

Sindre Børke

Data til statistikkbruk fra ulike kilder

© Statistisk sentralbyrå, august 2011 Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.	Standardtegn i tabeller	Symbol
ISBN 978-82-537-8159-4 Trykt versjon	Tall kan ikke forekomme	.
ISBN 978-82-537-8160-0 Elektronisk versjon	Oppgave mangler	..
ISSN 1891-5906	Oppgave mangler foreløpig	...
Emne: 00.90	Tall kan ikke offentliggjøres	:
Trykk: Statistisk sentralbyrå	Null	-
	Mindre enn 0,5 av den brukte enheten	0
	Mindre enn 0,05 av den brukte enheten	0,0
	Foreløpig tall	*
	Brudd i den loddrette serien	—
	Brudd i den vannrette serien	
	Desimaltegn	,

Forord

SSBs statistikk skal gi informasjon om helhet og sammenheng i samfunnet. Statistikken skal vise utviklingstrekk for befolkning og levekår, økonomi, miljø og næringsvirksomhet.

Datafangststrategien bygger på to hovedmål:

- Mer motiverte oppgavegivere
- Bedre kvalitet i datainnsamlingene

Utfordringen i datafangsten for SSB er å samle inn mer informasjon, med høyere kvalitet og til en lavere kostnad. Et viktig mål for datafangstarbeidet er bevisst og kontinuerlig arbeid for å begrense oppgavebyrden. Den teknologiske utviklingen gir tilgang til nye data og elektroniske tjenester. Mer automatisert datafangst bidrar til å begrense oppgavebyrden.

Dette notatet beskriver sammenhengen mellom ulike typer datakilder og innsamlingsmetoder. En rød tråd gjennom notatet er at planlegging av en datafangst i SSB må inkludere vurdering av flere alternative datakilder, og dermed en vurdering av ulike metoder og instrumenter som kan brukes i datafangsten.

Sammendrag

Dette notatet beskriver sammenhengen mellom ulike typer datakilder og innsamlingsmetoder. En rød tråd gjennom notatet er at planlegging av en datafangst i SSB må inkludere vurdering av flere alternative datakilder, og dermed en vurdering av ulike metoder og instrumenter som kan brukes i datafangsten.

Dataene som brukes til å lage statistikk kommer fra forskjellige kilder og gjennom forskjellige kanaler til SSB. I SSBs virksomhetsmodell er det definert aktiviteter som gjennomføres i denne sammenhengen. Dette strekker seg fra å vurdere om data allerede finnes i SSB, gjennom kartlegging av om andre kan ha samlet noen av de nødvendige dataene, til avklaring og planlegging av egen innsamling av data for det konkrete formålet.

I notatet beskrives sammenhengen mellom ulike typer datakilder og innsamlingsmetoder. En rød tråd gjennom notatet er at planlegging av en datafangst i SSB må inkludere vurdering av flere alternative datakilder, og dermed en vurdering av ulike metoder og instrumenter som kan brukes i datafangsten.

Dataene som benyttes til statistikk kan deles inn i tre forskjellige kategorier:

- Data i statistiske basisregistre, også omtalt som populasjonsregistre
- Data fra andre administrative eller statistiske registre
- Data fra utvalgsundersøkelser eller tellinger

Notatet omtaler registre i betydningen en komplett fortegnelse av enheter innenfor en gitt populasjon, og der hver enhet har en unik identifikasjon som gjør det mulig å oppdatere registret både med hensyn til enheter og variabler/-verdier. Notatet omtaler også administrative datasystemer som registre, i betydningen opplysninger som samles inn og oppbevares på en slik måte at de kan hentes fram til bruk i saksbehandlingen.

Kapittel 1 gir en kort beskrivelse av ulike typer data, med fokus på hvorfor og hvordan de blir tatt vare på der de oppstår. Kapittel 2 gir en grov framstilling av hvordan data kommer til SSB. Der synliggjøres også registrenes plass i datafangsten. Kapittel 3 gir en nærmere omtale av basisregistrene i det statistiske datasystemet. De andre administrative registrene omtales i kapittel 4, som er hovedkapitlet i dette notatet. Kapittel 4 går nærmere inn på ulike andre datakilder, hvor data blir lagret utenfor SSB, og hvordan disse dataene kan utnyttes i vår datafangst, spesielt i det som figur 1 omtaler som utvalgsundersøkelser. Kapittel 5 utdyper noen begreper som ofte brukes i datafangstsammenheng.

