

Godkjent i DM 23. oktober 2018

**Informasjonsforvaltning i SSB**

## Innhold

Informasjonsforvaltning i SSB.....	1
1. Innledning .....	3
1.1 Definisjoner.....	3
2. Formål .....	4
3. Målbilder for informasjonsforvaltningen .....	5
3.1 Prinsipper og nasjonale krav .....	5
3.2 Målbilde for mastersystemene .....	8
3.3 Hovedtilstander for data og tilhørende metadata .....	9
3.4 Krav til metadatastøtte i produksjonsprosessen. ....	11
4.Nasjonalt og internasjonalt samarbeid .....	12
4.1 Informasjonsforvaltning i offentlig sektor .....	12
4.2 Internasjonal samhandling .....	12

## 1. Innledning

*Hensikten med det foreliggende notatet er å gi en overordnet framstilling av kravene som stilles til SSBs informasjonsforvaltning (bl.a. via ulike prinsipper) og mastersystemenes rolle, og å definere dataenes hovedtilstander gjennom produksjonsprosessen. Kravene og de definerte hovedtilstandene vil være et viktig grunnlag i moderniseringsprosjektet, og f.eks. bidra til å beskrive hvilke data som skal lagres for å øke gjenbruksmulighetene.*

SSBs strategi fastslår at SSB skal prioritere tiltak som bidrar til bedre oversikt og kontroll med data som samles inn, bearbeides og formidles i tilknytning til statistikkproduksjonen og analysevirksomheten. Bedre kontroll og orden i datagrunnlaget er en vesentlig forutsetning for videre standardisering og effektivisering av statistikkproduksjonen. Felles definisjoner og beskrivelser av data er nødvendig for å kunne skape sammenheng i bearbeidingsprosessene, og danner grunnlaget for økt automatisering og derved effektivitet i statistikkproduksjonen. Samtidig vil dette være et vesentlig bidrag til å forenkle og effektivisere gjenbruk av SSBs data, statistikk og analyser, både internt og for eksterne brukere. Til sist er det viktig å påpeke at vi ved å styrke informasjonsforvaltningen for å skape kontroll og orden i våre data, også skaper forutsetningene for arbeidet med å ivareta informasjons- og datasikkerhet i SSB.

Vi velger å benytte begrepet «informasjonsforvaltning» om den helhetlige tilnærmingen som kreves for å lykkes på dette feltet, og notatet beskriver overordnede prinsipper og målbilder for informasjonsforvaltningen. «Informasjon» er i dette notatet avgrenset til informasjon som har med statistikkproduksjonen å gjøre. Informasjon knyttet f.eks. til virksomhetsstøtte (som økonomistyring) er ikke inkludert i informasjonsforvaltningsbegrepet.

Samtidig som SSB skal styrke egen informasjonsforvaltning, foregår et omfattende arbeid for å utvikle informasjonsforvaltning i offentlig sektor. SSB skal både bidra til å styrke offentlig informasjonsforvaltning, og samtidig dra nytte av det som utformes av felles løsninger, standarder og generell kompetanse på dette feltet.

### 1.1 Definisjoner

#### *Applikasjon<sup>1</sup>*

En applikasjon er et IT-system som er i drift, og som støtter forretningsfunksjoner og -tjenester. Applikasjoner bruker data og støttes av ulike teknologikomponenter, men er samtidig adskilt fra disse.

Nettlesere, e-post-programmer og tekstbehandlingsprogrammer er eksempler på generelle applikasjoner.

#### *Dataressurs*

Lagret informasjon satt sammen av et eller flere *datasett*.

#### *Forretningsfunksjon*

En ferdighet som en organisasjon har, og som krever en kombinasjon av organisasjon, personer, prosesser og teknologi for å utføres. Hovedprosessene, samt delprosessene i Virksomhetsmodellen, kan brukes for å gruppere *forretningsfunksjoner*. For eksempel kan «identifisere og undersøke outliers og kritiske verdier» grupperes som en *lavere nivå* *forretningsfunksjon* innenfor «Kontrollere på makronivå»-*forretningsfunksjonen*

#### *Informasjonsforvaltning<sup>2</sup>*

<sup>1</sup> Kilde: *Open Group Standard TOGAF® 9.1 Translation Glossary: English – Norwegian*  
<http://www.durusau.net/localcopy/C151.PDF>

<sup>2</sup> Difi-rapport 2013-10 «Informasjonsforvaltning i offentlig sektor»

Informasjonsforvaltning betyr et helhetlig syn på aktiviteter, verktøy og andre tiltak for å sikre best mulig kvalitet, utnytting og sikring av informasjon i en virksomhet. Organiseringen av informasjonen skal være systematisk og henge sammen med virksomhetens arbeidsprosesser.

