter og konsumenter av dataene, jf. oversikt over intervjuobjekter i Tabell 27 i vedlegg C, og vår forståelse av aktørenes problembeskrivelse.

3.1. Utfordringer på konsumentensiden

3.1.1. Forskere

Forskere opplever flere utfordringer knyttet til tilgang og bruk av kunnskapsdata. Før oppstart eller i oppstartsfasen av et forskningsprosjekt foretar forskere ofte en kartlegging av potensielle datakilder og hvilke data som eksisterer. Enkelte forskere, og særlig de som ikke har benyttet denne type data tidligere, opplever at det er vanskelig å navigere seg i et landskap med mange dataprodusenter og statistikk over samme type data som publiseres flere steder. Situasjonen gjør det vanskelig for forskerne å få oversikt over av hva som finnes av data og hva som er tilgjengelig til forsknings- og analyseformål. Det gjelder både registre, enkelte datasett, variabler og metadata. Mangelen på oversikt medfører at forskere kan bruke mye tid på å identifisere tilgjengelige variabler og at de ikke nødvendigvis finner tilgjengelige data. Manglende metadata er en utfordring når dataene tas i bruk ved at forskere bruker mye tid på å tolke og forstå dataene. En utbredt utfordring er manglende begrepsharmonisering for variabler som gjør at det forekommer ulike betydninger av samme begrep i ulike registre og databaser. I tillegg svekkes tilliten deres til dataene, som i ytterste konsekvens kan medføre at forskeren ikke benytter dataene.

Flere forskere opplever også at datakvaliteten på kunnskapsdata er svært varierende. Mens enkelte datasett kan ha høy kvalitet, inneholder andre datasett åpenbare feil og mangler. Dette medfører at forskere kan bruke tid på å «vaske» dataene før de kan benyttes til forskning.

Forskere som ønsker data fra ulike kilder som skal kobles, særskilt på individnivå, må i stor grad henvende seg til Statistisk sentralbyrå. Flertallet av de intervjuede forskerne, peker på at søknads- og tildelingsprosessen hos SSB tar lang tid og koster mye penger. En særskilt tidkrevende aktivitet i søknadsprosessen er avklaring av variabelbehov. Det henger bl.a. sammen med at mange forskere opplever manglende oversikt og beskrivelser av tilgjengelige variabler. Når de ikke klarer å definere variabelbehovet i første omgang, må de redefinere innhold og omfang og søke på nytt. De høye kostnadene er forbundet med SSB sitt arbeid med å behandle søknaden, sammenstille dataene og foreta selve uttrekket av dataene. Flere forskere opplever også at SSB har varierende kompetanse på kunnskapsdata, og at det virker som behandlingen er avhengig av hvem som behandler søknaden hos SSB.

Flere forskere trekker frem at aktører i kunnskapssektoren som utleverer data (f.eks. Utdanningsdirektoratet, Unit, NSD og IMDi) har ulik praksis for behandling av søknader om datatilgang. Dette gjør at søknadsprosessen blir tidkrevende. Forskere må både bruke tid på å sette seg inn i og forstå søknadsprosedyren, samt fylle ut ulike søknadsskjemaer. I tillegg er det mange dataprodusenter som gjør at man kan ende opp med å søke om tilgang til mange instanser.

Arbeidsnotatet «Koordinering av dataproduksjon og kopling av datakilder om forskning» gir en oversikt over dagens situasjon på produsentsiden innenfor forskningsområdet (Sivertsen & Gunnes, 2019). Notatet trekker frem at det er behov for å koordinere dataproduksjonen og legge til rette for kobling av data. Det argumenteres for at dette er bedre enn å erstatte nåværende systemer og aktører med nye systemer og fusjonerte aktører.

3.1.2. Politikktutviklere

Kunnskapsdepartementet og lokale politikere i kommune og fylkeskommune har som mål at politikktutvikling og styring i sektoren skal bygge på et solid kunnskapsgrunnlag, og at beslutninger skal være preget av faglighet. Politikktutviklere opplever imidlertid flere utfordringer ved uthenting og bruk av kunnskapsdata i sitt arbeid. Et eksempel er at mangelfulle beskrivelser av dataene gjør det krevende å vite hva som ligger i dataene, hvilke formål de har og hva de kan, og eventuelt ikke kan, benyttes til. Konsekvensen er unødvendig tidsbruk på innsamling og tolkning av dataene. Dette er problematisk da politikktutviklere til tider har korte tidsfrister på saker som må behandles, for eksempel faglige råd til politisk ledelse, som innebærer at de må ha rask tilgang til dataene. I tillegg medfører manglende metadata lavere tillit til dataene, som i sin tur kan føre til at de velger å ikke benytte dataene. I ytterste konsekvens kan også mangelfulle beskrivelser av data medføre at feil informasjonsgrunnlag benyttes i analyser.

Flere politikktutviklere trekker frem at en del aktører offentliggjør forskjellige tall på “samme” variabler. Dette gjelder for eksempel SSB og Utdanningsdirektoratet. Begge produserer statistikk over elevers gjennomføring av videregående utdanning, men med ulike definisjoner. Det gir ulike tall for gjennomføring. Et annet eksempel er fylkeskommunene som har ulike definisjoner på frafall i videregående skole, som gir ulike tall for frafall i statistikkene. Flere intervjuobjekter hevder at dette skyldes at produsentene i liten grad har brukt ressurser på å etablere felles definisjoner. Konsekvensen for konsumentene er at det gir mye merarbeid med å finne ut av hva de forskjellige variablene uttrykker, samtidig som det øker risikoen for feil forståelse og således ulik bruk av dataene. Innenfor enkelte områder svekker det tilliten til datagrunnlaget og statistikken.

Politikktutviklere etterspør kunnskapsdata fra leverandører i og utenfor sektoren til å lage statistikk og gjennomføre analyser. Dårlig datakvalitet gjennom varierende registrering av data og mangelfulle datasett gjør dette imidlertid vanskelig. En årsak som har blitt trukket frem er at produsentene i mindre grad bruker dataene selv. Det medfører at de ikke ser formålene med de rapportene, og da kan dataene ofte inneholde feil eller mangler. I tillegg påpeker enkelte av politikktutviklerne at kvaliteten på dataene bedres jo mer de brukes. Et eksempel er fraværstatistikken i videregående skole. Statistikken ble i mindre grad benyttet tidligere, utover oppfølging av den enkelte elev på skolene. I den senere tid har det kommet strengere krav til at fravær skal registreres og følges opp, og det er innført en fraværsgrense på 10-15 %. Dette har medført økt fokus og interesse for fraværstatistikk både på den enkelte skole, blant politikktutviklere og i media. Dette har gjort at kvaliteten på dataene har blitt bedre, og gir en mer presis oversikt over fraværet i videregående skole.

En annen utfordring politikktutviklerne opplever er at de har lav prioritet hos SSB. Det gjør det vanskelig å svare ut politikernes behov for å vurdere utvikling innenfor blant annet dagsaktuelle problemstillinger. Flere opplever også at Statistikkbanken til SSB har blitt mer rigid. Innenfor utdanningsområdet må enkelte politikktutviklere benytte opp til 15 ulike tabeller for å beskrive hele problemområdet. Det er for eksempel ingen kobling mellom utdannings- og arbeidsmarkedsdata, som gjør det krevende å analysere dataene og trekke ut innsikt.

3.1.3. Underliggende etater

De underliggende etatene opplever også utfordringer knyttet til tilgang og bruk av kunnskapsdata. En begrensning som flere av intervjuobjektene trekker frem, er at de ikke får tilgang til dataene de ønsker. For å utføre oppgavene på sine områder har de behov for å utføre mer omfattende analyser enn politikktutviklere, og således behov for tilgang til data på et lavere aggregeringsnivå. I motsetning til forskere får ikke de underliggende etatene utlevert individdata fra SSB. I tillegg opplever flere underliggende etater at produsentene av data er motvillige til å dele data. Det er et inntrykk fra intervjuene i kartleggingsfasen at denne «motvilligheten» har sammenheng med usikkerhet knyttet til hva som kan deles, og med hvem. For å være på den sikre siden, unnlater en da i enkelte tilfeller å dele data. Dette begrenser tilgangen til data ytterligere.

En annen utfordring som blir trukket frem er ulike definisjoner av variabler. Slik som politikktutviklere, opplever forvaltningsorganene å finne de samme variablene hos flere produsenter, men med ulike definisjoner og tallstørrelser. For eksempel er tallet på antall studenter i Norge forskjellig hos SSB, NSD og Unit. Selv om det i enkelte tilfeller er logiske og praktiske årsaker til at det er slik (f.eks. at SSB rapporterer til internasjonale statistikkprodusenter som benytter kalenderår, mens i nasjonale statistikker er studieåret det sentrale), bidrar manglende definisjonsforklaringer og veiledning til å skape usikkerhet hos konsumentene. Dette svekker troverdigheten og tilliten til dataene, noe som i sin tur øker terskelen for å ta dataene i bruk.

De underliggende etatene peker på at registreringen av data tidvis er mangelfull og inneholder feil. Dette er et problem som har ulike årsaker. Sentrale problemstillinger som trekkes frem er at de som registrerer dataene ikke selv ser nytten av dataene for sin virksomhet. Et annet eksempel er at data som samles inn av en aktør for et spesifikt formål som er viktig for denne aktøren, ikke nødvendigvis gir like mye verdi for andre. På lik linje med forskere, medfører dette at underliggende etater bruker en del tid på å forstå og «vaske» dataene før de kan analyseres. Flere opplever at kvalitetssikringen av data ikke blir tilstrekkelig prioritert, samt at den faglige kompetansen hos produsentene ikke alltid er like høy. I tillegg opplever flere at data som produsentene ikke bruker selv er mer fragmentert enn data som også produsentene selv bruker. Når konsumentene skal hente ut disse dataene vil de ikke få en like god oversikt over dataene de søker, og de må ofte bearbeide eller hente inn data fra flere kilder. Dette fører på den ene siden til betydelig merarbeid, og på den andre siden økt usikkerhet og uklarhet rundt datagrunnlaget. For forvaltningsorganene som skal bidra faglig til politikktutforming og styring er det svært viktig at datagrunnlaget de baserer seg på er pålitelig.

3.1.4. Tjenesteprodusenter

Tjenesteprodusentene opplever tilgangen til kunnskapsdata som utfordrende og begrensende for utvikling av sin egen virksomhet. Tjenesteprodusentene er primært interessert i data på et høyt detaljeringsnivå, og de ønsker data om tilsvarende institusjoner som seg selv - både for sammenlikning og analyser innenfor spesifikke fag- og virksomhetsområder, og på tvers av sektoren. Innenfor området høyere utdanning har UH-institusjonene kun tilgang til egne data på et tilfredsstillende aggregeringsnivå gjennom visualiseringsverktøy koblet mot Star. For data om andre UH-institusjoner får de kun data gjennom DBH som i stor grad publiseres med et høyt aggregeringsnivå. Det resulterer i at de ikke får gjort de analysene de ønsker, og de får heller ikke sammenlignet egen virksomhet opp mot sektoren på de nivåene de de ønsker. I tillegg må dataene om andre UH-institusjoner hentes ut manuelt, som gjør at prosessen blir både tid- og ressurskrevende.

Mange av intervjuobjektene opplever at dataene generelt er av god kvalitet, men at det samtidig er feil og mangler. Som forvaltningsorganene, opplever tjenesteprodusentene at kvalitetssikringen ikke alltid er tilstrekkelig, og at feilene skyldes både mangelfulle rutiner for rapportering og varierende faglig kompetanse hos dataprodusentene. Dette fører til at eventuelle feil eller mangler ikke oppdages og rettes. Tjenesteprodusentene trekker også frem utfordringer med ulike definisjoner og tallstørrelser på

“samme” variabel på linje med underliggende etater og politikkutviklere. Det er ikke alltid tydelig hvilke data som er relevante til en gitt problemstilling, som fører til at enkelte data anvendes med forsiktighet og potensielt ikke brukes. Flere av tjenesteprodusentene trekker frem at de tar i bruk den definisjonen som passer dem best, eller at de lager sin egen definisjon dersom de eksisterende ikke er dekkende nok. Dette resulterer i at det oppstår enda flere definisjoner og tallstørrelser. I enkelte tilfeller deles disse videre med andre konsumenter, som gjør at bildet blir enda mer uoversiktlig, og tilliten til dataene svekkes ytterligere.

Flere av tjenesteprodusenter innhenter data regelmessig fra dataprodusentene. Her kan det oppstå utfordringer når formålet for dataene eller betydningen av tall endres hos dataprodusentene. Dersom data f.eks. endres hos DBH vil ikke UH-institusjonene få beskjed om denne endringen. Dette gjør det vanskelig for institusjonene å ha oppdaterte eller korrekt forståelse av data til enhver tid. Dette innebærer at analysene baseres på feilaktig og utdatert tallmateriale som er misvisende og/eller lite relevant. Ofte er tjenesteprodusentene klar over at data ikke oppdateres automatisk og bruker derfor tid på å jevnlig å se etter oppdateringer. I enkelte tilfeller er dette så tidkrevende at tjenesteprodusentene velger å ikke ta i bruk dataene. For grensesnitt hvor det benyttes API-er er dette et mindre problem, siden det er strengere rutiner for å endre API-er. I stedet for å endre en eksisterende API er det vanlig å tilgjengeliggjøre en ny versjon av API-et, samtidig som den gamle versjonen lever videre til alle konsumentene har tatt i bruk siste versjon. NSD har utviklet et API mot DBH som vil redusere dette problemene etter hvert som institusjonene tar dette i bruk.

I mange tilfeller må tjenesteprodusentene benytte seg av en rekke ulike datakilder eller portaler for å danne seg et fullstendig bilde av virksomheten og for å sammenligne med andre tilsvarende virksomheter. Et eksempel er kommuner som følger opp sine egne skoler gjennom både lokale rapporteringer, GSI, SSB og eventuelt andre kilder. Mange av kildene er heller ikke koblet sammen. Dette gjør virksomhetsstyringen tidkrevende og analyser på tvers av ulike indikatorer hos tjenesteprodusentene krevende.

Flere tjenesteprodusenter ønsker å koble seg opp mot datavarehus eller ulike portaler fremfor direkte kobling til selve datakildene. Årsaken til dette er at de originale datakildene inneholder kompliserte datamodeller og for store mengder detaljerte data, som gjør det både tid- og ressurskrevende å hente ut og forstå data derfra. I tillegg mangler kildene ofte metadata. I datavarehusene er det i større grad “ryddet” i dataene og det er gode metadata, som gjør det enklere å forstå hva dataene uttrykker og hente ut dataene man trenger. Enkelte av intervjuobjektene mener at den store bruken av datavarehus er et symptom på at data ikke er tilstrekkelig tilrettelagt for viderebruk. Datavarehus behandler således et symptom, mens rotårsaken er at dataene egentlig ikke er godt nok tilrettelagt for formålene med viderebruk.

3.1.5. Næringslivet

Næringslivet opplever særlig utfordringer med å få tilgang til kunnskapsdata i kunnskapssektoren. Selv om mange registre er gjort tilgjengelig for bruk, har dataprodusentene fortsatt mye data som er egnet for næringslivet, men som ikke er tilgjengeliggjort. Dette inkluderer også muligheten til å kombinere data på tvers av områdene i sektoren og med data i andre sektorer. Nye kombinasjoner av data kan skape ny kunnskap og innsikt, som igjen kan åpne opp en rekke nye anvendelsesområder for innovative produkter og tjenester.

Flere av intervjuobjektene trekker frem at det er særlig i forbindelse med de større analysene at tilgangen til data utgjør en stor begrensning. Uten tilgang til tilstrekkelige mengder data har de ikke mulighet til å utføre de utredningene de tar på seg i oppdrag. Dette gjør utredningene mangelfulle og problemstillingene blir ikke svart ut grundig nok. Konsekvensen blir at næringslivet ikke kan påta seg disse oppdragene da datagrunnlaget de har tilgang til er for knapt.

I en del tilfeller konkurrerer private analysevirksomheter med offentlige forsknings- og analysevirksomheter om oppdrag for å gjennomføre analyser for forvaltningen. Enkelte av intervjuobjektene har påpekt at de offentlige virksomhetene har et fortrinn ved at de har enklere tilgang til data og dette gir en konkurransevridning.

Næringslivet opplever på lik linje med de andre konsumentgruppene utfordringer når det kommer til manglende beskrivelser av data. Ved tilfeller av mangelfulle metadata blir unødvendig mye tid brukt på å forstå og tolke dataene før de kan brukes. Resultatet er en høyere terskel for å ta i bruk dataene, og i enkelte tilfeller medfører det at dataene ikke benyttes.

3.1.6. Allmennheten

Allmennheten er en variert gruppe, med både behov for enkel visualisering av data og data som muliggjør mer omfattende analyser og utredninger. En av de største utfordringene for denne konsumentgruppen er å navigere i landskapet av portaler, nettsider og systemer for å hente ut data og informasjon de har behov for. Mye av dataene allmennheten har behov for er tilgjengelig, men det er krevende å finne data knyttet til temaet man leter etter. For å finne informasjon må man vite hvor man skal lete og i mange tilfeller oppsøke flere nettsider og portaler for å få en helhetlig oversikt innenfor området eller temaet man har interesse for. En årsak til dette er at produsentene i stor grad tilgjengeliggjør data de har lagret uten å samkjøre publisering av lignende data innenfor samme tema fra andre produsenter. Et eksempel er statistikken for antall lærere i grunnskolen, der Udir oppgir 69 079 lærere i 2018-2019, mens SSB opererer med 75 727 lærere for 2018.

Gjennom intervjuene har det kommet frem at data ikke alltid er enkelt og brukervennlig fremstilt. I flere tilfeller har data manglende beskrivelser, og fremstilling av tilsynelatende samme data varierer uten en god beskrivelse av årsaken. Kombinert med ulike portaler og nettsider med delvis overlappende informasjon, innebærer dette at det er krevende å forstå og tolke dataene på egenhånd. Dette kan i ytterste konsekvens føre til at allmennheten kan vegre seg for å lete etter og ta i bruk kunnskapsdata.

3.2. Utfordringer på produsentsiden

I dag produseres det mye kunnskapsdata fordelt på en rekke ulike aktører, som er illustrert i Figur 10. Data om aktivitetene registreres i ulike registre og databaser, og i mange tilfeller blir data overført videre til andre registre og databaser. De ulike registrene og databasene forvaltes av flere ulike aktører, og data tilgjengeliggjøres for konsumentene gjennom en rekke ulike nettsider, portaler, verktøytjenester og systemer. Dette landskapet har resultert i en rekke utfordringer på produsentsiden i kunnskapssektoren, og kan oppsummeres med følgende punkter:

- Uklar organisering i dataproduksjonen
- Manglende forståelse for bruk av data
- Svak kultur for deling
- Juridiske hindringer
- Varierende datakvalitet

3.2.1. Uklar organisering i dataproduksjonen

I dag er kunnskapsdata fordelt på en rekke ulike aktører. Dette omfattende bildet av ulike aktører og datakilder blir trukket frem som en utfordring av de ulike konsumentene i flere intervjuer. Konsumentene trekker frem at det er vanskelig å navigere og finne ut av hva som eksisterer av data,

hva som er tilgjengelig og hva det er mulig å hente ut til sitt formål. I tillegg opplever konsumentene at «samme» data kan være tilgjengelig hos flere aktører der informasjon kan være overlappende og ikke nødvendigvis helt lik. I tillegg er beskrivelsen av dataene ofte mangelfull, som videre medfører at prosessene rundt håndtering av dataene blir tid- og ressurskrevende.

En av de bakenforliggende årsakene til disse utfordringene ser ut til å være en uklar rolle- og ansvarsfordeling, manglende koordinering blant produsentene og manglende styring og myndighetsutøvelse i sektoren. I intervjuer med de ulike produsentene kommer det frem at de selv opplever å ha et tydelig definert ansvarsområde, men at det er mer uklart hvem som har ansvar for å dele data med hvem og hvordan dette skal organiseres, samt hvem som har ansvar for beskrivelser av dataene. Det later derfor til at ansvars- og rollefordelingen rundt det å dele data ikke er tilstrekkelig, og at det mangler styring fra underliggende etater.

En konsekvens av disse utfordringene er lav grad av harmonisering av data i kunnskapssektoren. Ingen aktør hos dataprodusentene har definisjonsmakt for ulike variabler på tvers av kunnskapssektoren, og dermed ingen mulighet til å etablere autoritative registre (kilde som til enhver tid har myndighet for et visst datasett) innenfor sine områder. I og med at det ikke finnes en definert autoritativ kilde for dataene fremstår en del data «fragmentert» for konsumenten med ulike registre for tilnærmet samme variabel.

En annen konsekvens av den uklare rolle- og ansvarsfordelingen og manglende styringen er at datainfrastrukturen er fragmentert. Situasjonen har ført til at flere av produsentene over tid har laget egne løsninger og systemer for registrering og lagring av data, samt egne verktøy for publisering og visualisering av data. I stedet for et fokus på å definere autoritative kilder og samordning av løsninger gjennom samarbeid om deling av data og verdikjeder på tvers, er det etablert egne løsninger for innhenting og registrering av data som kun løser «fragmenterte» behov som er oppstått hos en aktør. Dette medfører videre en stor grad av duplisering av data (i stedet for deling av autoritative data). Dupliseringen av data gjør det vanskelig å vite hvilke data som skal benyttes i og med at det finnes ulike varianter av de samme dataene. Et eksempel på dette er rapporteringen til Cristin på fjorårets vitenskapelige publikasjoner. Dette gjøres 1. april hvert år, og gir opphav til et fryst datasett. Rapporteringsinstruksen åpner for at forskerne kan rapportere en artikkel gitt at den har fått tildelt en identifisering av det elektroniske dokumentet (DOI), selv om den trykte utgaven ikke foreligger på rapporteringstidspunktet. Dette gjør at en artikkel som dukker opp på forskernes profil i Cristin kan ha forskjellige årstall fra det som ligger i datasettet knyttet til rapporteringen. Når det i tillegg rapporteres om mangelfulle metadata, blir det ineffektive prosesser for innhenting av data hos konsumentene. Det må likevel trekkes frem at fragmentert datainfrastruktur ikke er unikt for kunnskapssektoren. I boken «Digital Government Review of Norway» påpeker OECD at en av utfordringene Norge står overfor innenfor digitalisering av offentlig sektor er fragmentering. Innenfor dataproduksjon oppsummeres utfordringen i følgende overskrift: «*The challenge for Norway is to avoid fragmentation, leading to duplication and incoherence*» (OECD, 2017).

3.2.2. Manglende forståelse for andres bruk av data

Gjennom intervjuene er det avdekket at dataproduksjonen på flere områder ikke er tilstrekkelig tilpasset konsumentenes behov. En viktig årsak som blir trukket frem er at dataene i utgangspunktet er administrative registre som har blitt produsert i forbindelse med virksomhetsstyringen, samt oppfølging og utvikling av de offentlige tjenestene. Når disse dataene benyttes til andre formål, som for eksempel forskning, utredninger og politikkutforming, er dataene ofte ikke tilpasset denne bruken. Avstanden mellom produksjon og viderebruk av data er på flere områder for stor og det blir vanskelig for dataprodusentene å sette seg inn i og forstå hvilke behov konsumentene av dataene har.

I kunnskapssektoren er det flere tilfeller der data som i utgangspunktet er identiske er blitt definert og målt forskjellig. En viktig årsak til dette er at dataene har hatt forskjellige formål innenfor ulike verdikjeder, og dermed har de endt opp med forskjellige verdier i statistikker og registre. Det er derfor viktig at konsumentene som skal benytte dataene er innforstått med dataenes betydning og forutsetninger for bruk. Et eksempel på ulike definisjoner er «student», der kravene til rapportering av data i internasjonal sammenheng avviker fra nasjonale standarder. Eurostat etterspør antall studenter per kalenderår, mens det mest hensiktsmessige i Norge er å telle antall studenter per studieår ettersom årshjulet for studier i Norge er fra august til juli. Det at data har blitt målt forskjellig, kombinert med manglende beskrivelser av de ulike variablene, gjør at dataene fremstår som uklare for konsumentene. Dette gjør bruken av data vanskelig og svekker tilliten til dataene. Det er i dag satt i gang flere gode initiativer rundt harmonisering av begreper, f.eks. arbeidsgruppen som ble nedsatt av KD for å gjennomgå sentrale begreper og definisjoner knyttet til høyere utdanning og Difis rammeverk for «orden i eget hus». Selv om det er flere gode initiativer savnes det en sterkere styring for å få operasjonalisert rammeverk som dette.

Situasjonen der produsentene av data for viderebruk ikke alltid kjenner konsumentenes behov, forsterkes av at data som tilgjengeliggjøres for de ulike konsumentene i enkelte tilfeller kan ha et for lavt/høyt detaljeringsnivå i forhold til formålet. Når dataene er på et for høyt detaljeringsnivå, kreves det mer kunnskap om hva dataene uttrykker. Intervjuene viser at i slike tilfeller etableres det ofte datavarehus for å organisere dataene slik at de kan benyttes til det ønskede formålet. Prosessen som flytter data fra et eller flere registre over i et datavarehus (ETL-prosessen⁵) krever inngående kunnskap om dataenes betydning slik at dataene i datavarehuset gir mening for de som skal benytte og analysere dem. Etablering av et datavarehus for å løse et problem, kan eksemplifiseres med STAR. FS inneholder en mengde data fra universiteter og høyskoler knyttet til studieadministrasjonen ved den enkelte institusjon. STAR er et datavarehus som blant annet samler inn, bearbeider og strukturerer data fra FS for analyseformål. Ved å knytte datavariabler om f.eks. studiegjennomstrømning til andre kunnskapsdata, gjør STAR det mulig å gjøre større og dypere analyser på dataene. En av fordelene med STAR er at det er en felles løsning for alle institusjoner som benytter FS.

I noen tilfeller er det behov for data på et høyt aggregeringsnivå, eksempelvis når KD ønsker å se den samlede tilstanden i hele UH-sektoren. Dette løses ved å gjøre en aggregering av data fra relevante registre og databaser til et nivå som dekker makroperspektivet til KD. I disse dataene er ikke detaljene tatt med videre, og dersom andre konsumenter forsøker å benytte disse aggregerte dataene til formål som krever et høyere detaljeringsnivå, vil det ikke være mulig i og med at dataene ikke lar seg disaggregere. Et eksempel på aggregering av publiserte data er databasen DBH. DBH ble opprinnelig opprettet av Kunnskapsdepartementet for rapportering til departementet om tilstanden i UH-sektoren, og inneholder data om studenter, utdanning, forskning, personale, museum, areal, økonomi og selskapsdata. DBH fungerer fint til dette formålet. Utdanningsinstitusjonene har tilgang til data fra DBH om egen virksomhet, men DBH gir begrensede muligheter for å sammenligne egen virksomhet med resten av sektoren. Årsaken er at dataene som publiseres er aggregert på et nivå som ikke dekker institusjonenes behov for detaljinformasjon. Dette tilfellet er et eksempel på et forsøk om å benytte data til et annet formål enn hva som er hovedformålet, og dataene er dermed ikke tilrettelagt for dette formålet. På konsumentensiden kan dette oppfattes som dårlig datakvalitet, mens problemet egentlig er at konsumentene forsøker å benytte dataene til et formål de ikke er tilrettelagt for. Dette viser at det er behov for å legge til rette for at data kan benyttes til ulike formål der man i større grad enn i dag tar innover seg at dataene inngår i flere ulike verdikjeder.

⁵ En av de vanligste prosessene for å bygge et datavarehus er ETL-prosessen. ETL-prosessen består av Extract (hent ut data fra register/databaser), Transform (transformer dataene i henhold til organisering i datavarehuset) og Load (laste inn dataene i datavarehuset).

Eksemplene ovenfor illustrerer kompleksiteten ved å dele data gjennom hele verdikjeden fra produksjon til analyse på tvers av ulike verdikjeder. De illustrer også et annet viktig poeng med at konsumentene har ulike behov. I dag er erfaringsutvekslingen mellom produsenter og konsumenter av data i kunnskapssektoren tilfeldig og i liten grad strukturert. Flere intervjuobjekter argumenterer for en mer formell utveksling av erfaringer. En mulighet er at det etableres arenaer hvor produsenter og konsumenter møtes for å dele sine erfaringer. En annen mulighet er å etablere praksis der dataproducentene etablerer tjenester som benyttes både av interne og eksterne konsumenter. På denne måten vil organisasjonen som tilbyr data i større grad sette seg inn i konsumentenes perspektiv. I 2019 ble det gjennomført et initiativ for mer brukerfokus i kunnskapssektoren, der etater samarbeider på tvers for å tilby helhetlige brukerreiser (Kunnskapsdepartementet, 2019). En slik tankegang kan overføres til samarbeid om verdikjeder for data til ulike formål.

Konsekvensen av manglende forståelse av konsumentenes behov er flere. For det første oppleves dataene som mindre relevante for konsumentene, noe som bidrar til at bruken av eksisterende data i sektoren er lavere enn optimalt. Ved enkelte tilfeller kan det også føre til at dataene brukes feil. Feil bruk av data vil svekke tilliten til dataene eller resultatene fra for eksempel analysene der dataene er benyttet. For samfunnet vil dette til sammen kunne svekke beslutningsgrunnlaget for politikkutforming og gi en lavere kvalitet i tjenesteytingen i sektoren.

3.2.3. Svak kultur for deling

Intervjuene med aktører på produsent- og konsumentensiden viser at tilstanden for datadeling i kunnskapssektoren ikke er optimal. Inntrykket er at det ser ut til å være liten vilje til å dele data med konsumentene, noe som kan forklares av flere underliggende årsaker. For det første virker det å være en frykt for å eksponere data med tvilsom datakvalitet til konsumentene. For det andre er det ingen sterke pålegg om at offentlige etater skal dele data. Det årlige Digitaliseringsrundskrivnet legger føringer for at offentlige virksomheter skal legge til rette for samhandling og deling av data. Likevel finnes det ikke koordinering og styring hverken tverrsektorielt eller innenfor den enkelte sektor for å operasjonalisere dette på en fornuftig og kostnadseffektiv måte, som beskrevet i kapittel 3.2.1. For det tredje er det få eller ingen insentiver for å dele data med andre. Mange produsenter vurderer at gode løsninger for deling av data ikke er viktig nok til å forsvare kostnadene det medfører. Kostnadene for produsenten er vesentlig høyere enn gevinstene ved å eksempelvis automatisere en manuell prosess for deling av data, og det blir derfor ikke gjennomført. Samtidig er det en utfordring at hovedtyngden av kostnadene ved å gjøre data tilgjengelig for deling ligger hos dataproducenten, mens gevinstene i stor grad tas ut hos konsumentene. Disse faktorene medfører liten vilje for deling av data fra flere aktører.

En annen utfordring med deling av data er forpliktelsene deling med konsumentene medfører for produsenten. Ved å legge ut data åpent og tilgjengelig «forplikter» produsentene at data også vil være tilgjengelig i fremtiden, og produsentene føler et tap av kontroll over bruken av «sine» data. Det vil kunne oppstå et økt vedlikeholdsbehov ved å måtte støtte flere versjoner av de samme dataene, spesielt dersom man på et senere tidspunkt ser behov for å endre definisjoner. Produsentene kan ikke fjerne gamle data når andre konsumenter er avhengig av dem, noe som begrenser ønsket om å dele data i første omgang.

Konsekvensen av svak kultur for deling er blant annet at tilgangen til relevante data begrenses for konsumentene. Viktig informasjon blir dermed ikke tilgjengeliggjort på en tilfredsstillende måte slik at forskning, politikkutforming, tjenesteutvikling og innovasjon blir svekket i forhold til sitt potensial.

3.2.4. Juridiske utfordringer

Deling av data er regulert både i generelle lover og i en rekke særlover. Offentlige virksomheter er ikke lovpålagt å legge ut data til viderebruk, men dersom de velger å dele data er det mange lover og regler å forholde seg til. En særskilt juridisk utfordring som legger begrensninger for effektiv deling av data er regelverket knyttet til personvern. Dette regelverket er videre beskrevet i kapittel 4.1.1 under juridiske behov. Det strenge regelverket påvirker også de ulike konsumentene av data. Har de for eksempel behov for data som ikke ligger åpent tilgjengelig og inneholder sensitive personopplysninger må de forholde seg til en tidkrevende søknadsprosess.

Regelverket for deling av data setter begrensninger for en effektiv deling, men også usikkerhet knyttet til tolkning av lovverket er en utfordring. Som et eksempel medførte innføringen av ny personopplysningslov (GDPR) at flere dataprodusenter begynte å vegre seg for å dele data med andre aktører. Den nye loven økte usikkerheten rundt hva man kan dele, og flere produsenter har derfor valgt å ikke dele data for å unngå eventuelle lovbrudd. Konsekvensen av dette er at det har blitt mindre deling av data. Et eksempel på dette er hvordan DBH behandler statistikk basert på individdata. I tilfeller med individdata runder de av alle tall til nærmeste fem, og begrunner dette med å henvise til den nye personvernforordningen. Mange UH-institusjoner er uenige i denne praksisen, i og med at sensuren ikke gir mening når informasjonen i utgangspunktet er offentlig.

En annen utfordring som blir trukket frem i intervjuene med konsumentene er det økte behovet for data på individnivå, blant annet for å koble individdata over et livsløp. Dette utfordrer dagens lovverk og flere konsumenter etterspør derfor endringer i regelverket slik at eksisterende data kan benyttes i større grad enn i dag.

Konsekvensen av juridiske hindringer er at det blir vanskeligere for konsumentene å hente ut data og informasjon de trenger for å svare ut problemstillingene de står overfor.

3.2.5. Varierende datakvalitet

Intervjuene med konsumentene viser at kvaliteten på kunnskapsdata som tilgjengeliggjøres har varierende kvalitet. Det er særskilt to forhold på produsentsiden som bidrar til dårlig datakvalitet; mangelfull eller inkonsistent registrering og manglende metadata. I noen tilfeller er det også en utfordring at data er utdaterte i forhold til formålet når de skal benyttes.

Mangelfull registrering kan være forårsaket av både mangel på registrering, men også at data som er registrert er feil eller inkonsistente (eller en kombinasjon av disse). Konsekvensen av de nevnte forholdene er redusert tillit til data, som igjen kan medføre redusert bruk av data. Dette fremhever betydningen av gode rutiner for registrering og kvalitetssikring av data på produsentsiden, og en sterkere samhandling mellom konsumenter og dataprodusenter om formålet med bruken av dataene.

Manglende metadata om variablene trekkes også frem som en utfordring. Dette inkluderer blant annet mangelfull beskrivelse av variablene, hvem som eier dataene, hvor data kan lastes ned og om de er offentlig tilgjengelig. Denne informasjonen er sentral for konsumentene for å kunne benytte dataene på korrekt måte.

Som en forlengelse av de to forholdene nevnt over, er det i dag også lite fokus på dataenes «sporbarhet». I denne sammenhengen innebærer sporbarhet at det må være mulig å reprodusere data for å vite hva som er «gjort» med de underveis i verdikjeden. For å få til god sporbarhet må dataprodusentene derfor være tydelige og beskrive hvordan de har behandlet dataene. Mangel på sporbarhet i data, kan blant annet medføre at beslutninger tas på feil grunnlag fordi en ikke kjenner «historien» til dataene, og ikke vet om, og eventuelt hvordan de er bearbeidet gjennom verdikjeden frem til konsumenten.

Den varierende datakvaliteten kan også skyldes andre forhold enn dataene i seg selv. Dette er beskrevet i tidligere delkapittel. Dagens organisering av data oppleves som uoversiktlig for konsumentene med lav grad av samhandling og koordinering mellom produsentene, se kapittel 3.2.1. Produsentene har ingen tydelig fordeling av roller og ansvar mellom seg, som fører til en fragmentert datainfrastruktur. I tillegg trekkes det frem at det er for liten grad av samhandling mellom konsumentensiden og produsentsiden for tilrettelegging av data med basis i behovene konsumentene har. Situasjonen bærer preg av at «alle har nok med sitt» og det er ingen etablert kultur for, eller arena, hvor denne samhandlingen kan utøves. Det er heller ingen aktør som er overordnet ansvarlig for å sikre at verdikjedene for data i tilstrekkelig grad dekker behovene.

4. Behovsanalysen

Hensikten med *behovsanalysen* er å avdekke de reelle behovene for å utvikle en løsning for deling av data i kunnskapssektoren i fremtiden. Behovsanalysen søker et bredt perspektiv for å sikre at alle vesentlige behov inngår. Behovene er i henhold til KVVU-metodikken delt inn i normative behov, etterspørselsbaserte behov og interessentbaserte behov.

4.1. Normative behov

De normative behovene er behov som er forankret i politiske mål, lover og forskrifter og deles inn i:

- Politiske behov
- Juridiske behov

Politiske drevne behov knytter seg til politiske føringer og strategier. Disse føringene og strategiene er ofte formulert i ulike nasjonale og internasjonale politiske dokument som regjeringen stiller seg bak. De politiske behovene uttrykker et behov for økt digitalisering i utdannings- og integreringspolitikken med mål om å forbedre dataflyten og tilgjengeliggjøringen av data mellom aktørene i samfunnet. Nye behov drives også frem av juridiske endringer og føringer som blant annet kan omhandle lovverket om personvern.

4.1.1. Politiske behov

De politiske behovene belyses gjennom politiske føringer og strategier som tar for seg planlagte og ønskede endringer og tiltak i kunnskapssektoren og samfunnet for øvrig. Gjennom disse politiske strategiene og føringene kommer det frem hvilke behov kunnskapssektoren har i dag og i fremtiden. Det pågår allerede en rekke utviklings- og utredningsprosesser i Norge som omhandler det å tilrettelegge for deling av data. Disse kommer til uttrykk i både stortingsmeldinger, strategier og langtidsplaner:

- «Digital agenda for Norge — IKT for en enklere hverdag og økt produktivitet» (Kommunal- og moderniseringsdepartementet, 2016)
- IKT-sikkerhet — Et felles ansvar» (Justis- og beredskapsdepartementet, 2017)
- Digitaliseringsstrategien for universitets- og høyskolesektoren 2017-2021 (Kunnskapsdepartementet, 2017a)
- «Langtidsplan for forskning og høyere utdanning 2019-2028» (Kunnskapsdepartementet, 2018c)
- Difis tverrgående digitaliseringsstrategi (Direktoratet for forvaltning og IKT, 2019a)
- Digitaliseringsstrategi for grunnsopplæringen for 2017-2021 (Kunnskapsdepartementet, 2017b)
- Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata (Kunnskapsdepartementet, 2017d)
- Nasjonal strategi for digital sikkerhet (Justis- og beredskapsdepartementet, 2019a)
- Digitaliseringsstrategi for offentlig sektor 2019-2025 (Kommunal- og moderniseringsdepartementet, 2019)
- Regjeringens integreringsstrategi 2019–2022 (Kunnskapsdepartementet, 2018b)
- Nasjonal kompetansepolitisk strategi 2017-2021 (Kunnskapsdepartementet, 2017c)

Stortingsmeldingen «Digital agenda for Norge» legger premisset for regjeringen sin overordnede politikk for hvordan IKT skal utnyttes til samfunnets beste (Kommunal- og moderniseringsdepartementet, 2016). «Digital agenda for Norge» trekker frem at IKT-politikken må ta utgangspunkt i utfordringene næringsliv og offentlig sektor står overfor når det gjelder produktivitet, omstilling og effektivisering. I «Digital Agenda» har regjeringen høye ambisjoner om å fornye, forenkle og forbedre offentlig sektor gjennom økt bruk av IKT. Gjennom stortingsmeldingen viser regjeringen til fem hovedprioriteringer for å oppnå dette:

- Brukeren i sentrum
- IKT som en vesentlig innsatsfaktor for innovasjon og produktivitet
- Styrket digital kompetanse og deltakelse
- Effektiv digitalisering av offentlig sektor
- Godt personvern og god informasjonssikkerhet

I de senere årene har departementer og direktorater fulgt opp med egne strategier som omhandler digitaliseringsbehovet innenfor de ulike sektorene. I «Digitaliseringsstrategi for grunnskolesektoren 2017-2021» trekkes det frem konkrete tiltak for datainfrastrukturen i grunnskolesektoren som blant annet omhandler etablering av fellesløsninger for skytjenester, og bedre håndtering av informasjonssikkerhet og personvern (Kunnskapsdepartementet, 2017b).

I 2017 ga Kunnskapsdepartementet ut digitaliseringsstrategien for universitets- og høyskolesektoren for perioden 2017-2021 (Kunnskapsdepartementet, 2017a). I denne strategien trekkes det frem at både utdanning og forskning er områder der IKT-løsninger vil bidra mye i årene fremover. Gjennom digitalisering vil det skapes muligheter for nye og endrede lærings- og undervisningsprosesser, og nye organisasjons- og kommunikasjonsformer. I strategien ble det etablert et målbilde for data og infrastruktur:

- Data lagres én gang og tilgjengeliggjøres fra én kilde
- Data er gjenfinnbare, tilgjengelige, mulig å samhandle med og gjenbrukbare, iht. FAIR-prinsippet
- Infrastrukturen er fleksibel og legger til rette for mobilitet og utvikling
- Helhetlig styring og ledelse av informasjonssikkerhet er et fundament for digitalisering og strategiske satsinger, og bygger opp under sektorens målsetninger

I 2017 kom «Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata» som fremmer behovet for mer deling og gjenbruk av forskningsdata. Strategien trekker frem behovet for utvikling av selvbetjeningsløsninger for å redusere kostnader ved drift av datainfrastrukturer for forskning. Selvbetjeningsløsninger vil kunne forenkle prosesser for deponering av og tilgang til forskningsdata med tilhørende metadata der det er hensiktsmessig. I tillegg har strategien rettet fokus mot behovet for bedre teknisk tilrettelegging og bedre faglig samordning nasjonalt. Regjeringen har også satt seg som mål at alle norske vitenskapelige artikler skal være åpen og tilgjengelige innen 2024. Det trekkes frem at større åpenhet i forskningen er et viktig ledd i arbeidet med å forbedre praksis for å vurdere forskning basert på forskningskvalitet, og ikke de vitenskapelige tidsskriftenes innflytelsesfaktor (Kunnskapsdepartementet, 2017e).

Stortingsmeldingen «IKT-sikkerhet – et felles ansvar» omhandler regjeringens fokus på IKT-sikkerhet. Den trekker blant annet frem at regjeringen ønsker å legge til rette for en langsiktig oppbygging av kompetanse på IKT-sikkerhet gjennom en nasjonal kompetansestrategi. Strategien rettes mot å synliggjøre vårt felles ansvar innen IKT-sikkerhet, styrking av vår nasjonale evne til å avdekke og håndtere digitale angrep, samt å legge til rette for god informasjonsdeling og koordinert hendelseshåndtering (Justis- og beredskapsdepartementet, 2017).