Innhold

Forord	3
Sammendrag	4
Innledning	6
1. Data oppstår i samfunnet	7
1.1. Data som finnes i foretakene	7
1.2. Data fra personer og husholdninger	8
2. Datafangst til SSB	9
2.1. Gruppering av datakilder	9
2.2. Manuell og elektronisk datafangst	11
3. Statistiske basisregistre basert på administrative grunndataregistre	11
4. En gruppering av datakilder	13
4.1. Å definere og kommunisere SSBs databehov	14
4.2. Administrative registre	15
4.3. Tredje-parts-data	16
4.4. Foretakenes datasystemer	18
4.5. Data som ikke finnes elektronisk	21
4.6. En innsamling – flere kilder og innsamlings-instrumenter	21
5. Primært og sekundært i en statistikkdata-sammenheng	22
5.1. Ved innhenting av data	23
5.2. Bruk av innsamlede data	23
Referanser	25
Vedlegg 1 Registerforum - mandat	26
Vedlegg 2 EDAG – Elektronisk dialog med arbeidsgivere	28
Vedlegg 3 Seminar on Registers in Statistics - methodology and quality 21 - 23 May, 2007 Helsinki	30
Vedlegg 4 Nye metoder for prestasjonsmåling i byggenæringen	38
Vedlegg 5. Use of Administrative Data in Official Statistics – Past, Present and Future – With Special Reference to the Nordic Countries	44
Vedlegg 6 Datafangst lønnsstatistikk	51

Innledning

SSBs statistikk skal gi informasjon om helhet og sammenheng i samfunnet. Statistikken skal vise utviklingstrekk for befolkning og levekår, økonomi, miljø og næringsvirksomhet.

Datafangststrategien bygger på to hovedmål:

- Mer motiverte oppgavegivere
- Bedre kvalitet i datainnsamlingene

Utfordringen i datafangsten for SSB er å samle inn mer informasjon, med høyere kvalitet og til en lavere kostnad. Et viktig mål for datafangstarbeidet er bevisst og kontinuerlig arbeid for å begrense oppgavebyrden. Den teknologiske utviklingen gir tilgang til nye data og elektroniske tjenester. Mer automatisert datafangst bidrar til å begrense oppgavebyrden. SSB arbeider for økt elektronisk samhandling, både ved samarbeid mellom etatene om felles innrapportering, og for enklere bruk og gjenbruk av data. Altinn er myndighetenes løsning for innrapportering og dialog med næringslivet. Alle tjenester fra det offentlige skal gjøres tilgjengelig i Altinn. Framover heves blikket mot samhandling på tvers av aktører, med felles rapportering og tilbakemelding, tilpasset bedriftenes forretningsprosesser, uavhengig av myndighetenes inndeling i etater og forvaltningsområder.

Målsetningene om å effektivisere datafangsten og begrense oppgavebyrden er i datafangststrategien uttrykt gjennom følgende strategiske satsingsområder:

- Styrke satsingen på å utnytte administrative registre
- Øke rapportering direkte fra fagsystem
- Tilpasse datafangsten bedre for oppgavegiverne (mer systematisk kombinasjon av ulike innsamlingsmetoder)
- Forbedre datainnsamlingsinstrumentene
- Samordne utvalgsplaner

Dataene som brukes til å lage statistikk kommer fra forskjellige kilder og gjennom forskjellige kanaler til SSB. I virksomhetsmodellens¹ pkt 1.4. er det definert aktiviteter som gjennomføres i denne sammenhengen. Dette strekker seg fra å vurdere om data allerede finnes i SSB, gjennom kartlegging av om andre kan ha samlet noen av de nødvendige dataene, til avklaring og planlegging av egen innsamling av data for det konkrete formålet.

Dette notatet beskriver sammenhengen mellom ulike typer datakilder og innsamlingsmetoder. En rød tråd gjennom notatet er at planlegging av en datafangst i SSB må inkludere vurdering av flere alternative datakilder, og dermed en vurdering av ulike metoder og instrumenter som kan brukes i datafangsten.

Dataene som benyttes til statistikk kan deles inn i tre forskjellige kategorier, se figur 1:

- Data i statistiske basisregistre, også omtalt som populasjonsregistre
- Data fra andre administrative eller statistiske registre
- Data fra utvalgsundersøkelser eller tellinger

Dette notatet omtaler registre i betydningen en komplett fortegnelse av enheter innenfor en gitt populasjon, og der hver enhet har en unik identifikasjon som gjør det mulig å oppdatere registret både med hensyn til enheter og variabler/-verdier.² Notatet omtaler også administrative datasystemer som registre, i betydningen