### *Informasjonsobjekt*

Informasjonsobjekt brukes i dette notatet som en fellesbetegnelse på de ulike typer informasjon som forvaltes i SSB, f.eks. variabler, klassifikasjoner og kodelister. Informasjonsobjektene er beskrevet i SSBs informasjonsmodell (SSB IM)<sup>3</sup>.

### *Mastersystem*

Et mastersystem brukes i dette notatet om et system som er dedikert til å dokumentere og tilgjengeliggjøre bestemte typer av informasjonsobjekter, f.eks. tradisjonelle metadata som kodeverk (klassifikasjoner og kodelister). Mastersystemet er «fasit» for sin type informasjonsobjekter, og her skal en alltid finne gyldige informasjonsobjekter med endringshistorikk.

### *Metadata*

Metadata er data om data. *Statistikkmetadata* defineres som strukturert informasjon om statistikk. Dette omfatter informasjon som brukes for å produsere, formidle, forstå, finne eller (gjen)bruke statistikk.

## **2. Formål**

SSB har over flere år arbeidet med å standardisere statistikkproduksjonen i samsvar med funksjonelle krav basert på felles arbeidsprosesser, i henhold til SSBs virksomhetsmodell. Samtidig har det internasjonale samarbeidet mellom statistikkbyråer ledet til en felles utforming av to viktige, overordnede prosess- og informasjonsmodeller. Disse er GSBPM<sup>4</sup> (Generic Statistical Business Process Model) og GSIM (Generic Statistical Information Model)<sup>5</sup>. Formålet med disse referansmodellene er å legge grunnlaget for standardiserte løsninger til støtte for arbeidsprosessene, og at disse løsningene kan bygge på standarder for å strukturere og beskrive data som bearbeides i prosessene. Basert på disse modellene kan SSB utforme en overordnet virksomhetsarkitektur. Også her kan vi dra nytte av internasjonalt samarbeid, da tilsvarende arbeid både pågår i en rekke nasjonale statistikkbyråer, og er på dagsorden i det felles internasjonale samarbeidet både i regi av EU og FN (UNECE)<sup>6</sup>.

SSBs virksomhetsmodell samsvarer med GSBPM v5. Vår gjeldende metadatastrategi som er fra 2005, bygger ikke på GSIM, men det ble i 2017 utarbeidet en informasjonsmodell for SSB (SSB IM) som er basert på GSIM. I forlengelsen av GSIM-arbeidet har det internasjonale samarbeidet ledet til en forankring av SDMX standarden for utveksling av aggregerte data og DDI standarden for beskrivelser av surveydata opp mot GSIM. I arbeidet med å utvikle informasjonsforvaltningen i SSB skal vi bygge på det internasjonale arbeidet når det gjelder statistikkspesifikke modeller og standarder, mens vi vil bygge på det nasjonale arbeidet som pågår innenfor offentlig sektor når det gjelder mer generelle sider (f.eks. begrepsbeskrivelser) ved informasjonsforvaltningen.

<sup>3</sup> <https://wiki.ssb.no/display/MAS2/ORIG+SSB+Informasjonsmodell+-+Topppnotat>

<sup>4</sup> <https://statswiki.unece.org/display/GSBPM/GSBPM+v5.0>

<sup>5</sup> <https://statswiki.unece.org/display/gsim/Generic+Statistical+Information+Model>

<sup>6</sup> United Nations Economic Commission for Europe

Hensiktsmessig bruk av standarder for lagring og utveksling av data internt, setter oss også i stand til å tilpasse oss de mest anvendte standardene for å tilby gjenbruk av SSBs data til forskning eller som åpne data<sup>7</sup>. Dette sluttbrukerfokuset skal tas hensyn til i interne valg.

### 3. Målbilder for informasjonsforvaltningen

#### 3.1 Prinsipper og nasjonale krav

Følgende internasjonalt anbefalte prinsipper legges til grunn for arbeidet med informasjonsforvaltning i SSB<sup>8</sup>.

##### 3.1.1. Prinsipper

*Prinsipp 1: Informasjon skal behandles som en felles ressurs*

Forklaring: Informasjon er en viktig ressurs for virksomheten og må forvaltes i tråd med det.

*Prinsipp 2: Ivareta informasjonens livsløp*

Forklaring: All informasjon har et livsløp og må forvaltes slik at versjonering og pålitelig identifikasjon ivaretas.

*Prinsipp 3: Utvikle og ivareta krav til informasjonssikkerhet*

Forklaring: Alle persondata, samt konfidensielle og sensitive data, skal beskyttes, og data må behandles i henhold til dette. Det innebærer at kravene til konfidensialitet, integritet og tilgjengelighet må overholdes<sup>9</sup>.