I 2018 presenterte Difi sin tverrgående digitaliseringsstrategi (Direktoratet for forvaltning og IKT, 2019a). Den tverrgående strategien viser hvordan ambisjonene i stortingsmeldingen «Digital agenda for Norge» kan realiseres og består av et målbilde, fire strategiske satsingsområder og en handlingsplan. Under arbeidet ble det identifisert fire hovedutfordringer med tilhørende tiltak:

- Det er behov for sterkere styring, samordning og finansiering av tverrgående digitaliseringstiltak
- Det må utvikles et felles fundament som gjør det enklere å dele data og fellesløsninger
- Det må gis rom for nyskaping og utprøving
- Kompetanse og kultur innen digitalisering må styrkes

I den nyeste utgaven av langtidsplanen for forskning og høyere utdanning (2019-2028) er det blant annet lagt vekt på å tilrettelegge for enklere tilgang til forskningsdata og til offentlige data som brukes i forskning (Kunnskapsdepartementet, 2018c). Det skal i tillegg bli enklere for forskere å bygge på hverandres funn, og etterprøve hverandres resultater. Regjeringen ønsker å øke investeringene i høyere utdanning og forskning, og legge til rette for at kunnskapen tas i bruk. De vil stimulere til økt samspill og samarbeid mellom offentlige og private forskningsaktører, og trekker frem at dette krever en kombinasjon av mangfold og kvalitet.

I januar 2019 ble «Nasjonal strategi for digital sikkerhet» publisert. Strategien er en videreutvikling av tidligere nasjonale strategier og sikter mot styrking av offentlig-privat, sivil-militært og internasjonalt samarbeid. Strategien sikter mot å bedre evnen til å digitalisere, og samtidig bedre evnen til egenbeskyttelse mot uønskede digitale hendelser. Strategien er komplettert av «Nasjonal strategi for digital sikkerhetskompentanse» (Justis- og beredskapsdepartementet, 2019b).

I juni 2019 ble strategien «Én digital offentlig sektor» presentert. Dette er en ny digitaliseringsstrategi for offentlig sektor for tidsperioden 2019-2025 (Kommunal- og moderniseringsdepartementet, 2019). Retningen for digitaliseringsarbeidet i offentlig sektor er i tråd med «Digital Agenda for Norge», og strategien er en oppfølging av denne. Strategien retter seg mot en brukersentrert tjenesteutvikling og mer samordnet utnyttelse av felles IT-løsninger gjennom utvikling av felles økosystem for digital samhandling i offentlig sektor. Samhandlingen krever at offentlig sektor endrer arbeidsmåter og strukturer, har riktig kompetanse til å utnytte digitalisering, og at vi løser flere oppgaver digitalt. Det er spesielt tre mål i strategien som er relevant for denne utredningen:

- Offentlig sektor digitaliseres på en åpen, inkluderende og tillitsvekkende måte
- Offentlig sektor utnytter potensialet i deling og bruk av data til å lage brukervennlige tjenester, og for å bidra til verdiskapning for næringslivet
- Kommunale og statlige virksomheter bygger sine tjenester med utgangspunkt i et felles digitalt økosystem for samhandling

Tiltakene i strategien skal bidra til bedre ressursbruk og en mer effektiv forvaltning. Dette skal oppnås gjennom et styrket samarbeid mellom statlig og kommunal sektor. I tillegg til utvikling av sammenhengende tjenester knyttet til utvalgte livshendelser, økt datadeling og utvikling av et mer digitaliseringsvennlig regelverk, samt tilgang til nødvendig felles funksjonalitet og felles arkitekturer (felles økosystem).

Strategien trekker frem «orden i eget hus» som en viktig forutsetning for økt deling og bruk av data. «Orden i eget hus» er et pålegg statlige virksomheter har fått gjennom det årlige digitaliseringsrundskrivet, med tilhørende veileder fra Difi (Direktoratet for forvaltning og IKT, 2018). «Orden i eget hus» innebærer å beskrive egne data, begreper, informasjonsmodeller og API-er, og å dele data aktivt i tråd med nasjonale retningslinjer. Det mangler imidlertid mekanismer for å følge opp statlige virksomheters etterlevelse av felles instruksjoner gitt i digitaliseringsrundskrivet.

Digitaliseringsstrategien peker også på behovet for sammenhengende offentlige tjenester, økt deling av data og økt bruk av felles IT-løsninger. For å oppnå dette må samhandling og samordning på tvers av sektorer og mellom statlig og kommunal sektor styrkes. Det fremmes et behov for å se eksisterende og planlagte tjenesteutviklingsplattformer som et felles økosystem sammen med grunndataregistre. Dette skal sikre at offentlige virksomheter kan samhandle digitalt, og at de har tilgang til nødvendig fellesfunksjonalitet og felles IT-arkitekturer. For å nå disse målene må økosystemet styres på en samordnet måte, og innholdet som inngår i økosystemet må være samordnet og koordinert. Den mest grunnleggende forutsetningen for deling av data er å vite hvilke data som finnes hvor, hva de betyr og hvorvidt dataene kan deles. I tillegg er det behov for en overordnet plan for fordeling av ansvar for ulike typer data for å redusere fragmentering og overlappende datasett. Like viktig er det å få til en tettere samhandling mellom offentlige etater for å sikre gode verdikjeder for flyt av data for viderebruk.

Det foreligger i dag flere strategier på integrering- og etterutdanningsfeltet. Strategiene tar ikke for seg digitalisering konkret, men økt bruk av digitale verktøy er nevnt som mål i enkelte av strategiene.

Innen voksnes læring er det utarbeidet en nasjonal kompetansepolitisk strategi for perioden 2017-2021 (Kunnskapsdepartementet, 2017c). Denne strategien har et mål om å bidra til at enkeltmennesker og virksomheter har en kompetanse som gir Norge et konkurransedyktig næringsliv, en effektiv og god offentlig sektor, og gjør at færrest mulig står utenfor arbeidslivet. Målgruppen for strategien er hele den voksne befolkningen. For å sikre jobbtilknytning vil de fleste ha behov for å tilegne seg ny kompetanse i løpet av livet. For å nå det overordnede målet for strategien er det identifisert tre hovedinnsatsområder i kompetansepolitikken:

- Gode valg for den enkelte og for samfunnet
- Læring i arbeidslivet og god bruk av kompetanse
- Styrke kompetansen til voksne med svak tilknytning til arbeidslivet

I tillegg fremmes et mål om å styrke og videreutvikle den digitale kompetansen i hele arbeidsstyrken for å utnytte den teknologiske utviklingen og sikre nødvendig omstilling.

4.1.2. Juridiske behov

Regelverket for deling av data er komplisert. I «Digitaliseringsstrategi for offentlig sektor 2019-2025» trekkes det blant annet frem et behov for å tilpasse lovverket og lage et digitaliseringsvennlig regelverk (Kommunal- og moderniseringsdepartementet, 2019). I dag har enkelte departementer ansvar knyttet til tilgjengeliggjøring og deling av data, gjennom deres hovedansvar for de nasjonale grunndataregistrene. Fem av departementene (deriblant Kunnskapsdepartementet) har utarbeidet strategier for økt tilgjengeliggjøring av åpne data, fordi disse anses å være særlig samfunnsøkonomisk verdifulle. Med dette dannes et behov for et lovverk og juridiske føringer for å kontrollere tilgjengeliggjøring og deling av data. Deling av data er regulert i en rekke generelle lover:

- Forvaltningsloven
- Personopplysningsloven inkl. personvernforordningen (GDPR)
- Offentleglova
- Arkivloven
- Sikkerhetsloven

I tillegg er deling av data regulert i en rekke særlover. Særlovene regulerer hvorvidt dataene kan og/eller skal deles og brukes. Handlingsrommet for deling av data er avhengig av type data som skal deles og regulering i lovverket. Offentlige virksomheter er ikke lovpålagt å legge ut data til viderebruk, men dersom de velger å dele data er det mange lover og regler å forholde seg til. Offentleglova inneholder bestemmelser om retten til å få se (innsyn) i dokumenter i offentlig forvaltning og de generelle lovene har krav og prinsipper for hvordan data skal behandles. Formålet med offentliglova er å legge til rette for at offentlige virksomheter er åpne og transparente, i tillegg at det skal legges til rette for videre bruk av offentlig informasjon. Loven definerer enkelte unntak om data som ikke kan eller skal deles.

Personvern handler om retten til et privatliv og retten til å bestemme over egne personopplysninger. Denne rettigheten balanseres mot at legitime formål ofte krever behandling av personopplysninger. Bruk og deling av personopplysninger på individnivå vil være nødvendig for å sikre at det utvikles kunnskapsgrunnlag med kvalitet, legitimitet og relevans i kunnskapssektoren. Personopplysningsloven og GDPR åpner for gjenbruk av personopplysninger til arkivformål og forsknings- og statistikkformål som ikke ansees som uforenelig med de opprinnelige behandlingsformålene. Samtidig stiller personopplysningsloven tydelige krav til behandlingen av personopplysninger. For å kunne benytte data må det foreligge behandlingsgrunnlag gjennom for eksempel en lov hjemmel eller at den registrerte samtykker om bruken. Når formålet med

behandlingen er oppnådd, setter loven krav om at dataene skal slettes. Et prinsipp i lovverket er at data som er samlet inn til et bestemt formål ikke nødvendigvis kan benyttes til et annet formål. Arkivloven og pliktavleveringsloven skal sikre en helhetlig samfunnsdokumentasjon. Formålet med disse lovene er å ta vare på rettighetsdokumentasjon eller dokumentasjon som har verdi for forskning, kultur og forvaltning. Sammen med de utfyllende forskriftene skal arkivloven utgjøre et fullstendig juridisk rammeverk for alle arkivrelaterte spørsmål i offentlig forvaltning – helt fra dokumentet blir skapt som ledd i den daglige virksomheten, via arkivavgrensing og avlevering av bevaringsverdig arkivmateriale, til oppbevaring i depot og tilgjengeliggjøring for ettertiden.

Den nye sikkerhetsloven trådte i kraft 1. januar 2019 (Forsvarsdepartementet, 2018). Den nye loven er tilpasset den teknologiske utviklingen, og vil også sikre informasjon som sendes digitalt. Den nye loven legger til rette for økt samhandling mellom offentlige myndigheter, og mer samarbeid mellom offentlig og privat virksomhet, slik at det forebyggende sikkerhetsarbeidet blir bedre og mer helhetlig.

I juni 2019 ble det vedtatt en ny statistikklov. Den nye statistikkloven erstatter dagens regelverk fra 1989, og skal gi et godt rammeverk for offisiell statistikk og Statistisk sentralbyrå fremover. Loven bygger på statistikklovutvalgets utredning fra 2018, og legger større vekt på personvern sammenliknet med den gamle statistikkloven (NOU 2018: 7, 2018). Den nye loven understreker SSB sin faglige uavhengighet, og viderefører det todelte samfunnsoppdraget med både statistikk og forskning og analyse. Etter den nye loven vil et nasjonalt program for offisiell statistikk angi hva som er offisiell statistikk og hvem som er ansvarlig for å produsere den.

Digital kommunikasjon er nå hovedregelen innenfor forvaltningslovens virkeområde etter at endringer i forvaltningsloven og eForvaltningsforskriften trådte i kraft 7. februar 2014. En lovgivning som støtter digitalisering er et sentralt element i arbeidet for å fornye, forenkle og forbedre forvaltningen. Regjeringen nedsatte et utvalg som skulle revidere forvaltningsloven, og avga sin innstilling i februar 2019. En oppdatert forvaltningslov vil legge bedre til rette for bruk av IKT, samt en god og effektiv saksbehandling tilpasset vår tids forvaltningsoppgaver (NOU 2019: 5, 2019).

4.1.3. Oppsummering

De normative behovene viser at det er delvis motstridende hensyn knyttet til bruk og deling av data. På den ene siden ønsker regjeringen og offentlig sektor å legge til rette for fellesløsninger med mer åpne og tilgjengelige data med et mål om økt deling av data. På den andre siden skal personvern og informasjonssikkerhet ivaretas, noe som kan begrense datadelingen.

4.2. Etterspørselsdrevne behov

De etterspørselsbaserte behovene tar utgangspunkt i et misforhold mellom tilbudt kapasitet og etterspørsel. De er gjerne basert på observerte tilstander i dag og prognoser for framtidig utvikling. Et gap mellom tilbud og etterspørsel kan indikere et samfunnsbehov for å dempe etterspørselen eller skaffe mer kapasitet. En behovskartlegging med utgangspunkt i å løse et gitt problem eller utnytte en identifisert mulighet, vil være inkludert i etterspørselsbaserte behov. I denne utredningen er de viktigste etterspørselsbaserte behovene knyttet til kapasitet og åpenhet for å kunne ivareta dagens og fremtidens behov for data knyttet til utdanning, forskning og integrering. De etterspørselsbaserte behovene kan også omfatte andre forhold som kan ha betydning for infrastruktur, utforming og deling av data (for eksempel teknisk tilstand). De etterspørselsdrevne behovene er delt inn i:

- Teknologisk drevne behov
- Økonomisk drevne behov
- Samfunnsmessig drevne behov

Utredningen tar for seg faktorer som vil påvirke eksisterende og fremtidig tilbud og etterspørsel etter kunnskapsdata, for å få frem eventuelle misforhold mellom de to. Et sentralt spørsmål er om tilbudet utvikler seg tilsvarende som etterspørselen. Samlet sett er det en økende etterspørsel etter data i kunnskapssektoren, men tilbudet av data dekker ikke den økende etterspørselen. Dette er et tegn på at potensialet som ligger i deling og bruk av data ikke utnyttes godt nok i dag.

4.2.1. Teknologisk drevne behov

Ansvar for innsamling og forvaltning av kunnskapsdata er organisert i områder innenfor kunnskapssektoren og ligger hos den enkelte virksomhet. Behov knyttet til for eksempel kunstig intelligens og maskinlæring krever tilgang til store mengder data, og gjerne på tvers av områder og virksomheter. Teknologit utviklingen, fremveksten av stordata og mulighetene til å anvende disse for automatisert oppgaveløsning og analyseformål, vil derfor gi et økt behov for samhandling og tilgang på data på tvers av sektorer fremover. Hovedmålene i «Digital agenda for Norge» er en effektiv offentlig forvaltning, i tillegg til best mulig utnyttelse av muligheter for fornyelse av tjenester og prosesser (Kommunal- og moderniseringsdepartementet, 2016). Dette innebærer at for å kunne oppnå målene om økt deling og bruk av data står utnyttelse av teknologit utviklingen sentralt.

Et mye brukt ord innenfor teknologien de senere årene er digitalisering. Digitalisering omfatter alt fra å automatisere manuelle prosesser til å ta i bruk ny teknologi og skape nye tjenester. Utviklingen handler ikke kun om teknologi, men også om viljen og evnen til å endre seg. I tråd med digitaliseringen, øker også innbyggernes forventninger til kunnskapssektoren. Det er derfor viktig at kunnskapssektoren har en tydelig strategi for å møte forventningene om en digital verden.

En av strategiene som tar innover seg de overnevnte utfordringene er Difis tverrgående digitaliseringsstrategi (Direktoratet for forvaltning og IKT, 2019a). I digitaliseringsstrategien trekker Difi frem hvordan teknologien muliggjør nye måter å levere tjenester på gjennom blant annet automatiserte tjenester som gjør at innbygger sparer tid. I tillegg gir ny teknologi mulighet for å analysere store mengder data for å kunne forstå behovet bedre og dermed tilby økt kvalitet og samhandling. Samtidig legger Difi vekt på innbyggernes økende forventning om øyeblikkelig tilgang og enkle løsninger. En viktig del av utviklingsprosessen er en samordning av tjenester på tvers av sektoren med en helhetlig tilnærming. Difi har identifisert seks behovsområder i offentlig sektor:

- Nye og endrede finansieringsmodeller
- Mer kunnskap og økt kompetanse – kulturendring
- Mer innovasjon og samarbeid med markedet
- Felles rammeverk og løsninger
- Sterkere nasjonal styring, koordinering og samordning
- Et regelverk tilpasset digitalisering

I 2018 utarbeidet Difi en konseptvalgutredning om deling av offentlige data, og anbefaler en smidig tilnærming som tar hensyn til at offentlige virksomheter er på ulike modenhetsnivåer (Direktoratet for forvaltning og IKT, 2018). I KVVU-en trekkes det spesielt frem tre rapporter som omtaler bruk av data i offentlig sektor og understreker behovet for deling av data:

- Digital 21
- Kunstig intelligens: muligheter, utfordringer og en plan for Norge
- IT i praksis 2018

Digital21 ble opprettet av regjeringen for å fremme forslag til en bred og samlet strategi innenfor digitalisering på tvers av næringer og kompetansemiljøer. Det langsiktige målet for Digital21 er å bidra til økt digitalisering i næringslivet. Styringsgruppen for Digital21 har identifisert seks områder som har stor betydning for digitaliseringen. De seks områdene inkluderer blant annet muliggjørende teknologier, kompetanse, dataressurser og infrastruktur og offentlig rammeverk (Digital21, 2018a). I fremtiden vil det være viktig å gjøre dataressurser tilgjengelige for bruk, men også etablere et offentlig

rammeverk som stimulerer til innovasjon og digitalisering. Innenfor området «dataressurser og infrastruktur» trekkes det frem at vekstpotensialet ligger ikke bare i dataene i seg selv, men i en kombinasjon av tilgjengelige og strukturerte data, en infrastruktur som binder alt sammen, tilgang på datakraft og digitale teknologier og kompetanse til å utnytte helheten (Digital21, 2018b). Innenfor området «Dataressurser og infrastruktur» anbefales to viktige tiltaksområder:

- Implementere en aktiv strategi for datafangst og -deling
- Næringsrettet satsning på digital infrastruktur

Sentrale hovedelement som blir trukket frem er å etablere én dataportal for å gjøre offentlige datasett tilgjengelige, samt å sørge for åpne API-er. Samtidig er det viktig å rette fokus mot nasjonal kontroll for å sikre like vilkår for tilgang og bruk, dekning og sikkerhet.

I Teknologirådet sin rapport trekker de frem en rekke anbefalinger knyttet til teknologiutvikling og kunstig intelligens spesielt (Teknologirådet, 2018). Anbefalingene inkluderer åpne offentlige data og datadeling som tjener fellesskapet. Med dette legger de vekt på at data skal tilgjengeliggjøres, men at dersom data skal gjøres tilgjengelig bør fellesskapet få en merverdi i form av eksempelvis bedre offentlige tjenester eller næringsutvikling. I tillegg påpekes det at det er nødvendig at myndighetene etablerer juridiske rammer som gjør det mulig å utveksle data trygt og som sikrer at rettigheter, verdier og ansvar fordeles rettferdig og likeverdig.

I den årlige undersøkelsen «IT i praksis» kom det frem i 2018 at det har vært en stor økning av bruk av åpent tilgjengelig data fra de mest etablerte tilbyderne av data i offentlig sektor (Rambøll, 2018). Dette underbygger et behov for økt deling og åpenhet innenfor offentlig sektor.

4.2.2. Økonomisk drevne behov

De økonomisk drevne behovene omhandler kostnader som påløper i dagens situasjon der det kan være muligheter for å bruke ressursene på en bedre måte. Dette kan være kostnader som påløper ved søknad om tilgang til data, kostnader for analysekapasitet og utvikling av interne infrastrukturer for å løse utfordringer som oppstår. Det er behov for tiltak og løsninger som minimerer disse kostnadene mest mulig. Kostnadene for lagring av data er redusert til en brøkdel av hva det var for noen år tilbake, dette gjør det mulig å lagre, prosessere og utnytte data i et helt annet omfang i dag enn for bare få år siden (Digital21, 2018a).

En konsekvens av dagens situasjon er fragmenterte data og stor grad av dupliserende datasett innenfor flere områder, nærmere beskrevet i kapittel 3.2. En bedre dataflyt i kunnskapssektoren vil kunne gi store økonomiske gevinster. I dag har enkeltaktører høye kostnader knyttet til å lage interne løsninger som «kompenserende tiltak» for å løse utfordringene knyttet til manglende tilgang. Det er tidkrevende, dyrt og lite effektivt når flere aktører lager egne løsninger på bakgrunn av det samme behovet. Det er derfor behov for infrastruktur som åpner for enklere deling av data, slik at kostnader knyttet til tilgjengeliggjøring og uthenting av data kan reduseres og behovet for egne løsninger reduseres.

I dag er mange av prosessene knyttet til datadeling manuelle, der man klargjør utlevering som én flatfil av gangen. Dette skyldes liten bruk av maskinell deling av data gjennom API-er. Større bruk av API som grensesnitt for tilgang til data ville økt effektiviteten i datadelingen og redusert kostnader for både konsument og produsent. For å få til dette vil det være nødvendig med et arbeid knyttet til standardisering der data ikke endres uten å ta innover seg helheten og hele verdikjeden. Dette vil tilrettelegge for felles løsninger som igjen vil kunne redusere kostnader og øke effektiviteten.

4.2.3. Samfunnsmessig drevne behov

Samfunnsmessig drevne behov knytter seg blant annet til den økende etterspørselen etter kunnskapsdata og behovet for å utvikle bærekraftige løsninger for fremtiden. Etterspørselen etter data skaper et behov for mer tilgjengelig data fra kunnskapssektoren.

En økt etterspørsel etter data, og et tilbud som ikke møter denne etterspørselen tyder på at potensialet for bruk og deling av data ikke er fullt utnyttet. I 2018 poengterte Riksrevisjonen i sin gjennomgang av digitalisering i statlige virksomheter at statlige virksomheter kun delvis gjenbraker informasjon, selv om behovet for gjenbruk er stort (Riksrevisjonen, 2018). I undersøkelsen oppga også en stor andel av respondentene at informasjonen som er aktuell for gjenbruk ofte ikke eksisterer i et digitalt format. Dette viser at behovet for gjenbruk i mange tilfeller ikke blir dekket. En av årsakene som blir trukket frem er at deling av data er ressurskrevende, og prioriteres derfor ikke.

For å imøtekomme behovet for økt tilbud av data innenfor kunnskapssektoren vil det være behov for en kapasitetsøkning som må dekkes gjennom større grad av automatisering. I dag samles det inn store mengder data, og i fremtiden vil mengden data øke gjennom mer effektive metoder for innsamling og registrering. Dette krever en robust og skalerbar infrastruktur som kan samle et økende volum av data. Fleksible løsninger som kan tilpasses kapasitetsøkningen som kreves i de kommende årene er et viktig behov å dekke for kunnskapssektoren.

4.2.4. Oppsummering

Gevinster ved digitalisering blir i flere tilfeller ikke realisert. Det er derfor et behov for å implementere tiltak som kan føre til økt gevinstrealisering i kunnskapssektoren. Teknologiske, økonomiske og politiske drivere vil medføre økt etterspørsel etter kunnskapsdata i fremtiden, og økte muligheter for å utnytte dataene. I flere utredninger og strategier innenfor kunnskapssektoren understrekes behovet for å gjøre data mer åpne og tilgjengelige. Dette underbygger behovet om tiltak i kunnskapssektoren som kan bidra til å skape en harmonisert dataflyt. Dagens oppgaveløsning utfordres av økende informasjonsmengder, digitalisering og nye kanaler. Ved å ta i bruk teknologi og automatisering for å integrere nye løsninger og verktøy i de eksisterende løsningene er det et stort potensial for store besparelser og økt kvalitet på dataene.

I offentlig sektor er det et raskt voksende antall etater som registrerer data og kommuniserer med hverandre. Data som omhandler individer, andre offentlige data og øvrige dataressurser krever ulik forvaltning. Dersom data blir forvaltet riktig, med fokus på tilgjengelighet og sikkerhet, vil det danne et godt grunnlag for dataflyt i kunnskapssektoren.

4.3. Interessentbaserte behov

Identifisering av interessenter og aktører utføres for å kartlegge hvilke interessenter som berøres av utredningen og videre avdekke eventuelle interessekonflikter. Interessentanalysen skal ifølge Finansdepartementets veileder (2010) fokusere på interessenter som har betydning for behovet. I beskrivelsen av de interessentbaserte behovene er det fokusert på behovene til de ulike konsumentgruppene. Dette presenteres i de påfølgende delkapitlene.

4.3.1. Forskere

På bakgrunn av utfordringene beskrevet i kapittel 3.2, kan behovene til forskere oppsummeres i:

- Det er behov for bedre harmonisering av kunnskapsdata og bedre kvalitet på utleverte datasett.
- Det er behov for oversikt over tilgjengelig kunnskapsdata med utfyllende informasjon om dataene.

- Det er behov for tilgang til kunnskapsdata på individnivå, særskilt data knyttet til barnehage og grunnskole.
- Det er behov for en enhetlig behandling av søknader og tillatelser for data med begrenset offentlighet.
- Det er behov for å redusere ressursbruken knyttet til søknader om, behandling av og utlevering av data.

4.3.2. Politikktutviklere

På bakgrunn av utfordringene beskrevet i kapittel 3.2, kan behovene til politikktutviklere oppsummeres i:

- Det er behov for bedre beskrivelser av tilgjengelige kunnskapsdata.
- Det er behov for mer utfyllende datasett.
- Det er behov for bedre harmonisering av kunnskapsdata.
- Det er behov for økt vilje blant dataprodusentene til å dele kunnskapsdata.

4.3.3. Underliggende etater

På bakgrunn av utfordringene beskrevet i kapittel 3.2, kan behovene til de underliggende etatene oppsummeres i:

- Det er behov for tilgang til kunnskapsdata på individnivå.
- Det er behov for bedre beskrivelser av tilgjengelige kunnskapsdata.
- Det er behov for bedre harmonisering av kunnskapsdata.
- Det er behov for mer utfyllende datasett.
- Det er behov for økt vilje blant dataprodusentene til å dele kunnskapsdata.

4.3.4. Tjenesteprodusenter

På bakgrunn av utfordringene beskrevet i kapittel 3.2, kan behovene til tjenesteprodusentene oppsummeres i:

- Det er behov for bedre tilrettelegging av eksisterende data og økt tilgang til disaggregerte data.
- Det er behov for bedre beskrivelser av tilgjengelige kunnskapsdata.
- Det er behov for bedre harmonisering av kunnskapsdata.
- Det er behov for bedre rutiner og kvalitetssikring i dataprodusentenes rapportering.
- Det er behov for bedre formidling av oppdaterte data.

4.3.5. Næringslivet

På bakgrunn av utfordringene beskrevet i kapittel 3.2, kan behovene til næringslivet oppsummeres i:

- Det er behov for økt tilgang til kunnskapsdata.
- Det er behov for bedre beskrivelser av tilgjengelige kunnskapsdata.
- Det er behov for bedre harmonisering av kunnskapsdata.

4.3.6. Allmennheten

På bakgrunn av utfordringene beskrevet i kapittel 3.2, kan behovene til allmennheten oppsummeres i:

- Det er behov for bedre oversikt over tilgjengelige kunnskapsdata.
- Det er behov for enkle og brukervennlige fremstillinger av dataene.
- Det er behov for bedre beskrivelser av tilgjengelige kunnskapsdata.
-

4.3.7. Oppsummering

I Figur 13 er de ulike behovene til konsumentgruppene slått sammen og presentert på et mer overordnet nivå. Dette viser at det er mange overlappende behov mellom de ulike konsumentgruppene. Det er spesielt avdekket et behov for bedre datakvalitet og en bedre oversikt og beskrivelse av tilgjengelige kunnskapsdata.

Behov	 Forskere	 Politik- utviklere	 Underliggende etater	 Tjeneste- produsenter	 Næringsliv	 Allmennheten
Det er behov for bedre datakvalitet	✗	✗	✗	✗	✗	
Det er behov for bedre oversikt og beskrivelser av data.	✗	✗	✗	✗	✗	✗
Det er behov for økt tilgang til kunnskapsdata	✗		✗	✗	✗	
Det er behov for økt vilje blant dataprodusentene til å dele kunnskapsdata.		✗	✗	✗		
Det er behov for enhetlig og ressurseffektiv søknadsprosess knyttet til søknader om behandling av og utlevering av data.	✗		✗			
Det er behov for enkle og brukervennlige fremstillinger av data.						✗

FIGUR 13: KONSUMENTENES BEHOV

4.4. Prosjektutløsende behov

Et prosjektutløsende behov er det samfunnsbehovet som utløser gjennomføringen av et prosjekt på et bestemt tidspunkt. I utredningen er det prosjektutløsende behovet definert som følger på bakgrunn av problembeskrivelsen:

*Samfunnet har behov for bedre datakilder og at tilgangen blir raskere, rimeligere og sikrere. Dette innebærer at datakildene skal tilrettelegges for kobling til data og statistikk, samt publisering av statistikk og analyser. For å få til dette er det behov for **bedre rammebetingelser og felles datainfrastruktur** for deling av data i kunnskapssektoren.*

Behovet for bedre rammebetingelser og felles datainfrastruktur er knyttet til problembeskrivelsen som viser at roller og ansvar innenfor dagens dataforvaltning er uklare, jf. Figur 14.

I tillegg viser den at det er manglende styring og samordning av tverrgående digitaliseringstiltak. Konsekvensen er blant annet fragmentert datainfrastruktur og økte kostnader til databehandling. Videre er det behov for sterkere samhandling om datadeling for å sikre at data som deles er gjenspeilet i behov med bakgrunn i verdikjeder for deling av data.

Problembeskrivelsen viser også at det er varierende datakvalitet, som bidrar til feil bruk av data og svekket tillit til dataene. I tillegg er det avdekket at mange virksomheter har ikke en kultur for deling av data. Konsekvensen av dette er at mye relevant informasjon blir ikke tilgjengeliggjort.

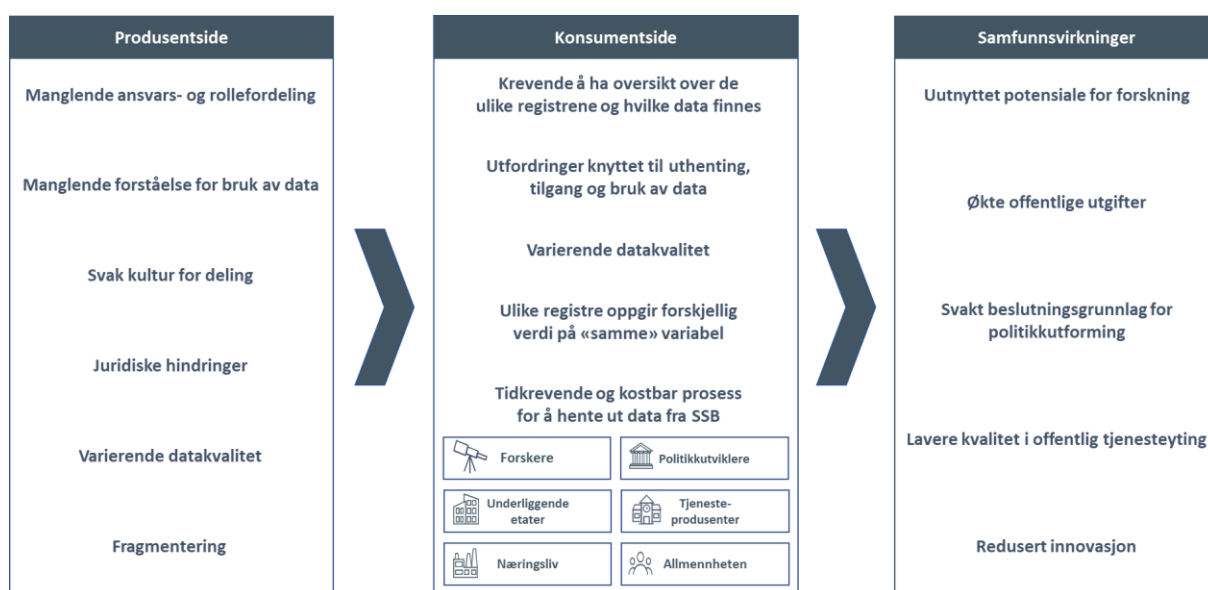
Problembeskrivelsen har avdekket at det må være flere rammebetingelser på plass for at fremtidens deling av data skal fungere tilfredsstillende. Oppsummert kan rammebetingelsene deles inn i:

- Det er behov for en tydeligere fordeling av roller og ansvar blant dataprodusentene.
- Det er behov for økt koordinering og samhandling mellom dataprodusentene.
- Det er behov for bedre styring og myndighetsutøvelse i hele sektoren.

- Det er behov for tettere samhandling mellom konsumenter og dataprodusenter for å oppnå et tydeligere fokus på nytteverdien av dataene for konsumentene.

Til sammen innebærer dagens rammebetingelser og datainfrastruktur for deling av data negative virkninger for samfunnet i form av:

- Uutnyttet potensiale for forskning
- Økte offentlige utgifter
- Svakt beslutningsgrunnlag for politikkutforming
- Lavere kvalitet i offentlig tjenesteyting
- Redusert innovasjon



FIGUR 14: OVERSIKT OVER PROBLEMBESKRIVELSEN OG SAMFUNNSVIRKNINGER

5. Mål for den fremtidige infrastrukturen

I kapittel 4.4 ble sentrale faktorer for å endre dagens situasjon for deling av data oppsummert i et prosjektutløsende behov. Med utgangspunkt i problembeskrivelsen, skal det defineres ett eller flere mål for hva som er en ønsket tilstand for den fremtidige løsningen for deling av data i kunnskapssektoren (Direktoratet for økonomistyring, 2018). I denne utredningen er det skilt mellom samfunns mål, effektmål og resultatmål. Kapittel 5.1 inneholder en nærmere beskrivelse av forskjellen på disse målene. De definerte målene for konseptutredningen er deretter presentert i de påfølgende delkapitlene.

5.1. Om mål

Samfunns målet gir uttrykk for den nytte og verdiskapning tiltaket skal føre til for samfunnet (Finansdepartementet, 2010). Målet skal vise intensjonen og ambisjonen med tiltaket, og samtidig reflektere det prosjektutløsende behov. Eksempler på samfunns mål er god folkehelse, at det oppleves trygt å oppholde seg i det offentlige rom i storbyer eller overgangen til en grønn økonomi.

Effektmål skal bygge opp under samfunns målet og er en konkretisering av hva som ønskes oppnådd for tiltakets målgruppe(r). Effektmål bør i størst mulig grad uttrykke ønskede virkninger av eventuelle tiltak på området. Hvor mange effektmål det er hensiktsmessig å sette opp, vil variere fra analyse til analyse. Dersom det utformes flere effektmål, bør de være innbyrdes konsistente.

Resultatmål refererer til alle tiltak som skal bidra til at effektmålene innfris. Hvis det ikke er en klar sammenheng mellom målene og problembeskrivelsen, er det en fare for at man ender opp med tiltak som i liten grad bidrar til å løse problemet. I tillegg er det viktig at målene ikke utformes så snevert at de virker begrensende på muligheten til å finne relevante tiltak. Derfor er det viktig å fokusere på hva man ønsker å oppnå med tiltaket, og ikke på selve løsningen. Samtidig er det viktig å ikke formulere målene på en så vag måte at det man ønsker å oppnå blir utydelig. Da blir det utfordrende å vurdere relevansen av de ulike tiltakene.

Målene som er formulert for den fremtidige løsningen for deling av data i kunnskapssektoren er oppsummert i Figur 15.



FIGUR 15: OVERSIKT OVER SAMFUNNSMÅL OG EFFEKTMÅL FOR FREMTIDIG INFRASTRUKTUR I KUNNSKAPSSEKTOREN

5.2. Samfunnsmål

Det formulerte samfunnsmålet for prosjektet er:

Fremtidig deling av data i kunnskapssektoren skal bidra til økt deltakelse i arbeids- og samfunnsliv, økt læringseffekt i utdanning og forskning, riktige prioriteringer, effektiv ressursbruk og høy verdiskapning

For samfunnet er det overordnede målet at den nye løsningen for deling av data i kunnskapssektoren skal bidra til at flere kommer i arbeid og deltar i samfunnslivet med en kompetanse som er tilpasset fremtidens behov. Høy yrkesdeltakelse er en av bærebjelkene i det norske velferdssamfunnet, samtidig som arbeid er viktig for helse, trivsel og livskvaliteten for hver enkelt innbygger. Et viktig grunnlag for deltakelsen i arbeids- og samfunnsliv, er at opplæring og utdanning har høy kvalitet og at det er en god integrering av innvandrere. Opplæring og utdanning omfatter alle former for læring gjennom hele livet, og er sentrale tjenester i kunnskapssektoren, i tillegg til integrering gjennom bosetting og opplæring i norsk, samfunnsfag m.m. En viktig faktor for å skape gode tjenester i kunnskapssektoren er tilgang på data av høy kvalitet. Økt tilgang til data av høy kvalitet vil også legge til rette for mer og bedre forskning, analyser og utredninger, som videre bidrar til et bedre beslutningsgrunnlag. Dette vil gi bedre grunnlag for å gjøre riktige prioriteringer og bruke ressursene på en effektiv måte, samt bidra til å utvikle tjenester og produkter som gir økt økonomisk verdiskapning og økt innovasjon, både i offentlig og privat sektor.

5.3. Effektmål

Med utgangspunkt i overnevnte målsetting er det utviklet fire effektmål.

Høy datakvalitet

Data av god kvalitet skal tilfredsstillende krav til korrekthet, kompletthet, tidsriktighet og konsistens. Høy datakvalitet innebærer også at dataene har evnen til å støtte formålene de skal brukes til, være tilgjengelig for konsumentene når de trenger det, og i et format de kan benytte. Dette henger også sammen med i hvor stor grad data oppfyller nåværende og fremtidige behov for informasjon hos konsumentene. Datakvaliteten skal bedres gjennom økt samhandling i kunnskapssektoren, og en økt teknisk og faglig kompetanse som skal være med å sikre bedre registrering av kunnskapsdata.

Rask tilgang til data

For å kunne ta datadrevne beslutninger er det viktig at datasettene er tidsaktuelle og relevante. Rask tilgang på data er derfor viktig for at beslutningsgrunnlaget er mest mulig oppdatert, og for at konsumentene skal ha en tiltro til at datasettene i størst mulig grad gjenspeiler dagens virkelighet. Det er også viktig med rask tilgang på data innenfor forskning og analyse. En lang utleveringstid av registerdata gjør at studenter og stipendiater sjeldent har mulighet til å benytte registerdata i sin forskning. Raskere tilgang på kunnskapsdata gir dermed muligheter for mer forskning og analyse for å kunne finne effekter og årsakssammenhenger som ligger i dataene.

Tilgjengelig data

For å utnytte potensialet i dataene i kunnskapssektoren er det avgjørende at kunnskapsdata gjøres lettere tilgjengelig for flere konsumenter enn i dag. Dette vil gjøre at mer kunnskapsdata kan brukes i forskning og analyse. Det at flere får tilgang til kunnskapsdata gjør at dataene kan benyttes i sammenhenger med andre typer data, og bidra til ny kunnskap. Dette skaper behov for et tilpasset og oppdatert regelverk som setter klarere grenser for hva som kan deles og hvordan. Et tilpasset og oppdatert regelverk vil være med på å sikre en økt samhandling mellom organisasjoner, og gi tekniske løsninger med felles forståelse av dataenes innhold innad i kunnskapssektoren. Data blir lettere å finne frem til, lettere å koble og lettere å hente ut.

Sikker datahåndtering

Innføring av ny personvernlovgivning og tilfeller både i Norge og verden hvor personer har fått uautorisert tilgang til data gjennom hackerangrep har skapt et økt fokus på informasjonssikkerhet. I den sammenheng er det viktig at kunnskapsdata i den fremtidige løsningen blir håndtert på en sikker måte for at folk skal ha tillit til både løsningen og institusjonene som behandler de registrertes personopplysninger. For å tilby en sikrere tilgang til kunnskapsdata er det derfor viktig at personopplysningene behandles tilstrekkelig sikkert ved bruk av egnede tekniske og organisatoriske tiltak. Dette vil si at opplysningene er vernet mot uautorisert eller ulovlig behandling og mot utilsiktet tap, ødeleggelse eller skade.

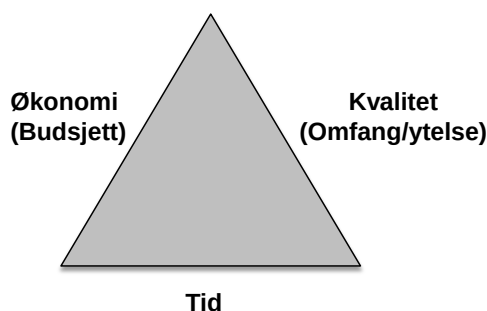
Ressurseffektiv forvaltning av data

Ved å sette klarere styringslinjer og en tydeligere rolle- og ansvarsfordeling i forvaltningen av data vil ressursene bli brukt på en hensiktsmessig måte. Det er viktig med tydelig definerte ansvarsområder for å unngå at alle lager egne løsninger som åpner opp for fragmentering og overlapp. Dette handler om å etablere en omforent informasjonsarkitektur for data i kunnskapssektoren, som vil skape en mer koordinert sektor hvor det er enklere å samhandle gjennom deling av data, og dermed skape en ressurseffektiv forvaltning.

5.4. Resultatmål

Tid, økonomi og omfang utgjør de sentrale styringsmessige parametre i prosjekter. Den såkalte prosjekttrekanten, vist i Figur 16, gjelder universelt for alle prosjekter og illustrerer at prosjektets resultatmål er innbyrdes avhengige av hverandre. Et press på økt omfang (bedre ytelse/høyere krav) skaper eksempelvis et press på økte kostnader. Et press på tid (forsering) skaper et press på økte kostnader. Et press på lavere kostnader skaper et press på lavere ytelser osv.

Disse sammenhengene skaper **styrbarhet** i prosjekter. Summen av måloppnåelsen på disse områdene vil til slutt fastsette hvor godt effekt- og samfunns mål oppnås. Ved fastsetting av detaljerte styringsmål innenfor hvert område er det derfor viktig å ha klart for seg denne innbyrdes sammenheng (som vist i Figur 16).



FIGUR 16: PROSJEKTTREKANTEN

For tiltaket anbefales at resultatmålene prioriteres som følger:

1. Kostnad
2. Tid
3. Kvalitet (Omfang/ytelse)

Begrunnelse

- En stram kostnadsstyring innenfor vedtatt omfang kan oppnås gjennom en hensiktsmessig fastsetting av styringsmål og styringsrammer på ulike organisatoriske nivåer. Midler til investeringer vil alltid være en knapp faktor, og kostnad er et uttrykk for hvor mye midler som eventuelt kunne vært benyttet på andre prosjekter. Ut fra dette vurderes kostnader som viktigere enn tid.
- Et nytt konsept for deling av data bør komme raskt i gang, men kostnadsstyring og funksjonsevnen er likevel viktigst. Fremdriften bør sees i sammenheng med andre initiativer som har grensesnitt til anbefalt konsept. Se mer omtale om dette i kapittel 9.
- Det vil være avgjørende for arbeidet med deling av data og dets langsiktige suksess, at tiltaket fungerer effektivt og godt i hele sin levetid.