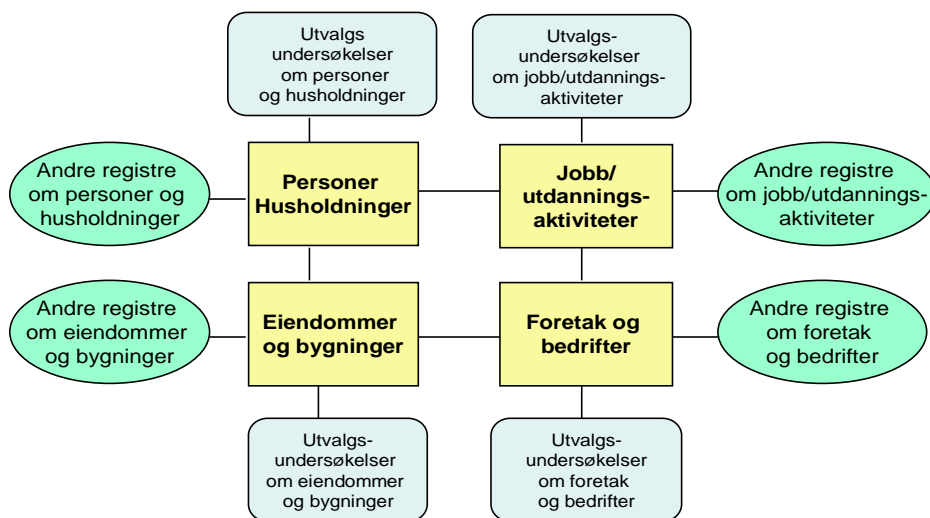
¹ Notater 2008/47

² Internt dokument 2006/2; Noen sentrale begreper knyttet til metadata.

opplysninger som samles inn og oppbevares på en slik måte at de kan hentes fram til bruk i saksbehandlingen.

Figur 1: Sammenhengen mellom tre kategorier datasett

Basisregistre, registre og utvalgsundersøkelser



Kapittel 1 gir en kort beskrivelse av ulike typer data, med fokus på hvorfor og hvordan de blir tatt vare på der de oppstår. Kapittel 2 gir en grov framstilling av hvordan data kommer til SSB. Der synliggjøres også registrenes plass i datafangsten. Kapittel 3 gir en nærmere omtale av basisregistrene i det statistiske datasystemet. De andre administrative registrene omtales i kapittel 4, som er hovedkapitlet i dette notatet. Kapittel 4 går nærmere inn på ulike andre datakilder, hvor data blir lagret utenfor SSB, og hvordan disse dataene kan utnyttes i vår datafangst, spesielt i det som figur 1 omtaler som utvalgsundersøkelser. Kapittel 5 utdyper noen begreper som ofte brukes i datafangstssammenheng.

1. Data oppstår i samfunnet

Statistikk produseres for å beskrive samfunnet i sammenheng og sett fra forskjellige vinkler. Grunnlaget for slike beskrivelser er data som skapes gjennom samfunnets virke. Hva som er data som kan benyttes i statistikkssammenheng er ikke gitt. Her gis en kort beskrivelse av hvilke data som samles eller finnes, og som er grunnlaget for datafangst og statistikkproduksjon i SSB.

1.1. Data som finnes i foretakene

Alle som driver en virksomhet har behov for og er pålagt å holde orden på en del opplysninger om virksomheten.

1.1.1. Lovpålagte rapporteringer til andre etater

De aller fleste virksomheter må føre og rapportere regnskap etter visse regler. Selvangivelse er et eksempel på en slik rapportering. I tillegg vil det komme andre pålagte krav for virksomheter med ansatte, der det skal rapporteres både selve ansettelsesforholdet og lønn m.v. Merverdiavgift skal innkreves og rapporteres for de som driver virksomhet der det er pålagt, og det finnes en god del bransjeavhengige krav til dokumentasjon, f.eks. ved omsetning av matvarer. Noen

av dataene som virksomhetene er pålagt å ta vare på er ikke pålagt innrapportering før eventuelle hendelser utløser dette, f.eks. innen HMS (helse, miljø og sikkerhet).

Disse pålagte rapporteringene utgjør hovedbasis for de administrative registrene som SSB i neste omgang får tilgang til som underlag for statistikk.

1.1.2. Plan- og styringssystemer/andre data i egen interesse

En større del av plan- og styringsdataene som finnes i norske virksomheter er pålagt. I praksis vil de fleste virksomheter også ha andre data i tillegg, med sikte på egen planlegging og styring av virksomheten. Hvor mye data som brukes til dette formålet vil avhenge av virksomhetens størrelse og kompleksitet.

Til en viss grad vil dataene som brukes være felles innenfor en bransje. Bruk av elektroniske hjelpemidler vil trekke i retning av standardisering, bl.a. gjennom felles programvare for å holde orden på og rapportere dataene internt. På den andre siden kan de økte mulighetene til å håndtere data øke forskjellene som følge av forskjeller i kapasitet og kompetanse til å nytte mulighetene.

Med tanke på data for mulig statistisk bruk, bør vi holde et øye med områder som ikke er lovpålagt, men likevel meget utbredt og standardisert. Blant slike vil vi finne data som kreves ved, og blir sendt inn i forbindelse med søknader om økonomisk støtte eller tillatelser til å drive og/eller utvide virksomheten. I praksis vil noen datasett basert på slike data kunne regnes som et administrativt register på linje med lovpålagte rapporteringer.