*Prinsipp 4: Bruk godkjente standarder og modeller*

Forklaring: All informasjon som brukes som input og output i forretningstjenester skal beskrives ved hjelp av en felles, forretningsorientert referansemodell<sup>10</sup>.

*Prinsipp 5: Dokumenter informasjon så tidlig som mulig*

Forklaring: Informasjon skal dokumenteres på en standardisert og strukturert måte så tidlig som mulig i statistikkproduksjonen. Dette skal gjøres for å sikre at informasjonen kan brukes i alle påfølgende prosesssteg.

*Prinsipp 6: Beskriv for å sikre gjenbruk*

<sup>7</sup> Difi definerer åpne data som «data som er fritt tilgjengelig for bruk og viderebruk av alle, både mennesker og maskiner. Dataeiere kan ikke kreve mer enn at man navngir og/eller deler videre på samme vilkår». I SSBs tilfelle vil de åpne datene være aggregerte data, f.eks. fra statistikker, samt klassifikasjoner og kodelister. <http://data.norge.no/document/del-og-skap-verdier-veileder-i-tilgjengeliggj%C3%B8ring-av-offentlige-data/hva-er-%C3%A5pne-data>

<sup>8</sup> De åtte første prinsippene bygger på CSPA v1.5 Chapter IV Information Architecture <https://statswiki.unece.org/display/CSPA/Information+Architecture>

<sup>9</sup> *Konfidensialitet* vil si at sensitiv eller gradert informasjon bare skal være tilgjengelig for autoriserte personer og prosesser; *informasjonskvalitet* og *–integritet* vil si at informasjonen er fullstendig, nøyaktig og gyldig, og et resultat av autoriserte og kontrollerte aktiviteter; og *tilgjengelighet* vil si oppfyllelse av krav til service, slik at alle berettigede krav om tilgang til informasjonen dekkes ved behov. (Kilde: Store norske leksikon)

<sup>10</sup> I SSB brukes Virksomhetsmodellen og SSB IM.

Forklaring: All informasjon skal beskrives på en måte som sikrer at informasjonen kan gjenbrukes i ulike prosesssteg. Gjenbruk skal redusere duplisering og feil og fremme automatiserte prosesser og selvbetjening.

*Prinsipp 7: Sørg for at det finnes en autoritativ kilde*

Forklaring: Informasjon som brukes og produseres i ulike prosesssteg, skal hentes fra og oppdateres i en autoritativ kilde. Informasjonen skal være konsistent gjennom alle relevant prosesssteg.

*Prinsipp 8: Ta vare på informasjonen som går inn i et prosesssteg.*

Forklaring: Informasjon som er input til et prosesssteg, må tas vare på for å hindre tap av informasjon.

*Prinsipp 9: Fødselsnummer skal pseudonymiseres så tidlig som mulig.*

Forklaring: Fødselsnummer skal pseudonymiseres så tidlig som mulig slik at enkeltpersoner ikke skal kunne identifiseres. Det skal imidlertid være mulig for autoriserte personer å få tilgang til fødselsnummer når dette er nødvendig i produksjonsprosessen.

*Prinsipp 10: Enhetsidentifikatorer<sup>11</sup> skal pseudonymiseres så tidlig som mulig<sup>12</sup>*

Forklaring: Enhetsidentifikatorer skal pseudonymiseres så tidlig som mulig slik at enkeltenheter ikke skal kunne identifiseres. Det skal imidlertid være mulig for autoriserte personer å få tilgang til identifikatorene når dette er nødvendig i produksjonsprosessen.

I den interne høringsrunden har det vært mange innsigelser til prinsipp 10. Hovedargumentet er at det i deler av produksjonsprosessen er nødvendig at enkelte medarbeidere har tilgang til identitet. Slik prinsippet er utformet, er det allerede lagt til rette for dette; «*Det skal imidlertid være mulig for autoriserte personer å få tilgang til identifikatorene når dette er nødvendig i produksjonsprosessen.*» Prinsippet innebærer at opplysninger skal lagres/håndteres uten direkte identifiserende kjennetegn for den allmenne bruker, men at autoriserte brukere skal kunne gis tilgang til enheten(es) identitet når det er nødvendig.

Det understrekes at prinsippet også følger av utkast til ny statistikklov<sup>13</sup> § 2-6, og at lovforslaget ikke er imøtegått i høringsrunden – verken av SSB eller andre høringsinstanser. Prinsipp 10 blir derfor beholdt.

### **3.1.2. Rammebetingelser i lover og reguleringer**

<sup>11</sup> Dette vil ikke gjelde alle enhetsidentifikatorer, men f.eks. organisasjonsnummer eller ISIN-internasjonalt verdipapir-identifikator (som brukes for finansielle transaksjoner). SSB må etter hvert identifisere hvilke enhetsidentifikatorer som omfattes av prinsippet

<sup>12</sup> Fødselsnummer er skilt ut fra de andre enhetsidentifikatorene siden personvern er omfattet av et eget lovverk.