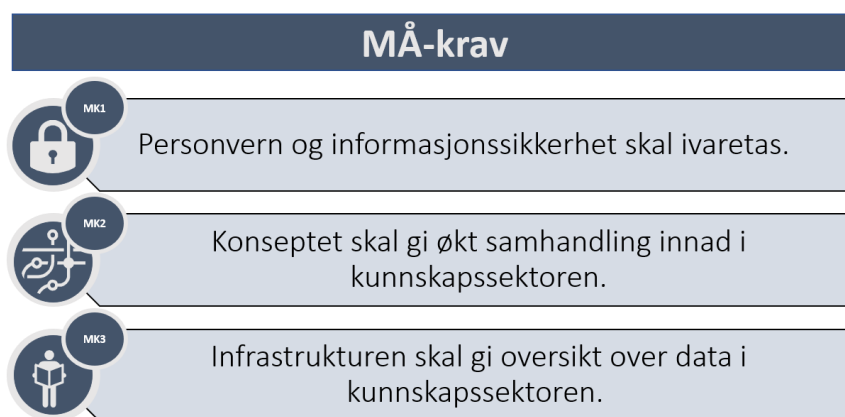
Prioritering av resultatmålene innebærer ikke en aksept av en manglende oppfyllelse av de resultatmålene som er prioritert som nummer 2 og 3 – kun en rangering for de tilfellene det blir konflikt mellom dem.

6. Krav til infrastrukturen

Kravene sammenfatter de betingelsene som skal oppfylles av den fremtidige løsningen for datadeling i kunnskapssektoren. De er i hovedsak identifisert på bakgrunn av de ulike behovene, både de normative-, etterspørselsdrevne- og interessentbaserte behov. Kravene fører til at mulighetsrommet for den fremtidige datadelingen blir avgrenset, noe som blir hensyntatt i utformingen av konseptene i kapittel 7. Kravene er definert i MÅ-krav og BØR-krav. MÅ-kravene er krav som skal oppfylles, mens BØR-krav er krav som er vurdert som sentrale, og som konseptene bør innfri i størst mulig grad.

6.1. MÅ-krav

Det er definert tre MÅ-krav til den fremtidige infrastrukturen. Hvert konsept som er vurdert i mulighetsstudien, må innfri alle tre kravene for at det skal utredes videre i analysen. Figur 17 gir en oversikt over de identifiserte MÅ-kravene.



FIGUR 17: MÅ-KRAV TIL LØSNING FOR DATAFLYT I KUNNSKAPSSEKTOREN

MK1: Personvern og informasjonssikkerhet skal ivaretas

Personvernlovgivningen stiller omfattende krav til behandling av direkte eller indirekte personidentifiserbare opplysninger. I tillegg gis de registrerte en rekke rettigheter, som innsyn i egne personopplysninger.

Det er behov for mekanismer som gir tillit til at alle aspekter ved informasjonssikkerhet er tilfredsstillende ivaretatt og det er en forutsetning at en fremtidig løsning skal etterleve det til enhver tid gjeldende regelverk for informasjonssikkerhet og behandling av personopplysninger.

Konseptene skal legge til rette for ivaretagelse av kravene i personvernlovgivningen og tilfredsstillende informasjonssikkerhet. Hvor omfattende tiltak som må iverksettes for slik ivaretagelse vil avhenge av det enkelte konsept og de generelle føringene som er omtalt i kapittel 7.2.1.

MK2: Konseptet skal gi økt samhandling innad i kunnskapssektoren

Konseptene skal legge til rette for økt samhandling i kunnskapssektoren. Norsk arkitekturrammeverk for samhandling er et rammeverk som skal understøtte samhandling. Dette er en norsk tilpasning av EUs European Interoperability Framework (EIF) som Norge har forpliktet seg til å implementere (Direktoratet for forvaltning og IKT, 2019b). I henhold til rammeverket så er det flere samhandlingsområder som må avklares for å oppnå en god og effektiv samhandling. Som vist i Figur 18 deler rammeverket samhandling opp i fem områder:

- *Juridisk samhandling* skal sikre at organisasjoner som arbeider under ulik lovgivning kan samhandle.
- *Organisatorisk samhandling* handler om å organisere arkitekturarbeid for å oppnå felles mål. Det omhandler hvordan organisasjonen styres og samhandler i praksis.
- *Semantisk samhandling* sikrer at formål og mening med informasjon er bevart og forstått gjennom en utveksling. Dette innebærer felles begrepsdefinisjoner.
- *Teknisk samhandling* sikrer at ulike systemer kan «snakke sammen».
- *Styring og forvaltning av integrerte offentlige tjenester* dekker samtlige av de andre samhandlingsområdene. Med styring og forvaltning menes det oppfølging av rammeverk, organisatoriske strukturer, roller og ansvar, retningslinjer, avtaler og annet som bidrar til samhandling på nasjonalt nivå.



FIGUR 18: SAMHANDLINGSOMRÅDER I NORSK ARKITEKTURRAMMEVERK FOR SAMHANDLING

Norsk arkitekturrammeverk for samhandling skal hjelpe offentlige virksomheter med å definere, designe, utvikle og forvalte digitale tjenester for utveksling av data i offentlig sektor. Ved at den fremtidige løsningen for deling av data skal forholde seg til arkitekturrammeverket innebærer det at samhandling innad i kunnskapssektoren styrkes med utgangspunkt i de definerte samhandlingsområdene.

MK3: Infrastrukturen skal gi oversikt over data i kunnskapssektoren

Kravet innebærer at det må skapes en bedre totaloversikt over hvilke data som er tilgjengelig for de formålene som konsumentene skulle ha. En slik oversikt må være lett tilgjengelig og dataene som inngår må beskrives med metadata som definerer hva dataene betyr, hvem som er ansvarlig for dataene, kriterier for bruk (om de er åpne, beskyttet gjennom lovhjemmel, krever samtykke, etc.), dataenes kvalitet og hvor ofte de oppdateres.

For å oppnå en slik totaloversikt må dataprodusentene etablere oversikt over hvilke data de håndterer og som de kan dele med andre. Produsentene må beskrive egenskapene ved dataene slik at informasjonen kan samles i en totaloversikt over data for hele sektoren. Bedre oversikt over data betyr at det blir enklere for konsumentene å finne frem til eksisterende kunnskapsdata, forstå dataene og ta dem i bruk.

6.2. BØR-krav

Det er definert fire BØR-krav der alle er vurdert som sentrale i konseptene for deling av data. BØR-kravene er presentert i Figur 19.



FIGUR 19: BØR-KRAV TIL LØSNING FOR DATAFLYT I KUNNSKAPSSEKTOREN

BK1: Det bør legges til rette for samhandling med andre sektorer

Samhandling med andre sektorer vil gi tilgang til et større utvalg av data, som vil gi nye og utvidede analysemuligheter rundt sammenhenger som involverer prosesser i flere sektorer. Det vil i tillegg bidra til å se data fra kunnskapssektoren i en større sammenheng, f.eks. å analysere hvordan effekter av ordninger i kunnskapssektoren slår ut i arbeidsmarkedet.

BK2: Det bør være mulig med samhandling om datadeling mellom offentlig og privat sektor

Samhandlingen om datadeling mellom det offentlige og private vil skape et tettere samarbeid, og dermed gjøre mer data fra kunnskapssektoren tilgjengelig for privat sektor, og motsatt. Dette vil gi utvidede analysemuligheter for de offentlige aktørene, og samtidig sørge for innovasjonsmuligheter i næringslivet ved at de kan utvikle tjenester som i enda større grad møter konsumentenes behov.

BK3: Det bør utprøves nye former for deling og bruk av data

Ved å prøve ut nye former for deling og bruk av data vil det legges til rette for at kunnskapssektoren er mer innovativ rettet, og dermed kunne skape et samarbeid med andre private og offentlige virksomheter. Utprøving av nye former gir også muligheter for raskt å ta i bruk nye og gode løsninger.

BK4: Det bør bygges videre på eksisterende teknisk infrastruktur og løsningskomponenter

Det å utvikle helt nye løsninger medfører en risiko da mange IT-prosjekter havarerer eller klarer ikke å oppnå sine fastsatte mål. Ved å bygge videre på eksisterende teknisk infrastruktur og løsningskomponenter unngår man samtidig at det blir en overlappende funksjonalitet i en eventuell ny løsning sammenliknet med det som allerede eksisterer og vil være tilgjengelig.

7. Mulighetsstudie

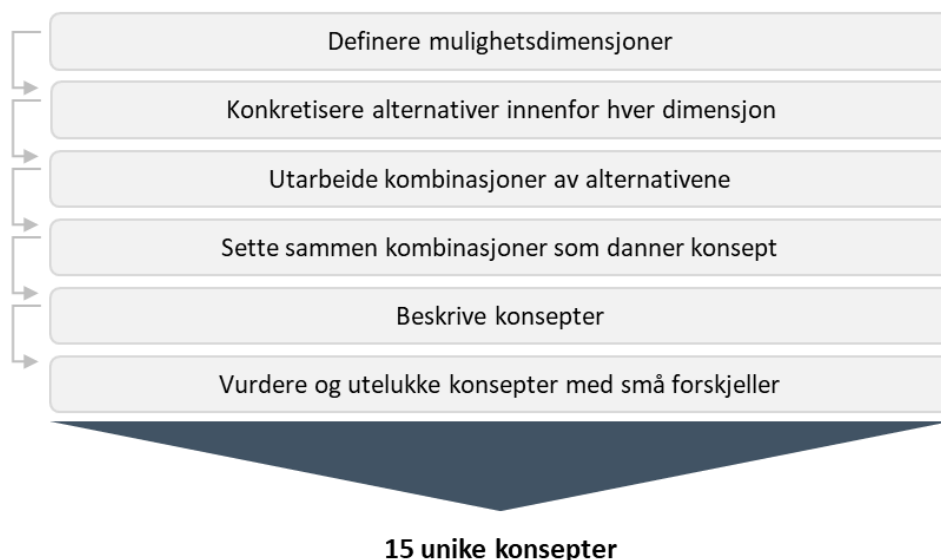
Hensikten med *mulighetsstudien* er å identifisere mulige tiltak for å forbedre infrastrukturen for deling av data i kunnskapssektoren. Grunnlaget for dette arbeidet er problembeskrivelsen i kapittel 3, behovsanalysen i kapittel 4, og målene for den framtidige infrastrukturen som er beskrevet i kapittel 5.

I analysen er det først gått bredt ut ved at det undersøkt flere måter å endre infrastrukturen på, dvs. at i henhold til KVVU-metodikken er *mulighetsrommet* blitt utforsket. Dette er gjort rede for i kapittel 7.1. På bakgrunn av dette er det konkretisert ulike modeller for den fremtidige infrastrukturen, som i henhold til KVVU-metodikken er kjent som *konsepter*. I denne fasen av analysen i en KVVU er det vanlig å gå bredt ut og utforme mange ulike konsepter for å vise spennet av ulike løsningsalternativer. Konseptene er presentert i kapittel 7.2. I neste steg er alle konseptene vurdert langs flere kriterier, bl.a. kravene til infrastrukturen som er beskrevet i kapittel 6. Denne prosessen er kjent som *nedvalg av konsepter* i KVVU-metodikken, og er gjort rede for i kapittel 7.3. På bakgrunn av vurderingen er enkelte av konseptene tatt med videre til en grundigere analyse av kostnader og gevinster. Denne analysen er kjent som *alternativanalysen*, og den er presentert i kapittel 8.

7.1. Mulighetsrommet

Mulighetsrommet er definert for å få oversikt over alle alternativer som kan være en mulig løsning på en gitt problemstilling, uavhengig av kompleksitet og sannsynlighet. For å utforme konsepter for fremtidig infrastruktur for deling av kunnskapsdata er det etablert en rekke mulighetsdimensjoner, der ulike valg i de enkelte mulighetsdimensjonene definerer fundamentale forskjeller mellom konseptene.

Figur 20 viser en overordnet gjengivelse av prosessen fra definering av mulighetsdimensjoner til utforming av 15 unike konsepter som er presentert i kapittel 7.2.



FIGUR 20: BESKRIVELSE AV PROSESSEN MED Å UTVIKLE KONSEPTER

De to første stegene i Figur 20 viser til arbeidet med å utforske mulighetsrommet. I utredningen er det totalt fire mulighetsdimensjoner; organisasjon, funksjonalitet, fagområde og konsumentgrupper. Innenfor hver mulighetsdimensjon er det flere ulike alternativer. Dette er illustrert i

Figur 21. De ulike alternativene varierer fra små endringer til mer komplekse alternativer for utforming av infrastrukturen. Nedenfor følger en mer detaljert beskrivelse av hver mulighetsdimensjon

Organisasjon	Funksjonalitet	Fagområde	Konsumentgrupper
Roller og ansvar som i dag	Metadatakatalog	Barnehage	Forskere
Tydeligere ansvars- og rollefordeling	Eksplorative tjenester	Grunnskole	Politikktviklere
Områdesentraler	Søknadstjeneste	Videregående skole	Underliggende etater
Ny enhet innenfor sektoren	Utlevering av flatfiler	Høyere utdanning	Tjenesteprodusenter
Ny enhet utenfor sektoren	Utlevering via API	Forskning	Næringslivet
	Statistikkbank	Voksnes læring	Allmennheten
	Nettsider og portaler	Integrering	
	Sikre analyserom		

FIGUR 21: MULIGHETSDIMENSJONER

Organisasjon

Mulighetsdimensjonen beskriver tiltak og endringer i organiseringen av dataproduksjonen i kunnskapssektoren. Dette innebærer alt fra mindre endringer som kun innebærer en tydeliggjøring av ansvarsområder og rollefordeling mellom dataprodusentene i kunnskapssektoren, til mer komplekse endringer for hvordan strukturen i kunnskapssektoren er bygd opp. Mulighetsdimensjonen inneholder fire ulike former for organisering: tydeligere ansvars- og rollefordeling, en deling av sektoren inn i områdesentraler for hvert fagområde, etablere en ny enhet *innenfor* kunnskapssektoren og etablere en fellessentral *utenfor* kunnskapssektoren med ansvar for all dataforvaltning. De mer komplekse alternativene der det opprettes nye enheter eller aktører vil også kreve avklaring av roller og ansvar.

Funksjonalitet

Denne mulighetsdimensjonen beskriver hvilken funksjonalitet som tilbys til konsumentene i de ulike konseptene. I denne sammenhengen omfatter funksjonalitet verktøy og tjenester knyttet til identifisering av databehov, søknader om utlevering av data, tilgjengeliggjøring av data og analyse av dataene. En beskrivelse av funksjonene som hører til de ulike stegene er vist i Figur 22. Konseptene kan inneholde en sammensetning av flere funksjonaliteter, hvor sammensetningen har betydning for grad av måloppnåelse for de ulike konsumentgruppene.

IDENTIFISERING AV DATA	SØKNAD OM UTEVERING AV DATA	TILGANG TIL DATA	ANALYSE AV DATA
Metadatakatalog – katalog som inneholder en oversikt over og beskrivelser av datasettene i katalogen Eksplorative tjenester – tjenester for å utforske analysemuligheter som ligger i tilgjengelige data	Søknadstjeneste – tjeneste som hjelper konsumentene med søknad til data med begrenset tilgang	Utlevering av flatfiler – fysisk utlevering av data Utlevering via API – maskinell utlevering av data Statistikkbank – samling av publisert data innenfor et statistikkområde Nettsider og portaler – publisering av data på web	Sikre analyserom – dataområder hvor det er mulig å gjøre analyser på informasjon som er sikret med hensyn til tilgang og muligheter for å kopiere eller flytte data ut av området

FIGUR 22: ALTERNATIVER TIL MULIGHETSDIMENSJONEN FUNKSJONALITET

Fagområder

Denne mulighetsdimensjonen omfatter de ulike fagområdene i kunnskapssektoren. Alternativene inkluderer derfor barnehage, grunnskole, videregående skole, forskning, høyere utdanning, voksnes læring og integreringsområdet. Et konsept kan fokusere på tiltak for deling og tilgjengeliggjøring av data innenfor ett eller flere områder.

Konsumentgrupper

Mulighetsdimensjonen omfatter konsumentene av kunnskapsdata i samfunnet. I denne utredningen er konsumentene gruppert i kategorier som benytter data til forskjellige formål, jf. Figur 11. Disse konsumentgruppene utgjør de ulike alternativene innenfor dimensjonen. Konseptene for den fremtidige løsningen for deling av kunnskapsdata kan være rettet mot én eller flere av konsumentgruppene. Dette innebærer at konseptene vil kunne prioritere de ulike konsumentgruppene ulikt, der noen konsumentgrupper får dekket sine behov for tilgang kunnskapsdata i større grad enn andre konsumentgrupper.

7.2. Konseptene

Med bakgrunn i mulighetsdimensjonene beskrevet over, er det identifisert og utarbeidet 15 unike konsepter for fremtidig deling av data i kunnskapssektoren, jf. prosessen vist i Figur 20. Konseptene er kategorisert etter hvor store endringer som kreves. De ulike konseptene er oppsummert i Figur 23.

Ingen endring		Mindre endringer		Områdefokus		Områdesentraler		Sentral i kunnskapssektoren		Sentral utenfor kunnskapssektoren	
K0	Null-alternativet	K1	Felles metadata-katalog	K6	Områdefokus integrering	K9	Frie område-sentraler	K12	Orden i kunnskaps-sektoren	K15	Norsk datasentral
		K2	Orden i eget hus	K7	Områdefokus UH-sektoren	K10	Koordinerte område-sentraler	K13	Kunnskaps-fordeler		
		K3	Forskning i sentrum	K8	Områdefokus skole	K11	Områdesentraler med ekstern utlevering	K14	Kunnskaps-sentralen		
		K4	Virksomhets-styring i sentrum								
		K5	Forvaltning i sentrum								

FIGUR 23: OVERSIKT OVER DE ULIKE KONSEPTENE DELT INN ETTER ENDRING SOM KREVES

Tabell 2 gir en beskrivelse av nullalternativet, som vil fungere som et sammenligningsalternativ for de andre konseptene som vurderes i utredningen.

TABELL 2: BESKRIVELSE AV KONSEPT K0

Nullalternativ	
K1 - Nullalternativet	 I nullalternativet er det ingen organisatoriske endringer og dataprodusentene har drifts- og forvaltningsansvar av data som i dag.
	 Konsumentene får ingen ekstra funksjonalitet sammenlignet med i dag.
	 I nullalternativet er alle områder i kunnskapssektoren like mye i fokus som i dag.
	 Alle konsumentgrupper er i konseptet like mye prioritert.

Tabell 3 gir en beskrivelse av konseptene som inngår i gruppen «Mindre endringer». Konseptene kjennetegnes ved at det gjøres små justeringer sammenlignet med K0.

TABELL 3: BESKRIVELSE AV KONSEPT K1 – K5

Mindre endringer	
K1 – Felles metadatakatalog	 I konseptet er det ingen organisatoriske endringer utover tydeliggjøring av roller og ansvar blant dataprodusentene for å adressere utfordringer knyttet til dataproduksjon.
	 I konseptet legger alle dataprodusentene inn informasjon om sine datakilder i Felles metadatakatalog hos Brønnøysundregistrene. Utover dette er det ingen endringer i funksjonalitet sammenlignet med nullalternativet.
	 I konseptet er alle områder i kunnskapssektoren like mye i fokus som i nullalternativet.
	 Alle konsumentgrupper er like mye prioritert i konseptet som i nullalternativet.
K2 – Orden i eget hus	 I konseptet er hver dataprodusent ansvarlig for å innføre «orden i eget hus». En aktør får ansvar for harmonisering av data, samt fordeling av roller og ansvar blant dataprodusentene. Aktøren har også ansvar for å etablere og forvalte en felles metadatakatalog for all kunnskapsdata, samt opprettelsen av et rådgivende forum for juridiske vurderinger rundt utlevering av data
	 I konseptet vil det være en metadatakatalog hos den ansvarlige aktøren. Utover metadatakatalogen vil funksjonaliteten for konsumentene være som i nullalternativet.
	 I konseptet er alle områder i kunnskapssektoren like mye i fokus som i nullalternativet.
	 Alle konsumentgrupper er like mye i fokus (som K1).

K3 – Forskning i sentrum		I konseptet er det ingen organisatoriske endringer utover nullalternativet.
		I konseptet legger alle dataprodusentene inn informasjon om sine datakilder i Felles datakatalog hos Brønnøysundregistrene (som i K1). I tillegg tilbys forskere en søknadsportal for tilgang til data. Utover dette er det ingen endringer i funksjonalitet sammenlignet med nullalternativet.
		I konseptet er alle områder i kunnskapssektoren like mye i fokus som i nullalternativet.
		I konseptet blir forskernes behov knyttet til søknad om data prioritert.
K4 – Virksomhetsstyring i sentrum		I konseptet er det ingen organisatoriske endringer utover nullalternativet.
		I konseptet legger alle dataprodusentene inn informasjon om sine datakilder i Felles datakatalog hos Brønnøysundregistrene (som i K1). Utover dette er det ingen endringer i funksjonalitet sammenlignet med nullalternativet.
		I konseptet er alle områder i kunnskapssektoren like mye i fokus som i nullalternativet.
		I konseptet blir konsumenter med behov for data til virksomhetsstyring prioritert.
K5 – Forvaltning i sentrum		I konseptet er det ingen organisatoriske endringer utover nullalternativet.
		I konseptet legger alle dataprodusentene inn informasjon om sine datakilder i Felles datakatalog hos Brønnøysundregistrene (som i K1). I tillegg etableres en statistikkbank med kunnskapsdata fra alle fagområder. Underliggende etater og politikktutviklere får tilgang til sikre analyser (microdata.no).
		Alle områder like mye i fokus.
		Konsumenter med behov for data til virksomhetsstyring, politikk og forvaltning prioriteres.

Tabell 4 gir en beskrivelse av konseptene som inngår i gruppen «Områdefokus». Konseptene i dette området skiller seg fra de andre gruppene ved de fokuserer på et bestemt fagområde innenfor kunnskapssektoren som integrering, UH-sektor eller barnehage, grunnskole og videregående skole.

TABELL 4: BESKRIVELSE AV KONSEPT K6 – K8

Områdefokus	
K6 – Områdefokus integrering	 I konseptet tydeliggjøres roller og ansvar blant dataprodusentene som behandler integreringsdata og hvordan dataene innenfor integrering organiseres. Kunnskapsdepartementet har ansvaret for endringene, men gir prosessansvaret til IMDi.
	 I konseptet legger alle dataprodusentene innenfor fagområdet integrering inn data i en metadatakatalog. Utover dette er funksjonalitet som i nullalternativet.
	 I konseptet er fagområdet integrering prioritert.
	 Konsumenter som benytter integreringsdata prioriteres (kan være alle konsumentgrupper).
K7 – Områdefokus UH-sektoren	 I konseptet tydeliggjøres roller og ansvar blant dataprodusentene som behandler forsknings- og UH-data og hvordan dataene innenfor UH-sektoren organiseres. Kunnskapsdepartementet har ansvaret for endringene, men gir prosessansvaret til Unit.
	 I konseptet legger alle dataprodusentene innenfor forskning og høyere utdanning inn data i en metadatakatalog. Utover dette er funksjonalitet som i nullalternativet
	 Konseptet fokuserer på fagområdet forskning og høyere utdanning.
	 Konsumenter som benytter data fra UH-sektoren prioriteres (kan være alle konsumentgrupper).
K8 – Områdefokus skole	 I konseptet tydeliggjøres roller og ansvar blant dataprodusentene som behandler data innenfor barnehage, grunnskole og videregående og hvordan dataene organiseres. Kunnskapsdepartementet har ansvaret for endringene, men gir prosessansvaret til Udir.
	 I konseptet legger alle dataprodusentene innenfor barnehage, grunnskole og videregående skole inn data i en metadatakatalog. Utover dette er funksjonalitet som i nullalternativet
	 Konseptet fokuserer på barnehage, grunnskole og videregående skole.
	 Konsumenter som benytter data fra barnehage, grunnskole og videregående skole prioriteres (kan være alle konsumentgrupper).

Tabell 5 gir en beskrivelse av konseptene som inngår i gruppen «Områdesentraler». Felles for konseptene i denne gruppen er at det etableres områdesentraler (noder) som får ansvar for dataproduksjon innenfor hvert sitt fagområde innenfor kunnskapssektoren.

TABELL 5: BESKRIVELSE AV KONSEPT K9 – K11

Områdesentraler	
K9 – Frie områdesentraler	 I konseptet etableres det områdesentraler (noder) som får ansvar for dataproduksjonen innenfor sitt område i kunnskapssektoren. Det innebærer at nodene har drifts- og forvaltningsansvar for all data innenfor sitt respektive område. Det gjøres ingen tiltak for koordinering på tvers av de ulike nodene med tanke på harmonisering og utlevering av data.
	 I konseptet etablerer områdesentralene metadatakatalog og statistikkbank for sitt område, i tillegg til egne nettsider. Utover dette vil det ikke være endringer fra dagens situasjon i funksjonaliteten for konsumentene.
	 Konseptet fokuserer like mye på hvert område.
	 I konseptet er alle konsumentgrupper like mye i fokus.
K10 – Koordinerte områdesentraler	 I konseptet etableres det områdesentraler (noder) som får ansvar for dataproduksjonen innenfor sitt område i kunnskapssektoren. Nodene innenfor hvert av områdene får drifts- og forvaltningsansvar for data innenfor sine respektive områder. Koordinering mellom nodene skjer gjennom en utnevnt aktør. Enheten har også prosessansvar for harmonisering av data og fordeling av roller og ansvar blant nodene.
	 I konseptet etablerer den koordinerende aktøren en felles metadatakatalog og søknadstjeneste. Hver node har egne statistikkbanker for sitt område, i tillegg til nettsider og portaler.
	 Konseptet fokuserer like mye på hvert område.
	 Alle konsumentgrupper like mye i fokus.
K11 – Områdesentraler med ekstern utlevering	 I konseptet etableres det områdesentraler (noder) som får ansvar for dataproduksjonen innenfor sitt område i kunnskapssektoren. Det innebærer at nodene har drifts- og forvaltningsansvar for all data innenfor sine respektive områder. En enhet utenfor kunnskapssektoren har ansvar for en felles metadatakatalog og søknadsportal, men ingen formell myndighet til å styre nodene.
	 I konseptet får konsumentene en felles metadatakatalog og søknadsportal hos aktøren utenfor kunnskapssektoren. Hver node er ansvarlig for statistikkbanker og nettsider/portaler til konsumentene.
	 Konseptet fokuserer like mye på hvert område.
	 Alle konsumentgrupper like mye i fokus.

Tabell 6 gir en beskrivelse av konseptene som inngår i gruppen «Sentral i kunnskapssektoren». Felles for konseptene i denne gruppen er at det etableres en aktør med ansvar for å samle data fra dataprodusentene for sekundærbruk i kunnskapssektoren. Konseptene varierer i hvilken grad dagens dataprodusenter organiseres.

TABELL 6: BESKRIVELSE AV KONSEPT K12 – K14

Sentral i kunnskapssektoren	
K12 – Orden i kunnskapssektoren	 I konseptet realiserer dataprodusentene «orden i eget hus» som beskrevet i konsept K2. En av KDs underliggende etater får ansvar for koordinering og samhandling mellom dataprodusentene. Dette inkluderer ansvar for prosessen for harmonisering av data, tydeliggjøring av roller og ansvar blant dataprodusentene, samt være en pådriver for orden i eget hus hos dataprodusentene.
	 I konseptet tilbyr den nye aktøren metadatatjenester, en felles søknadstjeneste og tilgang på data til forskning og analyse i sikre analyserom. Flere konsumentgrupper får tilgang til dataene, herunder politikkutviklere, underliggende etater og tjenesteprodusenter.
	 Konseptet fokuserer like mye på hvert område.
	 Alle konsumentgrupper like mye i fokus.
K13 - Kunnskapsfordeler	 I konseptet etableres det en aktør i kunnskapssektoren som samler inn all data fra dataprodusentene i sektoren. Den nye aktøren vil få kopier av all kunnskapsdata fra dataprodusentene og ha ansvar for å opprette et felles datavarehus for sektoren. All data som mottas vil kobles sammen og bli utlevert til konsumentene gjennom datavarehuset. Aktøren som opprettes er kun en mottaker av dataene, og har ikke et prosessansvar knyttet til tydeliggjøring av roller og ansvar i sektoren. Dette innebærer at det ikke vil være endringer i organisering for dataprodusentene.
	 I konseptet får konsumentene tilgang til data gjennom datavarehuset som etableres. Utover datavarehuset, tilbyr aktøren ikke annen funksjonalitet for konsumentene.
	 Konseptet fokuserer like mye på hvert område.
	 Alle konsumentgrupper like mye i fokus.
K14 - Kunnskapssentralen	 I konseptet konsolideres alle de eksisterende dataprodusentene til en aktør som har ansvar for forvaltningen av alle data i kunnskapssektoren. Kunnskapssentralen har ansvar for begrepsharmonisering, samt å legge til rette for tilgang av data og analysemuligheter som dekker konsumentenes behov.
	 Kunnskapssentralen tilbyr metadatatjenester, søknadstjenester og infrastruktur for analyse av kunnskapsdata i sikre analyserom. Offentlig forvaltning (politikkutviklere, underliggende etater og tjenesteprodusenter) og næringslivet får tilgang til kunnskapsdata i sikre rom gjennom sentralen.
	 Alle områder prioriteres like mye.
	 Alle konsumentgrupper i fokus.

Tabell 7 gir en beskrivelse av konseptene som inngår i gruppen «Sentral utenfor kunnskapssektoren». I gruppen inngår ett konsept, der det etableres en aktør utenfor kunnskapssektoren med ansvar for å forvalte data fra kunnskapssektoren.

TABELL 7: BESKRIVELSE AV KONSEPT K15

Sentral utenfor kunnskapssektoren	
K15 – Norsk datasentral	 I konseptet opprettes en aktør som ligger utenfor kunnskapssektoren med drifts- og forvaltningsansvar for alle data i kunnskapssektoren. Data fra tjenesteprodusentene går direkte fra tjenesteprodusentene og dagens dataprodusenter vil ikke motta data.
	 I konseptet får konsumentene stor funksjonalitet som utlevering av flatfiler, muligheter for analyser i sikre analyserom, statistikkbanker for hele kunnskapssektoren, nettsider og portaler.
	 Alle områder prioriteres like mye.
	 Alle konsumentgrupper i fokus.

7.2.1. Felles føringer

Ivaretagelse av kravene til personvern og informasjonssikkerhet er et MÅ-krav for samtlige konsepter, jf. kap 6. Dette medfører blant annet at roller og ansvar (behandlingsansvar og databehandler) må fastsettes tydelig for behandlingen av opplysninger. Videre må behandlingen være i samsvar med grunnleggende personvernprinsipper og det må foreligge et rettslig grunnlag både for innsamling, lagring og utlevering av personopplysningene, herunder ved behandling av særlige kategorier av personopplysninger. I tillegg er de registrerte personene gitt en rekke rettigheter etter personvernlovgivningen, blant annet innsyn i egne personopplysninger og retten til å kreve retting og sletting av opplysninger om seg selv.

Personvernlovgivningen stiller videre krav til at personopplysningene, ved bruk av egnede tekniske eller organisatoriske tiltak, behandles tilstrekkelig sikkert, herunder at opplysningene er vernet mot uautorisert eller ulovlig behandling og mot utilsiktet tap, ødeleggelse eller skade. Dette medfører blant annet krav om tilgangsstyring, sporing og logging av bruk av personopplysninger og at det legges til rette for innebygd personvern i eventuelle tekniske løsninger som utvikles.

7.3. Utvelgelse av konsepter

De ulike konseptene for fremtidig deling av data i kunnskapssektoren har gjennomgått en overordnet vurdering for å avdekke konseptenes egnethet, realisme og måloppnåelse. I første steg er konseptene vurdert opp mot de identifiserte må-kravene til den fremtidige løsningen (jf. kap. 6) og i andre steg er de gjenværende konseptene vurdert mot et utvalg av evalueringskriterier.

7.3.1. Vurdering opp mot MÅ-krav

I det første steget ble alle de identifiserte konseptene for fremtidig deling av data i kunnskapssektoren vurdert opp mot hvert av de tre MÅ-kravene.



Personvern og informasjonssikkerhet skal ivaretas

Det første MÅ-kravet innebærer at alle konseptene må tilfredsstillere kravene til informasjonssikkerhet og personvern i henhold til gjeldende regelverk, jf. kapittel 6.1.

Alle de identifiserte konseptene oppfyller dette kravet ved at det identifiseres felles føringer mht. personvern og informasjonssikkerhet som inkluderes i alle konseptene, jf. kapittel 7.2.1. Dette omfatter blant annet å gi innbyggere bedre kontroll over egne opplysninger og økt sikkerhet i de tekniske løsningene i konseptene. Andre tiltak som er lagt inn i ett eller flere konsepter er utlevering av individdata i sikre analyserom og felles søknadstjenester på tvers av dataprodusentene.



Konseptet skal gi økt samhandling innad i kunnskapssektoren

Det andre MÅ-kravet innebærer at organisasjoner og tekniske løsninger i kunnskapssektoren skal kunne samhandle i henhold til de fem samhandlingsområdene innenfor norsk arkitekturrammeverk for samhandling, jf. kapittel 6.1.

Det er flere av de identifiserte konseptene som ikke oppfyller dette kravet. I konsept K1 legges det opp til en tydeliggjøring av roller og ansvar blant dataprodusentene, noe som vil øke den organisatoriske samhandlingen i sektoren. Konseptet oppfyller likevel ikke kravet om økt samhandling innad i kunnskapssektoren da det ikke er planlagt tiltak knyttet til den juridiske, tekniske og semantiske samhandlingen.

Konseptene K3-K5 oppfyller heller ikke MÅ-kravet knyttet til økt samhandling. I konseptene er organiseringen i stor grad lik som i dag uten at dataprodusentene arbeider koordinert for å et felles mål. Samhandlingsevnen i kunnskapssektoren vil dermed ikke øke sammenlignet med i dag.

Konseptene K6-K8 oppfyller ikke MÅ-kravet fordi de kun fokuserer på enkelte fagområder innenfor kunnskapssektoren. Konseptene øker den semantiske og organisatoriske samhandlingen innenfor fagområdene de skal dekke, men sikrer ikke felles strukturer og avtaler på tvers av hele kunnskapssektoren. De vil dermed heller ikke bidra til en økt samhandling innad i kunnskapssektoren.

Konsept K9 oppfyller heller ikke MÅ-kravet. I konseptet legges det opp til selvstendige områdesentraler med ansvar for hver sine fagområder uten samhandling og koordinering på tvers av områdesentralene. Konseptet vil derfor ikke medføre økt samhandling innad i kunnskapssektoren.

K11 er det siste konseptet som ikke oppfyller dette MÅ-kravet. I konseptet legges det opp til at det opprettes områdesentraler som leverer all data til en aktør utenfor kunnskapssektoren uten en aktiv rolle i delingen av data innad i kunnskapssektoren. Opprettelsen av områdesentralene bidrar til en økt samhandling innad i områdene, men konseptet legger ikke opp til samhandling på tvers av områdene. Den utnevnte aktøren utenfor kunnskapssektoren har ingen myndighet til å styre områdesentralene, og det blir dermed ingen strukturer og avtaler som bidrar til en organisatorisk og semantisk samhandling innad i kunnskapssektoren i konseptet.

De identifiserte konseptene som oppfyller kravet er K2, K10 og K12-K15, totalt seks ulike konsept. Konseptene oppfyller kravet blant annet fordi det er lagt til rette for en tydeligere rolle- og ansvarsdeling innenfor hele kunnskapssektoren. Dette er en viktig forutsetning for styring av dataforvaltningen, inngåelse av avtaler, og muliggjør begrepharmonisering og felles mål. Konseptene vil dermed gi økt samhandling i hele sektoren og være et viktig bidrag for blant annet å øke datakvaliteten på de delte dataene i kunnskapssektoren.



Infrastrukturen skal gi oversikt over data i kunnskapssektoren

Det tredje MÅ-kravet innebærer at konsumentene i kunnskapssektoren har oversikt over hvilke data som er tilgjengelige, samt en beskrivelse av disse, jf. kap.6.2.

Tre av konseptene er vurdert til ikke å oppfylle kravet om å tilby en lett tilgjengelig oversikt over data i kunnskapssektoren. I konseptene K6, K7 og K8 er det kun en metadatakatalog for de spesifikke fagområdene omtalt i de respektive konseptene. Dette innebærer at konseptene ikke gir en felles oversikt over data i hele kunnskapssektoren og konseptene oppfyller dermed ikke kravet.

De andre identifiserte konseptene for fremtidig deling av data tilbyr en felles metadatakatalog for all data i kunnskapssektoren eller et datavarehus som gir oversikt over dataene. Konseptene oppfyller dermed kravet om at infrastrukturen skal gi en lett tilgjengelig totaloversikt over data i sektoren.

Oppsummering

Tabell 8 oppsummerer vurderingene av de identifiserte konseptene opp mot MÅ-kravene. Som det fremkommer av tabellen er følgende seks konsepter for fremtidig deling av data i kunnskapssektoren som oppfyller alle kravene:

- K2 Orden i eget hus
- K10 Koordinerte områdesentraler
- K12 Orden i kunnskapssektoren
- K13 Kunnskapsfordeler
- K14 Kunnskapssentralen
- K15 Norsk datasentral

Disse seks konseptene ble tatt med videre til det andre steget i utvelgelsen av konsepter til alternativanalysen.

TABELL 8: KONSEPTENE VURDERT MOT MÅ-KRAVENE

Konsept		Personvern og informasjonssikkerhet skal ivaretas	Det skal legges til rette for økt samhandling i sektoren	Infrastrukturen skal gi oversikt over data i kunnskapssektoren
K1	Felles metadatakatalog	✓	✗	✓
K2	Orden i eget hus	✓	✓	✓
K3	Forskning i sentrum	✓	✗	✓
K4	Virksomhetsstyring i sentrum	✓	✗	✓
K5	Forvaltning i sentrum	✓	✗	✓
K6	Områdefokus integrering	✓	✗	✗
K7	Områdefokus UH-sektoren	✓	✗	✗
K8	Områdefokus skole	✓	✗	✗
K9	Frie områdesentraler	✓	✗	✓
K10	Koordinerte områdesentraler	✓	✓	✓
K11	Områdesentraler med ekstern utlevering	✓	✗	✓
K12	Orden i kunnskapssektoren	✓	✓	✓
K13	Kunnskapsfordeler	✓	✓	✓
K14	Kunnskapssentralen	✓	✓	✓
K15	Norsk datasentral	✓	✓	✓

7.3.2. Vurdering opp mot evalueringskriterier

I steg 2 i utvelgelsen av konseptene er de seks konseptene som besto MÅ-kravene vurdert opp mot et sett av evalueringskriterier. Kriteriene som er benyttet i denne evalueringsprosessen er beskrevet i Tabell 9.

TABELL 9: EVALUERINGSKRITERIER FOR UTVELGELSE AV KONSEPTER

Kriterium	Beskrivelse
Funksjonelle behov	Funksjonelle behov beskriver i hvilken grad konseptet er i stand til å løse behov hos konsumentene.
Gjennomføringsrisiko	Gjennomføringsrisiko beskriver ulike risikoer som er knyttet til realiseringen av konseptet. Det ble definert aktuelle risikoer som kunne påvirke realiseringen, og konseptet ble vurdert opp mot disse.
Kostnad	Kostnad beskriver det forventede kostnadsbildet til konseptet. I første omgang ble det ikke gjennomført en grundig kostnadsanalyse.
Måloppnåelse	Måloppnåelse beskriver hvorvidt konseptet støtter opp de beskrevne effektmålene for den fremtidige løsningen for deling av data i kunnskapssektoren.

I vurderingen opp mot evalueringskriteriene er konseptene evaluert og scoret opp mot hvert kriterium. Ved å sammenligne de seks konseptenes score og konseptuelle likhet mot hverandre er følgende fire konsepter tatt med videre til alternativanalysen:

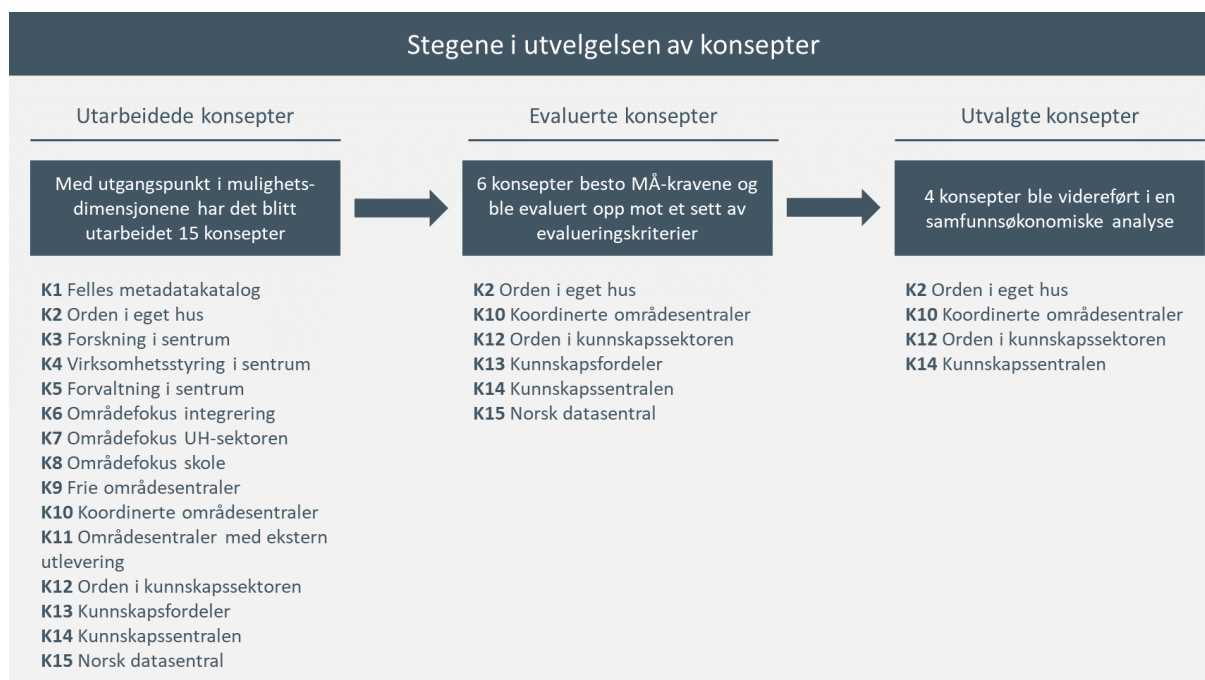
- K2 Orden i eget hus
- K10 Koordinerte områdesentraler
- K12 Orden i kunnskapssektoren
- K14 Kunnskapssentralen

Konseptet K13 Kunnskapsfordeler er ikke tatt med videre til alternativanalysen. Et datavarehus vil dekke konsumentenes behov. På den andre siden er det å legge inn alle data fra kunnskapssektoren i et datavarehus en ressurskrevende prosess. Det vil være nødvendig å legge inn data fra veldig mange fagsystemer og alle data må detaljeres på en form som er tilpasset datavarehuset. Prosessen vil derfor kreve mye arbeid med tilhørende høy kostnad, og tilsvarende høy gjennomføringsrisiko. Etter at datavarehuset er implementert vil det være en forvaltningskostnad forbundet med å vedlikeholde datavarehuset, i tillegg til kostnaden de eksisterende dataprodusentene fremdeles må bruke for å forvalte sine egne systemer. Dette gjør at forvaltningskostnaden i sektoren totalt sett vil øke, og konseptet scorer dermed lavt på måloppnåelse. Alle evalueringskriteriene tatt i betraktning gjør at konseptet ikke blir tatt med videre til alternativanalysen.