Også medlemskap i f.eks. bransjeorganisasjoner krever rapporteringer som gjør at virksomheter i bransjen tar vare på data i egne systemer for å rapportere til sin organisasjon. For SSB kan disse dataene være tilgjengelige hos organisasjonen, eller hos virksomheten selv, uten at det utløser stor oppgavebyrde.

1.1.3. Data til SSB

Selv med utstrakt bruk av data fra andre etater, gjennom deres innsamling etter lovpålagte eller søknadsbaserte rapporteringer, må SSB selv hente inn data. For foretak og bedrifter brukes statistikkloven³ som underlag for datainnsamling. Dataene skal dekke statistikkbehovene, og definisjonene må styres ut fra det. I mange tilfeller er definisjonene styrt av detaljerte internasjonale forordninger. Likevel er det viktig at det som etterspørres i størst mulig grad tilpasses de data vi kan forvente at virksomhetene har tilgjengelig. I de aller fleste tilfellene vil grunnlaget for de dataene vi spør om finnes i deres datasystemer (fagsystemer), mens den oppdelingen/aggregeringen vi ber om kanskje er annerledes enn de selv bruker (se 4.4.). I den grad det blir nødvendig å etterspørre data som det ikke er naturlig å ha et underlag for hos virksomhetene, er det viktig at dette varsles i så god tid at virksomhetene kan utvikle et opplegg for å ta vare på dataene i løpende virksomhet. Å komme i ettertid og be om data som virksomheten ikke har tilgjengelig medfører stor irritasjon og dårlig kvalitet på innrapporteringen, økt oppgavebyrde og virker lite gjennomtenkt fra SSBs side.

1.2. Data fra personer og husholdninger

Data fra personer og husholdninger vil nesten alltid være slike som vi regner med at personer selv kan frambringe eller gi uttrykk for "der og da". Det finnes få lovregler som pålegger privatpersoner å ta vare på data om sin virksomhet. Noen stønader som kan søkes forutsetter at personen kan dokumentere utgifter/inntekter, deltakelse i aktiviteter eller lignende, men i praksis vil dokumentasjonen bestå i utskrifter fra forretningsdrivendes eller offentliges systemer. Vi kan se på personopplysninger hos banker, forsikringsselskaper og aksjefond som tredjepartsopplysninger (se pkt 4.3.), men i praktisk datafangstarbeid vil vi regne

³ Lov av 16. juni 1984 nr 54: Lov om offisiell statistikk og Statistisk sentralbyrå, <http://www.ssb.no/omssb/statlov/>

selskapet som dataeier og dermed som oppgavegiver. Det samme gjelder administrative registre, vi ser det ikke som data personen er pålagt å ta vare på, men som data det offentlige har om personen.

Personer eller husholdninger som kommer med i våre utvalgsundersøkelser blir normalt spurt om ting de kan huske eller mene på det tidspunktet de blir spurt. Men i noen undersøkelser ber vi om at de tar vare på opplysninger om ting de foretar seg i en periode, for eksempel om penge- eller tidsbruk, for senere å rapportere dette. Til nå har vi kun tilbudt papirbaserte dagbøker til dette formålet, men utvikling av elektroniske løsninger på dette området er nærliggende. Det er rimelig å anta at det i all hovedsak blir basert på webløsninger utviklet av SSB.

Det er sjelden at personer eller husholdninger blir pålagt svarplikt med hjemmel i statistikkloven.