<sup>13</sup> I utkastet til statistikklov nevnes VoF som eksempel på et tilfelle der pseudonymisering av organisasjonsnummer ikke er nødvendig: «Eksempelvis bør grunndata om næringslivet i byråets virksomhets- og foretaksregister kunne behandles med organisasjonsnummer»

SSB må i sitt arbeid med informasjonsforvaltning forholde seg til flere nasjonale og internasjonale

lover og regelverk som regulerer ulike sider ved håndtering av data og informasjonen om dem, f.eks. Statistikkloven<sup>14</sup>, Personopplysningsloven<sup>15</sup>, Arkivloven<sup>16</sup>, digitaliseringsrundskrivet, eForvaltningsforskriften tildelingsbrevet, EU-forordning om europeisk statistikk<sup>17</sup>, The European Statistics Code of Practice<sup>18</sup>, GDPR- EUs nye personvernforordning som trer i kraft i 2018, og FNs prinsipper for offisiell statistikk<sup>19</sup>. Disse gir i liten grad detaljanvisninger for informasjonsforvaltningen, men definerer rammene vi må holde oss innenfor.

### 3.1.3 Krav til informasjonsforvaltning- nasjonale krav

I «Informasjonsforvaltning i offentlig sektor»<sup>20</sup> beskrives følgende som et minimum for å oppnå god informasjonsforvaltning:

*Virksomheten har oversikt over dataene som behandles.*

Rapporten sier at følgende krav må oppfylles for å oppnå dette:

1. Data i virksomheten er beskrevet
2. Oversikten over data er publisert
3. Tilgang til data er vurdert
4. Dataelementene er definert
5. Beskrivelsene (metadata) er strukturerte og maskinlesbare

Krav 1 sier at virksomheten i tilstrekkelig grad må dokumentere egne data. Detaljeringsgraden må vurderes ut fra hva som er formålet med dataene. Publisering av dataoversikter er en fordel (krav 2). Dersom oversikten ikke kan gjøres allment tilgjengelig, er det nyttig for samhandlingen i offentlig sektor om oversikten kan deles innenfor denne sektoren. Krav 3 sier at offentlige virksomheter skal vurdere hvem som skal ha tilgang til data internt, og hvem data kan deles med eksternt. Krav 4 sier at dataelementene skal være definert på en måte som gjør det enkelt å sette seg inn i hva de betyr (f.eks. hvordan inntekt er definert i dette datasettet), mens krav 5 sier at virksomhetene må sørge for at informasjonen er tilstrekkelig strukturert, på en standardisert måte, slik at maskiner kan bruke dokumentasjonen. God informasjonsforvaltning vil gi «orden i eget hus».

### 3.1.4 Informasjonssikkerhet

Informasjonsforvaltning kan også ses fra et informasjonssikkerhets-perspektiv der informasjonssikkerhet bl.a. skal sikre rett informasjon, på rett sted, til riktig tid, til riktig vedkommende. Et viktig element i informasjonssikkerheten er at en virksomhet vet hvilken informasjon den er ansvarlig for, og hvem som har tilgang til informasjonen. Dette krever oversikt over, og dokumentasjon av, det som forvaltes, både på data- og metadatasiden<sup>21</sup>. For å

<sup>14</sup> <https://lovdata.no/dokument/NL/lov/1989-06-16-54>

<sup>15</sup> <https://lovdata.no/dokument/NL/lov/2000-04-14-31>

<sup>16</sup> <https://lovdata.no/dokument/NL/lov/1992-12-04-126>

<sup>17</sup> <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32009R0223&qid=1399632056207>

<sup>18</sup> <http://ec.europa.eu/eurostat/web/quality/european-statistics-code-of-practice>

<sup>19</sup> <http://www.ssb.no/omssb/styringsdokumenter/lover-og-prinsipper/fns-prinsipper-for-offisiell-statistikk>

<sup>20</sup> Difi-rapport 2013:10 «Informasjonsforvaltning i offentlig sektor»

<http://www.difi.no/sites/difino/files/rapport-informasjonsforvaltning-i-offentleg-sektor-2013-10-10.pdf>

<sup>21</sup> ISO 27002 innebærer bl.a. at virksomheter har en oversikt over dataene de forvalter, med opplysninger bl.a. om verdiklassifisering, vurdering av taushetsplikt osv.

sikre at kun riktig vedkommende får tilgang på informasjon, kreves tilgangsstyring, noe som forutsetter identitetshåndtering.