Konseptet K15 Norsk datasentral er vurdert til å ha høye kostnader og stor gjennomføringsrisiko. Konseptet innebærer store og komplekse endringer i organiseringen av data i kunnskapssektoren ved at forvaltningsansvaret for alle data flyttes ut av kunnskapssektoren. KD vil derfor ha liten påvirkningskraft på forvaltningen av data om egen sektor. Dette innebærer derfor en potensiell lav måloppnåelse. Konseptet er også vurdert å være mer krevende å gjennomføre enn konsept K14 Kunnskapssentralen. Alle evalueringskriteriene tatt i betraktning blir derfor ikke konsept K15 tatt med videre til alternativanalysen.

7.3.3. Oppsummering

Gitt de vurderingene som er presentert ovenfor er utvelgelsen av konseptene oppsummert i Figur 24.



FIGUR 24: OPPSUMMERING AV UTVELGELSEN AV KONSEPTER

8. Alternativanalysen

I *alternativanalysen* utføres en samfunnsøkonomisk analyse av de utvalgte konseptene for deling av data i kunnskapssektoren. Analysen består av å identifisere og vurdere virkningene av de ulike konseptene. Virkningene kan forstås som de positive og negative effektene som oppstår som følge av et tiltak, og virkningene vurderes som endringer i konseptene sammenlignet med nullalternativet. I analysen skilles det mellom to ulike type virkninger; kostnad og nytte. Kostnadsvirkninger er ressursbruken som følger av konseptene, mens nytte er virkningene som øker velferden for en eller flere grupper i samfunnet eller samfunnet som helhet. I analysen vurderes kostnadene av konseptene i kroner, mens nytten av konseptene vurderes kvalitativt.⁶

Kapittelet begynner med beskrivelser av de utvalgte konseptene som inngår i alternativanalysen. Deretter vurderes virkningene av konseptene fordelt på kostnad og nytte. I tillegg gjøres det vurderinger av konseptenes gjennomføringsrisiko og realopsjon. Avslutningsvis gjøres en totalvurdering av konseptene, og det anbefales et konsept for fremtidig deling av data i kunnskapssektoren.

8.1. Beskrivelse av konseptene

8.1.1. Felles for alle konseptene

Det er noen felles egenskaper som ligger til grunn for alle konseptene som er beskrevet og vurdert i dette kapittelet. Disse er beskrevet her og gjelder for alle konseptene, inklusive Nullalternativet.

Felles datakatalog

Felles datakatalog er en sentral del av «Orden i eget hus» og alle data skal beskrives der i henhold til de krav til standardiserte metadata som er nedfelt i retningslinjene for orden i eget hus. Felles datakatalog har et minimumskrav til hva som skal registreres av metadata om det enkelte datasett. Felles datakatalog har funksjonalitet for å høste metadata om datasett fra dataprodusentene når dataprodusentene tilrettelegger for det. For de konseptene hvor det er introdusert en egen metadatakatalog for kunnskapssektorens data, er dette for å gi mulighet til å etablere mer avanserte data som gir rikere funksjonalitet, f. eks. eksplorative analyser kun ved hjelp av metadata. Eller for å støtte flere metadatastandarder for utveksling av data med eksterne aktører som benytter andre standarder enn det Felles datakatalog krever. Utveksling av data til microdata.no 2.0 vil være et eksempel på dette når denne funksjonaliteten etableres der.

Metadatakatalogene som inngår i konseptene nedenfor vil høste metadata om datasett fra dataprodusentene og videre eksponere subsett av metadata som Felles datakatalog trenger, og Felles datakatalog vil høste disse. Det vil derfor ikke være nødvendig å registrere data mer enn en gang for dataprodusentene.

⁶ Den samfunnsøkonomiske analysen er en forenklet nytte-kostnadsanalyse. I en nytte-kostnadsanalyse verdsettes alle positive og negative virkninger av et tiltak i kroner så langt det lar seg gjøre. I denne utredningen er kun kostnadene verdsatt i kroner, mens nytten er kvalitativt vurdert ved bruk av pluss-minusmetoden. Konseptene vurderes på bakgrunn av virkningene mht. kostnader og nytte for å komme frem til den samfunnsøkonomiske lønnsomheten for hvert konsept. Av de vurderte konseptene blir et konsept anbefalt til beslutningstaker. KVVU-metodikken er overordnet beskrevet i kap. 1.

Utveksling av data med andre sektorer

Ved å skape orden i eget hus hvor data er godt beskrevet med standardiserte metadata som publiseres gjennom metadatakataloger, blir det enklere å dele data på tvers av sektor- og landegrenser. Dette krever blant annet at en enes om hvilke metadata-standarder som benyttes slik at dataene tolkes likt av de som deler dataene. Gjennom at offentlige virksomheter støtter det nasjonale rammeverket for samhandling, vil alle aspekter ved deling av data adresseres og det blir enklere å realisere deling av data på tvers av sektorer. Rammeverket er basert på et europeisk rammeverk og vil forenkle deling av data med land som følger det samme rammeverket.

Selvforklarende datasett

Det er tidligere i rapporten pekt på at data som deles, i mange tilfeller ikke i tilstrekkelig grad dekker konsumentenes behov, eller at dataene er så detaljerte at det kreves kunnskap om den interne strukturen i kildene dataene er hentet fra, for å gi mening til dataene. Når data skal deles med andre må de organiseres som datasett på en måte som ikke krever andre forutsetninger for konsumentene enn de metadataene som beskriver datasettene. Felles for alle konseptene er at det i arbeidet med etablering av infrastrukturen er svært viktig å ha fokus på dette.

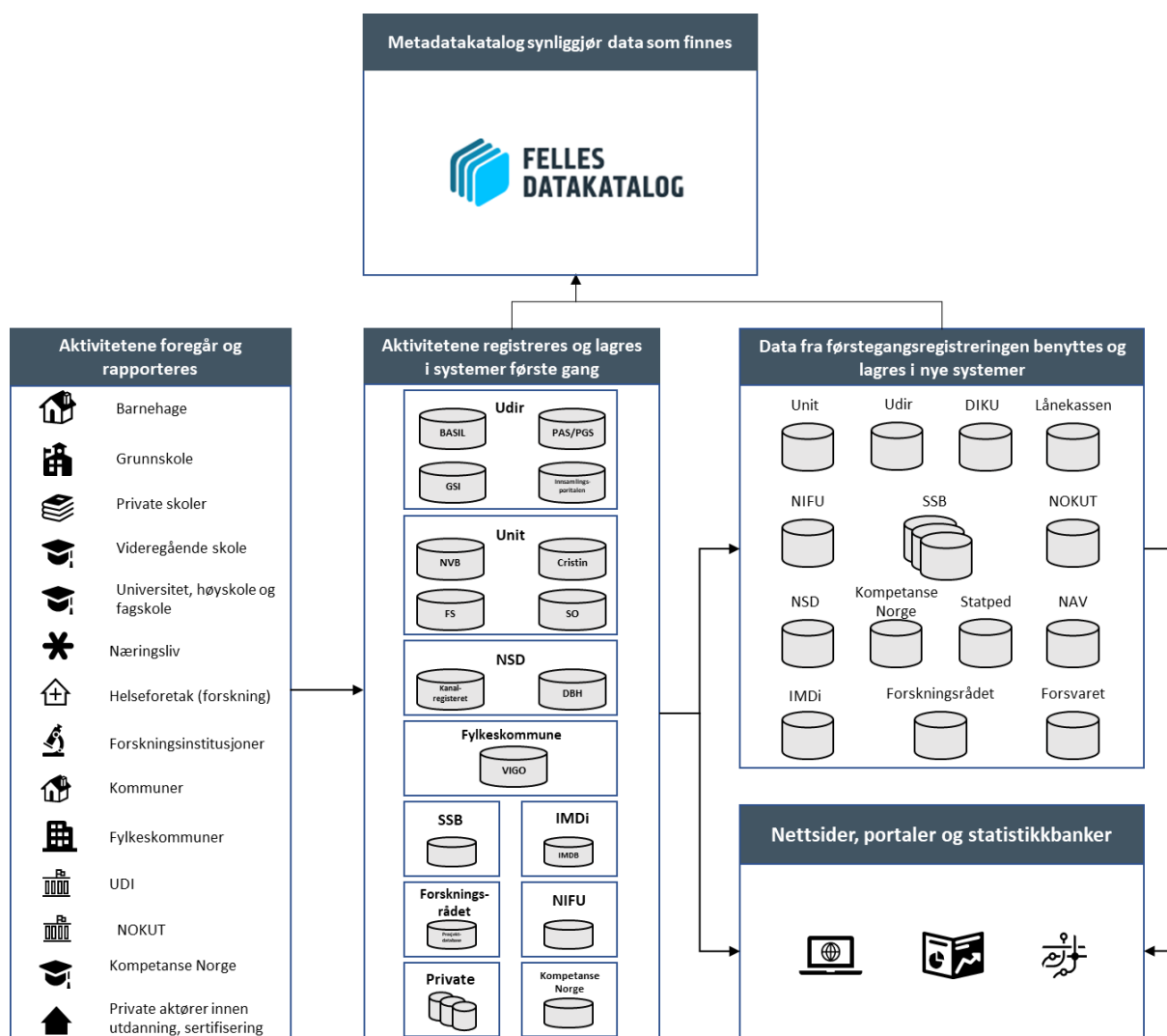
8.1.2. Nullalternativet

Nullalternativet beskriver dagens situasjon og forventet utvikling i organisering og forvaltning av data, funksjonalitet, tilgang og bruk av data, regelverk og finansiering. Nullalternativet benyttes som sammenligningsgrunnlag for å vurdere de utvalgte konseptene, dvs. K2, K10, K12 og K14.

Organisering og forvaltning av data

Forvaltningen av data i kunnskapssektoren er i dag desentralisert og spredt på mange ulike aktører, jf. Figur 25 og beskrivelse av dagens situasjon i kap. 2. Oppsummert skjer produksjonen av data i kunnskapssektoren i hovedsak der de offentlige tjenestene blir utøvd, dvs. i barnehager, skoler, universiteter osv. Utenfor sektoren skjer produksjonen av kunnskapsdata bl.a. hos private aktører, næringslivet, kommuner og fylkeskommuner. All data i kunnskapssektoren som blir produsert, blir samlet i et eller flere registre eller databaser hos en eller flere dataprodusenter. I Figur 3 er det delt mellom førstegangsregistreringen av dataene (f.eks. hos Udir, Unit og IMDi), og produsenter som benytter dataene videre og lagrer de i nye systemer (f.eks. Lånekassen og SSB). Flere av produsentene som viderebruker dataene, kobler disse sammen med nye og unike kunnskapsdata.

I nullalternativet har de ulike dataprodusentene drifts- og forvaltningsansvar for data innenfor sine ansvarsområder. Samtidig oppleves det uklart hvilke produsenter som har ansvar for å dele data med hvem og hvordan delingen skal organiseres, samt hvem som har ansvar for beskrivelser av dataene (metadata). Disse uklarhetene har blant annet medført liten grad av harmonisering av data og en fragmentert infrastruktur.



FIGUR 25: SKISSE OVER ORGANISERING I NULLALTERNATIVET

Funksjonalitet

Funksjonalitet omfatter verktøy og tjenester knyttet til de ulike stegene i konsumentenes brukerreise for å få tilgang til og analysere kunnskapsdata. Det vil si verktøy og tjenester til identifisering av data, søknader om utlevering av data, tilgjengeliggjøring av data og analyse av dataene. Konsumentenes brukerreise for å få tilgang til og analysere kunnskapsdata er illustrert i Figur 26.



FIGUR 26: KONSUMENTENES BRUKERREISE FOR Å FÅ TILGANG TIL OG ANALYSERE KUNNSKAPSDATA

For å **identifisere datakilder** og få oversikt over kunnskapsdata bruker konsumentene flere ulike kilder. Blant annet gir enkelte av dataprodusentene informasjon om tilgjengelige data på sine

nettsider. Et eksempel er NSD som gir oversikt over data i DBH. I tillegg gir noen av dataprodusentene informasjon om sine datasett i Felles datakatalog hos Brønnøysundregistret. I Felles datakatalog kan informasjonen bestå av beskrivelse av datasettet, type data, tilgangsnivå, kvalitet og evt. API-nøkler. Udir har f.eks. lagt ut informasjon om et datasett som inneholder offentlige og private videregående skoler i Norge med tilhørende kontaktinformasjon og opplysninger i Felles datakatalog. Det er per i dag ingen kontrollmekanismer knyttet til innholdet som legges i katalogen. Det betyr at datasett som omfatter samme type informasjon (dupliserende eller overlappende datasett) kan registreres mer enn en gang og det kan være ulik betydning av samme variabel. Synliggjøring av data i Felles datakatalog gjør det imidlertid lettere å oppdage dette problemet, men det er opp til dataprodusentene selv å fange det opp og eventuelt gjøre noe med det i samråd med andre dataprodusenter som samler samme type informasjon.

I nullalternativet har alle konsumentene uten videre tilgang til all kunnskapsdata som er offentlig publisert. Dersom konsumentene har behov for data som ikke er offentlig tilgjengelig på nettsider, portaler e.l., kan de henvende seg til dataprodusenter med forespørsel om enkeltuttrekk av data. I tilfeller der konsumentene har behov for data med begrenset tilgang må de **søke om tilgang** til dataene. Søknaden sendes direkte til dataprodusenten som har ansvar for dataene. Ved behov for data fra mer enn én dataprodusent må søknaden sendes til alle de aktuelle dataprodusentene. Konsumenter som har behov for kobling av kunnskapsdata med data fra andre sektorer, som ofte er forskere, henvender seg som regel til SSB for å søke om tilgang til dataene. SSB har mye data fra flere sektorer og har søknads- og behandlingstjenester knyttet til sammenkobling og utlevering av data.

I kunnskapssektoren blir **data tilgjengeliggjort** på flere måter. Flere av dataprodusentene har nettsider hvor de publiserer tall og statistikk innenfor sitt ansvarsområde, f.eks. Udirs barnehagefakta.no og IMDIs www.imdi.no/tall-og-statistikk. I tillegg har enkelte dataprodusenter statistikkbanker med større mengder data, f.eks. Database for høyere utdanning som forvaltes av NSD. Når det gjelder data til forskning, analyse eller allmenne formål, utleverer dataprodusentene datasett på flatfiler via e-post eller minnepinne. Enkelte dataprodusenter har opprettet API-er for deling av dataene. I nullalternativet forventes det at denne type deling av data vil øke i tiden fremover. En sentral aktør for deling av kunnskapsdata som befinner seg utenfor sektoren er SSB. SSB tilgjengeliggjør mye kunnskapsdata i sin statistikkbank og de kobler kunnskapsdata og data fra andre typer data ved enkelthenvendelser fra forskere o.l.

Til **analyse** av kunnskapsdata benytter konsumentene i hovedsak egne verktøy. Forskere har i tillegg mulighet til å søke om tilgang til data gjennom microdata.no, hvor de må benytte analyseverktøyet som ligger i løsningen.

Behov for regelverksutvikling

I nullalternativet vil det ikke være behov for endringer i regelverket sammenlignet med dagens situasjon. Personvernlovgivningen vil regulere behandlingen av individdata hos de ulike dataprodusentene. I henhold til Statistikkloven vil SSB ha ansvar for å koble individdata i kunnskapssektoren og på tvers av kunnskapssektoren og andre sektorer. Ny forvaltningslov er nå på høring og vil bedre kunne legge til rette for bruk av IKT til en god og effektiv saksbehandling. Dette vil kunne ha relevans for bruk av data fra kunnskapssektoren fremover.

Personvernkonsekvenser

Det er et grunnleggende personvernprinsipp å ha kontroll over egne personopplysninger. Nullalternativet gir ikke den registrerte muligheten til å få en helhetlig oversikt over hvor egne opplysninger er registrert, hvem som har tilgang til opplysningene og til hvem og hvilke formål opplysningene utleveres. Den registrerte må forholde seg til mange aktører

og en kompleks organisering. Manglende oversikt gjør det utfordrende for den registrerte å ivareta sine rettigheter, som blant annet retten til innsyn, retting, sletting osv.

Utlevering av data skjer i dag på ulike måter, men ofte utleverer dataprodusentene datasett på flatfiler via e-post eller minnepinne. Det er liten kontroll med informasjonssikkerheten etter utlevering, og det er risiko for at opplysningene kan komme på avveie.

8.1.3. K2 Orden i eget hus

Konsept K2 fokuserer på at hver dataprodusent har bedre oversikt over egne data og realiserer Difis strategi «orden i eget hus». Dette gir en bedre oversikt og enklere tilgang til data i kunnskapssektoren, og dataprodusentenes ansvarsområder blir tydeliggjort. En av KDs underliggende etater gis ansvar for koordinering og samhandling på tvers av alle dataprodusenter. Det samlede ambisjonsnivået i konseptet mht. behov og utfordringer på konsument- og produsentsiden er relativt lavt. Samtidig inneholder konseptet tiltak som er sentrale byggesteiner i de andre konseptene.

Organisering og forvaltning av data

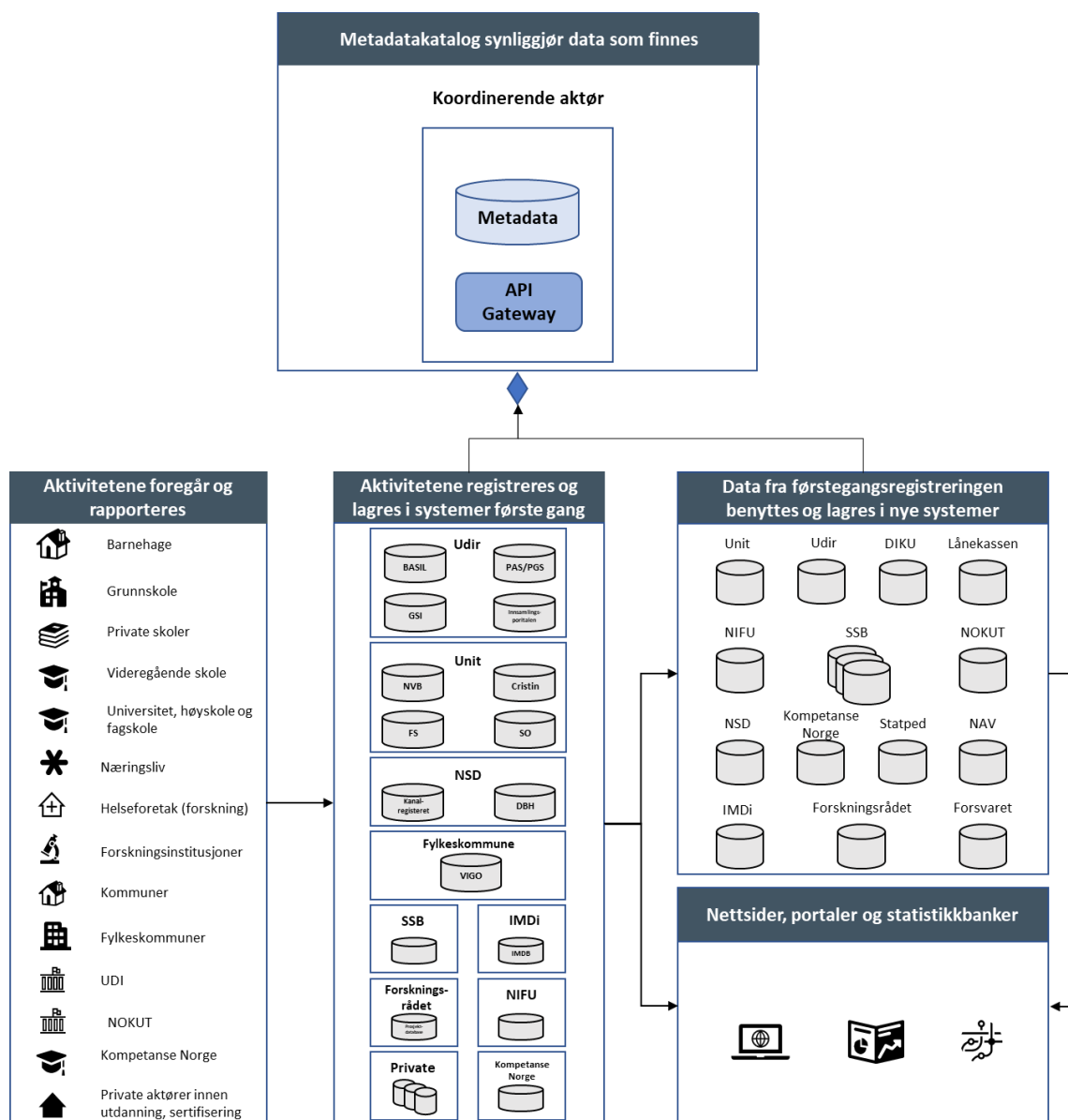
I konsept K2 er forvaltningen av kunnskapsdata desentralisert og spredt på mange ulike aktører, hvor aktørene (dataprodusentene) har drifts- og forvaltningsansvar for dataene innenfor sine ansvarsområder, jf. Figur 27. Konseptet innebærer derfor ingen betydelige organisatoriske endringer fra nullalternativet.

Hver dataprodusent har imidlertid ansvar for å realisere «orden i eget hus» i henhold til Difis retningslinjer for deling av data innenfor sitt område (Direktoratet for forvaltning og IKT, 2018). «Orden i eget hus» innebærer at den enkelte virksomhet har oversikt over hvilke data den håndterer, hva dataene betyr, hva de brukes til, hvilke prosesser de inngår i, og hvem som kan bruke dem. I rapporten *Informasjonsforvaltning i offentlig sektor* (Direktoratet for forvaltning og IKT, 2013) konkretiseres fem punkter som virksomhetene må realisere for å oppnå orden:

1. Datasettene i virksomheten er beskrevet i en oversikt
2. Oversikten over datasettene er publisert
3. Tilgang til data er vurdert og beskrevet
4. Dataelementene er beskrevet (vha. såkalt «metadata»)
5. Beskrivelsene er strukturerte og maskinlesbare

Overordnet vil «orden i eget hus» bidra til at det finnes gode metadata om dataene i Kunnskapssektoren. Gode metadata gjør det vesentlig enklere for konsumentene å få oversikt over tilgjengelig kunnskapsdata, og raskere og enklere ta dataene i bruk.

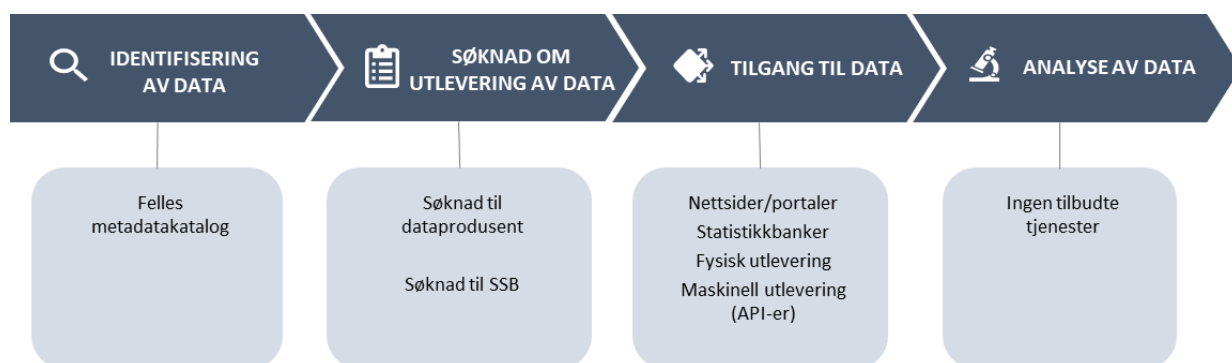
I konsept K2 har en av KDs underliggende etater ansvar for koordinering og samhandling på tvers av alle dataprodusenter i sektoren. Ansaret gjelder på flere områder. For det første har aktøren ansvar for en helhetlig forvaltning av data i sektoren. Et sentralt tiltak er å etablere en felles metadatakatalog for all kunnskapsdata og synliggjøre og forvalte katalogen i etterkant (se nærmere omtale av katalogen under funksjonalitet). I tillegg har aktøren ansvar for å følge opp dataprodusentenes arbeid med «orden i eget hus» og bidra til å tydeliggjøre roller og ansvar i dataforvaltningen, som blant annet vil redusere fragmenteringen av data. Det innebærer blant annet at det er tydelig hvilke produsenter som har ansvar for å dele hvilke data og hvordan delingen organiseres. Videre har aktøren prosessansvar for harmonisering av data og gi retningslinjer for klassifisering av dataene. For det andre har aktøren ansvar for samordning i forbindelse med utlevering av data til konsumenter. Dette omfatter blant annet å opprette et rådgivende forum for juridiske vurderinger rundt utlevering av data. Det kan både gi en mer enhetlig utleveringspraksis og økt juridisk kompetanse knyttet til deling av data hos dataprodusentene.



FIGUR 27: SKISSE OVER ORGANISERING I KONSEPT K2

Funksjonalitet

Figur 28 oppsummerer funksjonaliteten i konsept K2 knyttet til de ulike stegene i konsumentens brukerreise.



FIGUR 28: FUNKSJONALITET I KONSEPT K2

For å **identifisere datakilder** og få oversikt over kunnskapsdata benytter konsumentene i hovedsak den nye felles metadatakatalogen. Dataprodusentene har ansvar for å publisere datasettene de besitter i metadatakatalogen for sitt område ved å beskrive dem og hva som kreves for å få tilgang til dataene. Den koordinerende aktøren har ansvar for å bestemme strukturen på katalogen og gi retningslinjer til dataprodusentene. Katalogen synliggjøres via en nettside, f.eks. slik som Felles datakatalog til Brønnøysundregisteret.

I konsept K2 har alle konsumentene tilgang til kunnskapsdata som er offentlige og publisert gjennom metadatakatalogen. For data med begrenset offentlighet, kan de kan **søke om tilgang** til data direkte til dataprodusentene som i nullalternativet. I konseptet tilbys det ingen felles søknadstjenester i sektoren, men det gjennomføres tiltak for å sikre en mer enhetlig søknadsprosess, blant annet med støtte i et juridisk rådgivende forum. I tilfeller der konsumentene har behov for å koble kunnskapsdata med andre typer data, f.eks. sosioøkonomiske data, må konsumentene henvende seg til SSB som i nullalternativet.

Når det kommer til **tilgjengeliggjøring av data**, medfører realiseringen av «orden i eget hus» at dataprodusentene gjør kunnskapsdata tilgjengelig for konsumentene gjennom API-er. En mulig tilnærming for kontroll på tilgangsstyringen til API-ene er API-gateway hos den koordinerende aktøren.⁷ Dette vil gi konsumentene raskere og enklere tilgang til kunnskapsdata sammenlignet med nullalternativet. Det vil fortsatt være mulig å få utlevert koblede kunnskapsdata på flatfiler dersom konsumentene har behov for dette, men behovet er mindre enn i nullalternativet som følge av økt synliggjøring og enklere tilgang på data gjennom metadatakatalogen og API-er. Utover dette, publiserer dataprodusentene tall og statistikk på nettsider og i statistikkbanker som i nullalternativet.

I tilfeller der konsumenter har behov for data med begrenset offentlighet må SSB eller de ulike dataprodusentene behandle søknadene om tilgang, evt. foreta selve koblingen av datasettene og utlevere dataene. Konsumentene vil i disse tilfellene få en fysisk utlevering som i nullalternativet. Utlevering av data hos SSB skjer både gjennom fysisk utlevering og gjennom microdata-løsningen som i nullalternativet.

⁷ API-gateway, også kalt API-manager, er en tjeneste for å kontrollere og styre tilganger til API-er. Tjenesten sikrer at konsumenter er de de utgir seg for å være (autentisering) og holder rede på de tilganger konsumentene skal ha (autorisasjon). Kildens webservice (tjeneste eller system som tilbyr lese- og/eller skrivetilgang til tjenesten eller systemet) tilbyr en rekke datauttrekk i form av et API. API-manager/gateway er tjenesten som tilgangsstyrrer uttrekkene slik at konsumentene kun har tilgang til avtalte deler av API-et.

I **analysefasen** benytter konsumentene egen analyseinfrastruktur eller -verktøy uavhengig om de henter data direkte på nettsider/portaler, hjemmesider eller gjennom API-er. Ved tilgang til SSBs microdata-løsning må konsumentene benytte analyseverktøyet som er innebygd i løsningen.

Behov for regelverksutvikling

Konseptet forutsetter at det ikke er behov for endringer i regelverket. For å styrke prosessene rundt deling av data opprettes det et rådgivende forum for dataproducentene for å avklare juridiske spørsmål knyttet til deling av data både internt i kunnskapssektoren og med andre. Det rådgivende forumet utarbeider retningslinjer for deling av data med basis i tolkning av gjeldende regelverk.

Personvernkonsekvenser

Konseptet legger opp til at dataproducentene fortsatt ivaretar behandlingsansvaret for behandling av personopplysninger og at eventuelle mottakere av personopplysninger etter en utlevering vil være behandlingsansvarlig for de disse.

Ved opprettelsen av et rådgivende forum må det avklares hvorvidt dette skal ha en rolle som også vil innebære behandling av personopplysninger i forbindelse med rådgivningen, herunder om dette innebærer behov for inngåelse av databehandleravtaler med dataproducentene. Opprettelse av et rådgivende forum vil kunne gjøre det enklere for dataproducentene å etterleve dagens regelverk for utveksling av personopplysninger.

Opprettelse av en felles metadatakatalog som er offentlig tilgjengelig vil kunne bidra til at den registrerte enklere kan skaffe seg oversikt over hvilke aktører som behandler personopplysninger om vedkommende.

Utlevering av datasett på flatfiler via e-post eller minnepinne, og tilhørende liten kontroll rundt informasjonssikkerheten vil også i dette konseptet kunne utgjøre en risiko.

8.1.4. K10 Koordinerte områdesentraler

Konsept K10 har fokus på konsolidering av eksisterende dataproducenter og tilgjengeliggjøring av data for analyse i et sikret analyserom. Konseptet innebærer at dagens dataproducenter er fordelt i et hensiktsmessig antall områdesentraler. Hver områdesentral realiserer «orden i eget hus» innenfor sitt ansvarsområde. En koordinerende aktør får ansvar for koordinering og samhandling på tvers av områdesentralene som i K2. Den koordinerende aktøren har i tillegg ansvar for å tilby en infrastruktur for analyse av kunnskapsdata i sikre analyserom. Det samlede ambisjonsnivået i konseptet mht. behov og utfordringer på konsument- og produsentsiden er høyt.

Organisering og forvaltning av data

I konsept K10 blir det etablert områdesentraler som består av en eller flere av dagens dataproducenter. Hver områdesentral har drifts- og forvaltningsansvar for data innenfor sine ansvarsområder jf. Figur 29.⁸ Konseptet innebærer derfor større organisatoriske endringer sammenlignet med konsept K2.

Hver områdesentral har ansvar for å realisere «orden i eget hus» i henhold til Difis retningslinjer for deling av data innenfor sitt område (Direktoratet for forvaltning og IKT, 2018). «Orden i eget hus»

⁸ I figuren er det illustrert fem områdesentraler. På nåværende tidspunkt er ikke antall områdesentraler avklart, som innebærer at det hensiktsmessige antallet og fordelingen av fagområder hos de ulike områdesentraler, må utredes i et evt. forprosjekt.

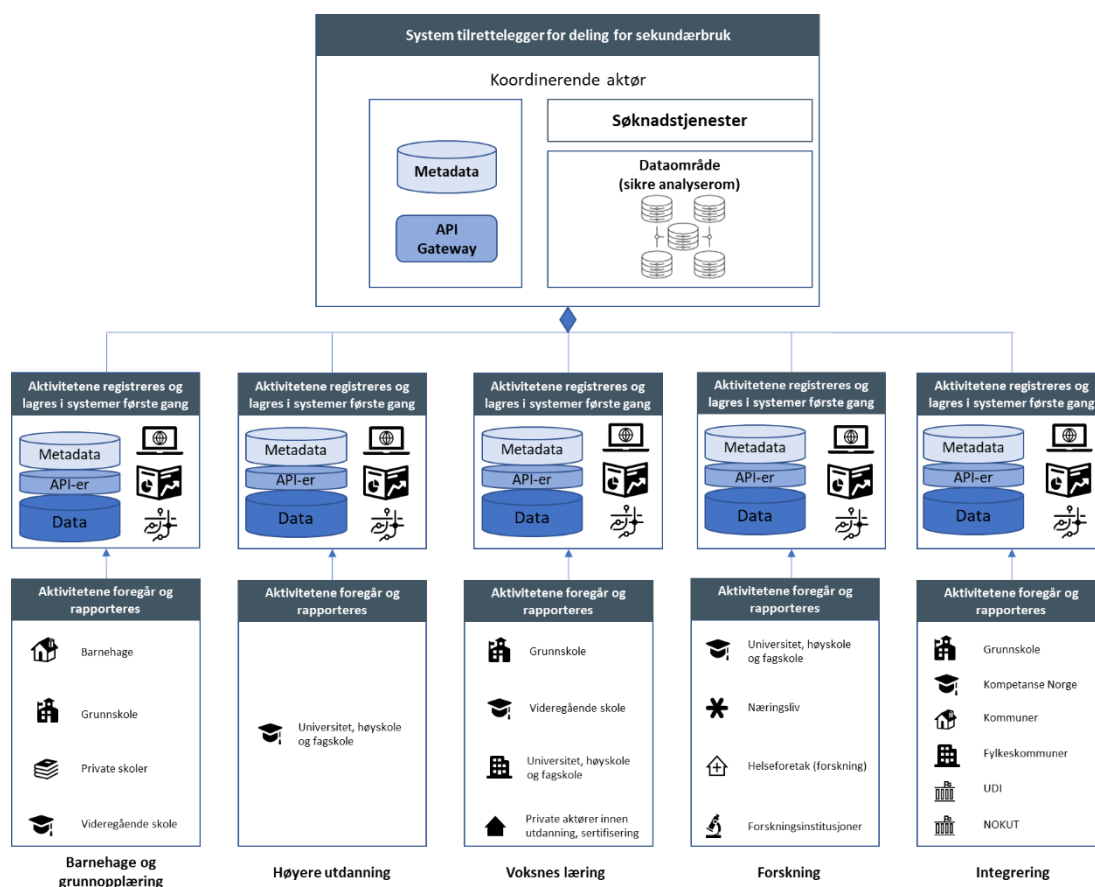
innebærer at den enkelte virksomhet har oversikt over hvilke data den håndterer, hva dataene betyr, hva de brukes til, hvilke prosesser de inngår i, og hvem som kan bruke dem. I rapporten *Informasjonsforvaltning i offentlig sektor* (Direktoratet for forvaltning og IKT, 2013) konkretiseres fem punkter som virksomhetene må realisere for å oppnå orden:

1. Datasettene i virksomheten er beskrevet i en oversikt
2. Oversikten over datasettene er publisert
3. Tilgang til data er vurdert og beskrevet
4. Dataelementene er beskrevet (såkalt «metadata»)
5. Beskrivelsene er strukturerte og maskinlesbare

Overordnet vil «orden i eget hus» bidra til at det finnes gode metadata om dataene i Kunnskapssektoren. Gode metadata gjør det vesentlig enklere for konsumentene å få oversikt over tilgjengelig kunnskapsdata, og raskere og enklere ta dataene i bruk.

I konsept K10 har en av KDs underliggende etater ansvar for koordinering og samhandling på tvers av områdesentralene. Den koordinerende aktøren får ansvar for å tilby en infrastruktur for analyse av kunnskapsdata i sikre analyserom. I tillegg har den koordinerende aktøren ansvar for å tilby en felles søknadstjeneste som koordinerer søknadsprosessen på tvers av områdesentralene for konsumenten som søker om tilgang til data.

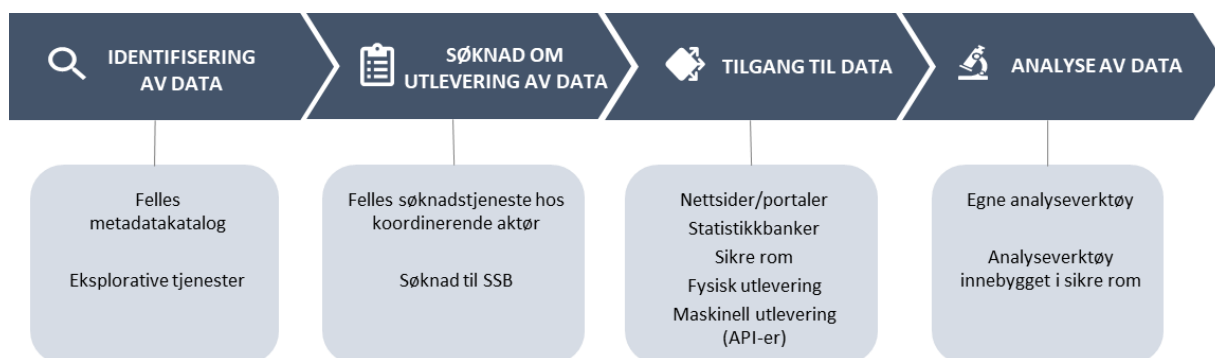
Den koordinerende aktøren har også en rekke andre ansvarsområder for å sørge for koordinering og samhandling på tvers av områdesentralene. Ett av ansvarsområdene aktøren har ansvar for, er en helhetlig forvaltning av data i sektoren. Et sentralt tiltak til dette er å etablere en felles metadatakatalog for all kunnskapsdata, og synliggjøre og forvalte katalogen i etterkant (se nærmere omtale av katalogen under funksjonalitet). I tillegg har aktøren ansvar for å følge opp områdesentralenes arbeid med «orden i eget hus» og bidra til å tydeliggjøre roller og ansvar i dataforvaltningen, som blant annet vil redusere fragmenteringen av data. Det innebærer blant annet at det er tydelig hvilke dataprodusenter som har ansvar for å samle inn og dele hvilke data og hvordan delingen organiseres. Videre har aktøren prosessansvar for harmonisering av data, i tillegg til å gi retningslinjer for klassifisering av dataene. Et annet ansvarsområde til aktøren er ansvaret for samordning i forbindelse med utlevering av data til konsumenter. Dette omfatter blant annet å opprette et forum for juridiske vurderinger rundt utlevering av data. Det kan både gi en mer enhetlig utleveringspraksis og økt juridisk kompetanse knyttet til deling av data mellom områdesentralene.



FIGUR 29: SKISSE OVER ORGANISERING I KONSEPT K10

Funksjonalitet

Figur 30 oppsummerer funksjonaliteten i konsept K10 knyttet til de ulike stegene i konsumentenes brukerreise.



FIGUR 30: FUNKSJONALITET FOR KONSUMENT I KONSEPT K10

For å **identifisere datakilder** og få oversikt over kunnskapsdata benytter konsumentene i hovedsak den nye felles metadatakatalogen. Dataprodusentene har ansvar for å publisere datasettene de besitter i metadatakatalogen for sitt område ved å beskrive dem og hva som kreves for å få tilgang til dataene. Den felles metadatakatalogen samler inn (høster) referanser til metadatabeskrivelser fra områdekatalogene og gir totaloversikt over sektorens data. Den koordinerende aktøren har ansvar for

å bestemme strukturen på katalogen og gi retningslinjer til områdesentralene. Katalogen synliggjøres via en nettside, f.eks. slik som Felles datakatalog til Brønnøysundregisteret. Den koordinerende aktøren tilbyr også eksplorative tjenester gjennom avanserte metadata. Dette gir konsumentene mulighet til å utforske analysemuligheter som ligger i tilgjengelige data, f.eks. om det er mulig å koble ulike datasett via koblingsvariabler.

I konsept K10 har alle konsumentene tilgang til kunnskapsdata som er offentlige og publisert gjennom metadatakatalogen. For data med begrenset offentlighet, kan de **søke om tilgang** til data gjennom søknadstjenesten som den koordinerende aktøren tilbyr. Dette er en felles søknadstjeneste på tvers av områdesentralene som gjør det enkelt for konsumentene å få tilgang til data med begrenset offentlighet dersom de oppfyller vilkårene for bruk av dataene. I tillegg gjøres det tiltak for å sikre en mer enhetlig søknadsprosess ved å opprette et juridisk rådgivende forum. I tilfeller der konsumentene har behov for å koble kunnskapsdata med andre typer data, f.eks. sosioøkonomiske data, må konsumentene henvende seg til SSB som i nullalternativet.

Når det kommer til **tilgjengeliggjøring av data**, medfører realiseringen av «orden i eget hus» at områdesentralene gjør kunnskapsdata tilgjengelig for konsumentene gjennom API-er. En mulig tilnærming for kontroll på tilgangsstyringen til API-ene er API-gateway hos den koordinerende aktøren. Dette vil gi konsumentene raskere og enklere tilgang til kunnskapsdata sammenliknet med nullalternativet. Det vil fortsatt være mulig å få utlevert koblede kunnskapsdata på flatfiler dersom konsumentene har behov for dette, men behovet er mindre enn i nullalternativet som følge av økt synliggjøring og enklere tilgang på data gjennom metadatakatalogen og API-er. Utover dette, publiserer områdesentralene tall og statistikk på nettsider og i statistikkbanker som i konsept K2. Gjennom tydelig ansvarsfordeling mellom områdesentralene er det imidlertid et mindre antall nettsider og statistikkbanker i sektoren sammenliknet med nullalternativet.