2. Datafangst til SSB

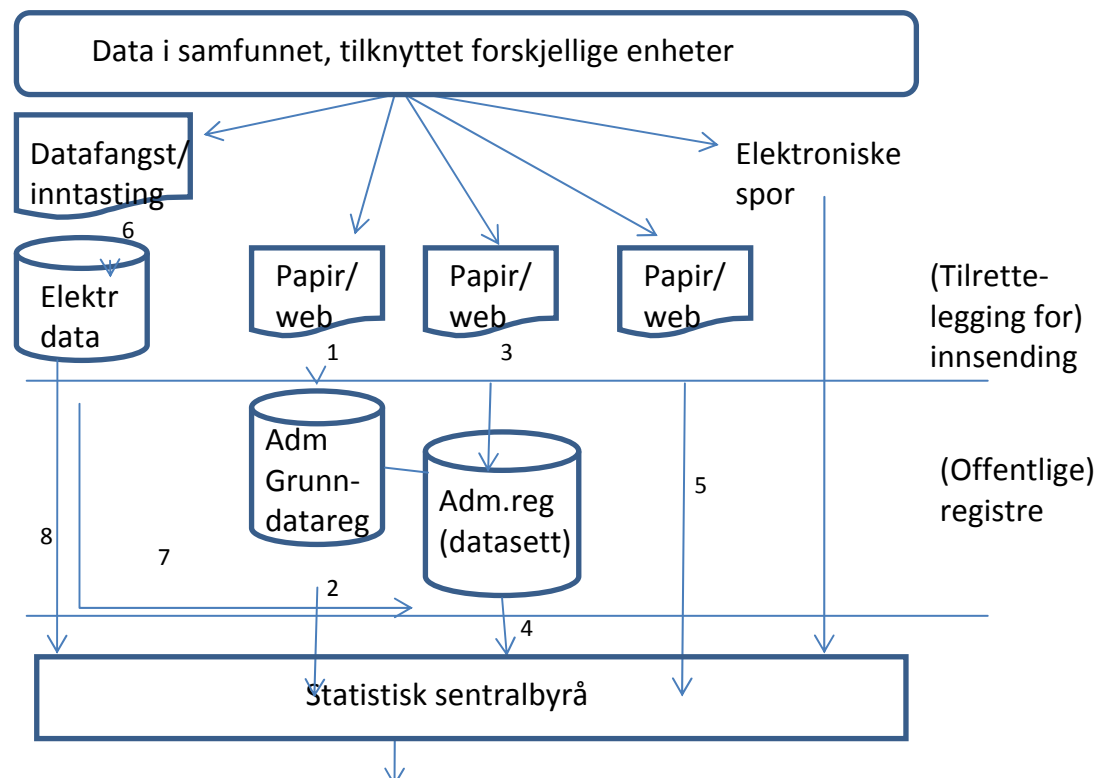
Dette kapitlet beskriver ulike løp fra data oppstår til de kommer til SSB, og synliggjør noen ”mellomstasjoner”. I kapittel 4 blir mellomstasjonene behandlet som selvstendige kilder og mer utfyllende omtalt.

Statistikken skal speile samfunnet. Dataene vi skal lage statistikk av oppstår i samfunnet. I noen tilfeller samler vi inn direkte fra kildene, i andre tilfeller henter vi data som andre har samlet inn, slik at vi får en større samling data fra ett sted. Begrepet **datafangst** omfatter begge måter å skaffe oss data på.

2.1. Gruppering av datakilder

Figur 2 gir en oversikt over ulike måter å fange data på.

Figur 2: En skisse over hvordan data kommer til SSB



Data kommer til SSB når de passerer den nederste horisontale linjen i figur 2. Mellom de horisontale linjene finner vi (offentlige) administrative registre. Disse er delt i to grupper ut fra forskjellig funksjon og bruk både i offentlig administrasjon og i statistikksammenheng. Under omtales de ulike datafangstløpene (angitt som piler i figuren):

(1 og 2): De **administrative grunndataregistrene** er helt sentrale for datafangsten og statistikkproduksjonen i SSB. Vedlikehold (1) av disse gjøres ved meldinger fra personer, næringslivet og offentlig virksomhet/myndigheter og gjennom å sette sammen informasjon fra ulike kilder. SSB henter data fra disse registrene (2), og gir generelle tilbakemeldinger om kvaliteten i registrene. For Enhetsregistret (Brønnøysundregistrene) bidrar SSB også med oppdateringsdata på enhetsnivå. Kapittel 3 omtaler nærmere bruken av de administrative grunndataregistrene i statistikkarbeidet.

(3 og 4): **Administrative registre** er en hovedkilde for data til SSB. Overføringene (4) mellom registreier og SSB (SSBs datafangst) skjer ved elektronisk overføring av elektroniske data. Datainnsamling til de administrative registrene (3) gjøres som for SSBs innsamlinger. Både i datainnsamlingen fra personer og fra bedrifter foregår det en overgang fra papirskjema til webskjema. Saksbehandlere i registersammenheng kan sammenlignes med intervjuerne i statistikksammenheng. De kan registrere data som den det gjelder har gitt ved frammøte eller i telefon. (Se punkt 4.2.)

(5): Datainnsamling fra personer og husholdninger gjennomføres med **skjema (på papir og/eller web)**, eller med bruk av intervjuere (ved telefon eller besøk). SSBs datainnsamling fra foretak/bedrifter har i utgangspunktet vært via skjema eller ved intervju. Alle skjema har nå en elektronisk variant. I overkant av 60 % av skjemaene kommer inn via webvarianten. Pil 5 i figur 2 representerer direkte innhenting med skjema eller intervjuere. (Se punkt 4.5.)

(6, 7 og 8): Datainnsamling gjennom **elektronisk overføring** av data som finnes elektronisk i næringslivet har vært i bruk i mange år, både til SSB (8) og til andre etater (7). Vi snakker som regel om "filoverføringer" eller "fagsystemløsninger". De senere årene er datasystemene i næringslivet blitt både bedre utbygd og mer brukt i deres egen virksomhet. Det innebærer at næringslivet foretar en manuell operasjon ("skjema" i en eller annen forstand) for å gjøre dataene elektronisk tilgjengelig for seg selv i sitt eget system (6). Deretter åpner dette for utvidet bruk av elektronisk overføring/rapportering til andre, enten det er et administrativt register (7) eller SSB (8). Selv om dette ikke er noe grunnleggende nytt, er det kanskje på dette området den mest markerte endringen av SSBs datainnsamling kan skje de nærmeste årene. (Nærmere om dette i punkt 4.4.)