Personopplysningsloven sier også at den behandlingsansvarlige og databehandleren, gjennom planlagte og systematiske tiltak, skal sørge for tilfredsstillende informasjonssikkerhet med hensyn til konfidensialitet, integritet og tilgjengelighet ved behandling av personopplysninger. *Konfidensialitet* betyr at informasjonen kun er tilgjengelig for autoriserte personer og/eller systemer og at det på forhånd er foretatt en gyldig identifisering og autentisering. Med *integritet* menes at informasjonen ikke blir endret eller slettet på uautorisert måte, og *tilgjengelighet* betyr at informasjon eller dataressurser er til stede, stabile og anvendelige for brukere etter behov. For å oppnå tilfredsstillende informasjonssikkerhet skal den behandlingsansvarlige og databehandleren dokumentere informasjonssystemet og sikkerhetstiltakene.

SSB vil også forholde seg til Personopplysningsloven (som består av nasjonale regler samt GDPR- EUs personvernforordning) i sitt arbeid med informasjonssikkerhet.

### 3.2 Målbilde for mastersystemene

Informasjonsobjekter fra mastersystemene skal være enkelt og raskt tilgjengelig for alle brukere, lagres og vedlikeholdes ett sted (mastersystemet), og gjenbrukes overalt der det er relevant. Med enkelt tilgjengelig mener vi at de ulike brukersystemene<sup>22</sup> må kunne kommunisere med mastersystemer på en enhetlig måte. Dette betyr bl.a. at det ikke skal lages spesifikke grensesnitt mellom de ulike mastersystemene og brukersystemene.

Informasjonsobjektene bør i tillegg i størst mulig grad vedlikeholdes av statistikkprodusentene, og systemene må derfor være brukervennlige. Mastersystemene skal gjøres tilgjengelige for alle prosessene i virksomhetsmodellen, og systemene skal selv kunne varsle brukerne (maskin eller menneske) når det skjer endringer i innholdet. Varslingen forutsetter en abonnementsordning e.l.

Informasjonsobjekter oppstår og endres i ulike delprosesser i produksjonsløpet. Vi må ha systemer som kan dokumentere livsløpet til informasjonsobjektene, både over tid og mellom hovedprosessene.

Informasjonsobjekter skal i størst mulig grad hentes direkte fra mastersystemene og inn i produksjonsprosessene. Dette skal sikre at alle bruker korrekt versjon av informasjonsobjektene, og at disse har riktig kvalitet, dvs. at det er de «offisielle» informasjonsobjektene. Det kan gjøres unntak, f.eks. at det tillates å benytte en lokal kopi av informasjonsobjekter dersom dette gjør prosessen mer effektiv, men forutsetningen er at kravene til korrekt versjon og riktig kvalitet oppfylles.

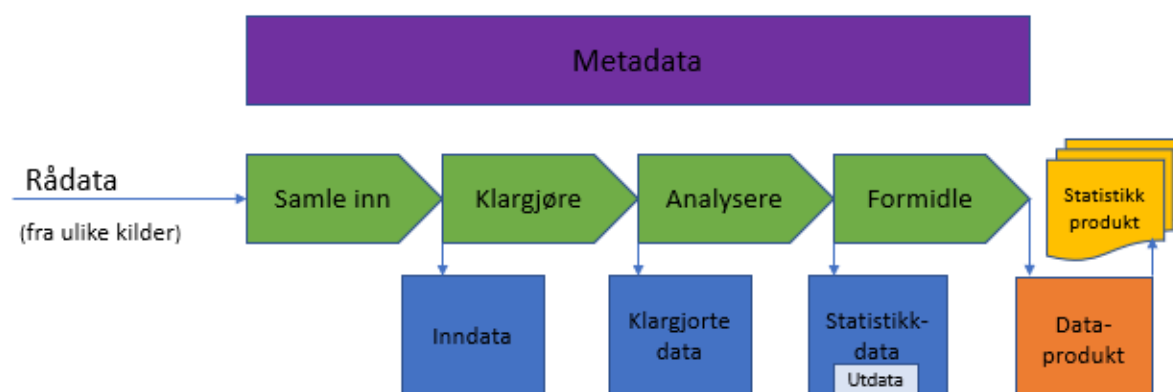
I målbildet for informasjonsplattformen<sup>23</sup> i Moderniseringsprosjektet inngår mastersystemer både i «Kataloger-metadatas» og «Populasjonsregistre».

<sup>22</sup> Med brukersystemer menes her systemer som brukes i statistikkproduksjonen, f.eks. datavarehuset og ISEE.

<sup>23</sup> <https://wiki.ssb.no/pages/viewpage.action?pageId=76939449>



### 3.3 Hovedtilstander for data og tilhørende metadata



Figur 1 Målbilde for hovedtilstandene (blå bokser) av data og tilhørende metadata i produksjons- prosessen

Vi finner det hensiktsmessig å foreslå fire hovedtilstander i dataenes livsløp, noe som bl.a. er foreslått i et tidligere notat i SSB<sup>24</sup>. Dette vil bidra til at produksjonsprosessene kan gjenskapes og kontrolleres når det er nødvendig, og gi grunnlag for lovbestemt ivaretagelse av arkiververdige data, og tilbud om individbaserte data til forskningsmiljøene.