I tilfeller der konsumenter har behov for data med begrenset offentlighet må SSB eller den koordinerende aktøren behandle søknadene om tilgang, evt. foreta selve koblingen av datasettene og utlevere dataene. Utlevering av data fra den koordinerende aktøren skjer ved at konsumentene får tilgang til sikre rom hos aktøren, i motsetning til fysisk utlevering som i nullalternativet. De sikre rommene har mekanismer som forhindrer at data kan kopieres eller flyttes ut til et område utenfor det sikre rommet. Hvordan selve dataene synliggjøres for konsumentene i rommene er avhengig av hva konsumentene har fått innvilget tilgang til. I konseptet legges det både til rette for løsninger hvor konsumentene kan se selve dataene (f.eks. som i TSD-løsningen til UiO) og løsninger med anonymiserende brukergrensesnitt (f.eks. som i microdata-løsningen, jf. boks 2 i kap. 2.1). Utlevering av data hos SSB skjer både gjennom fysisk utlevering og gjennom microdata-løsningen som i nullalternativet.

I **analysefasen** benytter konsumentene egen analyseinfrastruktur eller -verktøy uavhengig om de henter data direkte på nettsider/portaler, hjemmesider eller gjennom API-er. Ved utlevering av individdata til sikre rom hos den koordinerende aktøren, må konsumentene benytte analyseverktøyet som er innebygd i løsningen.

Behov for regelverksutvikling

I konseptet er det behov for regelverksendringer for å sikre hjemmel for oppgaver hos den koordinerende aktøren og områdesentralene. Konseptet innebærer et behov for å gjennomgå den koordinerende aktørens og områdesentralenes rolle, samt utrede en løsning for plassering av behandlingsansvaret.

Behovet for å etablere en egen regulering for de sikre analyserommene hos den koordinerende aktøren vil antagelig øke hvis den koordinerende aktøren og områdesentralene skal ha en stor grad av selvstendig ansvar og selvstendige formål med behandlingen som skjer i analyserommene som

etableres. Egen regulering vil redusere behovet for bruk av databehandleravtaler for behandling av personopplysninger i de sikre analyserommene hos den koordinerende aktøren.

Det opprettes et rådgivende forum for å avklare juridiske spørsmål på tvers av områdesentralene knyttet til deling av data internt i kunnskapssektoren og med andre. Forumet utarbeider retningslinjer for deling av data med basis i tolkning av gjeldende regelverk.

Personvernkonsekvenser

Konseptet legger opp til at områdesentralene ivaretar behandlingsansvaret for behandling av personopplysninger og at eventuelle mottakere av personopplysninger etter en utlevering vil være behandlingsansvarlig for disse. Det vil derfor være behov for å avklare roller og ansvar etter personvernregelverket.

Konseptet vil tidvis innebære at det lagres kopier av personopplysninger i de sikre analyserommene, og at det må inngås databehandleravtaler mellom den ansvarlige dataprodusenten og de øvrige dataprodusentene som skal benytte de sikre analyserommene for sin behandling av opplysninger. For den registrerte vil det kunne anses som en risiko at det behandles kopier av personopplysninger og det må derfor legges til rette for tilfredsstillende slette- eller anonymiseringsfunksjonalitet i tråd med kravene til innebygd personvern.

Ved opprettelsen av et rådgivende forum må det avklares hvorvidt dette skal ha en rolle som også vil innebære behandling av personopplysninger i forbindelse med rådgivningen, herunder om dette innebærer behov for inngåelse av databehandleravtaler med områdesentralene. Opprettelse av et rådgivende forum vil kunne gjøre det enklere for dataprodusentene å etterleve dagens regelverk for utveksling av personopplysninger.

Opprettelse av en felles metadatakatalog som er offentlig tilgjengelig vil kunne bidra til at den registrerte enklere kan skaffe seg oversikt over hvilke aktører som behandler personopplysninger om vedkommende.

Risikoen knyttet til fysisk utlevering av datasett med indirekte personidentifiserbare opplysninger vil reduseres gjennom at konseptet tilbyr en løsning med mulighet for behandling av slike datasett i sikre analyserom.

8.1.5. K12 Orden i kunnskapssektoren

Konsept K12 har i likhet med K2 fokus på at hver dataprodusent har bedre oversikt over egne data og realiserer Difis strategi «orden i eget hus». Dette gir en bedre oversikt og enklere tilgang til data i kunnskapssektoren, og dataprodusentenes ansvarsområder blir tydeliggjort. Som i K2, får en av KDs underliggende etater ansvar for koordinering og samhandling på tvers av alle dataprodusentene. Til forskjell fra K2 får den koordinerende aktøren i K12, i likhet med K10, i tillegg ansvar for å tilby en infrastruktur som tilrettelegger for analyse av kunnskapsdata i sikre analyserom. Det samlede ambisjonsnivået i konseptet mht. behov og utfordringer på konsument- og produsentsiden er middels.

Organisering og forvaltning av data

I konsept K12 er forvaltningen av kunnskapsdata desentralisert og spredt på mange ulike aktører, hvor aktørene (dataprodusentene) har drifts- og forvaltningsansvar for dataene innenfor sine ansvarsområder, jf. Figur 31. Konseptet innebærer derfor ingen betydelige organisatoriske endringer fra nullalternativet.

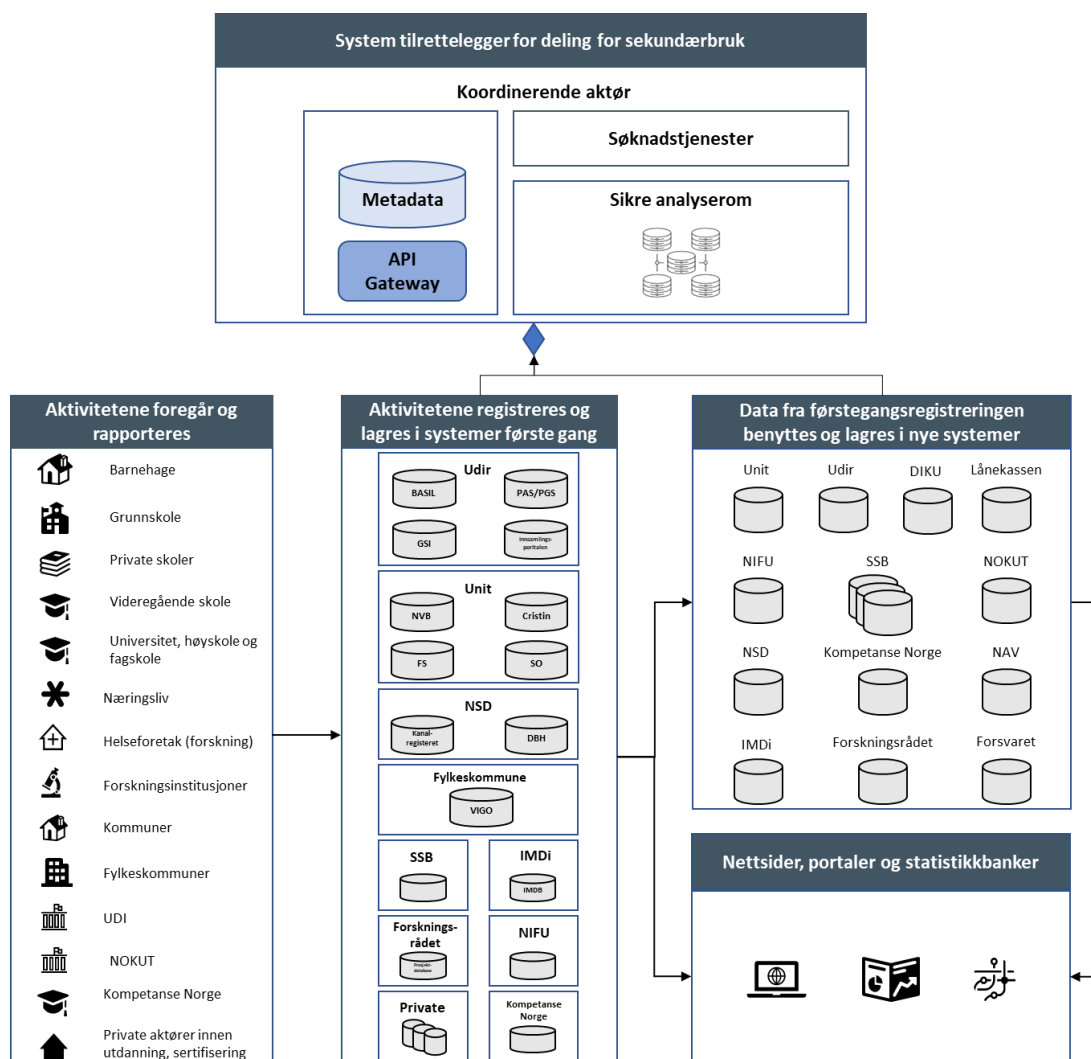
Hver dataprodusent har imidlertid ansvar for å realisere «orden i eget hus» i henhold til Difis retningslinjer for deling av data innenfor sitt område (Direktoratet for forvaltning og IKT, 2018). «Orden i eget hus» innebærer at den enkelte virksomhet har oversikt over hvilke data den håndterer, hva dataene betyr, hva de brukes til, hvilke prosesser de inngår i, og hvem som kan bruke dem. I rapporten *Informasjonsforvaltning i offentlig sektor* (Direktoratet for forvaltning og IKT, 2013) konkretiseres fem punkter som virksomhetene må realisere for å oppnå orden:

1. Datasettene i virksomheten er beskrevet i en oversikt
2. Oversikten over datasettene er publisert
3. Tilgang til data er vurdert og beskrevet
4. Dataelementene er beskrevet (såkalt «metadata»)
5. Beskrivelsene er strukturerte og maskinlesbare

Overordnet vil «orden i eget hus» bidra til at det finnes gode metadata om dataene i Kunnskapssektoren. Gode metadata gjør det vesentlig enklere for konsumentene å få oversikt over tilgjengelig kunnskapsdata, og raskere og enklere ta dataene i bruk.

I konsept K12 har, i likhet med K2 og K10, en av KDs underliggende etater ansvar for koordinering og samhandling på tvers av alle dataprodusenter. Unikt i konsept K12 er at den koordinerende aktøren får ansvar for å tilby en infrastruktur for analyse av kunnskapsdata i sikre analyserom. I tillegg har den koordinerende aktøren ansvar for å etablere en felles søknadstjeneste som koordinerer søknadsprosessen på tvers av dataprodusentene for konsumenten som søker om tilgang til data.

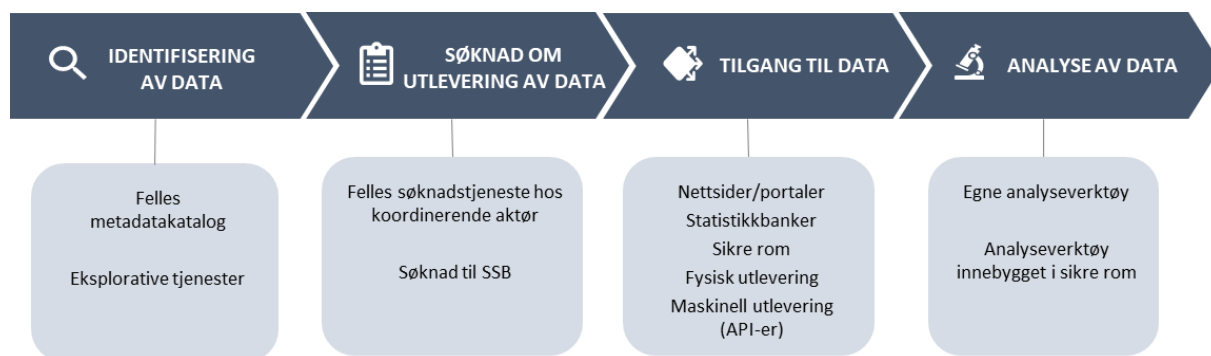
Tilsvarende K2 og K10, har den koordinerende aktøren i K12 også en rekke andre ansvarsområder. For det første har aktøren ansvar for en helhetlig forvaltning av data i sektoren. Et sentralt tiltak til dette er å etablere en felles metadatakatalog for all kunnskapsdata, og synliggjøre og forvalte katalogen i etterkant (se nærmere omtale av katalogen under funksjonalitet). I tillegg har aktøren ansvar for å følge opp dataprodusentenes arbeid med «orden i eget hus» og bidra til å tydeliggjøre roller og ansvar i dataforvaltningen som blant annet vil redusere fragmenteringen av data. Det innebærer blant annet at det er tydelig hvilke produsenter som har ansvar for å samle inn og dele hvilke data og hvordan delingen organiseres. Videre har aktøren prosessansvar for harmonisering av data og å gi retningslinjer for klassifisering av dataene. Et annet ansvarsområde for aktøren er ansvaret for samordning i forbindelse med utlevering av data til konsumenter. Dette omfatter blant annet å opprette et rådgivende forum for juridiske vurderinger rundt utlevering. Det kan både gi en mer enhetlig utleveringspraksis og økt juridisk kompetanse knyttet til deling av data hos dataprodusentene.



FIGUR 31: SKISSE OVER ORGANISERING I KONSEPT K12

Funksjonalitet

Figur 32 oppsummerer funksjonaliteten i konsept K12 knyttet til de ulike stegene i konsumentenes brukerreise.



FIGUR 32: FUNKSJONALITET I KONSEPT K12

For å **identifisere datakilder** og få oversikt over kunnskapsdata benytter konsumentene i hovedsak den nye felles metadatakatalogen. Dataprodusentene har ansvar for å publisere datasettene de besitter i metadatakatalogen for sitt område ved å beskrive dem og hva som kreves for å få tilgang til dataene. Den koordinerende aktøren har ansvar for å bestemme strukturen på katalogen og gi retningslinjer til dataprodusentene. Katalogen synliggjøres via en nettside, f.eks. slik som Brønnøysundregisterets Felles datakatalog. Den koordinerende aktøren tilbyr også eksplorative tjenester gjennom avanserte metadata. Dette gir konsumentene mulighet til å utforske analysemuligheter som ligger i tilgjengelige data, f.eks. om det er mulig å koble ulike datasett via koblingsvariabler.

I konsept K12 har alle konsumentene tilgang til kunnskapsdata som er offentlige. For data med begrenset offentlighet, kan de **søke om tilgang** til data gjennom søknadstjenesten som den koordinerende aktøren tilbyr. Dette er en felles søknadstjeneste på tvers av dataprodusentene som gjør det enkelt for konsumentene å få tilgang til data med begrenset offentlighet dersom de oppfyller vilkårene for bruk av dataene. I tillegg gjøres det tiltak for å sikre en mer enhetlig søknadsprosess ved å opprette et juridisk rådgivende forum. I tilfeller der konsumentene har behov for å koble kunnskapsdata med andre typer data, f.eks. sosioøkonomiske data, må konsumentene henvende seg til SSB som i nullalternativet.

Når det kommer til **tilgjengeliggjøring av data**, medfører realiseringen av «orden i eget hus» at dataprodusentene gjør kunnskapsdata tilgjengelig for konsumentene gjennom API-er. En mulig tilnærming for kontroll på tilgangsstyringen til API-ene er at det etableres en API-gateway hos den koordinerende aktøren. Dette vil gi konsumentene raskere og enklere tilgang til kunnskapsdata sammenlignet med nullalternativet. Det vil fortsatt være mulig å få utlevert koblete kunnskapsdata på flatfiler dersom konsumentene har behov for dette, men behovet er mindre enn i nullalternativet som følge av økt synliggjøring og enklere tilgang på data gjennom metadatakatalogen og API-er. Utover dette, publiserer dataprodusentene tall og statistikk på nettsider og i statistikkbanker som nullalternativet og konsept K2.

I tilfeller der konsumenter har behov for data med begrenset offentlighet må SSB eller den koordinerende aktøren behandle søknadene om tilgang, evt. foreta selve koblingen av datasettene og utlevere dataene. Utlevering av data fra den koordinerende aktøren skjer ved at konsumentene får tilgang til sikre rom hos aktøren, i motsetning til fysisk utlevering som i nullalternativet. De sikre rommene har mekanismer som forhindrer at data kan kopieres eller flyttes ut til et område utenfor det sikre rommet. Hvordan selve dataene synliggjøres for konsumentene i rommene er avhengig av hva konsumentene har fått innvilget tilgang til. I konseptet legges det både til rette for løsninger hvor konsumentene kan se selve dataene (f.eks. som i TSD-løsningen til UiO) og løsninger med anonymiserende brukergrensesnitt (f.eks. som i microdata-løsningen, jf. boks 2 i kap. 2.1). Utlevering av data hos SSB skjer både gjennom fysisk utlevering og gjennom microdata-løsningen som i nullalternativet.

I **analysefasen** benytter konsumentene egen analyseinfrastruktur eller -verktøy uavhengig om de henter data direkte på nettsider/portaler, hjemmesider eller gjennom API-er. Ved utlevering av individdata til sikre rom hos den koordinerende aktøren, må konsumentene benytte analyseverktøyet som er innebygd i løsningen.

Behov for regelverksutvikling

Konseptet innebærer et behov for å gjennomgå den koordinerende aktørens rolle, samt utrede en løsning for plassering av behandlingsansvaret. Behovet for å etablere en egen regulering, f.eks. lovhjemmel, for de sikre analyserommene vil antagelig øke hvis den koordinerende aktøren skal ha en stor grad av selvstendig ansvar og selvstendige formål med behandlingen som skjer i analyserommene. Egen regulering vil redusere behovet for bruk av databehandleravtaler for behandling av personopplysninger i analyserommet.

For å styrke prosessene rundt deling av data opprettes det et juridisk råd for dataprodusentene. Dette er et råd som bistår med å avklare juridiske spørsmål knyttet til deling av data både internt i kunnskapssektoren og med andre. Rådet utarbeider retningslinjer for deling av data med basis i tolkning av gjeldende regelverk.

Personvernkonsekvenser

Konseptet legger opp til at dataprodusentene fortsatt ivaretar behandlingsansvaret for behandling av personopplysninger og at eventuelle mottakere av personopplysninger etter en utlevering vil være behandlingsansvarlig for disse.

Konseptet vil tidvis innebære at det lagres kopier av personopplysninger i de sikre analyserommene, og at det må inngås databehandleravtaler mellom den ansvarlige dataprodusenten og de øvrige dataprodusentene som skal benytte de sikre analyserommene for sin behandling av opplysninger. For den registrerte vil det kunne anses som en risiko at det behandles kopier av personopplysninger og det må derfor legges til rette for tilfredsstillende slette- eller anonymiseringsfunksjonalitet i tråd med kravene til innebygd personvern.

Ved opprettelsen av et juridisk forum må det avklares hvorvidt dette skal ha en rolle som også vil innebære behandling av personopplysninger i forbindelse med rådgivningen, herunder om dette innebærer behov for inngåelse av databehandleravtaler med dataprodusentene. Opprettelse av et rådgivende forum vil kunne gjøre det enklere for dataprodusentene å etterleve dagens regelverk for utveksling av personopplysninger.

Opprettelse av en felles metadatakatalog som er offentlig tilgjengelig vil kunne bidra til at den registrerte enklere kan skaffe seg oversikt over hvilke aktører som behandler personopplysninger om vedkommende.

Risikoen knyttet til fysisk utlevering av datasett med indirekte personidentifiserbare opplysninger vil reduseres gjennom at konseptet tilbyr en løsning med mulighet for behandling av slike datasett i sikre analyserom.

8.1.6. K14 Kunnskapssentralen

I konsept K14 konsolideres alle de eksisterende dataprodusentene til en aktør som har ansvar for forvaltningen av alle data i kunnskapssektoren. I tillegg tilbyr kunnskapssentralen en infrastruktur for analyse av kunnskapsdata i sikre analyserom. Konseptet innebærer at Kunnskapssentralen realiserer Difis strategi «orden i eget hus» i hele kunnskapssektoren. Det samlede ambisjonsnivået i konseptet mht. behov og utfordringer på konsument- og produsentsiden er høyt.

Organisering og forvaltning av data

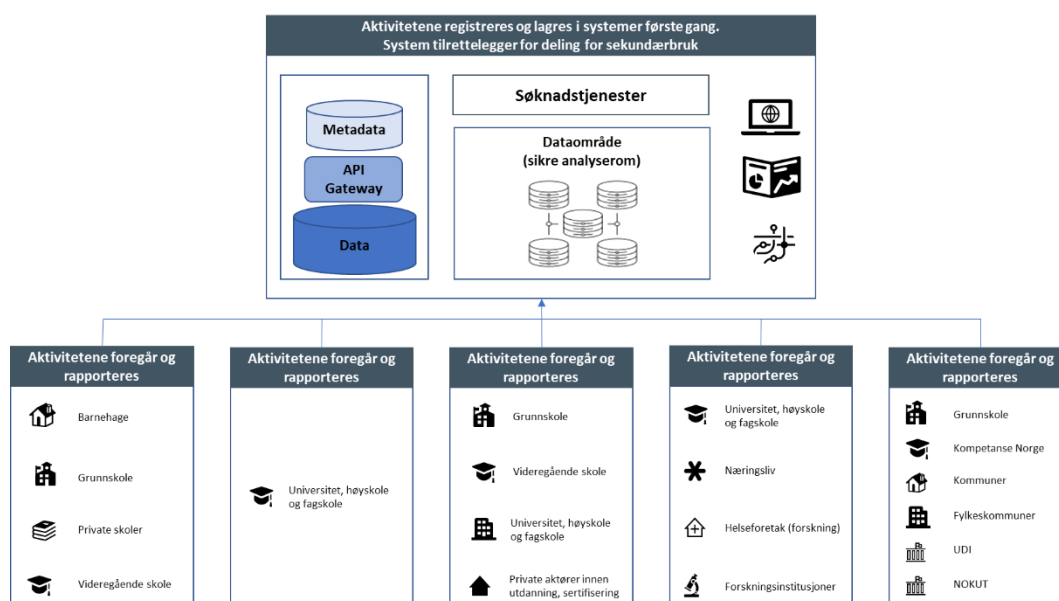
I konsept K14 er forvaltningen av kunnskapsdata sentralisert og kunnskapssentralen har drifts- og forvaltningsansvar for data i hele kunnskapssektoren, jf. Figur 33. Konseptet innebærer derfor større organisatoriske endringer sammenlignet med nullalternativet. Etableringen av kunnskapssentralen innebærer en sammenslåing av alle eksisterende dataprodusenter.

Kunnskapssentralen har ansvar for å realisere «orden i eget hus» i henhold til Difis retningslinjer for deling av data innenfor sitt område (Direktoratet for forvaltning og IKT, 2018). «Orden i eget hus» innebærer at den enkelte virksomhet har oversikt over hvilke data den håndterer, hva dataene betyr, hva de brukes til, hvilke prosesser de inngår i, og hvem som kan bruke dem. I rapporten *Informasjonsforvaltning i offentlig sektor* (Direktoratet for forvaltning og IKT, 2013) konkretiseres fem punkter som virksomhetene må realisere for å oppnå orden:

1. Datasettene i virksomheten er beskrevet i en oversikt
2. Oversikten over datasettene er publisert
3. Tilgang til data er vurdert og beskrevet
4. Dataelementene er beskrevet (såkalt «metadata»)
5. Beskrivelsene er strukturerte og maskinlesbare

Overordnet vil «orden i eget hus» bidra til at det finnes gode metadata om dataene i Kunnskapssektoren. Gode metadata gjør det vesentlig enklere for konsumentene å få oversikt over tilgjengelig kunnskapsdata, og raskere og enklere ta dataene i bruk.

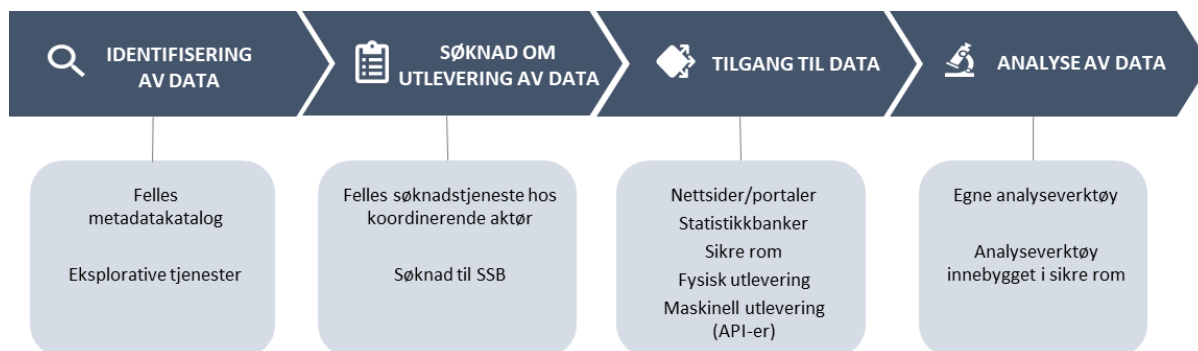
I konsept K14 har kunnskapssentralen ansvar for en helhetlig forvaltning av data i sektoren gjennom å innføre «orden i eget hus og» og etablere og forvalte en felles metadatakatalog for all kunnskapsdata. Videre har kunnskapssentralen ansvar for harmonisering av data, klassifisering av data og tilrettelegging av data for analyseformål. Som eneste aktør som utleverer data til konsumentene, har kunnskapssentralen en enhetlig utleveringspraksis og like juridiske vurderinger. Kunnskapssentralen har, tilsvarende den koordinerende aktøren i K10 og K12, ansvar for å tilby en infrastruktur for analyse av kunnskapsdata i sikre analyserom. I tillegg tilbyr kunnskapssentralen en søknadstjeneste for konsumentene som søker om tilgang til data.



FIGUR 33: SKISSE OVER DATAINFRASTRUKTUREN I KONSEPT K14

Funksjonalitet

Figur 34 oppsummerer funksjonaliteten i konsept K14 knyttet til de ulike stegene i konsumentenes brukerreise.



FIGUR 34: FUNKSJONALITET FOR KONSUMENT I KONSEPT K14

For å **identifisere datakilder** og få oversikt over kunnskapsdata benytter konsumentene i hovedsak den nye felles metadatakatalogen. Kunnskapssentralen har ansvar for å publisere datasettene de besitter i metadatakatalogen for sitt område ved å beskrive dem og hva som kreves for å få tilgang til dataene. Katalogen synliggjøres via en nettside, f.eks. slik som Felles datakatalog til Brønnøysundregisteret. Den koordinerende aktøren tilbyr også eksplorative tjenester gjennom avanserte metadata. Dette gir konsumentene mulighet til å utforske analysemuligheter som ligger i tilgjengelige data, f.eks. om det er mulig å koble ulike datasett via koblingsvariabler.

I konsept K14 har alle konsumentene uten videre tilgang til kunnskapsdata som er offentlige og publisert gjennom metadatakatalogen. For data med begrenset offentlighet, kan de **søke om tilgang** til data gjennom søknadstjenesten som den koordinerende aktøren tilbyr. Dette er en felles søknadstjeneste for alle dataprodusentene som gjør det enkelt for konsumentene å få tilgang til data med begrenset offentlighet dersom de oppfyller vilkårene for bruk av dataene. I tilfeller der konsumentene har behov for å koble kunnskapsdata med andre typer data, f.eks. sosioøkonomiske data, må konsumentene henvende seg til SSB som nullalternativet.

Når det kommer til **tilgjengeliggjøring av data**, medfører realiseringen av «orden i eget hus» at kunnskapssentralen gjør kunnskapsdata tilgjengelig for konsumentene gjennom API-er. En mulig tilnærming for kontroll på tilgangsstyringen til API-ene er API-gateway hos den koordinerende aktøren. Dette vil gi konsumentene raskere og enklere tilgang til kunnskapsdata sammenlignet med i nullalternativet. Det vil fortsatt være mulig å få utlevert koblede kunnskapsdata på flatfiler dersom konsumentene har behov for dette, men behovet er mindre enn i nullalternativet som følge av økt synliggjøring og enklere tilgang på data gjennom metadatakatalogen og API-er. Utover tilgjengeliggjøring gjennom API-er og fysisk utlevering, publiserer kunnskapssentralen tall og statistikk på nettsider og i statistikkbanker. Da kunnskapssentralen er den eneste dataprodusenten i konseptet er det bare én statistikkbank i sektoren.

I tilfeller der konsumenter har behov for data med begrenset offentlighet må SSB eller den koordinerende aktøren behandle søknadene om tilgang, evt. foreta selve koblingen av datasettene og utlevere dataene. Utlevering av data fra kunnskapssentralen skjer ved at konsumentene får tilgang til sikre rom hos kunnskapssentralen, i motsetning til fysisk utlevering som i nullalternativet. De sikre rommene har mekanismer som forhindrer at data kan kopieres eller flyttes ut til et område utenfor det sikre rommet. Hvordan selve dataene synliggjøres for konsumentene i rommene er avhengig av hva konsumentene har fått innvilget tilgang til. I konseptet legges det både til rette for løsninger hvor konsumentene kan se selve dataene (f.eks. som i TSD-løsningen til UiO) og løsninger med anonymiserende brukergrensesnitt (f.eks. som i microdata-løsningen, jf. boks 2 i kap. 2.1). Utlevering av data hos SSB skjer både gjennom fysisk utlevering og gjennom microdata-løsningen som i nullalternativet.



I **analysefasen** benytter konsumentene, ved analyser på åpne data, egen analyseinfrastruktur eller -verktøy uavhengig om de henter data direkte på nettsider/portaler, hjemmesider eller gjennom API-er. Ved utlevering av individdata til sikre rom som tilbys gjennom den koordinerende aktøren, må konsumentene benytte analyseverktøyet som er innebygd i løsningen.

Behov for regelverksutvikling

I konsept K14 er det behov for regelverksendringer for å sikre hjemmel for oppgaver som legges til kunnskapssentralen og å gjennomgå kunnskapssentralens rolle, samt utrede en godløsning for plassering av behandlingsansvaret. Egen regulering vil redusere behovet for bruk av databehandleravtaler for behandling av personopplysninger i den nye infrastrukturen og de sikre analyserommene. Videre er det behov for regelverksendringer for å sikre hjemmel for oppgaver hos kunnskapssentralen.

Personvernkonsekvenser

I konsept K14 er det behov for å gjennomgå kunnskapssentralens rolle, samt utrede en god løsning for plassering av behandlingsansvaret. En egen regulering vil redusere behovet for bruk av databehandleravtaler for behandling av personopplysninger i den nye infrastrukturen og de tilhørende sikre analyserommene.

Konseptet legger opp til en løsning der kunnskapssentralen vil ha selvstendige formål med behandlingen av data som skjer i den nye infrastrukturen. Dette betyr at det er behov for regelverksutvikling. Løsningen vil forhindre en situasjon med dupliserte kilder og fragmentert behandlingsansvar, samt muliggjøre en helhetlig tilnærming til sikkerheten ved behandlingen når vurderingene i stor grad vil gjøres sentralt.

8.1.7. Oppsummering

TABELL 10: SAMMENLIGNINGSTABELL FOR KONSEPTENE

	Null	K2	K10	K12	K14
Organisering	Ingen organisatoriske endringer.	Ingen organisatoriske endringer for dataproducenter En aktør får ansvar for koordinering og samhandling på tvers av alle dataproducentene.	Antall dataproducenter reduseres gjennom fusjonering til områdesentraler. En aktør får ansvar for koordinering og samhandling på tvers av de nye områdesentraler.	Ingen organisatoriske endringer for dataproducentene. En aktør får ansvar for koordinering og samhandling på tvers av alle dataproducentene.	En aktør får ansvar for hele dataproduksjonen i kunnskapssektoren.
Forvaltning av data	Dataproducentene har drifts- og forvaltningsansvar som i dag.	Dataproducentene har drifts- og forvaltningsansvar, og gjennomfører «orden i eget hus» innenfor sitt ansvarsområde.	De nye områdesentralene har drifts- og forvaltningsansvar, og gjennomfører «orden i eget hus» innenfor sitt ansvarsområde.	Dataproducentene har drifts- og forvaltningsansvar, og gjennomfører «orden i eget hus» innenfor sitt ansvarsområde.	Den nye aktøren får drifts- og forvaltningsansvar for all data i sektoren, og gjennomfører «orden i eget hus» i hele sektoren.
Funksjonalitet					
Identifisering av data	Oversikt over data som i dag.	Oversikt over data i enkel metadatakatalog.	Oversikt over data i avansert metadatakatalog. Eksplorative tjenester	Oversikt over data i avansert metadatakatalog. Eksplorative tjenester	Oversikt over data i avansert metadatakatalog. Eksplorative tjenester
Søknadsstøtte	Ingen	Ingen	Felles søknadstjeneste	Felles søknadstjeneste	Felles søknadstjeneste

	Null	K2	K10	K12	K14
Tilgang til data	Dataproducenter tilbyr: Nettsider/portaler Statistikkbanker Fysisk utlevering Maskinell utlevering (API-er)	Dataproducenter tilbyr: Nettsider/portaler Statistikkbanker Fysisk utlevering Maskinell utlevering (API-er)	Områdesentraler tilbyr: Nettsider/portaler Statistikkbanker Fysisk utlevering Maskinell utlevering (API-er) Koordinerende aktør tilbyr: Sikre rom	Dataproducenter tilbyr: Nettsider/portaler Statistikkbanker Fysisk utlevering Maskinell utlevering (API-er) Koordinerende aktør tilbyr: Sikre rom	Kunnskapssentralen tilbyr: Nettsider/portaler Statistikkbanker Fysisk utlevering Maskinell utlevering (API-er) Sikre rom
Analyse av data	Egne analyseverktøy	Egne analyseverktøy	Egne analyseverktøy Analyseverktøy innebygget i sikre rom	Egne analyseverktøy Analyseverktøy innebygget i sikre rom	Egne analyseverktøy Analyseverktøy innebygget i sikre rom
Juridisk	Ingen endringer	Ingen endringer	Regelverksendring for å sikre hjemmel for oppgaver hos koordinerende aktør og områdesentraler. Opprette juridisk forum for å sikre lik tolkning av lovverk.	Regelverksendring for å sikre hjemmel for oppgaver hos koordinerende aktør. Opprette juridisk forum for å sikre lik tolkning av lovverk.	Regelverksendring for å sikre hjemmel for oppgaver hos kunnskapssentral.

8.2. Vurdering av kostnadsvirkninger

Kostnadsvirkninger er ressursbruken som følger av konseptene. Kostnadene som er inkludert i analysen er alle kostnadene som er forventet å påløpe i analyseperioden. Med utgangspunkt i dette er det utformet en gruppering av kostnader tilpasset denne utredningen, jf. figur 36. Alle kostnadene er verdsatt i kroner og vurdert relativt til nullalternativet. Dette innebærer at kostnadene er det som vil påløpe i tillegg til kostnadene som inntreffer i nullalternativet.



FIGUR 35: OVERSIKT OVER KOSTNADER

8.2.1. Prosjektnedbrytningsstruktur (PNS)

Estimeringen av kostnadene i konseptene er basert på en generell projektnedbrytningsstruktur (PNS) som er identisk for alle konseptene og viser hvilke delelementer kostnadene består av.

Vedlegg D: Projektnedbrytningsstruktur illustrerer nedbrytningsstrukturen, hvor alle kostnadselementer som inngår i konseptene også er presentert.

8.2.2. Forutsetninger

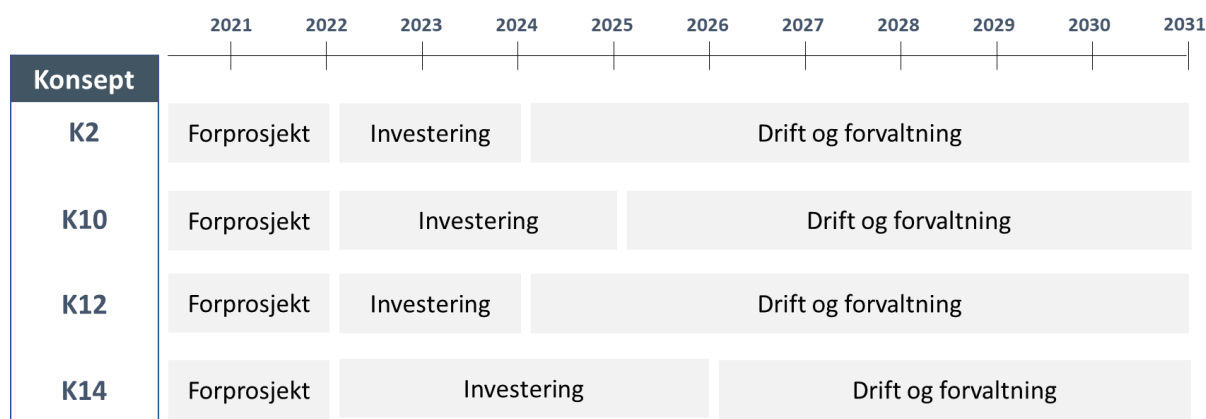
Forutsetningene som er lagt til grunn for verdsettingen av kostnadene er nærmere beskrevet under.

Analyseperiode

Analyseperioden er perioden kostnadene og nytten (jf. kap. 8.3) av et tiltak beregnes for. Det påløper kostnader i forbindelse med forprosjekt fra 2020. I denne utredningen har vi lagt til grunn en forprosjektfase på to år for alle konseptene. I tillegg er det en periode på 10 år som inneholder både investeringer og drift og forvaltning av det innførte konseptet. Oppstartstidspunktet for investeringsperioden er 2022 ettersom de første kostnadene knyttet til investeringer påløper dette året. Det vil si at kostnadene estimeres og vurderes i perioden 2020-2031.

Periodisering av kostnader i analysen

Forprosjektfasen skal legge grunnlaget for en god implementeringsprosess og det valgte konseptet detaljeres med nøyaktig hvilke tiltak som må til for innføringen av konseptet. Overgangen fra investeringsperiode til drift og forvaltning varierer mellom de ulike konseptene som følge av at enkelte konsepter er mer komplekse enn andre. Det er antatt at ny infrastruktur er på plass fra og med 2024 i K2 og K12. Dette betyr at drifts- forvaltningskostnader påløper fra og med dette tidspunktet. I og med at K10 og K14 har en lengre investeringsperiode enn de overnevnte konseptene påløper drifts- og forvaltningskostnader fra henholdsvis 2025 og 2026 for konsept K10 og K14.



FIGUR 36: PERIODISERING AV KOSTNADER I ANALYSEN

Årsverkskostnad

I kostnadsberegningene brukes antall årsverk som estimat for ressursbehovet for de ulike fasene og konseptene. Her inngår både interne og eksterne ressurser. Årskostnaden for en intern ressurs er satt til 1 MNOK eks. mva. og årskostnaden for en ekstern ressurs er 2 MNOK eks. mva.

Uspesifiserte kostnader

I kostnadsestimatene er det et 7 prosent påslag for uspesifiserte kostnader fordi det i en konseptfase er vanskelig å anslå det totale omfanget av løsningen og det vil kunne påløpe kostnader en ikke har forutsett.

Diskontering og kalkulasjonsrente

Diskontering benyttes for å kunne sammenligne alternativer som oppstår på ulike tidspunkt. Dette gjøres ved å omregne virkningene til en verdi et gitt år ved å benytte en diskonteringsrente (kalkulasjonsrente). Kalkulasjonsrenten settes til fire prosent i henhold til Finansdepartementets rundskriv (Finansdepartementet, 2014).

Merverdiavgift, arbeidsgiveravgift og skattefinansieringskostnad

For merverdiavgiften (MVA) benyttes den generelle satsen som ligger på 25 prosent. MVA inngår i alle kostnadsestimatene. For arbeidsgiveravgiften benyttes den generelle satsen som ligger på 25 prosent. Samfunnsøkonomisk kostnad beregnes uten MVA og arbeidsgiveravgift. Arbeidsgiveravgiften varierer i ulike deler av landet. I utredningen brukes et estimat på 13 prosent som tilsvarende landsgjennomsnittet. Skattefinansiering av offentlige tiltak innebærer at kostnader for samfunnet skal inkluderes i den samfunnsøkonomiske analysen. Det fremgår av Finansdepartementets rundskriv R-109/2014 at skattekostnaden bør settes til 20 øre per krone og at denne satsen skal benyttes av alle sektorer.⁹ Skattefinansieringskostnaden inngår i den samfunnsøkonomiske kostnaden.

Under kommer en nærmere beskrivelse av de tre hovedelementene, jf. Figur 35, og hvilke kostnadselementer som inngår i dem. Dette etterfølges av begrunnelsen som er lagt til grunn for kostnadsestimeringen og konseptene sett opp mot hverandre.

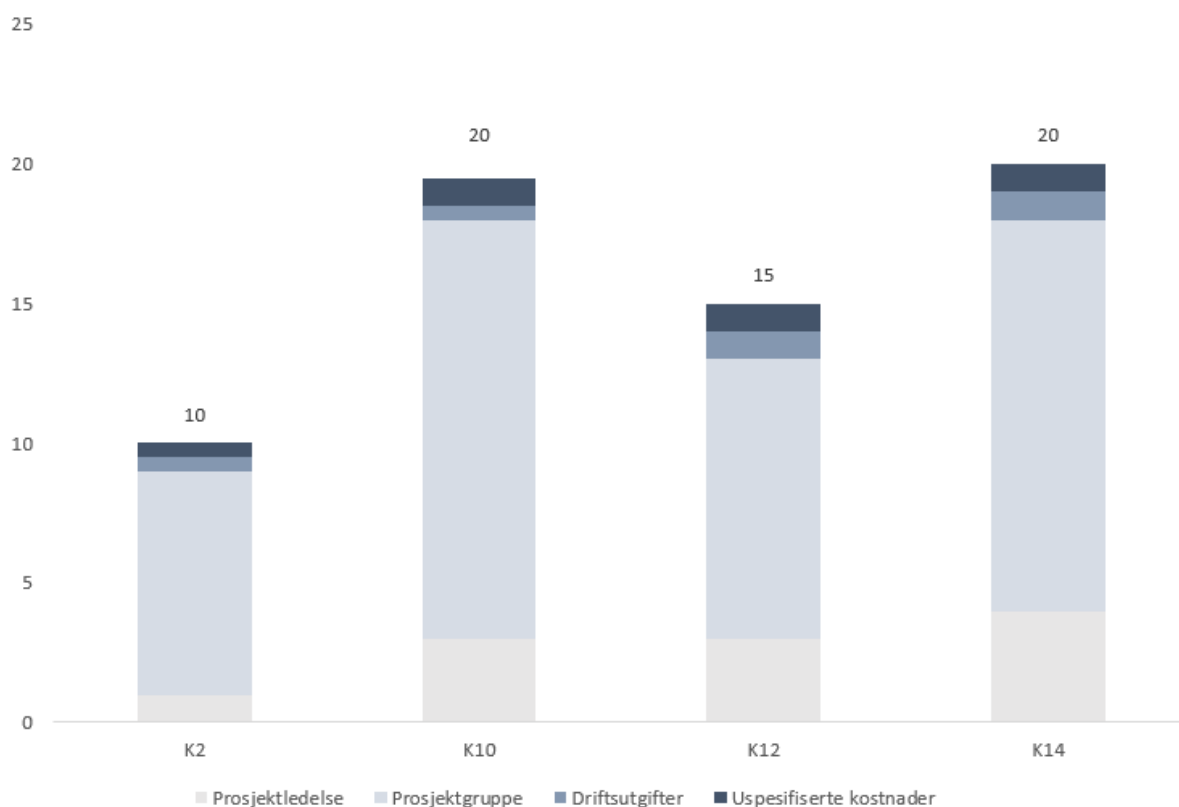
⁹ Alle investerings- og driftskostnader lenger som presenteres i kapittelet vises som faktiske kostnader kunnskapssektoren står overfor. Det vil si at merverdiavgift og arbeidsgiveravgift er inkludert i tallene. For å beregne skattefinansieringskostnader må vi derfor trekke ut merverdiavgift og arbeidsgiveravgift.