Et viktig område i denne sammenhengen vil være de såkalte "3.parts-datene", eller data som lagres hos en annen enn den de gjelder (se pkt 4.3.). Da er det skjedd en datainnsamling hos en aktør som ikke lager et register i den betydningen vi bruker det her, og SSB gjør sin datafangst fra denne samlingen av data.

Prinsipielt er det ikke noen forskjell, men i praksis finnes det ikke tilsvarende elektroniske data for personer og husholdninger hos dem selv eller tredjeparter. Dette har utgangspunkt i at de aller fleste spørsmålene vi retter til dem omhandler fakta, vurderinger eller meninger det ikke er naturlig å ta vare på- på den måten.

I en del sammenhenger oppstår det elektroniske data andre steder enn hos de tradisjonelle oppgavegiverne. Disse dataene er ikke nødvendigvis knyttet til de tradisjonelle enhetene slik vi kjenner dem og bruker dem i datainnhenting til SSB, men de er spor etter virksomhet som vi kartlegger for å lage statistikk. Vi må vurdere fortløpende om noe av dette kan brukes i statistikkarbeidet, som erstatning for nåværende datafangst eller som tillegg til denne. Når vi snakker om å utnytte

elektroniske spor, vil det si å hente data uten at den de gjelder tar del i det. Utfordringene både personvernmessig, strategisk og metodisk virker å være store. Disse mulighetene omtales ikke nærmere i dette notatet, men notatet gir i vedlegg 5 et utdrag av en bearbeidet versjon av et innlegg Svein Nordbotten holdt på IAOS-konferansen i 2008⁴, og som omtaler slike data og muligheter.

2.2. Manuell og elektronisk datafangst

Data som ikke kommer inn elektronisk blir normalt overført til elektronisk medium kort etter ankomsten til SSB. Men om vi ser på denne figuren, blir de fleste data overført til elektroniske medier andre steder og i andre sammenhenger (1, 3 og 6 i figur 2). SSB vurderer kvaliteten på datasettene, men har sjelden analysert eller overvåket den "opprinnelige" datainnsamlingen og tilretteleggingen. Vi har arbeidet mye med å lage gode skjema for egen datainnhenting (5). Denne påpekingen antyder at vi må ta i betraktning at all datafangst baserer seg på en datainnsamling, og at mye skjemaarbeid med betydning for vår datafangst foregår andre steder enn i SSB. En konsekvens av dette er at det er svært viktig med en god dialog med våre samarbeidspartnere.

Å bruke webskjema istedenfor papir krever tilleggskompetanse i utforming av skjemaløsningen. Vurdering av kvaliteten i et webskjema vil også bli noe annerledes. Webskjema åpner for kontroller direkte knyttet til innsamlingen av data. Det er imidlertid etablering av elektroniske data i foretakenes egne systemer (eller tredjepartssystemer) som åpner for størst endring i datafangstopplegg. (Overgangen fra pil 5 til pil 6-8 og fra pil 3 til pil 6-7 i figur 2)

I spesielle tilfeller kan vi tenke oss at SSB får adgang til selv å hente ut data fra et administrativt register. Men rent prinsipielt skal det være slik at en oppgavegiver (også registreier) skal **levere** data til SSB.

3. Statistiske basisregistre basert på administrative grunndataregistre

Et knippe av de administrative registrene har en spesiell betydning for statistikkvirksomheten. De benyttes som underlag for et sett av **statistiske basisregistre** i Statistisk sentralbyrå.

*"Et basisregister er et statistisk register av stor betydning for registersystemet. Basisregistre definerer viktige objekttyper, viktige objektsett eller standardiserte populasjoner, og har link til objekter i andre basisregistre."*⁵

I tillegg til at det må være basert på stabile og kvalitetssterke administrative kilder, må et basisregister ha følgende karakteristikk:

- Definere viktige objekttyper
- Definere viktige objektsett eller standardiserte populasjoner
- Inneholde linker til objekter i andre basisregistre
- Inneholde linker til andre registre som omfatter de samme objekttypene
- Være viktig for systemet som et hele – med tilhørende krav om god kvalitet og god dokumentasjon
- Være viktig som en trekkeramme for utvalgsundersøkelser
- Kunne nyttes til statistikk om tilgang og avgang – med tilhørende krav om datering av hendelser

⁴ International Association of Official Statistics, Shanghai 14. oktober 2008

⁵ Wallgren og Wallgren: Registerbased Statistics, Administrative Data for Statistical Purposes

Gjennom Samarbeidsforum for Enhetsregistret samarbeider sentrale registreiere om vedlikehold og distribusjon av informasjon i Enhetsregistret. *Registerforum* (se vedlegg 1) arbeider for økt registersamarbeid og samspill for sentrale registre, der innholdet regnes som grunndata i samfunnet. Registrene i dette samspillet kalles **grunndataregistre**, men omtales også som mastersystemer, og det er utstrakt rett og plikt til å bruke dem i norsk forvaltning. I utgangspunktet er det lovgivningen (rett og plikt til å samle og bruke) som avgjør hva som er grunndata, men Registerforum har en sentral rolle. Fra 2011 er Registerforum oppløst som eget forum og inngår i SKATE (Styring og koordinering av tjenester i eforvaltning). Framover forventer vi at Brønnøysundregistrene, gjennom metadatasystemet SERES⁶ står for den praktiske gjennomføringen.