En hovedtilstand er det informasjonsobjektet som i GSIM på overordnet nivå kalles informasjonsressurs, og som består av et eller flere informasjonssett (som f.eks. kan være datasett). De fire hovedtilstandene av data som skal lagres, dokumenteres og versjoneres, er følgende: inndata, klargjorte data, statistikkdata og utdata.

Dataene som kommer inn til SSB, kommer imidlertid i form av **rådata**. Dette er ikke en stabil datatilstand med felles datastruktur og kvalitetskrav, og rådataene har ikke gjennomgått noen transformasjoner internt i SSB. Det gjøres imidlertid kontroll av om dataene er «mottakelige», dvs. kan leses automatisk, i Koblingspunktet.

Det finnes også flere mellomtilstander som er viktige i dagens statistikkproduksjon, f.eks. tilstanden mellom inndata og klargjorte data der inndata transformeres til formatet for klargjorte data før man kan kjøre dagens klargjøringsprosesser. Dette notatet vil imidlertid ikke gå inn i disse mellomprosessene. De vil behandles nærmere, sammen med resten av produksjonsprosessen, i moderniseringsprosjektet.

De fire hovedtilstandene fra figur 1 beskrives som følger:

**Inndata:** Inndata er resultatet av hovedprosessen «Samle inn», og anses som mottatt av SSB. Vi bruker inndata om data som er samlet inn fra eksterne kilder. Dette er rådata som er transformert slik at de oppfyller SSBs krav til lagringsstruktur (format), og har de nødvendige metadata. De oppfyller også kvalitetskravene som er definert i leveranseavtalen. De samme inndata kan inngå i grunnlaget til flere utdata, og alle inndata skal kunne brukes i alle SSBs systemer.

Mange statistikker baserer seg imidlertid ikke kun på eksterne data, men statistikk A kan f.eks. også gjenbruke klargjorte data og/eller statistikkdata fra statistikk B og C<sup>25</sup>, eller tidligere årganger av egen statistikk. Vi har valgt å kalle samlingen av inndata og gjenbrukte interne

<sup>24</sup> «Elektronisk dialog med arbeidsgiver (EDAG): Felles databegrep», Zhangs notat fra 2013

<sup>25</sup> Disse dataene har allerede vært gjennom klargjøring og ev. analyse og er dermed kvalitetssikret, bl.a. ved at de har riktig lagringsstruktur og tilstrekkelig med metadata.

data for **startdata**. Disse startdataene er de som vil være utgangspunktet for klargjøringsprosessen til statistikk A.

**Klargjorte data** Klargjorte data er resultatet av hovedprosessen «Klargjøre» («Klargjøre» i utvidet forstand, se neste avsnitt). Her kan en f.eks. koble til data som har sin opprinnelse i ulike kilder, og utføre ulike former for kontroller og editering. Hovedmålet med prosessen «Klargjøre» er å forbedre dataenes nøyaktighet. Sluttresultatet er tilstrekkelig kvalitetssikrede og korrekt kodede data til videre bruk.

At klargjorte data er resultatet av hovedprosessen «Klargjøre», betyr ikke at prosesser som i Virksomhetsmodellen er beskrevet under hovedprosess «Analysere», ikke skal brukes for å komme fram til klargjorte data. I praksis vil en bevege seg fram og tilbake mellom prosessene under de to hovedprosessene, for til slutt å tilfredsstille kravene som settes til klargjorte data, dvs. at de er «tilstrekkelig kvalitetssikrede og korrekt kodede data til videre bruk».

Statistikkdata (omtalt nedenfor) fra produksjonsløpet for en statistikk kan også brukes som klargjorte data og/eller startdata for en annen statistikk. SSBs statistikkregistre<sup>26</sup> er typiske eksempler på klargjorte data. Klargjorte data karakteriseres ved sin gjenbruksverdi, dvs. som en fellesressurs (jfr. informasjonsforvaltningsprinsipp 1) for hele SSB.