8.2.3. Kostnader knyttet til forprosjekt

I forprosjektet fattes det beslutninger som har betydning for utvikling av konseptet som skal innføres, og skal legge grunnlaget for en god implementeringsprosess. Her detaljeres konseptet i høyere grad og nøyaktig hvilke tiltak som må til for innføringen av konseptet.

Hvert konsept har en forprosjektfase som varer i to år. Forprosjektet består av en prosjektgruppe som planlegger gjennomføringen av konseptet. Kostnadene inkluderer et team av prosjektleder(e) og prosjektdeltakere(e), i tillegg til andre driftsutgifter.

I K2 er det lagt til grunn færrest årsverk ettersom det er det minst komplekse konseptet. I de øvrige konseptene er det lagt til grunn flere årsverk, hvor driftsutgiftene øker med antall prosjektmedlemmer. Figur 37 oppsummerer de totale kostnadene (inkl. mva) knyttet til forprosjekt for alle konseptene. K2 og K12 har kostnader i forprosjektfasen på henholdsvis 10MNOK og 15 MNOK. K10 og K14 har de høyeste kostnadene knyttet til denne fasen, begge på 20 MNOK.



FIGUR 37: OPPSUMMERING AV KOSTNADER KNYTTET TIL FORPROSJEKT I MNOK

8.2.4. Investeringskostnad

Investeringskostnader inkluderer alle kostnader knyttet til etablering og innføring av det valgte konseptet. Investeringsperioden varer i to år i K2 og K12, tre år for K10 og fire år for K14. De ulike komponentene som er lagt til grunn i kostnadsestimeringen beskrives nedenfor.

Prosjektledelse og administrasjon

Prosjektledelse og administrasjon inkluderer alle prosjektspesifikke kostnader i forbindelse med investeringen, deriblant prosjektleder, prosjektdeltakere og kommunikasjon, i tillegg til øvrige kostnader knyttet til prosjektet.

Teknisk utvikling

Kostnadsposten teknisk utvikling inkluderer alle kostnader knyttet til etablering av nye tekniske løsninger for forvaltning og deling av kunnskapsdata og alle kostnader som påløper i forbindelse med videreutvikling av allerede eksisterende løsninger. Dette er kostnader knyttet til utvikling av tekniske løsninger i form av videreutvikling av eksisterende infrastruktur eller etablering av en ny infrastruktur som for eksempel sikre analyserom. Kostnadene knyttet til teknisk utvikling omfatter innkjøp og/eller etablering av:

- Ny infrastruktur som sikre analyserom
- Analyseverktøy som skal inngå i infrastrukturen
- Søknadstjenester
- Sentrale integrasjoner
- Lokale integrasjoner
- Oppgradering av dagens løsning

Organisasjonsutvikling

Innenfor kostnadsposten «organisasjonsutvikling» ligger alle kostnader som påløper i forbindelse med organisatoriske endringer som innføres i kunnskapssektoren, for eksempel ny organisering i kunnskapssektoren og hver aktørs ansvar for å skape «orden i eget hus». Organisasjonsutvikling innebærer også kostnader som påløper som konsekvens av at det gjøres organisatoriske endringer eller innføres tiltak som krever omstilling. Dette inkluderer kostnader knyttet til etablering av roller og ansvar, systemendringer, kodeverk og terminologi, opplæring og teknisk bistand til innføring av infrastrukturen og regelverksutvikling. Det påløper kostnader i alle konseptene for regelverksutvikling og det er behov for en gjennomgang og mulige endringer av regelverk knyttet til personvern og informasjonssikkerhet. I tillegg inkluderer noen av konseptene kostnader til opprettelse av et rådgivende juridisk forum.

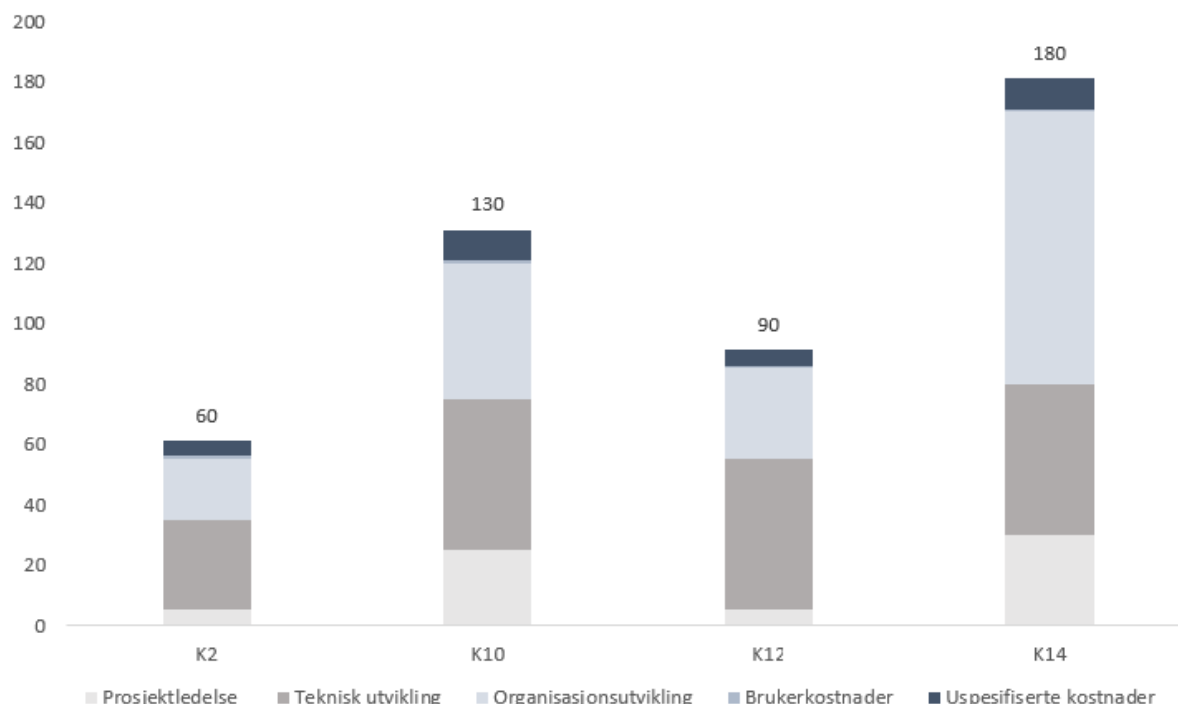
Konsumentkostnader

Ved endringer i infrastruktur og oppbygning av kunnskapssektoren for deling av kunnskapsdata må også konsumentene forholde seg til endringer som blir gjort. Kostnadene som påløper i forbindelse med dette er konsumentkostnadene. Dette omfatter kostnader i forbindelse med innføring av nye systemer og ressurser som kreves for opplæring av konsumentene i disse systemene.

Totale investeringskostnader

Samlet varierer investeringskostnadene betydelig mellom konseptene. K2 har lavest investeringskostnader ettersom det skjer mindre organisatoriske endringer og kun oppgradering av eksisterende systemer. K12 har høyere investeringskostnader enn K2. Det skyldes i hovedsak at den koordinerende aktøren her får en utvidet rolle i forbindelse med harmonisering og tydeliggjøring av roller og ansvar, samt at det i dette konseptet tilbys flere tjenester til konsumentene. K10 krever større organisatoriske og tekniske endringer enn K12. I dette konseptet gjøres det betydelige endringer i organisasjonsstrukturen som også medfører større tekniske endringer fra dagens situasjon, og investeringskostnadene øker deretter. I K14 øker kompleksiteten ytterligere på grunn av en total omorganisering av dataprodusentene i sektoren til en virksomhet. Dette gir økte kostnader knyttet til investering, og konseptet har de høyest investeringskostnadene blant konseptene.

Totale investeringskostnader for de ulike konseptene er vist i Figur 38. Kostnadene er estimert til 60 MNOK i K2, 90 MNOK i K12, 130 MNOK i K10 og 180 MNOK i K14.



FIGUR 38: TOTALE INVESTERINGSKOSTNADER PER KONSEPT I MNOK

8.2.5. Drift- og forvaltningskostnader

Kostnader til drift og forvaltning er estimert fra tidspunktet konseptet anses som implementert og løper frem til analyseperiodens slutt i 2031. Drifts- og forvaltningskostnader inkluderer teknisk drift og forvaltning av infrastrukturen. I tillegg påløper det kostnader for drift av organisasjonen som etableres i noen av konseptene for å forvalte eierskapet til løsningene og yte tjenester til konsumentene. Drift- og forvaltningskostnadene kan deles i kostnader som faller på dataprodusenter og den koordinerende aktøren.

Dataprodusenter

Dataprodusentene får økte drifts- og forvaltningskostnader som følge av konseptene. I kostnadsestimatene for konsept K2 og K12 legges det til grunn at samme antall dataprodusenter som i nullalternativet blir berørt av økte kostnader til drift og forvaltning. I K10 får de fem områdesentralene økte drifts- og forvaltningskostnader og i K14 gjelder dette kun kunnskapssentralen.

Disse kostnadene inkluderer:

- Vedlikeholde «orden i eget hus»
- Vedlikeholde nødvendig kompetanse
- Tilgjengeliggjøre nye datasett ved behov
- Søknadsvurdering knyttet til tilgang på datasett som følge av økt etterspørsel etter data

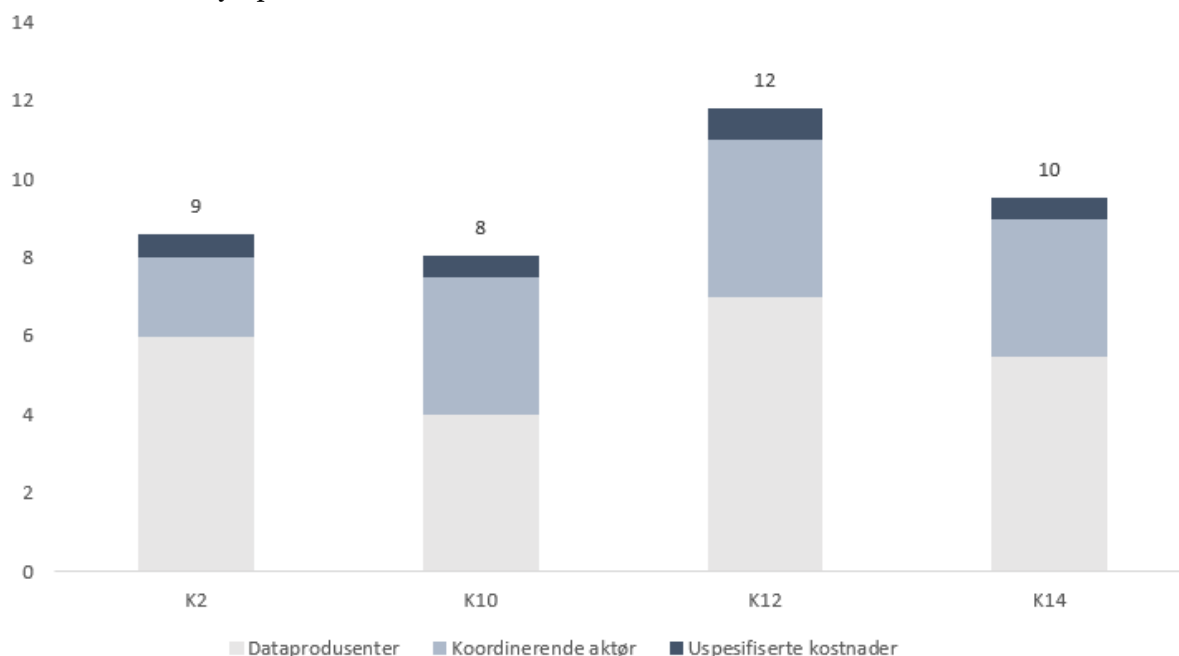
Koordinerende aktør

Koordinerende aktør har ansvar på et mer overordnet nivå enn dataprodusentene, og vil ha kostnader knyttet til det overordnede ansvaret i sektoren samt kostnader for funksjonaliteten som tilbys konsumentene. I kostnadene inngår forvaltning av metadatakatalogen, søknadstjenestene og lisenskostnader knyttet til sikre analyserom.

Totale drifts- og forvaltningskostnader

Drift- og forvaltningskostnadene varierer mellom de ulike konseptene. K10 har de laveste kostnadene knyttet til drift og forvaltning. Dette skyldes at kostnadene i fordeles på de fem områdesentralene. K2 har liten økning i drift og forvaltningskostnader fra K10. Her gjøres det få endringer fra nullalternativet, men samtidig koster det sektoren mye at flere dataprodusenter skal vedlikeholde de oppgraderte løsningene. K14 krever enda høyere drift- og forvaltningskostnader enn de to overnevnte konseptene. I dette konseptet er det kun kunnskapssentralen som har behov for å vedlikeholde «orden i eget hus», men endringene som gjøres er mer omfattende enn i K2 og K10 og dette konseptet har derfor høyere drift- og forvaltningskostnader. En annen årsak er kostnader til den koordinerende aktøren for å vedlikeholde tiltakene som er innført i forbindelse med etableringen av kunnskapssentralen. K12 har de høyeste drift- og forvaltningskostnadene. Konseptet har en mer avansert metadatakatalog og tilbyr bedre konsumenttjenester enn iblant annet K2 som gir økte kostnader.

Figur 39 viser at årlige drift- og forvaltningskostnadene (inkl. mva) varierer fra 8 MNOK i K10 til 12 MNOK i K12 i analyseperioden.



FIGUR 39: ÅRLIGE DRIFTS- OG FORVALTNINGSKOSTNADER PER KONSEPT I MNOK

8.2.6. Basiskostnader

Tabell 11 viser en sammenstilling av basiskostnadene for konseptene som inngår i alternativanalysen. K2 har lavest basiskostnad og K14 har høyest basiskostnad. Hovedårsaken til forskjellen mellom konseptene er ulik grad av kompleksitet og gjennomføringstid.

TABELL 11: BASISKOSTNADER I MNOK

Kostnad	K2	K10	K12	K14
Forprosjekt	10	20	15	20
Investering	60	130	90	180
Sum	70	150	105	200

Tabell 12 viser årlige og totale drifts- og forvaltningskostnader av konseptene. K10 har de laveste drifts- og forvaltningskostnadene totalt, mens K12 har de høyeste kostnadene. I K12 gjøres det oppgraderinger av infrastrukturen samtidig som dagens dataprodusenter opprettholdes. Sammenliknet med K10 og K14 blir flere aktører berørt av høyere drifts- og forvaltningskostnader enn i de andre konseptene.

TABELL 12: BASISKOSTNAD FOR DRIFT OG FORVALTNING I MNOK

Kostnad	K2	K10	K12	K14
Årlige drifts- og forvaltningskostnader	9	8	12	10
Totalt drifts- og forvaltningskostnader for hele analyseperioden	70	55	95	60

8.2.7. Usikkerhetsvurderinger

Estimeringen av kostnadene er på et overordnet nivå og gjennomføringen av det valgte konseptet vil skje tidligst om 2-3 år. Jo lenger frem i tid, desto mer usikkerhet er det knyttet til blant annet teknologisk utvikling, som igjen påvirker kostnadene for de tekniske løsningene. Kostnadsestimatene er dermed beheftet med usikkerhet.

For å få belyst usikkerheten på et overordnet nivå er det valgt å gjennomføre en kvalitativ usikkerhetsvurdering av kostnadsestimatene. I denne vurderingen skilles det mellom optimistisk, basis og pessimistisk kostnadsestimat. Optimistisk kostnadsestimat uttrykker en situasjon der prosjektgjennomføringen går bedre enn forventet, omfanget av nye tekniske løsninger er enklere og organisasjonsendringene og omstillingene er mindre enn forventet. Basis kostnadsestimat tilsvarer kostnadsestimatene som er beskrevet over. Pessimistisk kostnadsestimat innebærer det motsatte av optimistisk kostnadsestimat.

Optimistisk kostnadsestimat

Tabell 13 under viser det optimistiske kostnadsestimatet som ligger om lag 10 prosent under basisestimatet.

TABELL 13: OPTIMISTISK KOSTNADSESTIMAT I MNOK

Kostnad	K2	K10	K12	K14
Forprosjekt	10	20	15	20
Investering	55	115	80	160
Sum	65	135	95	180

For **forprosjektfasene** er det ikke lagt til grunn endringer fra basisestimatet. Det vil imidlertid være tiltak og hendelser som vil kunne bidra til å redusere kostnadene i denne fasen. For eksempel vil prosjektledelsen kunne drive gjennom forprosjektet raskere, eksempelvis ved at planleggingen og detaljeringen av prosjektet er enklere enn antatt. Dette gjør forprosjektet enklere og rimeligere. Tilsvarende vil et tydelig mandat og styring fra Kunnskapsdepartementet bidra til å forenkle arbeidet med forprosjektet. Blant annet ved at forankringsarbeidet med sektorens aktører gjøres enklere.

I **gjennomføringsfasen** er kostnadsestimatene redusert med rundt 10 prosent. Reduksjonen er litt større for K14 enn for de andre konseptene. Reduksjonen bygger på en vurdering av at forprosjektet bidrar til å gi tydelige føringer for gjennomføring av prosjektet, slik at det for eksempel gjennomføres raskere enn antatt i basisestimatet. I K10 og K14 er det behov for organisasjonsendringer. Omstilling og organisasjonsutvikling er krevende, men med god planlegging og riktige ressurser med bred erfaring er det mulig å gjennomføre denne delen av prosjektet mer effektivt enn antatt. Tilsvarende vil det være for arbeidet med å utvikle den tekniske infrastrukturen. For eksempel kan arkitekturen som skal etableres vise seg å være mindre kompleks enn antatt i basisestimatet, slik at den kan utarbeides med færre fagressurser. Parallelt med dette tiltaket er det andre prosesser som har grensesnitt til de løsninger som foreslås blant annet i K10 og K14. Et eksempel på dette er arbeidet med Microdata.no 2.o. Dersom det er mulig å utnytte resultatene fra dette arbeidet, for eksempel i forbindelse med å tilby sikre analyserom, vil det være mulig å redusere ressursbruken knyttet til den tekniske infrastrukturen.

Pessimistisk kostnadsestimat

Tabell 14 under viser pessimistisk kostnadsestimatet som ligger om lag 20 prosent over basisestimatet som er beskrevet over.

TABELL 14: PESSIMISTISK KOSTNADSESTIMAT I MNOK

Kostnad	K2	K10	K12	K14
Forprosjekt	12	24	18	24
Investering	70	155	110	215
Sum	82	179	128	239

For **forprosjektfasene** er det lagt til grunn en økning på rundt 20 prosent sammenlignet med basisestimatet. Mulige årsaker til de økte kostnadene er at det tar lengre tid å mobilisere prosjektet og/eller at prosjektet tar lengre tid å gjennomføre enn det som er antatt i basisestimatene. Dette kan skyldes at forankringen er mer krevende og tar lengre tid. Beskrivelsen av organisering, tekniske løsninger, infrastruktur, inkludert dataflyt mellom aktørene kan også ta lengre tid og øke kostnaden.

I **gjennomføringsfasen** er kostnadsestimatene økt med rundt 20 prosent. Økningen er noe større for K14 enn for de andre konseptene. Økningene i de pessimistiske kostnadsestimatene bygger på en vurdering om at arbeidet med «orden i eget hus» hos dataprodusentene blir mer krevende enn antatt. En annen faktor som kan legges til grunn er at organisering og omstilling blir mer krevende enn antatt i basisestimatet. Konsekvensen er økt ressursbruk, for eksempel av fagressurser innen endringsledelse og organisasjonsutvikling. En tredje faktor som bidrar til å øke ressursbruken er at behovet for regelverksendringer er mer krevende enn antatt. For eksempel kan det være behov for å revidere lovverket i tillegg til forskriftsendringer som ligger til grunn i basisestimatet. Avslutningsvis nevnes også muligheten for at forprosjektet ikke bidrar i tilstrekkelig grad til å gi tydelige føringer for gjennomføring av prosjektet. Dette vil komplisere gjennomføringen og øke ressurspådraget ytterligere.

Samlet vurdering

Samlet vurdert viser den kvalitative usikkerhetsanalysen at det er mer sannsynlig at ressursbruken blir større enn at den blir mindre, sammenlignet med basisestimatet. Kostnadsanslagene som er gjort er, der hvor dette har vært mulig å skaffe, basert på erfaringstall innhentet fra andre etater, som er sammenstilt og vurdert. Kostnadsanslag for noen elementer er basert på realisering av lignende, men ikke helt tilsvarende, løsninger i andre sektorer. Konseptene innebærer en teknologisk kompleksitet som automatisk gir en viss usikkerhet med hensyn til omfanget på investeringene. I tillegg er konseptene på et overordnet nivå, noe som gjør usikkerheten større. Detaljeringsgraden i konseptene medfører også at teknologi og arkitektur ikke er spesifisert på et detaljert nivå. Det er også vanskelig å få oversikt over omfanget av nødvendig regelverks- og organisasjonsutvikling på nåværende tidspunkt,

noe som bidrar til ytterligere usikkerhet i estimatene. Både større regelverksendringer og økt kompleksitet knyttet til teknisk utvikling og organisasjonsendringer vil kunne medføre at prosjektet blir forsinket og at investeringskostnaden øker.

8.2.8. Samfunnsøkonomisk kostnad

For å beregne de samfunnsøkonomiske kostnadene er det tatt utgangspunkt i kalkylenes basiskostnader eksklusiv MVA og arbeidsgiveravgift¹⁰. Tabell 15 viser konseptene neddiskontert for analyseperioden 2021 - 2031.

TABELL 15: NEDDISKONTERTE KOSTNADER I MNOK

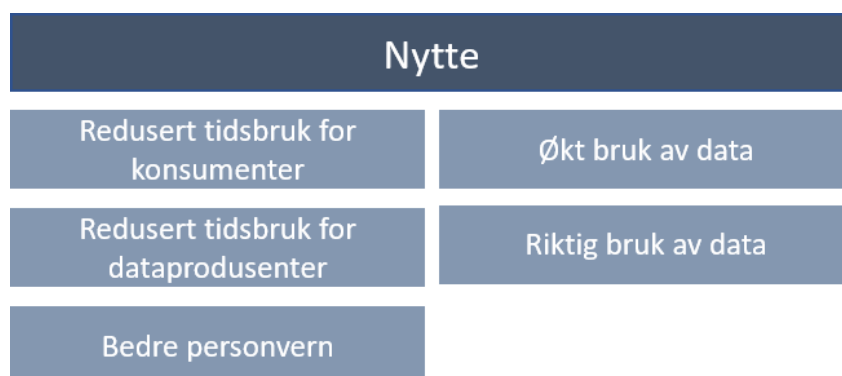
Kostnad	K2	K10	K12	K14
Forprosjekt	5	10	10	15
Investerings	50	100	70	125
Skattefinansieringskostnad	10	20	15	30
Sum	65	130	95	170
Rangering	1	3	2	4

Kostnadsanalysen viser at K2 er det samfunnsøkonomisk rimeligste alternativet. Forskjellen mellom konseptene skyldes hovedsakelig forskjeller i investeringskostnader, som er vesentlig lavere i K2 enn i de andre konseptene. K2 rangeres derfor høyest. K12 har en samfunnsøkonomisk kostnad på 95 MNOK, og med en lavere samfunnsøkonomisk kostnad enn de andre to konseptene er dette rangert som nummer to. Konsept K10 rangeres som nummer tre med en samfunnsøkonomisk kostnad på 130 MNOK. Videre gir K14 en samfunnsøkonomisk kostnad på 170 MNOK og rangeres derfor som nummer fire.

8.3. Vurdering av nyttevirksomheter

Nyttevirksomheter er de positive eller negative effektene av tiltaket vurdert som en endring sammenlignet med nullalternativet. Nytte bidrar til å øke velferden for en eller flere grupper i samfunnet eller samfunnet som helhet. Nyten av de ulike konseptene er oppsummert i Figur 40.

¹⁰ Se kap 8.2.6



FIGUR 40: VURDERING AV NYTTEN AV KONSEPTENE

I alternativanalysen er nyttevirkningene av konseptene verdsatt kvalitativt ved bruk av pluss-minusmetoden, som er anbefalt metode for å vurdere ikke-prissatte virkninger i Finansdepartementets R-109/2014 (Finansdepartementet, 2014). Det første trinnet i pluss-minus metoden er å vurdere **betydningen** de ulike virkningene har for samfunnet. Vurderingen gjøres ved bruk av en tredelt skala; liten, middels og stor betydning. Dersom virkningen har betydning for hele samfunnet beskrives virkningen som «stor betydning for samfunnet». Et vanlig eksempel på en virkning med stor betydning for samfunnet er forsvarsevne, siden denne virkningen inkluderer alle innbyggere, dvs. hele samfunnet. Dersom virkningen har betydning for deler av samfunnet vurderes virkningen til middels. I tilfeller der virkningen kun har betydning for en liten del av samfunnet eller gruppe har virkningen liten betydning for samfunnet. Et eksempel på dette er virkninger som kun berører forskernes arbeidsprosesser. Forskere kan spare tid som følge av bedre datainfrastruktur, men dette har liten samfunnsmessig betydning når virkningene for hele samfunnet legges til grunn.

Det neste steget innebærer å vurdere **omfanget** av virkningen i konseptene sammenlignet med nullalternativet. Skalaen for å vurdere virkningens omfang er kvalitativ. Skalaen skiller mellom intet omfang, litt, middels og stort omfang, samt et skille mellom positivt og negativt omfang. Dersom virkningen gir store forbedringer for *gruppen av samfunnet som vurderes*, vil omfanget være av stort positivt omfang. Et eksempel på en slik virkning er dersom konseptet bidrar til at forskere ikke trenger å gå gjennom en lang og omfattende prosess for å få utlevert data til forskning, men få dem utlevert gjennom automatiserte prosesser. Dersom virkningen kun er en liten forbedring sammenlignet med nullalternativet, vil omfanget være av litt positivt omfang.

Det er den samlede vurderingen av betydning og omfang som avgjør hvordan man vurderer **konsekvensen** av nyttevirkningene. Konsekvensen av tiltaket oppgis ved bruk av en skala med ni kategorier, fra meget stor positiv konsekvens (++++) til meget stor negativ konsekvens (----). Eksempelvis vil en kombinasjon av lav betydning og stort positivt omfang resultere i en middels positiv konsekvens (++). Hele skalaen for pluss-minusmetoden er illustrert i Figur 41.

		Betydning		
Omfang		Liten	Middels	Stor
	Stort positivt	++ Middels positiv konsekvens	+++ Stor positiv konsekvens	++++ Meget stor positiv konsekvens
	Middels positivt	+ Liten positiv konsekvens	++ Middels positiv konsekvens	+++ Stor positiv konsekvens
	Litt positivt	o Ubetydelig konsekvens	+ Liten positiv konsekvens	++ Middels positiv konsekvens
	Intet	o Ingen konsekvens	o Ingen konsekvens	o Ingen konsekvens
	Litt negativt	o Ubetydelig konsekvens	- Liten negativ konsekvens	-- Middels negativ konsekvens
	Middels negativt	- Liten negativ konsekvens	-- Middels negativ konsekvens	--- Stor negativ konsekvens
	Stort negativt	-- Middels negativ konsekvens	--- Stor negativ konsekvens	---- Meget stor negativ konsekvens

FIGUR 41: TABELL FOR PLUSS-MINUSMETODEN

8.3.1. Redusert tidsbruk for konsumentene

Virkningen *Redusert tidsbruk for konsumentene* omfatter all tidsbruk knyttet til innhenting og analyse av data for alle *konsumentene* sett under ett. I nullalternativet bruker konsumentene mye tid på å hente inn informasjon om kunnskapsdata fra de ulike dataprodusentene. Dette skyldes blant annet mangel på en samlet oversikt over data i kunnskapssektoren. Etter å ha funnet riktig datasett, må konsumentene bruke tid på å søke om tilgang, innhente og sammenstille kunnskapsdata. Dette gjelder særlig når data benyttes til forskning og analyser. I de ulike konseptene er det gjort endringer i funksjonaliteten som reduserer tidsbruken for konsumentene. Funksjonaliteten inkluderer blant annet metadatatjenester, søknadstjenester, sikre analyserom og automatisert innhenting av data via API-er, i tillegg til visualiseringer av data på nettsider og ulike portaler dataprodusentene tilbyr. Betydningen og omfanget av den reduserte tidsbruken er redegjort for i de påfølgende avsnittene.

At konsumenter av kunnskapsdata sparer tid til innhenting og analyse, vurderes til å ha **middels samfunnsmessig betydning**. Årsaken er at hele samfunnet ikke har behov for eller bruker kunnskapsdata i stor grad. Likevel er det en såpass stor gruppe av samfunnet som til enhver tid bruker tid på innhenting av kunnskapsdata til at virkningen er mer enn liten. Virkningens betydning for samfunnet vurderes som lik i alle konseptene.

I konsept **K2** får konsumentene av kunnskapsdata oversikt over tilgjengelig data i sektoren og informasjon om dataene gjennom en enkel, felles metadatakatalog hos den koordinerende aktøren. Metadatakatalogen gir oversikt over hvilke data som er tilgjengelig, hvor de finnes og hvordan konsumentene kan få tilgang til dataene. I sum gir dette konsumentene en bedre beskrivelse av kunnskapsdata sammenlignet med nullalternativet. En forbedret oversikt med bedre beskrivelser samlet på et sted gir en tidsbesparelse for konsumentene ved at de bruker mindre tid på leting etter

kunnskapsdata og tid på å forstå beskrivelsene av de ulike datasettene, spesielt i tilfeller der to variabler tilsynelatende er like. Dette gir i sum en tidsbesparelse for konsumentene. I K2 får ikke konsumentene en forbedret søknadstjeneste for tilgang til kunnskapsdata og de må fremdeles søke de ulike dataprodusentene eller SSB for tilgang på individdata fra kunnskapssektoren og eventuelle koblinger med data fra andre sektorer. I tillegg må konsumentene fortsatt forholde seg til et høyt antall ulike dataprodusenter som innebærer mange ulike nettsider, portaler og statistikkbanker. Omfanget av endringen for konsumentene vurderes derfor til **litt positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles er den samlede samfunnsøkonomiske konsekvensen for K2 av virkningen *redusert tid for konsumentene* en liten **positiv konsekvens (+)**.

I konsept **K10** får konsumentene, i likhet med K2, oversikt over tilgjengelige data i kunnskapssektoren gjennom en felles metadatakatalog. I konsept K10 er metadatakatalogen mer avansert enn i K2 og inneholder gode beskrivelser av data som finnes i kunnskapssektoren. I tillegg får konsumentene tilgang på eksplorative tjenester. Dette gir konsumentene mulighet til å utforske analysemuligheter som ligger i tilgjengelige data, blant annet om det er mulig å koble ulike datasett via koblingsvariabler.

Sammenlignet med nullalternativet gir K10 en bedre beskrivelse av data på grunn av like krav for hvordan dataene skal beskrives på tvers av områdesentralene. Bedre oversikt og beskrivelse av dataene reduserer konsumentenes tidsbruk knyttet til leting, i tillegg til lavere usikkerhet knyttet til hvilke data som er korrekte for sitt behov. Gjennom begrepsharmonisering er konsumentene også trygge på hvilke datasett som er autoritative i enhver situasjon. Ved søknad om data skjer dette koordinert på tvers av de ulike områdesentralene, og dermed sparer konsumenten tid i søknadsprosessen. I tillegg medfører det juridiske forumet at dataprodusentene blir mer koordinerte i vurderingene av regelverket slik at søknader tar kortere tid sammenlignet med nullalternativet. Data for forskning og analyse tilgjengeliggjøres via sikre analyserom, som potensielt kan medføre at konsumentene bruker mindre tid på å klargjøre data («vaske») på egenhånd. Ved kobling av data fra andre sektorer enn kunnskapssektoren må dette fortsatt gjøres av SSB, som vil innebære like mye ventetid på dataene som i nullalternativet. I K10 er det økt bruk av API-er sammenlignet med i dag, som fører til spart tid ved oppdateringer av kunnskapsdata man jevnlig har behov for. API-er gir mulighet for å automatisere oppdateringer av data sammenlignet med nullalternativet. Omfanget av endringen er vurdert til **middels positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K10 en **middels positiv konsekvens (++)**.

For konsumenten er konsept **K12** tilnærmet lik K10 ved at begge konseptene gir tilgang på en avansert metadatakatalog, eksplorative tjenester, søknadstjeneste og tilgang til data via sikre analyserom. Omfanget av endringen i K12 er derfor vurdert lik K10 med **middels positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K10 en **middels positiv konsekvens (++)**.

I konsept **K14** får konsumentene, som i K10 og K12, bedre funksjonalitet sammenlignet med nullalternativet. Funksjonaliteten er lik som K10 og K12, men det vil bare være en aktør i sektoren. Dette vil blant annet medføre at søknader om tilgang til kunnskapsdata blir vurdert uten å involvere andre dataprodusenter. I tillegg vil all kunnskapsdata og statistikk være publisert et sted. Samlet sett er derfor endringen vurdert til å være bedre enn K10 og K12, og derfor **stort positivt omfang**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K14 en **stor positiv konsekvens (+++)**.

Vurderingene er oppsummert i Tabell 16.

TABELL 16: VURDERING AV REDUSERT TIDSBruk FOR KONSUMENTENE

Betydning	Middels			
	K2	K10	K12	K14
Omfang	Litt positivt	Middels positivt	Middels positivt	Stort positivt
Konsekvens	+	++	++	+++

8.3.2. Redusert tidsbruk for dataproducentene

Virkningen *Redusert tidsbruk for dataproducentene* omfatter tidsbruken knyttet til forvaltning av deling av data. I nullalternativet består kunnskapssektoren av flere ulike dataproducenter som deler data for videre bruk. Tidsbruken til dataproducentene er knyttet til forvaltning og tilrettelegging av data som tilgjengeliggjøres for konsumentene. I de ulike konseptene kan dataproducentenes tidsbruk påvirkes av endringer iblant annet antall dataproducenter, koordinering blant dataproducentene og hvordan de ulike systemene hos dataproducentene snakker sammen.

I et samfunnsperspektiv har dataproducentene samlet sett liten betydning for samfunnet i sin helhet. Dataproducentenes tidsbesparelse er vurdert til å ha **liten samfunnsmessig betydning**. Betydningen er vurdert lik for alle konseptene.

I konsept **K2** gir bedre koordinering og samhandling på tvers av alle dataproducenter i kunnskapssektoren økt grad av autoritative registre og en større del av dataflyten automatiseres gjennom bruk av API-er, der dataproducentene i nullalternativet må klargjøre datasett manuelt. Dette gir tidsbesparelser for dataproducentene. I tillegg innebærer konseptet mindre behov for å tilrettelegge data til konsumentene, siden konsumentene har bedre informasjon om de ulike datasettene gjennom den felles metadatakatalogen og i større grad benytter seg av API-er. Endringene i K2 er imidlertid er ikke store sammenlignet med nullalternativet. Selv med en tydeligere rolle- og ansvarsdeling, vil det fortsatt være mange dataproducenter i sektoren og en kompleks dataflyt der mange dataproducenter er involvert ved datadeling. Omfanget av endret tidsbruk for dataproducentene i K2 er derfor vurdert til **litt positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K2 en **ubetydelig konsekvens** (0) for samfunnet.

I konsept **K10** er dataproducentene konsolidert i områdesentraler, som reduserer antallet dataproducenter i kunnskapssektoren. I konseptet tilbys det også sikre analyserom, noe som innebærer at dataproducentene bruker mindre tid på å tilrettelegge data for utlevering. Den felles søknadstjenesten vil spare tid totalt sett for dataproducentene, både gjennom koordinering og at det i flere tilfeller vil være færre dataproducenter involvert i en søknadsprosess sammenlignet med nullalternativet. Samlet sett vurderes endringen i tidsbruken for dataproducentene til å være **stort positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K10 en **middels stor konsekvens** (++).

I likhet med K2 er det i konsept **K12** fortsatt mange dataproducenter i kunnskapssektoren. Dataproducentene vil, som i K2 og K10, oppleve en reduksjon i tidsbruken sammenlignet med nullalternativet. I likhet med K10 tilbys det sikre analyserom for deling og bruk av data hos den

koordinerende aktøren som reduserer behovet for å koble og levere ut data til konsumenter. I K12 må fortsatt hver dataprodusent vurdere søknader om tilgang til kunnskapsdata. Selv om den koordinerende aktøren tilbyr en felles søknadstjeneste for konsumentene, vil ikke nødvendigvis dataprodusentene spare tid knyttet til vurdering av søknader. Samlet er tidsbesparelsen for dataprodusentene vurdert til å være lavere i K12 enn i K10, og omfanget av endret tidsbruk for dataprodusentene er derfor vurdert til **middels positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K12 en **liten positiv konsekvens** (+) for samfunnet.

I konsept **K14** reduseres antallet dataprodusenter i kunnskapssektoren til én aktør som får ansvar for å samle inn og tilgjengeliggjøre alle data fra sektoren. Den tekniske løsningen og funksjonaliteten som tilbys til konsumentene og tidsbesparelsen for dataprodusentene vil likevel være relativt lik K10. Omfanget av endret tidsbruk for dataprodusentene er derfor, i likhet med K10, vurdert til å være **stort positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K14 en **middels stor konsekvens** for samfunnet (++).

Vurderingene er oppsummert i Tabell 17.

TABELL 17: VURDERING AV REDUSERT TIDSBRUK FOR DATAPRODUSENTENE

Betydning	Liten			
	K2	K10	K12	K14
Omfang	Litt positivt	Stort positivt	Middels positivt	Stort positivt
Konsekvens	0	++	+	++

8.3.3. Økt bruk av data

Virkingen *Økt bruk av data* omhandler volumet av data som deles i kunnskapssektoren og dermed gjøres tilgjengelig for konsumentene i de ulike konseptene. Årsakssammenhengen for å forklare økt bruk av kunnskapsdata i de ulike konseptene er sammensatt. For det første får konsumentene bedre oversikt over data som finnes i sektoren i alle konseptene. Dataene bli lettere tilgjengelig for konsumentene gjennom forbedrede metadatatjenester som blant annet gir informasjon om hvordan konsumentene kan få tilgang til dataene. I tillegg vil bedre søknadstjenester og økt bruk av API-er medføre økt bruk. Den økte bruken av data vil gi bedre beslutningsgrunnlag i offentlig sektor, som igjen kan medføre økt kvalitet på tjenestene eller en bedre fordeling av samfunnets ressurser.

Økt bruk av data gir bedre kunnskaps- og beslutningsgrunnlag i offentlig sektor. Kunnskap er viktig for hele samfunnet og virkingen er derfor vurdert til å ha **stor samfunnsmessig betydning**. Betydningen er vurdert lik for alle konseptene.

I konsept **K2** får konsumentene bedre oversikt over tilgjengelig data i kunnskapssektoren gjennom metadatakatalogen – som igjen vil kunne medføre økt bruk av data. I tillegg medfører en tydeligere rolle- og ansvarsdeling økt grad av autoritative registre som gir konsumentene økte insentiver til å automatisere innhenting av kunnskapsdata via API-er. Som i nullalternativet, er mange av datasettene fortsatt tilgangsstyrt, og konseptet tilbyr ikke sikre analyserom for å gi konsumentene tilgang til kunnskapsdata på individnivå. Prosessen for tilgang vil fortsatt være ressurskrevende, spesielt for konsumenter som har behov for individdata til forsknings- og analyseformål. Imidlertid er ikke

endringene store sammenlignet med nullalternativet. Blant annet vil det fortsatt være mange dataprodusenter som må koordinere på tvers ved datadeling. Omfanget av endret tidsbruk for dataprodusentene i K2 er derfor vurdert til **litt positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K2 en **middels konsekvens** (++).

I konsept **K10** får konsumentene, i likhet med K2, bedre oversikt over tilgjengelig kunnskapsdata i metadatakatalogen. I konseptet tilbys også eksplorative tjenester konsumentene kan benytte for å finne data de har behov for, og dermed øke bruken av kunnskapsdata. Det vil også være en reduksjon i antallet nettsider og portaler hvor data og statistikk publiseres ved at dataprodusentene er organisert som områdesentraler, som vil gi færre brukergrensesnitt og portaler å forholde seg til for konsumentene. Dette taler for økt bruk av data sammenlignet med nullalternativet. Den felles søknadstjenesten vil også bidra til å senke terskelen for å søke om tilgang til data som krever forhåndsgodkjenning. Dette argumentet underbygges av etableringen av forumet for juridiske vurderinger på tvers av områdesentralene, som vil bidra til at lovverket vurderes likt på tvers. I konsepter er det også rimelig å anta at API-er vil kunne senke terskelen for innhenting av data, ved at det er lettere å hente inn data når tilgang først er gitt. Videre gir den koordinerende aktøren tilgang til sikre analyserom slik at flere konsumentgrupper enn forskere potensielt kan få tilgang til data til analyseformål. Dette kan også medføre at flere konsumenter etterspørre kunnskapsdata. Imidlertid vil kunnskapsdata som skal kobles med data fra andre sektorer fortsatt kun være tilgjengelig via SSB gjennom microdata.no eller tradisjonell utlevering. For denne bruken av data er det få endringer sammenlignet med nullalternativet. Basert på beskrivelsen over er omfanget av endret bruk av data i kunnskapssektoren vurdert til **middels positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles er den samlede samfunnsøkonomiske konsekvensen for K10 en **stor positiv konsekvens** (+++).

I konsept **K12** får konsumentene funksjonalitet som vil være relativt lik som i K10. Den største forskjellen fra K10 er at det fortsatt eksisterer mange ulike nettsider og portaler hvor data og statistikk publiseres. Dette gir ulike visualiseringer og brukergrensesnitt på tvers av dataprodusenten og innebærer derfor ingen endring sammenlignet med nullalternativet. Forskjellen mellom K10 og K12 er likevel relativt liten og omfanget av endret bruk av data i kunnskapssektoren er derfor vurdert til **middels positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K12 en **stor positiv konsekvens** (+++).