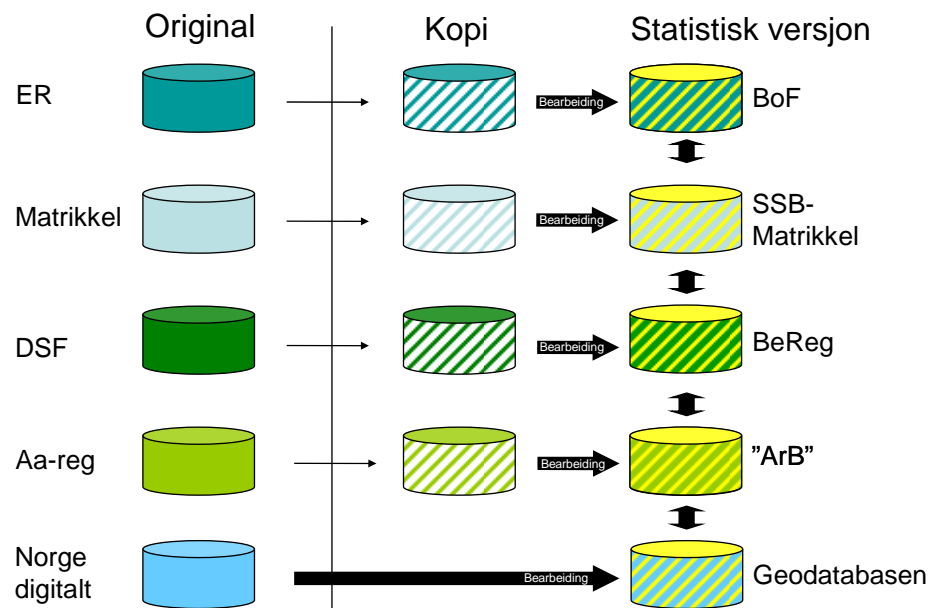
Det er basisregistrene og deres samspill med tilhørende administrative mastersystemer som står sentralt i populasjonsarbeidet. Disse er;

1. Befolkningsregisteret (BeReg)
Bygger på Det sentrale folkeregister (DSF) i Skattedirektoratet (SKD)
2. Bedrifts- og foretaksregisteret (BoF)
Bygger på Enhetsregisteret (ER) i Brønnøysundregistrene (BR)
3. SSB-Matrikkel
Bygger på Registeret over grunneiendommer, adresser og bygninger inkludert Eiendomskart (Matrikkel) i Statens kartverk (SK)
4. "ArB" dvs. jobb/arbeidsforhold/ansatte
Bygger på NAV Aa-register i NAV
5. Geodata
Bygger på Eiendomskart og andre geodata i hovedsak gjennom Norge Digitalt

Arbeidet med de statistiske populasjonsregistrene har som mål å gi det kvalitetsmessig beste grunnlaget for utarbeiding av offisiell statistikk og skal være heldekkende, oppdaterte, kvalitetssikrede og dokumenterte. Sammenheng på mikronivå er spesielt viktig for å sikre kvaliteten i registerbasert statistikk, og at framtidige folketellinger kan gjennomføres uten egen datainnhenting.

Samlingen av populasjonsregistrene gir nye bruksområder og ny og kvalitativt bedre informasjon om populasjoner og egenskaper.

⁶ SEmantisk Register for Elektronisk Samhandling

Figur 3: Fra administrative grunndataregistre til statistiske basisregistre.

I et aktivt samspill mellom registreierne og SSB holdes populasjonsregistrene løpende à jour, og er ramme for de aller fleste statistikkproduksjonene. De er trekkeregistre for utvalgsundersøkelser, men også basis for populasjonsavgrensinger i registerbaserte undersøkelser. Enhetene i registrene har unike identifikasjoner. Registeropplysningene kan kobles og utnyttes på forskjellige måter. For eksempel vil en person (DSF) ha et bosted (DSF/Matrikkel) og kanskje en jobb, som i så fall er knyttet til et foretak (ER) og et arbeidssted (en bedrift/arbeidsgiver, Aa-reg/Matrikkel). Det arbeides kontinuerlig med vedlikehold og kvalitetsheving av dette samspillet mellom basisregistrene. Det benyttes flere datakilder i kvalitetssikring av grunndataregistrene (mastersystemene), og ved omforming fra vår kopi til et statistisk basisregister.