**Statistikkdata:** Statistikkdata er resultatet av hovedprosessen «Analysere» i virksomhetsmodellen, og grunnlaget for data- og statistikkproduktene. Statistikkdata er en blanding av klargjorte data fra egen produksjonsprosess og påkoblede klargjorte data (og/eller statistikkdata) produsert i andre produksjonsprosesser. En prosess' statistikkdata kan igjen inngå som klargjorte data i en annen prosess, til et annet formål (statistisk behov), da typisk i sammenheng med nye klargjorte data. Transformasjonen fra klargjorte data til statistikkdata krever analyse, tolkning, forklaring, osv. samt nødvendig tilrettelegging for datautveksling

Skillet mellom klargjorte data og statistikkdata er ikke tydelig i dagens statistikkproduksjon. Det blir derfor arbeidet videre med definisjonen av disse (og de andre hovedtilstandene) i moderniseringsprosjektet. En ønsker bl.a. å se om mer presise definisjoner kan gi mer automatisering ved lagring av hovedtilstandene, og gjøre valg av data enklere for de som skal gjenbruke andres. Nærmere beskrivelse av dette arbeidet, som er mer «implementeringsnært», hører imidlertid ikke hjemme innenfor rammene av dette notatet.

**Utdata:** Utdata er den delmengden av statistikkdata som er klare for formidling ved at konfidensielle data er undertrykt. Dette innebærer f.eks. aggregeringer og prikking av celler.

I tillegg til de fire hovedtilstandene for data, har vi **dataprodukt** der utdata er bearbeidet til ulike produkter for ulike informasjonskonsumenter (f.eks. allmennheten og internasjonale organisasjoner). Dataprodukt er resultatet av hovedprosessen «Formidling», og kjernen i de fleste statistikkprodukter. Den delmengden av dataprodukt som presenterer data på aggregert nivå i ulike formater og utforminger (f.eks. tabeller, grafer/figurer, tekstlig), er statistikkprodukter.

Forskningsdata<sup>27</sup> på individnivå vil også være et dataprodukt, men ikke et statistikkprodukt. Forskningsdata vil som oftest basere seg på klargjorte data, ikke utdata med undertrykte data.

Statistikkprodukt trenger imidlertid ikke være et dataprodukt, f.eks. vil SSB tilby klassifikasjoner og kodelister som statistikkprodukt.

I arbeidet med å identifisere informasjonen som skal dokumentere de fire hovedtilstandene, får «Avklare behov»- og (særlig) «Planlegge»-prosessen i Virksomhetsmodellen stor betydning. For å oppnå målet om en «heldekkende» informasjonsstøtte i produksjonsprosessen, må statistikkprodusentene (forretningssiden) definere hvordan statistikkdata skal se ut allerede i

<sup>26</sup> Et statistikkregister (også kalt populasjonsregister) er et register som definerer og identifiserer grunnleggende statistiske enheter (VOF, BeReg, SSB-Matrikkel). (Fra Noen sentrale begreper knyttet til metadata <http://www.byranettet.ssb.no/Artikler/1147349621.21>)

<sup>27</sup> Gjelder data både til interne og eksterne forskere

delprosessene Avklare behov/Planlegge, identifisere informasjonsobjekter som skal brukes fra mastersystemene og eventuelt oppdatere/dokumentere nye informasjonsobjekter i mastersystemene. Disse informasjonsobjektene, samt informasjonsobjekter som eventuelt oppstår i andre deler av produksjonsprosessen, skal så kunne følge dataene gjennom hele produksjonsløpet til formidling. SSB mangler i dag et standardisert opplegg for hvordan dette skal gjøres. Det er heller ingen formelle krav knyttet til disse to hovedprosessene.

Det vil være noen krav til de fire hovedtilstandene: (presisering av prinsipp 6 og 8 i kap. 4.1.1)

1. Hovedtilstandene skal lagres, dokumenteres og versjoneres
2. Det må være felles formelle og strukturelle krav til hvordan dataene skal lagres.
3. Det må være felles formelle og strukturelle krav til hvilke metadata som skal lagres, og hvordan.
4. Det må nyttes kvalitetskrav til innholdet for hver hovedtilstand
5. Det må være regler for hvordan data i en hovedtilstand skal kunne endres, og hvordan endringer skal varsles.
6. Metadata for de ulike hovedtilstandene må håndteres på en konsistent måte gjennom virksomhetsmodellen slik at en kan se hvordan de endres mellom ulike hovedtilstander.

I samarbeid med Arkivverket har vi definert at klargjorte data utgjør datagrunnlaget for SSBs statistikkprodukter og analyser. De er derfor våre arkivverdige<sup>28</sup> data. I enkelte tilfeller kan det også være hensiktsmessig for gjenbruk og forståelse av data at SSB beholder rådata og inndata. Dette gjelder tilfeller der SSB selv skaper data gjennom egen innsamling vha. skjema og intervjuer. Når det gjelder statistikkdata og utdata, så er det transformasjonsprosessene som beskriver hvordan disse dataene er fremkommet.

I tillegg til å lagre dataenes hovedtilstander, må vi også ta vare på transformasjonsprosessene som skjer mellom hovedtilstandene.