I konsept **K14** vil konsumentene oppleve tilgangen på kunnskapsdata lik som i K10 og K12. Omfanget av økt bruk av data i kunnskapssektoren er derfor vurdert til **middels positivt**. I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K14 en **stor positiv konsekvens** (+++).

Vurderingene er oppsummert i Tabell 18.

TABELL 18: VURDERING AV ØKT BRUK AV DATA

Betydning	Stort			
	K2	K10	K12	K14
Omfang	Litt positivt	Middels positivt	Middels positivt	Middels positivt
Konsekvens	++	+++	+++	+++

8.3.4. Riktig bruk av data

Virkningen *Riktig bruk av data* er knyttet til hvordan konsumentene benytter data som tilgjengeliggjøres i kunnskapssektoren. Riktig bruk av data forutsetter data som er godt beskrevet og tilrettelagt for formålet det skal benyttes til. Videre gir riktig bruk av kunnskapsdata bedre beslutningsgrunnlag i offentlig sektor, som igjen kan medføre økt kvalitet på tjenestene eller en bedre fordeling av samfunnets ressurser. Riktig bruk av data i konseptene blir påvirket av blant annet metadatatjenester og «orden i eget hus» blant dataprodusentene.

Riktig bruk av data gir bedre kunnskaps- og beslutningsgrunnlag i offentlig sektor. Kunnskap er viktig for hele samfunnet og virkningen er derfor vurdert til å ha **stor samfunnsmessig betydning**. Betydningen er vurdert lik for alle konseptene.

I konsept **K2** gir metadatakatalogen bedre beskrivelser av dataene i kunnskapssektoren sammenlignet med nullalternativet. I tillegg medfører økt koordinering blant dataprodusentene mer omforent bruk av begrepene og økt grad av autoritative registre i sektoren. Til sammen gjør dette data mer egnet for deling med konsumentgrupper som har ulike formål med sin databruk. Samtidig er konsumentene trygge på at data brukes riktig gjennom beskrivelsene i metadatakatalogen. På den andre siden er det fortsatt mange ulike statistikkbanker og ulike visualiseringer av data hos de ulike dataprodusentene i K2. Konsumentene må derfor i stor grad sammenstille data selv, noe som øker sannsynligheten for at feil bruk og tolkning av kunnskapsdata kan inntreffe. Omfanget av endret tidsbruk for dataprodusentene i K2 er derfor vurdert til **litt positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K2 en **middels konsekvens** (++).

I konsept **K10** får konsumentene, i likhet med K2, bedre oversikt over data gjennom metadatakatalogen. Metadatakatalogen gir imidlertid bedre beskrivelser av dataene sammenlignet med K2 og bedre mulighet for å utforske dataene og deres muligheter – som øker sannsynligheten for at data blir benyttet riktig av konsumentene. Ved at dataprodusentene er samlet i områdesentraler er koordineringen mellom dataprodusentene enklere og en omforent bruk av begrepene øker mulighetene for flere autoritative registre sammenlignet med K2. I tillegg har områdesentralene i større grad mulighet til å utarbeide bedre tilrettelagte data for konsumentene – som igjen øker sannsynlighet for riktig bruk.

Basert på beskrivelsen over er omfanget av endret bruk av data i kunnskapssektoren vurdert til **middels positiv**. I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K10 en **stor positiv konsekvens** (+++).

I konsept **K12** får konsumentene, i likhet med i K2 og K10, oversikt over data gjennom metadatakatalogen. I likhet med K10 er metadatakatalogen mer avansert og den koordinerende aktøren tilbyr bedre tjenester til konsumentene til å utforske og bli kjent med dataene, noe som understøtter riktig bruk av data. I likhet med K2 er det fortsatt mange ulike statistikkbanker og ulike visualiseringer av data, som trekker litt ned. Omfanget av endret bruk er likevel vurdert til **middels positiv**. I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K12 en **stor positiv konsekvens** (+++).

Konsept **K14** har samme egenskaper som K10 og K12. Konsumentene får bedre oversikt over data, bedre beskrivelser av data og bedre tilrettelagte data for deling. I og med at all kunnskapsdata er samlet i kunnskapssentralen gir konseptet høyest grad av autoritative registre – som øker sannsynlighet for at konsumentene benytter seg av riktig data til sitt formål.

Basert på beskrivelsen over er omfanget av endret bruk av data i kunnskapssektoren vurdert til **stort positivt**. I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K14 en **meget stor positiv konsekvens** (++++).

Vurderingene er oppsummert i Tabell 19.

TABELL 19: VURDERING AV RIKTIG BRUK AV DATA

Betydning	Stort			
	K2	K10	K12	K14
Omfang	Litt positivt	Middels positivt	Middels positivt	Stort positivt
Konsekvens	++	+++	+++	++++

8.3.5. Bedre personvern

Et grunnleggende prinsipp er at personopplysninger bare skal behandles for spesifikke, uttrykkelig angitte og berettigede formål. Det er i tillegg et krav at opplysningene ikke skal behandles til andre formål som er uforenlig med det opprinnelige. I slike tilfeller er det nødvendig med eget rettsgrunnlag, vanligvis samtykke. Dette prinsippet gjelder for nullalternativet og de fire konseptene som inngår i alternativanalysen.

Personvern er viktig for alle innbyggere med personlige data i ulike registre i kunnskapssektoren. Registrene omfatter data fra barnehage, grunnopplæring og videregående skole, høyere utdanning, voksnes læring og integrering. Til sammen er antall personer som er registrert i kunnskapssektoren stort og utgjør en betydelig andel av samfunnet. Virkningens betydning for samfunnet er derfor vurdert som **stor**. Virkningenes betydning er vurdert som lik for alle konseptene.

I **K2** er det liten endring i forvaltningen av data i kunnskapssektoren sammenlignet med nullalternativet. Metadatakatalogen som etableres inneholder informasjon over hvilke data som er tilgjengelige, hvor de finnes og hvem som er behandlingsansvarlig. Katalogen inneholder ikke beskyttelsesverdige personopplysninger, men vil hver innbygger vedre oversikt over hvilke data som er lagret om hver enkelt i ulike registre. Innbyggere vil også få styrket kontroll som følge av «orden i eget hus» hos dataprodusentene, der blant annet dataminimering vil begrense innsamling av personopplysninger til det som er nødvendig for innsamlingsformålet.

Personopplysningene er som i nullalternativet lagret hos hver enkelt dataprodusent. Funksjonalitet for utlevering av data vil også være identisk med nullalternativet, men økt bruk av API-er vil likevel kunne redusere tradisjonell utlevering og medføre at personvernet blir litt bedre ivaretatt i K2. Opprettelsen av et forum for juridiske vurderinger der dataprodusentene avklarer juridiske spørsmål knyttet til deling av data internt i kunnskapssektoren og med andre sektorer bidrar til likebehandling og muliggjør samordning av vurderinger for deling av data. Totalvurderingen av omfang knytte til bedre personvern i K2 er på bakgrunn av en helhetsvurdering satt til **litt positivt**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den K2 vurdert til å **middels konsekvens** (++).

I **K10** er et styrket personvern knyttet til etableringen og forvaltningen av en felles metadatakatalog for alle kunnskapsdata, samt opprettelsen av sikre analyserom som reduserer bruken av fysisk utlevering av data til konsumentene. En felles metadatakatalog vil gi hver innbygger bedre oversikt over data som er lagret om hver enkelt i ulike registre, og hvilken områdesentral som er

behandlingsansvarlig for dataene. At dataene er samlet hos færre aktører taler for et styrket personvern. Det legges til grunn at tilgang til de sikre analyserommene kun gis til konsumenter og konsumentgrupper med tjenstlig behov for tilgang. Som i K2 vil felles søknadstjeneste på tvers av områdesentralene og et rådgivende forum for juridiske vurderinger knyttet til deling av data, taler også for et styrket personvern. Basert på beskrivelsen over er omfanget av endringen i personvernet derfor vurdert til **middels positiv**.

I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K10 en **stor positiv konsekvens** (+++).

I **K12** styrkes personvernet, i likhet med K10, gjennom arbeidet med etableringen og forvaltningen av en felles datakatalog for hele sektoren. I likhet med nullalternativet og K2 er data fortsatt fordelt på mange ulike dataprodusenter, som medfører mange ulike databehandlingsansvarlige og at personopplysninger er fordelt hos mange ulike aktører. I likhet med K10, vil den koordinerende aktøren tilby en infrastruktur for analyse av kunnskapsdata i sikre analyserom, som vil redusere behovet for tradisjonell utlevering av data. En felles søknadsprosess og et felles forum for juridiske vurderinger knyttet til utlevering av data vil innebære like vurderinger på tvers av dataprodusentene og dermed bedre etterlevelse av regelverket knyttet til personvern. Lovverket sikrer at konsumenter ikke får uautorisert tilgang til beskyttelsesverdige personopplysninger.

Basert på beskrivelsen over er omfanget av endringen i personvernet vurdert til **middels positiv**. I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K12 en **stor positiv konsekvens** (+++).

Konsept **K14** har samme funksjonalitet som K10 og K12. I og med at all kunnskapsdata samles inn og tilgjengeliggjøres av én aktør i kunnskapssentralen, gir konseptet høyest grad av styrket personvern. Sentralisering av dette ansvaret gjør det enklere å etterleve regelverket for informasjonssikkerhet og personvern. Sammenlignet med nullalternativet og de andre konseptene vil fragmenteringen av data reduseres og det oppnås bedre oversikt over data i sektoren som omfatter beskyttelsesverdige personopplysninger.

Basert på beskrivelsen over er omfanget av endret bruk av data i kunnskapssektoren vurdert til **stort positiv**. I henhold til Figur 41, hvor virkningens betydning og omfang sammenstilles, er den samlede samfunnsøkonomiske konsekvensen for K14 en **meget stor positiv konsekvens** (++++).

Vurderingene er oppsummert i Tabell 20.

TABELL 20: VURDERING AV BEDRE PERSONVERN

Betydning	Stort			
	K2	K10	K12	K14
Omfang	Litt positivt	Middels positivt	Middels positivt	Stort positivt
Konsekvens	++	+++	+++	++++

8.3.6. Oppsummering

Tabell 21 oppsummerer vurderingen av nytten i de ulike konseptene. Alle konseptene vil samlet sett bidra positivt til samfunnet og styrke delingen av data. Felles for alle fire er derfor at de vil redusere samfunnets ressursbruk i forbindelse med deling og bruk av data i sektoren. Gjennom tilrettelegging for økt og riktig bruk av data vil utredninger, analyser og forskning styrkes. Av de ulike konseptene er

K14 konseptet som verdsettes høyest.¹¹ K14 vurderes som det beste konseptet innenfor virkningene *redusert tidsbruk for konsument, riktig bruk av data og bedre personvern*. K2 er verdsatt lavest av konseptene innenfor alle nyttevirkningene.

Merk at de ikke-prissatte virkningene samlet sett er beheftet med usikkerhet og påvirkningen av konseptenes rangering er derfor også usikker. Usikkerheten regnes imidlertid som lik for de ulike konseptene, og det er derfor naturlig å anta at endringer i de ikke-prissatte virkningene ikke virker inn på den samfunnsøkonomiske rangeringen.

TABELL 21: SAMLET OVERSIKT OVER VURDERINGENE AV NYTTEN

Nytte	Konsept			
	K2	K10	K12	K14
Redusert tidsbruk for konsument	+	++	++	+++
Redusert tidsbruk for dataprodusenter	0	++	+	++
Økt bruk av data	++	+++	+++	+++
Riktig bruk av data	++	+++	+++	++++
Bedre personvern	++	+++	+++	++++
Rangering av nytte	4	2	3	1

8.4. Andre vurderinger

Utover virkningene av konseptene knyttet til kostnader og nytte, er konseptenes gjennomføringsrisiko og realopsjon vurdert i denne utredningen.

8.4.1. Gjennomføringsrisiko

Gjennomføringsrisiko er risikoen knyttet til konseptenes gjennomføring, dvs risikoen fra oppstart av prosjektet og frem til prosjektet er fullført slik det er beskrevet i konseptbeskrivelsen.

Gjennomføringsrisikoen er delt inn i organisatorisk og teknisk utvikling. Gjennomføringsrisikoen i nullalternativet er ikke vurdert.

I konsept **K2** er det mindre organisatoriske endringer for dataproducentene i kunnskapssektoren. Endringene skal gjennomføres innenfor eksisterende systemer. Dermed vil dette konseptet antagelig være relativt enkelt å implementere, sammenlignet med K10, K12 og K14. Imidlertid vil det være knyttet gjennomføringsrisiko til hver enkelt dataproducentens evne til å realisere «orden i eget hus» og tydeliggjøringen av roller og ansvar i dataforvaltningen på tvers av de ulike dataproducentene. Som en

¹¹ Det er viktig at tabellen under ikke leses slik at antall pluser og minuser telles. Vurderingen som er beskrevet over er en kvalitativ vurdering av de ulike virkningene, og det er den samlede vurderingen som gir grunnlag for å rangere konseptene.

forlengelse av dette, er det knyttet risiko til den koordinerende aktørens evne til å skape enighet blant dataprodusentene og forvaltningen av metadata. Det er også knyttet risiko til hvordan dataprodusentene vil oppnå en mer enhetlig utleveringspraksis og om de vil rette seg etter de juridiske retningslinjene fra det rådgivende forumet. Det er ikke større teknisk utvikling i konseptet og teknologien vil være basert på kjent og moden teknologi med lav grad av nyutvikling. Gjennomføringsrisikoen i K2 er derfor vurdert til å være **lav**.

I konsept **K10** er det større organisatoriske endringer enn i K2 ved at eksisterende dataprodusenter konsolideres i områdesentraler. Etableringen av områdesentraler vil innebære en større gjennomføringsrisiko, sammenlignet med K2. I konseptet er også gjennomføringen avhengig av at alle områdesentralene klarer å koordinere innsatsen og at roller og ansvar mellom de ulike områdesentralene blir definert på en god måte. Det er også risiko knyttet til den koordinerende aktørens evne til å skape enighet blant områdesentralene og etablere en enhetlig dataforvaltning på tvers av områdesentralene. Det vil også være nødvendig å skape enighet og avklaringer knyttet til søknad- og utleveringspraksis på tvers, inkl. hvordan søknadstjenesten hos den koordinerende aktøren skal fungere og hvordan søknader vurderes på tvers og hva det juridiske rådets rolle skal være.

I K10 vil det også være risiko knyttet til teknologien når dataprodusentene konsolideres til områdesentraler som betyr at det må etableres en ny infrastruktur for alle områdesentralene. Det kreves en koordinering mellom områdesentralene for å sikre en infrastruktur som bygger på samme designprinsipper og helst også felles teknisk plattform. Basert på disse utfordringene er gjennomføringsrisikoen i K10 vurdert til å være **medium/høy**.

I konsept **K12** er det færre organisatoriske endringer sammenlignet med K10 og gjennomføringsrisikoen er derfor lavere. Konseptet er likevel avhengig av at den koordinerende aktøren får til samhandling og koordinering på tvers av et høyt antall dataprodusenter. I likhet med K2, er konseptet avhengig av at alle dataprodusentene realiserer «orden i eget hus». Kravene til dataprodusentene er høyere i K12 enn K2, i og med at metadatakatalogen som etableres i kunnskapssektoren er mer avansert for å muliggjøre eksplorative tjenester og det sikre analyserommet på en tilfredsstillende måte. Forvaltningen av metadatakatalogen blir også mer krevende grunnet det høye antallet dataprodusenter. I tillegg må dataprodusentene bli enige om en felles søknads- og utleveringspraksis, som øker gjennomføringsrisikoen ytterligere.

Teknisk kompleksitet er høyere i konsept K12 enn i K2. Det er ulik modenhet blant dataprodusentene og det er behov for å løfte kompetansenivået. Gjennomføringsrisikoen i K12 er derfor vurdert til **medium** risiko.

Konsept **K14** har høyest gjennomføringsrisiko av konseptene. De organisatoriske endringene i konseptet er større enn i K2, K10 og K12 ved at det etableres en aktør for dataforvaltning i kunnskapssektoren. Det er mange avklaringer som må på plass, blant annet knyttet til lovverk. Kunnskapsentralen vil overta et høyt antall ulike registre og datamodeller med ulik modenhet på teknologien, og må alene realisere «orden i eget hus» i hele kunnskapssektoren. Gjennomføringsrisikoen er i konseptet er derfor vurdert til **høy** risiko.

Gjennomføringsrisikoen i de ulike konseptene er oppsummert i Tabell 22.

TABELL 22 VURDERING AV GJENOMFØRINGSRISIKO

	K2	K10	K12	K14
Gjennomføringssrisiko	L	M/H	M	H

8.4.2. Realopsjoner

Realopsjoner innebærer at det er fleksibilitet og valgfrihet slik at det kan gjøres endringer underveis. Ved å integrere slike opsjoner i tiltakene, økes mulighetene for å tilpasse beslutningene underveis dersom risikofaktorer, ny informasjon eller kunnskap foreligger, og disse har betydning for innretning av konseptene. Den viktigste realopsjonen er som oftest knyttet til muligheten til å innføre tiltak trinnvis, etter hvert som det tilkommer mer informasjon om hvordan de ulike tiltakene fungerer.

Konseptene som inngår i konseptutredningen har ulik grad av realopsjoner. K2 har størst grad av fleksibilitet dersom det over tid viser seg at tiltaket ikke gir de forventede nyttevirksomheter. Videre vurderes både K2 og K12 ha større realopsjoner enn K10 og K14. I konseptene K2 og K12 vil det være mulig å utarbeide og implementere tjenester suksessivt. For eksempel er det mulig å starte med å etablere «orden i eget hus», for deretter etablere en metadatakatalog som gir oversikt over data i kunnskapssektoren. Årsaken til at konseptene K10 og K14 har lavere realopsjoner er at det vil være en omfattende prosess og gå tilbake til en mer desentralisert organisering dersom utviklingen viser at en sentralisert organisering ikke fungerer i henhold til målsettingen.

Samlet sett, vurderes realopsjonene i K2 og K12 som størst. Felles for begge disse konseptene er endringene ikke er omfattende. I K10 er realopsjonen lavere enn for K2 og K12, og K14 vurderes til å ha lavest realopsjoner. Dette knyttes i hovedsak til de store organisatoriske endringene. Vurderingen av realopsjonene i de ulike konseptene er oppsummert i Tabell 23.

TABELL 23 VURDERING AV KONSEPTENES REALOPSJONER

	K2	K10	K12	K14
Realopsjon	H	M	H	L

8.5. Samlet vurdering av konseptene

I dette kapittelet sammenstilles vurderingene av kostnadsvirkninger og nyttevirksomheter, samt andre virkninger (gjennomføringsrisiko og realopsjoner).

Konsept K2 er det beste konseptet dersom kun **kostnadsestimatene** vurderes. I K2 gjøres det små endringer sammenlignet med nullalternativet og de totale kostnadene knyttet til forprosjekt og investering er beregnet til 65 MNOK. Konsept K12 er 5 MNOK dyrere enn K2 som følge av at den koordinerende aktøren her får en utvidet rolle i forbindelse med harmonisering og tydeliggjøring av roller og ansvar, samt at det i dette konseptet tilbys flere tjenester til konsumentene. K10 og K14 har samme funksjonalitet for konsumentene som K12. I tillegg innebærer konseptene større strukturendringer i kunnskapssektoren og har dermed vesentlig høyere kostnadsestimater, henholdsvis 130 MNOK og 170 MNOK.

Alle konseptene vil samlet sett bidra positivt til samfunnet og styrke delingen av data i kunnskapssektoren. De vil også bidra til å redusere samfunnets ressursbruk i forbindelse med deling og bruk av data i sektoren. I tillegg vil økt og riktig bruk av data styrke analyser og forskning som igjen kan medføre økt kvalitet på offentlige tjenester eller en bedre fordeling av samfunnets ressurser. Av konseptene er K14 konseptet som verdsettes høyest innenfor **nyttevirksomhetene** *reduisert tidsbruk for konsument, riktig bruk av data og bedre personvern*. Innenfor *økt bruk av data* vurderes K10, K12 og K14 som like bra, mens K10 og K14 er vurdert best innenfor *reduisert tidsbruk for dataprodusenter*. K14 vurderes derfor som det beste konseptet innenfor nytte, men hverken K10 eller K12 er vurdert som vesentlig mye dårligere. K2 er verdsatt lavest av konseptene innenfor alle nyttevirksomhetene.

Gjennomføringsrisikoen varierer mellom konseptene. I K2 gjøres det små endringer sammenlignet med i dag, og gjennomføringsrisikoen er dermed lav. K10 innebærer større organisatoriske endringer enn i K2 ved at eksisterende dataprodusenter konsolideres i områdesentraler. Dette skaper en større gjennomføringsrisiko enn i K2. I tillegg er det risiko knyttet til koordineringen mellom områdesentralene. I konsept K12 er det færre organisatoriske endringer sammenlignet med K10 og K14 og gjennomføringsrisikoen er derfor lavere. Teknisk kompleksitet økes fra K2 til de øvrige konseptene og bidrar til økt gjennomføringsrisiko. Konsept K14 har høyest gjennomføringsrisiko av konseptene. De organisatoriske endringene i konseptet er større enn i K2, K10 og K12 ved at det etableres en aktør for dataforvaltning i kunnskapssektoren. Det er mange avklaringer som må på plass, blant annet knyttet til lovverk, som bidrar til å øke gjennomføringsrisikoen i dette konseptet.

Vurderingen av **realopsjon** viser at konseptene har ulik grad muligheter til å utsette valg av løsning. Samlet sett har K2 og K12 større realopsjoner enn K10 og K14 fordi konseptene åpner opp for å utarbeide og implementere tjenester suksessivt. Årsaken til at konseptene K10 og K14 har lavere realopsjoner er at det vil være en mer krevende jobb å gjøre endringer etter disse konseptene er implementert dersom utviklingen viser at konseptene ikke fungerer i henhold til målsettingen. Samlet sett vurderes realopsjonene i K2 og K12 som størst. Dette skyldes at organisasjonsendringene i de to konseptene ikke er omfattende. Og det vil være mulig å utarbeide og implementere tjenester suksessivt

Sett under ett vurderes K12 som det beste konseptet, jf. Tabell 24. Investeringsbehovet i K12 er vesentlig lavere enn i K10 og K14. K12 sørger for at det på en kostnadseffektiv måte etableres et rammeverk som sikrer økt og riktig bruk av data. Samtidig oppnås mye av den samme nytten som de beste alternativene (K10 og K14). I tillegg er gjennomføringsrisikoen lav og fleksibiliteten høy. Som en konsekvens av dette forbedres utredninger, analyser og forskning, i og om kunnskapssektoren. Til sammen styrker dette beslutningsgrunnlaget for tiltak og tjenesteutvikling i kunnskapssektoren.

TABELL 24: SAMLET VURDERING AV KONSEPTENE

	K2 Orden i eget hus	K10 Koordinerte områdesentraler	K12 Orden i kunnskaps- sektoren	K14 Kunnskaps- sentralen
Forprosjekt	5	10	10	15
Investering	50	100	70	125
Skattefinansieringskostnad	10	20	15	30
Sum (*)	65	130	95	170
Årlig drift- og forvaltning	9	8	12	10
Rangering kostnad	1	3	2	4
Redusert tidsbruk for konsument	+	++	++	+++
Redusert tidsbruk for dataprodusenter	0	++	+	++
Økt bruk av data	++	+++	+++	+++
Riktig bruk av data	++	+++	+++	++++
Bedre personvern	++	+++	+++	++++
Rangering nytte	4	2	3	1
Gjennomføringsrisiko	Lav risiko	Medium/ høy risiko	Medium risiko	Høy risiko
Realopsjon	Høy	Medium	Høy	Lav
Samlet vurdering	3	2	1	4
*Den samfunnsøkonomiske kostnaden er ikke identisk med kostnad fra kostnadsanalyse da den er neddiskontert, uten mva og arbeidsgiveravgift og inkluderer skattefinansieringskostnad.				

9. Veien videre

I dette kapittelet gis det en beskrivelse av det videre arbeidet for å realisere det anbefalte konseptet K12. Det er viktig å påpeke at konseptet skal utvikles samtidig som andre viktige initiativ knyttet til deling av data i offentlige sektor gjennomføres. Kapittelet gir derfor en kort oversikt over de mest relevante tiltakene hvor grenseflatene til anbefalt konsept vil være relativt store. Koordineringen med disse tiltakene bør gjøres i en avklaringsfase. I tillegg gis det en overordnet beskrivelse av hvordan en avklaringsfase bør innrettes. Avslutningsvis gir det en beskrivelse av suksesskriterier og fallgruver for det videre arbeidet.

9.1. Tiltak med grenseflater til anbefalt konsept

En rekke tiltak og initiativ vil ha tydelige grenseflater til anbefalt konsept K12. Nedenfor gis en overordnet oversikt over tiltak hvor grenseflatene til anbefalt konsept vil være relativt store.

9.1.1. «Orden i eget hus»

Det anbefalte konseptet legger til grunn at alle dataprodusentene i sektoren operasjonaliserer Difis «Orden i eget hus». Det er ingen grunn til å vente med å starte denne jobben. Retningslinjer ved tilgjengeliggjøring av offentlige data (Kommunal- og moderniseringsdepartementet, 2017) som blant flere tiltak også omfatter «orden i eget hus», har eksistert i nærmere tre år. Digitaliseringsrundskrivnet for 2019 pålegger i prinsippet offentlige virksomheter å gjøre sine data tilgjengelig i tråd med viderebruksbestemmelsene i Offentleglova og disse retningslinjene.

Det vil forenkle arbeidet og øke kvaliteten på resultatet ved at dette gjøres gjennom en koordinert innsats slik som anbefalt konsept legger opp til. I stedet for at de enkelte virksomhetene skal gjøre dette isolert, bør en utnytte mulighetene som ligger i en koordinert prosess med tett samhandling for å rydde opp i informasjonsarkitekturen i sektoren. Dette omfatter harmonisering av databegreper, og tydeliggjøring av ansvar og roller som bidrar til å redusere den fragmenterte dataforvaltningen vi ser i dag. Avklaringer av juridiske hindre for deling av data bør også fokuseres i dette arbeidet.

Dette taler for at den koordinerende aktøren som har en sentral rolle i det anbefalte konseptet pekes ut så snart som mulig og kan involveres i forprosjektet.

9.1.2. Felles datakatalog

Felles datakatalog er en oversikt over hvilke data de ulike offentlige virksomhetene har, hvordan dataene henger sammen og hva dataene betyr. I første trinn er Felles datakatalog en felles dataoversikt og en felles forståelse av informasjon i form av begreper. Neste trinn er allerede igangsatt, og har fokus på tilgjengeliggjøring av informasjon i en felles API-katalog. I tredje trinn ligger alt til rette for at gjenbruk av data er hovedregel og for prinsippet om kun én gang.

I K12 etableres en ny felles metadatakatalog for kunnskapssektoren. Denne tilgjengeliggjøres via en egen nettside, for eksempel slik som i Felles datakatalog. Grenseflaten og synergier mellom den nye felles metadatakatalogen og Brønnøysundregisterets katalog bør derfor avklares.

9.1.3. Deling av data i regi av Difi

Difi har startet arbeidet med forprosjekt for deling av data, slik at offentlig data blir enklere tilgjengelig og enklere å bruke. Forprosjektet er startet på bakgrunn av en KVVU som ble gjennomført i 2018 (Direktoratet for forvaltning og IKT, 2018). I det anbefalte tiltaket fra KVVU-en tas det hensyn til at offentlige virksomheter er på ulike modenhetsnivåer. Konseptet anbefaler at det utvikles felles tekniske løsninger og kompetansetiltak for de virksomhetene som ikke har kommet så langt. De modne virksomhetene gjennomfører sine planer, men vil fremdeles være viktige bidragsytere for å

sikre tverrgående løsninger. Etablering av datafordeler, datasjø og et nasjonalt ressurscenter for deling av data er sentrale anbefalinger i utredningen.

I det anbefalt konseptet K12 i denne utredningen, er en av tiltakene at det etableres et rådgivende forum for juridiske vurderinger knyttet til deling av data. Fordeling av ansvar mellom dette forumet og det nasjonale ressurscenteret for deling av data hvor jus inngår som ett av flere kompetanseområder, må avklares. En hensiktsmessig løsning kan være at det nasjonale ressurscenteret spiller på juridiske ressurser i de ulike sektorene med spisskompetanse innenfor sine sektorer. Det rådgivende, juridiske forumet i konsept K12 vil besitte slik spisskompetanse om kunnskapssektoren.

Konseptet i Difis konseptvalgutredning kjennetegnes ved at det synliggjør en utviklingsretning. Tiltakene inneholder ikke alt som må gjennomføres og kan justeres og tilpasses basert på læring underveis. Det legges videre vekt på smidig gjennomføring og det legges opp at initiativer knyttet til deling av data i offentlig sektor skal utnyttes og styrke valgt konsept.

Anbefalt konsept K12 må sees i lys av Difis arbeid. I planleggingen av det videre arbeidet bør det avklares hvordan samhandlingen med Difis forprosjekt kan ivaretas.

9.1.4. Andre eksisterende løsninger

Det finnes i dag allerede løsninger i kunnskapssektoren som ivaretar utvalgte konsumentgruppers behov for analyser på sensitive data i sikre analyserom. Eksempler på dette er UiO og UiB sine etablerte tjenester – Tjenester for Sensitive Data (TSD) og Sikker Adgang til Forskningsdata og E-infrastruktur (UiO og UiB) - som gjør det mulig å samle inn, lagre og forske på sensitive data. Disse tjenestene har grenseflater med utredningens anbefalte konsept K12 for hvordan dataene fra kunnskapssektoren kan overføres til UiOs og UiBs tjenester mest mulig smidig for konsumentene av data. I tillegg bør det vurderes om og hvordan disse tjenestene kan benyttes for å dekke behovene for sikre analyserom sett opp mot andre tilsvarende tjenester, vil være et aktuelt tema i avklaringsfasen. I dag overføres dataene fra kunnskapssektoren i all hovedsak via SSB.

9.1.5. Microdata.no 2.0

Microdata.no er en tjeneste som samler og sammenstiller data fra kunnskapssektoren og andre sektorer. I dag er det 7-8 variabler i microdata.no knyttet til utdanning. Spørsmålet om flere utdanningsvariable i løsningen er ikke avhengig av prosjektet direkte, men mer knyttet opp mot en etterspørselsdrevet prioritering av SSBs arbeid med løpende tilrettelegging av nye data. I begynnelsen av 2020 starter arbeidet med Microdata.no 2.0. Prosjektet har en varighet på fire år, dvs. innenfor samme tidsperioden som anbefalt konsept planlegges for. Prosjektet planlegger løsninger som forenkler og automatiserer tilrettelegging og innlasting av nye data. Omfanget av data vil gradvis økes, med første utvidelse ved utløpet av 1. kvartal 2020 hvor det også vil bli lagt inn noen flere variabler fra utdanningssektoren.

Det er ingen hindre i lovverk for å åpne for nye konsumentgrupper så lenge personvern hensynet fortsatt ivaretas i løsningen. Prioriteringen av nye konsumentgrupper er mer et praktisk spørsmål, hvor samarbeid om nye data og nye konsumenter vil drives av behov og interesse, men fortsatt i kontrollerte former. Det er et mål i prosjektet at microdata.no også skal være en løsning for at andre register-/dataeiere skal kunne gjøre dataene tilgjengelig for kobling og analyse i tjenesten. Kreftregisteret inngår som samarbeidspartner for å utvikle konsepter og løsninger for dette formålet. Per i dag er det vanskelig for prosjektet å forskuttere når en slik tjeneste vil være moden nok for utrulling.

Arbeidet med Microdata.no 2.0 har tydelige grenseflater med anbefalt konsept og sektorens arbeid med deling av data. Etterhvert som microdata.no-prosjektet når målet sitt med å bli en løsning for at andre register-/dataeiere kan gjøre dataene tilgjengelig på microdata.no 2.0 vil dette møte mange av

de behovene som er identifisert i denne utredningen. Om microdata.no 2.0 blir tilgjengelig for flere konsumentgrupper enn kun forskere som i dag, som for eksempel saksbehandlere i kommuner, ansatte i underliggende etater i kunnskapssektoren og analysemiljøer i privat sektor, vil deler av behovet for tilgang til data til forsknings- og analyseformål være dekket.

For det videre arbeidet med anbefalt konsept K12 er det derfor viktig å avklare samarbeidet og samhandlingen med NSD og SSB og deres arbeid med microdata.no 2.0. Hvordan samarbeidet innrettes vil ha direkte føring for hvordan arbeidet med detaljeringen av K12 skal være.

9.1.6. Livslang læring

Kompetansereformen «Lære hele livet» har ført til en rekke utredninger og tiltak som skal sikre at «ingen går ut på dato» og at flere skal kunne stå i jobb lenger. Begrepet livslang læring oppstod gjennom St.meld. nr. 42 (1997-98) Kompetansereformen og defineres som «den opplæring og tilegnelse av kunnskaper som mennesker opplever hele livet».

Livslang læring forventes å påvirke rammene for utdanningsområdet og kan potensielt påvirke både organiseringen av kunnskapssektoren og skape et større behov for å følge enkeltindivider gjennom det livslange læringsløpet. For å følge enkeltindivider er det behov for å koble data fra alle deler av læringsløpet via gjennomgående identifikatorer for enkeltindivider. I dag eksisterer data på individnivå kun for deler av dette livsløpet. Utdanningssektoren har behov for en gjennomgående, anonymiserende identifikator som gjør det mulig å følge enkeltindivider uten at identiteten avsløres. Mulig realisering er foreløpig på idé-stadiet og det må avklares både juridisk og politisk. En underliggende identifikator vil ikke gi full anonymisering da noen må ha ansvar for en kobling av identifikatoren til et konkret individ. Det ligger imidlertid et stort potensial i en slik identifikator for å kunne følge opp effektene av livslang læring og gi et kunnskapsgrunnlag for kontinuerlig forbedring av tiltakene.

I det videre arbeidet med å realisere det anbefalte konseptet K12 er det viktig å involvere seg i å utforske mulighetene for å innføre en slik gjennomgående identifikator.

9.2. Avklaringsfasen

Tiltakene som er beskrevet over, er også i utvikling. En konsekvens av dette er tiltakene på nåværende tidspunkt ikke er tilstrekkelig definerte for å identifisere blant annet synergier. Det kan derfor være hensiktsmessig å gjennomføre en avklaringsfase for anbefalt konsept hvor grenseflatene med andre tiltak undersøkes nærmere. Avklaringsfasen vil kunne avklare eventuelle synergier med de ovennevnte tiltakene som kan innebære gjenbruk i stedet for at tjenester som dekker samme behov utvikles flere ganger. I en avklaringsfase vil det også være hensiktsmessig å utarbeide en mer detaljert gjennomføringsplan. Dette innebærer blant annet å beskrive gjennomføringsstrategi for forprosjektfasen, herunder fastsette hvilke tiltak som kan gjennomføres relativt raskt og hvilke tiltak som krever detaljert planlegging i et eget eller egne forprosjekt. I praksis innebærer dette at resultatet fra avklaringsfasen vil være en «plan for planen», dvs. en omforent gjennomføringsstrategi forprosjektfasen for K12 med berørte aktører og interessenter.

Avklaringsfasen bør startes opp raskest mulig, slik at det blir mulig «å hekte seg på» andre tiltak som har tydelige grenseflater til anbefalt konsept. Innenfor deling av data skjer utviklingen raskt og dersom det ventes for lenge vil man kunne risikere at deler av resultatene fra utredningen blir utdatert, og man kan risikere å måtte starte forfra igjen.

Avklaringsfasen inngår ikke i basiskostnadene for konsept K12. Årsaken er at en avklaringsfase normalt ikke er neste trinn etter en konseptutredning, men i denne utredningen en anbefaling for å ivareta avhengighetene til andre viktige initiativ knyttet til deling av data i offentlige sektor

9.3. Organisering av avklaringsfasen og forprosjektet

Det anbefales at forprosjektet blir gjennomført som et prosjekt i regi av en av dataprodusentene i kunnskapssektoren. Forprosjektet skal utarbeide et sentralt styringsdokument. Dette innebærer blant annet å beskrive gjennomføringsstrategi, organisering og struktur av arbeidet, detaljere tiltakene, estimere og legge disse ut i tid med mer. Dette arbeidet bør følge den nye veilederen for digitaliseringsprosjekter i Statens prosjektmodell.

9.4. Suksessfaktorer og fallgruver

Kritiske suksessfaktorer er forhold som prosjektet må lykkes med for å oppnå gode resultater sett opp mot målsettingene. Primært vil disse være knyttet til resultatmålene (Kostnad, Tid og Kvalitet), som bidrar til å oppnå effektmålene og skape nytteverdi. Suksessfaktorene er således prosjektspesifikke, og bygger på det overordnede usikkerhetsbildet fra prosjektets usikkerhetsanalyse. Suksessfaktorene gjelder i utgangspunktet alle prosjektets aktører og gir et bilde over hva som kreves av aktørene for at prosjektet skal bli en suksess. Ut ifra suksessfaktorene vil en vurdere de mest relevante tiltak som optimaliserer mulighetene å oppnå prosjektets målsettinger.

Det er identifisert en rekke kritiske suksessfaktorer, som i ulik grad vil treffe ett eller flere resultatmål og dernest effektmålene:

- Det må sikres nødvendig fleksibilitet i løsningen slik at den er levedyktig over hele prosjektets levetid. Dette innebærer løsninger som er fleksible for utvidelser og endringer i teknologi, organisatoriske endringer mv.
- Godt forankrede brukerbehov med alle brukergrupper for å sikre riktig funksjonalitet.
- God prosjekteierstyring for å sikre effektive beslutningsprosesser og nødvendig ressurstilgang. Dette innebærer forankring av arbeidet på tvers kunnskapssektoren.
- Prosjektledelse og prosjektorganisasjon må bemannes med tilstrekkelig ressurser med nødvendig kompetanse, jf. den nye veilederen for digitaliseringsprosjekter. Prosjektets egenart og risiko tilsier at prosjektet krever betydelige ressurser mht. planlegging, styring, ledelse og gjennomføring.

Fallgruver er det motsatte av suksessfaktorer og viser til forhold som kan bidra til at prosjektets målsettinger ikke innfris. I denne utredningen er ikke fallgruver vurdert ut over at de vil uttrykkes som det motsatte av ovenstående suksessfaktorer. Slike forhold søkes unngått gjennom relevante tiltak og vil spesifiseres i forprosjektet.

10. Vedlegg

10.1. Vedlegg A: Oversikt over dataprodusenter og datakilder

Tabell 25 gir oversikt over dataprodusenter og deres respektive datakilder.

TABELL 25: OVERSIKT OVER DATAPRODUSENTER OG DATAKILDER

Aktør	Data-kilde	Nettportal	Innhold
Direktoratet for internasjonalisering og kvalitetsutvikling (DIKU)	Espresso		Dikus søknads- og rapporteringssystem Espresso blir brukt til å søke støtte fra, og rapportere på de fleste programmer og ordninger som Diku administrerer, med unntak av Erasmus+.
Integrerings- og mangfoldsdirektoratet (IMDi)	NIR		Register over enkeltpersoners deltakelse i opplæring i norsk og samfunnskunnskap og introduksjonsprogram etter introduksjonsloven. NIR omfatter også personer som er i målgruppa for opplæring i norsk og norsk kultur og norske verdier for beboere i mottak (asylsøkere).
Fylkeskommunene	VIGO		Vigo IKS (interkommunalt selskap) ivaretar utvikling av fylkeskommunens IT-systemer innen videregående opplæring.
Kompetanse Norge / Udir		Utdanning.no	Utdanning.no er den nasjonale nettportalen for informasjon om utdanning og yrke, med oversikt over det norske utdanningstilbudet og lærlingeplasser. Ansvar for portalen blir overført fra Udir til Kompetanse Norge 1.1.2020.
Nasjonalt organ for kvalitet i utdanningen (NOKUT)		NOKUT-portal	NOKUT-portal inneholder indikatorer som er relevante og viktige for å kunne vurdere kvaliteten i høyere utdanning, og utgjør en viktig del av datagrunnlaget NOKUT trenger i tilsynsarbeidet sitt. Portalen bidrar til bedre og lettere tilgjengelige indikatorer for kvalitet i utdanning og forskning, og synliggjør og effektiviserer arbeidet med kvalitetssikring innenfor høyere utdanning.
		Studie-barometeret.no	Portal inneholder resultater fra de nasjonale studentundersøkelsene innen høyere utdanning og fagskoleutdanning og den nasjonale underviserundersøkelsen.

<i>Aktør</i>	<i>Data-kilde</i>	<i>Nettportal</i>	<i>Innhold</i>
<i>Norsk forskningsråd (NFR)</i>		<i>Prosjektbanken.no</i>	Informasjon om prosjekter som finansieres av og følges opp av NFR. NFR har også lagt ut et datasett i Difis datahotell som omfatter mer data og som det er mulig å hente ut gjennom et API.
<i>Nordisk institutt for studier av innovasjon, forskning og utdanning (NIFU)</i>		<i>nifu.no/fou-statistiske/indikatorrapporten/</i>	Det norske forsknings- og innovasjonssystemet – statistikk og indikatorer gir en samlet oversikt over status og utvikling av virksomheten innenfor forskning, innovasjon, vitenskap og teknologi og bygger bl.a. på resultater fra de FoU-statistiske undersøkelsene. Rapporten inneholder beskrivelser og analyser av forsknings- og innovasjonsaktiviteten i Norge, internasjonale sammenligninger og en større tabelldel.
		<i>foustatistikkbanken.no</i>	FoU-statistikkbanken omfatter statistikk over ressurser til FoU i Norge og andre land. Her gis konsumenter mulighet til selv å velge tabeller og data
<i>Norsk senter for forskningsdata (NSD)</i>	<i>DBH</i>		NSDs Database for statistikk om høgre utdanning (DBH) inneholder et bredt spekter av informasjon om den tertiære utdanningen i Norge som inkluderer universiteter, høyskoler og fagskoler. DBH inneholder data om studenter, utdanning, forskning, personale, museum, areal, økonomi og selskapsdata. Datatilbudet er åpent tilgjengelig for forskning på sektoren, for planlegging både sentralt i departementer og lokalt ved hver enkelt institusjon.
	<i>DBH-F</i>		DBH-F samler informasjon om utdanningstilbud og studentdata ved fagskoler i Norge. DBH-F ble etablert på initiativ fra Kunnskapsdepartementet og fagskolesektoren og var i drift fra oktober 2011.