Det er viktig å se SSBs informasjon i sammenheng med forretningsprosessene der den behandles, metodene som brukes, applikasjonene/løsningene som inngår og sikkerheten som kreves. Dette ivaretas gjennom **virksomhetsarkitekturen** som skal sikre at virksomhetens informasjon og ulike løsninger ses som del av en helhet, og ikke betraktes hver for seg. Dermed kan en f.eks. unngå at det etableres løsninger som ikke snakker sammen, eller såkalte siloer. Nærmere beskrivelse av virksomhetsarkitekturen (som bl.a. inkluderer informasjonsarkitektur) finnes andre steder<sup>29</sup>.

### 3.4 Krav til metadatastøtte i produksjonsprosessen.

- SSB skal, så sant de er tilgjengelige, hente og bruke eierens metadata når vi benytter data fra andre. Metadata fra felleskomponentene brukes når det er mulig.
- Metadata skal være tilgjengelige, og kunne oppdateres, i de ulike hovedprosessene i virksomhetsmodellen. Metadata fra mastersystemene skal inkluderes allerede fra Avklare behov- og Planlegge-hovedprosessene, og

<sup>28</sup> Arkivverdige data er data som arkiveres (lagres) ihht. en lovhjemmel. Andre data, f.eks. statistikkdata, skal også lagres, men vi sier ikke at de arkiveres, siden lagringen av disse ikke har utgangspunkt i en lovhjemmel.

<sup>29</sup> F.eks. <http://www.prosjektveiviseren.no/bibliotek/tema/virksomhetsarkitektur>, <https://wiki.ssb.no/display/VIR/Om+virksomhetsarkitektur>, <https://wiki.ssb.no/display/VIR/Beskrivelse+av+tjenestetilbudet+til+seksjon+for+Virksomhets+og+IT-arkitektur>

suppleres underveis. Eventuelle endringer i metadata skal også dokumenteres. Dokumentasjonen skal automatiseres i den grad det er mulig.

- Det skal lages et tjenestelag (grensesnitt) for metadata som de andre systemene i SSB kan høste metadata via. Tjenestelaget vil hente metadata fra de ulike mastersystemene, og vil fungere som en felles metadataportal for maskiner. I tillegg vil mastersystemene ha spesifikke grensesnitt som kan brukes bl.a. til vedlikehold av de respektive informasjonsobjektene.
- Metadata fra SSBs mastersystemer skal gjenbrukes overalt der det er relevant.
- Metadata skal lagres sammen med tilhørende data i de fire hovedtilstandene.
- SSB skal levere metadata sammen med sine dataprodukter og statistikkprodukter
- Metadata som støtter statistikkproduksjon (håndbøker, retningslinjer etc.) skal være enkelt tilgjengelige for alle trinn i virksomhetsmodellen.

## **4.Nasjonalt og internasjonalt samarbeid**

### **4.1 Informasjonsforvaltning i offentlig sektor**

Difi og Brønnøysundregistrene har over flere år ledet et nasjonalt prosjekt som utarbeider et rammeverk<sup>30</sup> for informasjonsforvaltning i offentlig sektor. Dette rammeverket innebærer bl.a. veiledere, standarder og spesifikasjoner. Rammeverkets visjon er at det skal gi tilstrekkelige føringer og støtte slik at offentlige virksomheter kan utveksle og dele data og beskrivelser av data, også maskinelt. Rammeverket utvikles kontinuerlig ut fra virksomhetenes behov, og SSB har deltatt i arbeidet, bl.a. med oppdatering av Standard for begrepsbeskrivelser,

### **4.2 Internasjonal samhandling**

For å effektivisere statistikkproduksjonen, er det viktig også å kunne trekke veksler på samarbeid med andre statistikkbyråer. Deling av tjenester og utveksling av informasjon forutsetter standardisering knyttet til beskrivelse av informasjon. UNECEs High Level Group for the Modernisation of Statistical Production and Services (HLG-MOS) er f.eks. nå ansvarlig for CSPA (Common Statistical Production Architecture), et initiativ som skal bidra til at ulike statistikkbyråer er mer på linje når det gjelder virksomhetsarkitektur. Dette skal oppnås ved å utvikle et felles rammeverk, inkludert prinsipper, prosesser og retningslinjer. Rammeverket vil gjøre det enklere å samarbeide om utvikling av og/eller deling av f.eks. tjenester eller prosesskomponenter, og dermed på sikt bidra til å redusere kostnader hos de enkelte statistikkbyråer. SSB deltar i CSPA, og rammeverket herfra vil påvirke vår interne informasjonsforvaltning.

---

<sup>30</sup> <https://www.difi.no/fagomrader-og-tjenester/digitalisering-og-samordning/nasjonalarkeitektur/informasjonsforvaltning/rammeverk-informasjonsforvaltning>