<i>Aktør</i>	<i>Data-kilde</i>	<i>Nettportal</i>	<i>Innhold</i>
<i>Statistisk sentralbyrå (SSB)</i>	<i>KOSTRA</i>	<i>ssb.no/offentlig-sektor/kostra</i>	KOSTRA står for Kommune-Stat-Rapportering og gir styringsinformasjon om ressursinnsatsen, prioriteringer og måloppnåelse i kommuner, bydeler og fylkeskommuner. Det finnes tall om f.eks. pleie- og omsorgstjenester, barnehagedekning og saksbehandlingstid, og kommuner kan sammenlignes med hverandre, med regionale inndelinger og med landsgjennomsnittet.
	<i>Statistikk banken</i>	<i>ssb.no/statbank</i>	Statistikkbanken er SSBs publiseringskanal for offisiell statistikk og inneholder detaljerte tabeller med tidsserier. Statistikk finnes innen en rekke tematiske klassifiseringer relatert til økonomi, befolkning og samfunn på nasjonalt, regionalt og lokalt nivå. Brukere kan lage egne utvalg av tallene og lagre disse i flere filformater. Det finnes også et API mot Statistikkbanken.
	<i>SSB register-data</i>	<i>microdata.no</i>	Tjenesten gir forskere og studenter ved godkjente forskningsinstitusjoner tilgang til å bruke registerdata fra SSB. Data om befolkning, utdanning, inntekt, arbeidsmarked og trygd er tilgjengelig. Løsningen har et anonymiserende grensesnitt med innebygd personvern.
<i>Statped</i>	<i>Fabris</i>		Statpeds fagsystem som også omfatter register over brukere av Statpeds tjenester og hvilke tjenester de mottar, rapporter og statistikk. Dette er pr. i dag kun til internt bruk, men fra tid til annen er det forespørsel om å få ut data fra systemet som utveksles gjennom flatfiler.
<i>Utdanningsdirektoratet (Udir)</i>	<i>BASIL</i>	<i>basilrapport.udir.no</i>	I BASIL kan barnehagene rapportere inn informasjon om årsmeldinger og resultatregnskap. I tillegg kan påloggede brukere finne og sammenstille nøkkeltall om barnehager.

<i>Aktør</i>	<i>Data-kilde</i>	<i>Nettportal</i>	<i>Innhold</i>
	<i>Grep</i>		Den nasjonale databasen for fag, læreplaner og opplæringstilbud i grunnsopplæringen. Alle fastsatte læreplaner i Kunnskapsløftet legges inn i Grep. I tillegg finnes kodeverk og informasjon om fag i grunnskole og videregående opplæring, inkludert vurderingsordninger, samt fag- og vitnemålsmerknader til bruk i dokumentasjon av opplæringen. Tilgang via API.
	<i>GSI</i>	<i>gsi.udir.no</i>	Samler inn en omfattende mengde data om grunnskolen i Norge. Det samles inn drøyt 1000 opplysninger om hver eneste grunnskole. Du finner informasjon om følgende tema i GSI: elevtall, årstimer, ressurser, spesialundervisning, språklige minoriteter, målform, fremmedspråk, fysisk aktivitet, leksehjelp, SFO, valgfag og PPT.
	<i>NxR</i>		Fellestjeneste (API) for nedlastning av data fra Nasjonalt skoleregister (NSR), Nasjonalt Barnehageregister (NBR), Nasjonalt register for opplæringskontorer (NOR) og Nasjonalt register for lærebedrifter (NLR). Tilgang via API.
	<i>PAS</i>		System til gjennomføring av nasjonale prøver, forkortelse for prøveadministrasjonssystem.
		<i>Udir.no/statistikk</i>	Portal som inneholder statistikk og analyser om barnehager og grunnsopplæringen i Norge.
		<i>Barnehagefakta.no</i>	Portal som inneholder data om barnehager. Portalen er utformet med tanke på foreldre som skal søke barnehageplass, og inneholder et fåtall av indikatorer. Data i portalen er også tilgjengelig via API.
		<i>Skoleporten.no</i>	Skoleporten er et verktøy for vurdering av kvalitet i grunnsopplæringen. Målet med Skoleporten er at skoler, skoleeiere, foresatte, elever og andre interesserte skal få tilgang til relevante og pålitelige nøkkeltall for grunnsopplæringen. Det er ikke mulig å hente ut data via API.

Aktør	Data-kilde	Nettportal	Innhold
Direktoratet for IKT og fellestjenester i høyere utdanning og forskning (Unit)	Cristin		Cristin (Current research information system in Norway) er et nasjonalt forskningsinformasjonssystem som samler og tilgjengeliggjør informasjon om norsk forskning, forenkler forskningsadministrative oppgaver ved å legge til rette for gjenbruk av forskningsinformasjon, følger opp rapporteringen av vitenskapelige publikasjoner til Kunnskapsdepartementet og Helse- og omsorgsdepartementet (NVI-rapportering).
	FS		FS er et studieadministrativt system for universiteter og høyskoler.
	NVB		Nasjonal vitnemålsdatabase (NVB) er en database med vitnemål for fullført og bestått videregående opplæring i Norge. Vitnemål blir sendt fra de skoleadministrative systemene som datafiler, slik at de blant annet kan brukes til kvalifisering og poengberegning for søkere som skal søke opptak til høyere utdanning gjennom Samordna opptak.
	SO		Samordna opptak (SO) ble opprettet i 1994 som et service- og koordineringsorgan for opptak til grunnutdanninger ved universiteter og høyskoler. I dag koordinerer SO opptaket til 27 universiteter og høyskoler og man kan søke blant mer enn 1 300 studier i samme søknad.
	STAR		STAR er et datavarehus med studentdata, og da i hovedsak eksamensdata og gjennomstrømningsdata. STAR bygger på data fra FS.
		vitnemalsportalen.no	Vitnemålsportalen er en webapplikasjon hvor en autentisert bruker kan se sine vitnemål fra høyere utdanning og dele dem med andre på en sikker måte. Portalen henter inn vitnemål fra de aktuelle UH- og/eller fagskole-institusjonene når konsumenten logger seg inn. Dataene (vitnemålene) persisteres ikke i portalen, men slettes ved utlogging og hentes inn på nytt fra primærkilden(e) neste gang brukeren logger seg inn.

10.2. Vedlegg B: Sentrale begreper med forklaring

Tabell 26 viser en oversikt over begreper med tilhørende definisjoner brukt i konseptvalgutredningen.

TABELL 26: SENTRALE BEGREPER MED FORKLARING

Begrep	Forklaring
<i>Alternativanalyse</i>	Analyse utarbeidet på bakgrunn av behovsanalyse, strategi, overordnede krav og mulighetsstudien. Inneholder minst to alternative konsept som tilfredsstiller kravene, samt et nullalternativ for sammenligning. Konseptene vurderes på bakgrunn av kostnads- og nyttevirksomheter, og evt andre relevante vurderinger. Formålet med analysen er å velge den samlet sett beste løsningen for samfunnet.
<i>Analyseverktøy</i>	Verktøy brukt for å analysere data. Dette kan være programmatisk verktøy som R og Python eller grafiske verktøy som SAS Visual Studio, Tableau eller Microsoft Power BI.
<i>API</i>	Forkortelse for det engelske begrepet Application Programming Interface. Er et grensesnitt i en programvare som gjør at deler av programvaren kan aktiveres («kjøres») fra en annen programvare. I utredningen benyttes API til å definere et grensesnitt mot en tjeneste for å lese data fra en datakilde (lesetilgang) og eventuelt oppdatere data i datakilden (skrivetilgang). API-et realiserer et teknisk kommunikasjonsgrensesnitt mellom en dataprodusent og én eller flere datakonsumenter. API-et definerer semantikken («språket») for dataene som utveksles slik at de forstås likt av aktørene på begge sider av API-et, og er å betrakte som en kontrakt mellom dataprodusenten og datakonsument(e).
<i>API-gateway</i>	Også kalt API-manager. Tjeneste for å kontrollere og styre tilganger til API-er. Sikrer at konsument er hvem de utgir seg for å være (autentisering) og holder kontroll på tilganger konsumentene skal ha (autorisasjon).
<i>Autoritativt register</i>	Et register som er gitt autoritet som kilde (også kjent som masterdatakilde) for et spesifikt datasett. Sikrer felles bruk (gjenbruk) slik at ulike datakonsumenter benytter den samme, oppdaterte informasjonen. Et eksempel er Folkeregisteret (Det sentrale folkeregisteret) for navn og fødselsdato for norske statsborgere.
<i>Behandlingsgrunnlag</i>	Behandling av personopplysninger krever et rettslig grunnlag for å være lov, omtalt som behandlingsgrunnlag. Det er nødvendig med behandlingsgrunnlag av hver enkelt personopplysning til hvert enkelt formål.
<i>Data</i>	Data er det materielle (eller fysiske) grunnlaget for å tilføre mennesker informasjon. Data er avhengig av at de knyttes til en kontekst for å gi de mening som kan tilføre kunnskap. Denne konteksten defineres gjennom metadata - i vid forstand. Et eksempel er at tallene i en telefonkatalog er et telefonnummer, men kan bety noe annet i en annen kontekst.
<i>Database</i>	Samling av data lagret i et elektronisk medium. Datasamlingen er organisert og strukturert etter en bestemt strategi eller modell – omtalt som en databasemodell. Dataene brukes og vedlikeholdes gjennom et bestemt programsystem som kalles et databasesystem eller databasehåndteringssystem.

Begrep	Forklaring
<i>Databehandler</i>	Den som behandler personopplysninger på vegne av den [data]behandlingsansvarlige. Kan være en fysisk eller juridisk person, offentlig myndighet, institusjon eller ethvert annet organ som behandler personopplysninger på vegne av den [data]behandlingsansvarlige.
<i>Databehandler-avtale</i>	Avtale mellom en databehandlingsansvarlig og en databehandler om behandling av opplysninger på vegne av den databehandlingsansvarlige. En databehandleravtale skal sikre at personopplysninger blir behandlet i samsvar med regelverket og setter en klar ramme for hvordan databehandleren kan behandle opplysninger.
<i>[Data]behandlingsansvarlig</i>	En fysisk eller juridisk person, en offentlig myndighet, en institusjon eller ethvert annet organ som alene eller sammen med andre bestemmer formålet med behandlingen av personopplysninger og hvilke midler som skal benyttes; når formålet med og midlene for behandlingen er fastsatt i unionsretten eller i medlemsstatenes nasjonale rett, kan den behandlingsansvarlige, eller de særlige kriteriene for utpeking av vedkommende, fastsettes i unionsretten eller i medlemsstatenes nasjonale rett.
<i>Dataflyt</i>	Beskriver hvordan data flyter mellom systemer gjennom dataenes livssyklus og brukes for et eller annet formål av de systemene som er involvert i dataflyten. I mange tilfeller tilføres mer informasjon til dataene i løpet av denne dataflyten. Se verdikjede for data.
<i>Datainfrastruktur</i>	Digital infrastruktur for organisering av data og datakilder. Datainfrastrukturen definerer hvordan data gjøres tilgjengelig for deling, og hvordan dataflyten (forsyningskjeden for data) foregår. Implementasjonen til datainfrastrukturen har store konsekvenser for driftskostnadene og effektiviteten til forsyningskjedene.
<i>Datakonsument</i>	En datakonsument er en aktør som ønsker tilgang til data fra en dataprodusent). Data underlagt betingelser for deling krever at konsumenten dokumenterer tilstrekkelig behandlingsgrunnlag for å få tilgang til dataene.
<i>Datakilde</i>	Betegnelse på en kilde som inneholder informasjon av en eller annen form. En datakilde kan være alt fra en fil til store databaser med store mengder data. I mange tilfeller sikter datakilde til et system eller en tjeneste som samler data innenfor et spesifikt område. Et eksempel på en slik kilde er GSI som samler data om grunnskolen i Norge.
<i>Datakvalitet</i>	Datakvalitet favner ulike egenskaper ved data. God datakvalitet innebærer at dataene har evnen til å støtte de informasjonsformål de er tiltenkt, og at de oppfyller formålets krav til: • Korrekthet (accuracy) • Kompletthet (completeness) • Tidsriktighet (timeliness) • Konsistens (consistency)
<i>Datamodell</i>	Modell som beskriver hvordan data er strukturert. Et eksempel er en personprofil bestående av fornavn, etternavn, adresse og telefonnummer. Datamodell er nært knyttet til begrepet informasjonsarkitektur, men ofte i en mer spesifikk kontekst.

Begrep	Forklaring
<i>Dataproducent</i>	Aktør med data som kan deles med andre aktører (datakonsumenter) under gitte betingelser. Åpne data kan deles uten betingelser. Data med begrenset tilgang krever at datakonsument kan dokumentere tilstrekkelig behandlingsgrunnlag for de opplysninger som ønskes utlevert fra dataproducent.
<i>Datavarehus</i>	En organisering av data i en database på en tematisk måte. Hensikten er å tilrettelegge dataene for analytisk behandling, til forskjell fra databaser som er tilrettelagt for transaksjoner knyttet til arbeidsprosesser. Et datavarehus er organisert i fakta (nøkkeltall) som kan ses i sammenheng med et sett av dimensjoner. Et eksempel er studiegjennomstrømning (fakta) sett i forhold til studieprogram, karakterer fra videregående skole og finansiering (dimensjoner).
<i>Deling av data</i>	Deling av data innebærer at data som er samlet inn gjøres tilgjengelig for andre for gjenbruk eller viderebruk. Et viktig formål med deling av data er å unngå at mange aktører hver for seg samler inn og lagrer samme informasjon. Deling av data gir grunnlag for betydelige gevinster ved digitalisering.
<i>Eksplorative tjenester</i>	Tjenester for å utforske analysemuligheter som ligger i tilgjengelige data, med basis i metadata som beskriver dataenes betydning og egenskaper, f.eks. om det er mulig koble ulike datasett via koblingsvariabler.
<i>ETL-prosessen</i>	ETL står for Extract-Transform-Load. Brukes om prosessen for lasting av data inn i et datavarehus gjennom at data kopieres ut av produksjonssystemer (f. eks. et økonomisystem) og bearbeides til informasjon som kan benyttes til analyseformål før informasjonen lastes inn i datavarehuset. ETL er ofte ressurskrevende prosesser som krever inngående kunnskap om forretningslogikk knyttet til de ulike datakildene for å kunne omsette data fra produksjonssystemene til annen kunnskap.
<i>Etterspørselsdrevne behov</i>	Etterspørselsbaserte behov skal synliggjøre prognoser for fremtidig etterspørsel etter tjenester eller leveranser fra et prosjekt, samt hvordan etterspørselen har påvirkning på andre faktorer. Blir dokumentert som et gap mellom dagens situasjon og fremtidig forventet situasjon, og danner utgangspunkt for målsettinger og krav som skal sikre at alternativene som senere utredes best mulig kan støtte opp under behovene.
<i>FAIR-prinsippene</i>	FAIR er et akronym for ordene <i>findable</i> , <i>accessible</i> , <i>interoperable</i> og <i>reusable</i> . Prinsippene bak FAIR er at forskningsdata skal være enkelt tilgjengelig, ha overføringsverdi mellom fagfelt og av en kvalitet som muliggjør gjenbruk. Prinsippene stiller krav til at forskningsdata beskrives gjennom gode metadata.
<i>Formål (med data)</i>	Et uttrykk for hvilke «spørsmål» dataene skal besvare hos sluttkonsumenten av dataene. Et eksempel på et formål med data er å analysere frafall i videregående skole. Formålet vil avlede behov til hvilke data som må registreres og eventuelt foredles gjennom verdikjeden fra produksjonssiden til konsumenten som skal få besvart sine «spørsmål»

Begrep	Forklaring
<i>Forskere</i>	Forskere er i denne sammenheng personer knyttet til forskningsinstitusjoner som arbeider med forskning tilknyttet kunnskapssektoren og annen forskning som benytter kunnskapsdata i forskningen. Forskere inkluderer, men er ikke begrenset til, prosjektledere og prosjektmedlemmer i forskningsprosjekter.
<i>Forskning</i>	Skapende arbeid på systematisk grunnlag med henblikk på å øke den vitenskapelige og tekniske viten, inkludert viten om mennesker, kultur og samfunn – samt utnyttelse av eksisterende viten til å anviser nye praktiske anvendelser.
<i>GDPR</i>	GDPR er en forkortelse for General Data Protection Regulation. På norsk blir det kalt personvernforordningen, og den ble vedtatt den 27. april 2016 av EU, med gyldighet fra 25. mai 2018. GDPR er en forordning om beskyttelse av individer ved behandling av personopplysninger og om fri flyt av slike opplysninger. Forordningens formål er å sørge for en god beskyttelse av personopplysninger samtidig som personopplysninger skal kunne utveksles fritt innenfor EØS-området.
<i>Harmonisering</i>	Samordne, få til å stemme overens. Momenter kan trekke i ulik retning, og det er da en harmonisering av momentene når det blir gjort en avveining mellom dem. Utredningen omtaler harmonisering av begreper og datamodeller, der hensikten er å fjerne eller redusere forskjeller mellom dem slik at organisasjoner og IT-løsninger som benytter disse kan samhandle bedre.
<i>Informasjonsarkitektur</i>	Informasjonsarkitektur er en fagdisiplin innenfor modellering og strukturering av informasjon. Formålet med informasjonsarkitektur er å oppnå en fornuftig informasjonsstruktur som gjøre det enkelt å finne og forstå informasjon. Et utvidet aspekt ved informasjonsarkitektur er knyttet til å definere autoritative kilder (se også «autoritativt register») som gjenbrukes i stedet for at kopier av de samme dataene ligger flere steder i informasjonsarkitekturen.
<i>Informasjons-sikkerhet</i>	Fagområde knyttet til at sikring av informasjon og informasjonssystemer bare er tilgjengelig for de som skal ha tilgang (konfidensialitet), at dataene er korrekte, gyldige og fullstendige (integritet) samt tilgjengelige innenfor fastsatte krav. Et netts eller informasjonssystemets evne til, på et bestemt sikkerhetsnivå, å stå imot utilsiktede hendelser eller ulovlige eller skadelige handlinger som svekker tilgjengeligheten, autentisiteten, integriteten og konfidensialiteten til lagrede eller overførte personopplysninger, samt sikkerheten i beslektede tjenester som tilbys av eller er gjort tilgjengelige via nevnte nett og systemer, av offentlige myndigheter, enheter for IT-beredskap (CERT), enheter for IT-sikkerhetshendelser (CSIRT), leverandører av elektroniske kommunikasjonsnett og -tjenester og leverandører av sikkerhetsteknologier og -tjenester. Dette kan f.eks. omfatte å hindre ulovlig tilgang til elektroniske kommunikasjonsnett og spredning av skadelige koder og å stoppe ««tjenestenektangrep»» og skade på datasystemer og elektroniske kommunikasjonsystemer.

Begrep	Forklaring
<i>Informasjons-system</i>	Et system for innsamling, lagring, behandling, overføring og presentasjon av informasjon. I prinsippet kan et informasjonssystem være helt manuelt, men ordet brukes oftest om systemer som er basert på informasjons- og kommunikasjonsteknologi (IKT). Til forskjell fra et datasystem, som er rent teknisk, inkluderer informasjonssystemet også de menneskelige brukerne av systemet. Det kan sies at et informasjonssystem består av ett eller flere datasystem sammen med deres brukere.
<i>Innovasjon</i>	Innovasjon er en betegnelse på menneskeskapt endring av verdiskapende aktiviteter.
<i>Kobling</i>	I kontekst av denne utredningen er begrepet kobling knyttet til en sammenstilling av datasett gjennom en variabel (koblingsvariabel), f.eks. et organisasjonsnummer (orgnr.) for en virksomhet. Det betyr at en kan sammenstille datasett som beskriver ulike dimensjoner ved en virksomhet og se disse i sammenheng dersom datasettene inneholder orgnr. Dette er i prinsippet det som gjøres i et datavarehus.
<i>Konseptvalg-utredning (KVU)</i>	Faglig utredning for store offentlige prosjekter som gjennomføres for å vurdere alternative måter å løse brukernes behov på («konsepter»).
<i>Kunnskapsdata</i>	I utredningen benyttes begrepet kunnskapsdata om data i kunnskapssektoren. Det innebærer data om opplæring, utdanning, forskning og integrering.
<i>Kunnskapssektoren</i>	Alle enheter under Kunnskapsdepartementet, dvs. statlige virksomheter (direktorater, etater etc. med særskilte fullmakter m.m.) statlige selskaper, stiftelser m.m.
<i>Kunstig intelligens</i>	Kunstig intelligens (KI) er en teknikk man bruker for å gi datamaskiner og dataprogrammer en mest mulig intelligent atferd. De intelligente datasystemene blir ofte kalt for intelligente agenter. En intelligent agent er et system som er i stand til å løse problemer og lære av egen atferd. De observerer sitt miljø og tar avgjørelser for å maksimere sin egen suksess.
<i>Maskinlæring</i>	Maskinlæring er en gren av kunstig intelligens. Se også «kunstig intelligens». Maskinlæring er utvikling av algoritmer som gjør datamaskiner i stand til å lære fra og utvikle atferd basert på empiriske data. Et hovedfokus innen forskning på maskinlæring er automatisk å lære gjenkjenning av komplekse mønstre og gjøre intelligente beslutninger basert på data. En læringsalgoritme bruker et sett treningsdata for å utvikle eller forbedre en atferd. Den må også være i stand til å generalisere og finne løsninger på problemer den ikke har observert eksempler på tidligere.
<i>Metadata</i>	(Fra gresk meta «om» og latin data «opplysninger».) Data som tjener til å definere eller beskrive andre data. Metadata, data om data, er informasjon som beskriver annen informasjon. Metadata inneholder typisk et emneord, tittel og tidspunkt for opprettelse og endring av data som beskrives.
<i>Metadatakatalog</i>	En katalog med samling av metadata som typisk beskriver en eller flere datakilder. Dette kan inkludere beskrivelser som navn, kodeverk og bruksområde.

Begrep	Forklaring
<i>Mulighetsdimensjon</i>	En mulighetsdimensjon definerer et aspekt innenfor mulighetsrommet som varierer mellom konseptene. Et eksempel på en mulighetsdimensjon er konsumentgrupper. De ulike konseptene kan ha fokus på ulike konsumentgrupper. Dimensjonene opprettes for å skille konseptene fra hverandre.
<i>Mulighetsrommet</i>	Mulighetsrommet viser spennet av alle alternativer/konsepter som er mulig innenfor prosjektets rammer. Formålet med mulighetsrommet er å synliggjøre alle muligheter som finnes, alt fra de minste endringene til de mer komplekse. Behovene, målene og kravene som defineres i fasen før mulighetsrommet setter de ytre rammene for mulighetsrommet.
<i>Normative behov</i>	Det beskriver problemstillingens forankring i relevante politiske strategier, lover og forskrifter. De tar for seg politiske mål, lover og forskrifter på et overordnet nivå. Disse behovene utledes av lover, forskrifter, rundskriv og annet relevant regelverk, samt fra overordnede politiske mål uttrykt i nasjonale og internasjonale politiske dokumenter som regjeringen stiller seg bak.
<i>Orden i eget hus</i>	Begrepet «orden i eget hus» benyttes om et konsept for god forvaltning av egen informasjon. Sentralt i konseptet er å oppfylle strategien om deling av data og prinsippet «kun én gang». Dette prinsippet sier at informasjon som offentlige virksomheter samler inn om personer og virksomheter skal gjøres kun én gang. Deretter skal informasjonen kunne gjenbrukes på tvers av de offentlige virksomhetene gjennom deling av data. Dette i stedet for at alle som har behov for denne informasjonen samler inn den selv. Difi har utarbeidet en veileder for «orden i eget hus» som beskriver hvordan data i offentlig sektor skal forvaltes for å oppnå målsetningene nedfelt i ulike digitaliseringsstrategier. Statlige virksomheter er pålagt å innføre «orden i eget hus» gjennom Digitaliseringsrundskrivet.
<i>Personopplysninger</i>	Enhver opplysning om en identifisert eller identifiserbar fysisk person («den registrerte»); en identifiserbar fysisk person er en person som direkte eller indirekte kan identifiseres, særlig ved hjelp av identifikator, f.eks. et navn, et identifikasjonsnummer, lokaliseringsopplysninger, en nettidentifikator eller ett eller flere elementer som er spesifikke for nevnte fysiske persons fysiske, fysiologiske, genetiske, psykiske, økonomiske, kulturelle eller sosiale identitet.
<i>Personvern</i>	Retten til et privatliv og retten til å bestemme over egne personopplysninger. Alle mennesker har en ukrenkelig egenverdi. Som enkeltmenneske har du derfor rett på en privat sfære som du selv kontrollerer, hvor du kan handle fritt uten tvang eller innblanding fra staten eller andre mennesker. Innebygd personvern betyr at det tas hensyn til personvern i alle utviklingsfaser av et system eller en løsning.
<i>Politikkutviklere</i>	Personer eller interesseorganisasjoner som påvirker utforming av politikken. Eksempler på politikkutviklere er politikere og rådgivere i departementer, kommuner og fylkeskommuner.

<i>Begrep</i>	<i>Forklaring</i>
<i>Register</i>	En samling av data på om et bestemt emne. Register kan brukes om en fil eller tabell bestående av objekter eller poster. Samtidig kan register brukes om en samling av tabeller og filer. I det tilfellet er det synonymt med database.
<i>Registerdata</i>	Data hentet ut fra et register.
<i>Sporbarhet av data</i>	Sporbarhet er egenskapen som muliggjør dokumentasjon av hvordan analysedata er skapt og kan gjenskapes. Den fremmer transparens og er en forutsetning for å skape troverdighet til analyseresultater. Sporbarhet skapes gjennom å dokumentere metoden brukt for å lede frem til et datasett.
<i>Sikre (analyse)rom</i>	Dataområder hvor det er mulig å gjøre analyser på informasjon som er sikret med hensyn til tilgang og muligheter for å kopiere eller flytte data ut av området
<i>Standardisering</i>	Det å fastsette standarder; ensretting.
<i>Statistikkbank</i>	En statistikkbank er en samling av data innenfor et statistikkområde, som f.eks. sosioøkonomiske data, data om helse og sykdom, osv. I en statistikkbank kan konsumenten lage egne statistikktabeller og rapporter ved å velge variabler, verdier og tidsserier. Eksempler på statistikkbanker er SSBs Statistikkbanken og Udirs Skoleporten.
<i>Stordata (big data)</i>	Stordata er betegnelsen på store mengder digital data. Det er datasett som er så komplekse at de er vanskelige å analysere med vanlige dataprosesseringsverktøy. Det engelske uttrykket «big data» blir ofte brukt. Hva som ligger under betegnelsen «stor» er varierende fra tid til annen, men avhenger blant annet av prosesseringskraft, lagringskapasitet, kommunikasjonskapasitet og standardidentifikatorer.
<i>Tilgjengeliggjøring</i>	Databehandler/databehandlingsansvarlig gjør data tilgjengelig for analyse ved utlevering (distribusjon av data) eller ved å gi bruker tilgang til data i et analysemiljø.
<i>Tjenesteprodusenter</i>	Tjenesteprodusenter er i denne utredningen benyttet om alle aktører som utfører tjenester til innbyggere, som barnehager, skoler, universiteter, forskningsinstitusjoner og asylmottak.
<i>Underliggende etater</i>	Underliggende etater er statlige virksomheter som har som hovedoppgave å treffe eller forberede forvaltningsvedtak. De er en del av staten som juridisk person og ikke egne rettssubjekter. Typiske underliggende etater er regjeringen, departementer, direktorater, fylkeskommuner og kommuner. De underliggende etatenes virksomhet er underlagt forvaltningsloven.
<i>Variabel og variabelverdi</i>	Et symbol som representerer et vilkårlig tall eller element i en mengde. Eksempel: Kjønn. Elementene i denne mengden kalles verdiene til variabelen eller variabelverdien. Eksempel for variabelen 'Kjønn': 1 Mann, 2 Kvinne, 9 Ukjent.
<i>Viderebruk av data</i>	Viderebruk av data innebærer at data brukes i en «videre» sammenheng, dvs. til andre formål enn det de ble samlet inn for. Dette er i motsetning til gjenbruk av data, hvor data samlet inn av en aktør gjenbrukes til det samme formål av en annen aktør.
<i>Visualisering (av data)</i>	I kontekst av denne utredningen brukes dette begrepet om hvordan data presenteres for konsumentene for at de effektivt skal skaffe seg verdifull innsikt, som kan gi gode og velbegrunnede beslutninger.

<i>Begrep</i>	<i>Forklaring</i>
<i>Økosystem</i>	Et økosystem er et sett med gjensidig avhengige aktører som deler en digital plattform for å oppnå en gjensidig nytte. Det vil si at eier av plattformen, tjenesteleverandører og brukere er avhengig av hverandre. Digitale økosystemer har tjenester fra flere leverandører.

10.3. Vedlegg C: Oversikt over gjennomførte intervjuer

Tabell 27 gir en oversikt over intervjuer som har blitt gjennomført ifm. konseptutredningen.

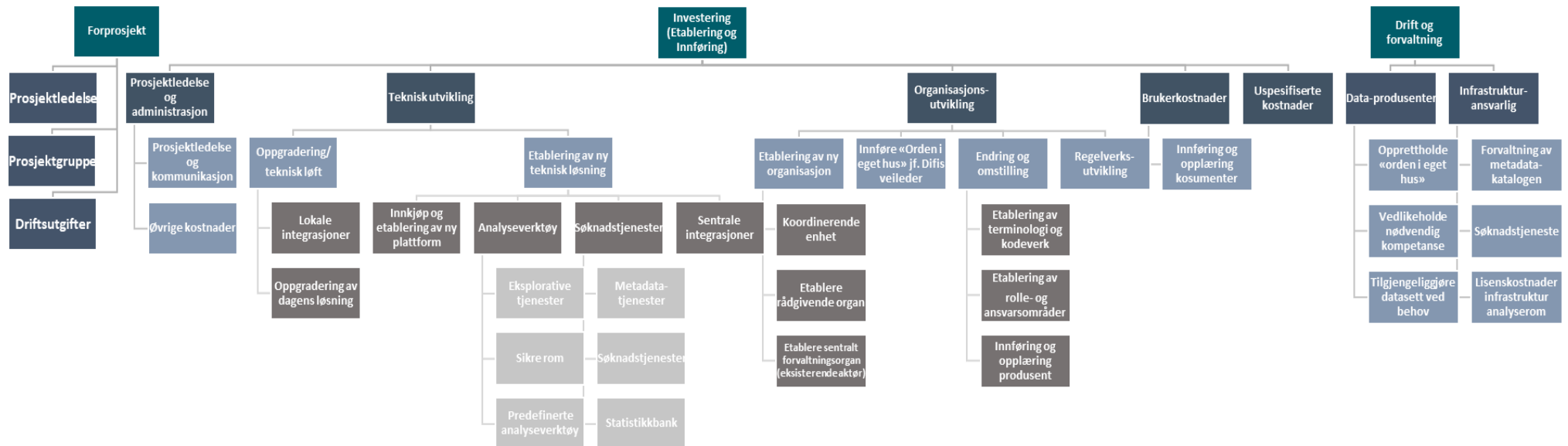
TABELL 27: OVERSIKT OVER GJENNOMFØRTE INTERVJUER

<i>Institusjon</i>
Bærum Kommune
Buskerud Fylkeskommune
DIKU
Fafo
Forskningsrådet (NFR)
Frischsenteret
Gol Kommune
Ideas 2 evidence
IMDi
Kunnskapsdepartementet (KD)
Kompetanse Norge
Kommunenes sentralforbund (KS)
Lånekassen
Microdata.no
NAV
Norges Handelshøyskole (NHH)
NIFU
NOKUT
NSD
NTNU
NTNU samfunnsforskning
Oslo kommune
Oslo MET
Rambøll

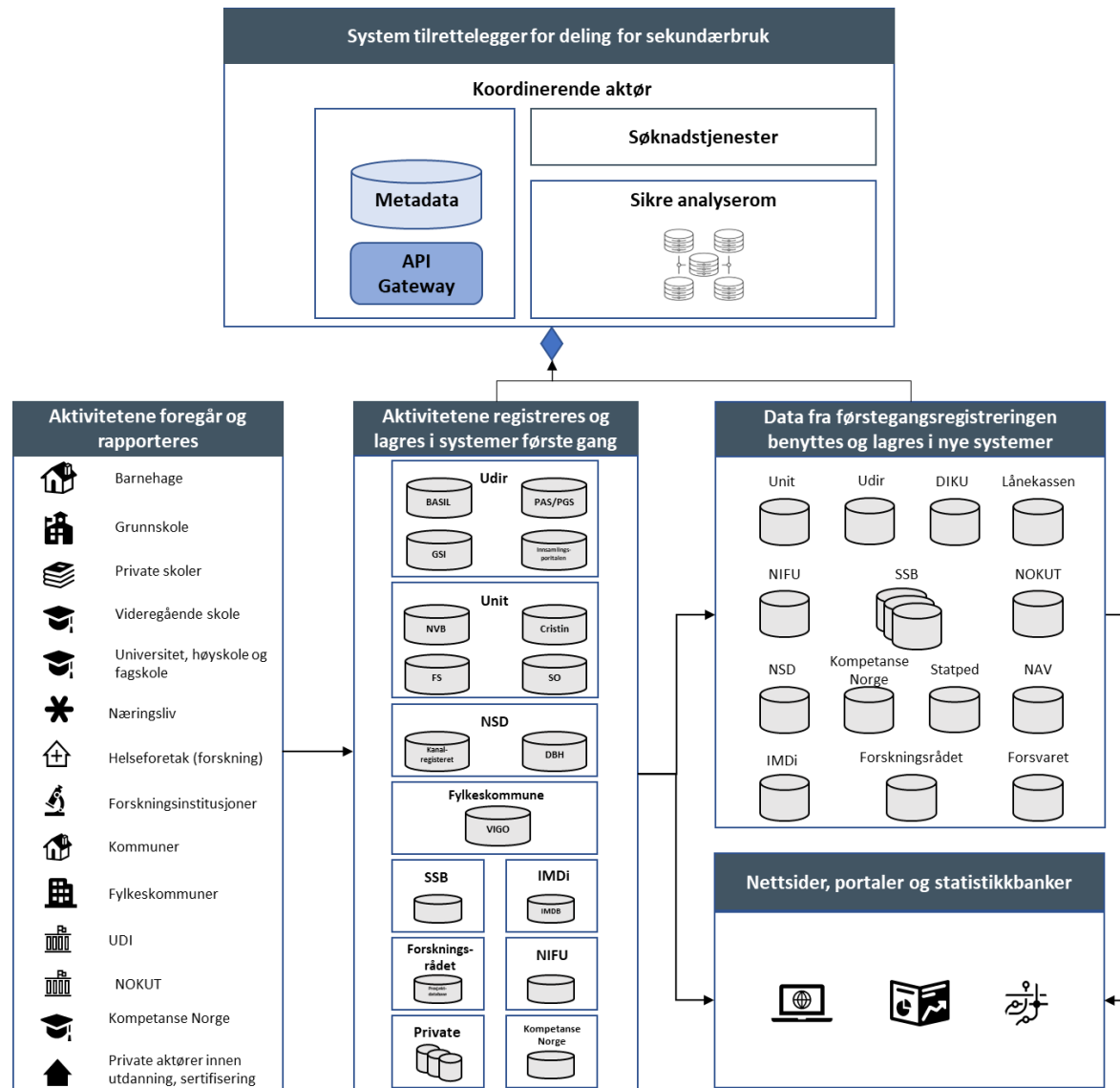


Samordna opptak
SSB
Statped
Udir
Unit
Utdanning.no
Universitet i Stavanger (UiS)
VIGO IKS
Viken fylkeskommune

10.4. Vedlegg D: Prosjektnedbrytningsstruktur



10.5. Konseptuell skisse av anbefalt konsept K12 Orden i kunnskapssektoren



11. Referanser

- Braathen, F. (2019). *Nytt datasystem til 560 millioner kroner er tre år forsinket. Lærere fortviler over problemer med alt fra timeplaner til registrering av fravær*. Hentet fra Aftenposten: <https://www.aftenposten.no/norge/i/XgJznm/nytt-datasystem-til-560-millioner-kroner-er-tre-aar-forsinket-larere-fortviler-over-problemer-med-alt-fra-timeplaner-til-registrering-av-fravaer>
- Digital21. (2018a). *Digitale grep for norsk verdiskaping*. Hentet fra https://digital21.no/wp-content/uploads/2018/09/Digital21_strategi_2018.pdf
- Digital21. (2018b). *Innspill fra ekspertgruppe for dataressurser og infrastruktur*. Hentet fra https://digital21.no/wp-content/uploads/2018/09/EG4_Dataressurser_og_infrastruktur_Digital21_2018.pdf
- Direktoratet for forvaltning og IKT. (2013). *Informasjonforvaltning i offentlig sektor*. Oslo: Direktoratet for forvaltning og IKT.
- Direktoratet for forvaltning og IKT. (2018). *Deling av data - konseptvalgutredning*. Oslo: Direktoratet for forvaltning og IKT.
- Direktoratet for forvaltning og IKT. (2018). *Veileder for orden i eget hus*. Hentet fra <https://doc.difi.no/data/veileder-orden-i-eget-hus/>
- Direktoratet for forvaltning og IKT. (2019a). *Difis tverrgående digitaliseringsstrategi*. Hentet fra https://www.difi.no/sites/difino/files/difis_tverrgaende_digitaliseringsstrategi_v1.2.pdf
- Direktoratet for forvaltning og IKT. (2019b). *Norsk arkitekturrammeverk for samhandling*. Hentet fra <https://www.difi.no/fagomrader-og-tjenester/digitalisering-og-samordning/nasjonal-arkitektur/arkitekturrammeverk-samhandling>
- Direktoratet for økonomistyring. (2018). *Utredningsinstruksen*.
- Finansdepartementet. (2010). *Utarbeidelse av KVVU/KL dokumenter*.
- Finansdepartementet. (2014). *R-109/14: Prinsipper og krav ved utarbeidelse av samfunnsøkonomiske analyser mv.*. Hentet fra https://www.regjeringen.no/globalassets/upload/fin/vedlegg/okstyring/rundskriv/faste/r_109_2014.pdf?id=2220435
- Forsvarsdepartementet. (2018). *Ny sikkerhetslov skal gjøre Norge tryggere*. Hentet fra <https://www.regjeringen.no/no/aktuelt/ny-sikkerhetslov/id2623522/>
- Justis- og beredskapsdepartementet. (2017). *IKT-sikkerhet – Et felles ansvar (Meld. St. 38 (2016–2017))*. Hentet fra <https://www.regjeringen.no/>
- Justis- og beredskapsdepartementet. (2019a). *Nasjonal strategi for digital sikkerhet*. Oslo: Justis- og beredskapsdepartementet.
- Justis- og beredskapsdepartementet. (2019b). *Nasjonal strategi for digital sikkerhetskompetanse*. Oslo: Justis- og beredskapsdepartementet.
- Kommunal- og moderniseringsdepartementet. (2016). *Digital agenda for Norge – IKT for en enklere hverdag og økt produktivitet (Meld. St. 27 (2015–2016))*. Hentet fra <https://www.regjeringen.no/>
- Kommunal- og moderniseringsdepartementet. (2017). *Retningslinjer ved tilgjengeliggjøring av offentlige data*. Oslo: Kommunal- og moderniseringsdepartementet.
- Kommunal- og moderniseringsdepartementet. (2019). *En digital offentlig sektor*. Oslo: Kommunal- og moderniseringsdepartementet.
- Kompetanse Norge. (2019). *Om Kompetanse Norge*. Hentet fra <https://www.kompetansenorge.no/om-kompetanse-norge/>
- Kunnskapsdepartementet. (2017a). *Digitaliseringsstrategi for universitets- og høyskolesektoren 2017–2021*. Oslo: Kunnskapsdepartementet.
- Kunnskapsdepartementet. (2017b). *Framtid, fornyelse og digitalisering*. Oslo: Kunnskapsdepartementet.

- Kunnskapsdepartementet. (2017c). *Nasjonal kompetansepolitisk strategi*. Oslo: Kunnskapsdepartementet.
- Kunnskapsdepartementet. (2017d). *Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata*. Oslo: Kunnskapsdepartementet.
- Kunnskapsdepartementet. (2017e). *Nasjonale mål og retningslinjer for åpen tilgang til vitenskapelige artikler*. Hentet fra <https://www.regjeringen.no/no/dokumenter/nasjonale-mal-og-retningslinjer-for-apen-tilgang-til-vitenskapelige-artikler/id2567591/>
- Kunnskapsdepartementet. (2018a). *Et velfungerende kunnskapssystem*.
- Kunnskapsdepartementet. (2018b). *Integrering gjennom kunnskap*. Oslo: Kunnskapsdepartementet.
- Kunnskapsdepartementet. (2018c). *Langtidsplan for forskning og høyere utdanning 2019–2028 (Meld. St. 4 (2018-2019))*. Hentet fra <https://www.regjeringen.no/>
- Kunnskapsdepartementet. (2019). *Tjenestekjeder i kunnskapssektoren 2019*. Hentet fra <https://www.regjeringen.no/>
- NOU 2018: 7. (2018). *Ny lov om offisiell statistikk og Statistisk sentralbyrå*. Oslo: Finansdepartementet.
- NOU 2019: 5. (2019). *Ny forvaltningslov*. Oslo: Justis- og beredskapsdepartementet.
- OECD. (2017). *Digital Government Review of Norway: Boosting the Digital Transformation of the Public Sector*. Paris: OECD Publishing.
- Rambøll. (2018). *IT i praksis 2018*. Hentet fra <https://no.ramboll.com/presse/publikasjoner/it-i-praksis>
- Riksrevisjonen. (2018). *Riksrevisjonens undersøkelse av digitalisering i statlige virksomheter*. Bergen: Fagbokforlaget AS.
- Sivertsen, G., & Gunnes, H. (2019). *Koordinering av dataproduksjon og kopling av datakilder om forskning*. Oslo: Nordisk institutt for studier av innovasjon, forskning og utdanning.
- Statistisk sentralbyrå. (2016). *Befolkningsframskrivninger, 2016-2100*. Hentet fra Statistisk sentralbyrå: <https://www.ssb.no/befolkning/statistikker/folkfram/aar/2016-06-21>
- Teknologirådet. (2018). *Kunstig intelligens - muligheter, utfordringer og en plan for Norge*. Oslo: Allkopi.
- Utdanningsdirektoratet. (u.d.). *Grunnskolen Informasjonssystem*. Hentet fra <https://gsi.udir.no/>

UNIT

DIREKTORATET FOR IKT OG FELLESTJENESTER
I HØYERE UTDANNING OG FORSKNING