Ved siden av identifikasjonsopplysningene inneholder basisregistrene noen kjernevariabler eller klassifikasjonsvariabler, som ofte er **grunndata** som er tatt med fra det administrative grunndataregistret, men det kan også være viktige variabler tilført fra andre administrative registre. Grunndata i det statistiske registret er altså mer enn grunndata fra det administrative registret. Disse utgjør variabler som er ofte i bruk og av mange, og er ofte sentrale for avgrensing av forskjellige populasjoner i statistikksammenheng. For eksempel er Utenlandsrelasjoner i BOF regnet som grunndata i SSB, men inngår ikke i ER.

4. En gruppering av datakilder

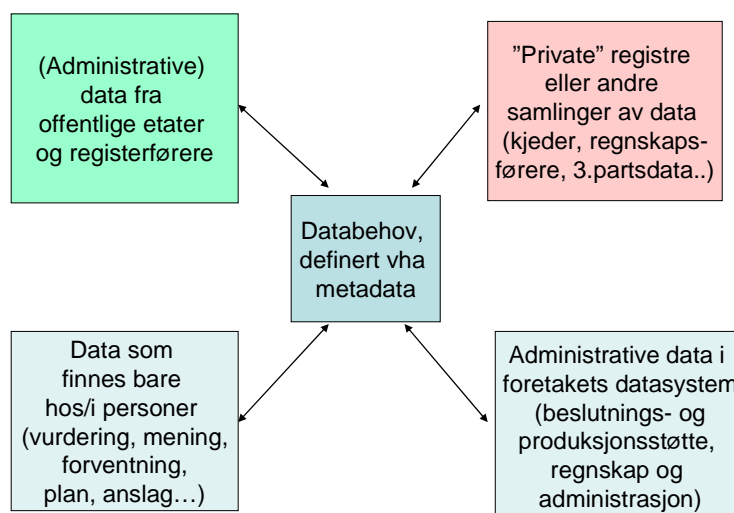
De dataene som ikke finnes i registre, og som vi har behov for å samle inn om personer og husholdninger, vil vi i praksis samle inn fra personene eller husholdningene selv. I næringsliv og offentlig sektor er bildet mer variert.

Foretakene har egne behov og lovpålagte plikter til å registrere, rapportere og ta vare på data. Hvordan og hvor de gjør dette kan variere. Når SSB etterspør data, kan disse finnes hos en annen enn den dataene gjelder for. Administrative registre er en slik mulighet. Men det finnes andre; Om et foretak bruker en ekstern regnskapsfører, vil denne ha data om foretaket og oftest en rekke andre lignende foretak. Som medlem av en kjede, en "sentral" for felles booking og lignende, eller

en (bransje-) organisasjon vil foretaket overlate enkelte data til denne sammenslutningen. Å kartlegge slike mulige datakilder er en viktig del av planleggingsarbeidet forut for datafangsten.

Figur 4: Fire grupper av datakilder

Kilder til datafangst



Ved innsamling av data om/fra personer eller husholdninger er det i praksis bare de to boksene til venstre i figur 4 som brukes.

Administrative data i foretakenes datasystemer (eller hos tredjepart som f.eks. regnskapsfører) er ikke nødvendigvis elektroniske. Modellen omfatter i prinsippet alle data som foretaket tar vare på og benytter i drift og planlegging av sin virksomhet. Det er likevel nettopp at de i stor grad er blitt elektroniske som gjør dem interessante i denne sammenheng.

Når data hentes fra registre, omfatter det alle enheter som er i registret. Ofte er det hele populasjonen som skal inngå i statistikkene. Når vi bruker de andre kildene, vil vi få med bare en eller noen av enhetene vi er ute etter fra hvert sted (ved hver henvendelse til en oppgavegiver). Det blir derfor som regel nødvendig å utvikle innsamlingsopplegg for alternative kilder til samme datainnsamling. Datafangststrategien omtaler dette som "mer systematisk kombinasjon av ulike innsamlingsmetoder".

4.1. Å definere og kommunisere SSBs databehov

Alle data som skal fanges må være definert, de må ha et sett metadata. Dersom dataene kommer fra et register, vil metadataene komme fra registreier sammen med dataene. I noen grad kan det være nødvendig å supplere disse med egne nedtegninger, og med informasjon om hvordan dataene er bearbeidet etter mottaket. Vi kan i noen tilfeller påvirke registreier mht hvordan metadata formes, men som oftest vil det være registreier som bestemmer både datadefinisjoner/-innhold og metadatautforming. Internt vil vi som regel ønske å legge metadataene over i våre egne systemer.

En forutsetning for samhandling med næringslivet og andre statsetater er **felles utvikling** av, eller **sammenknytning** av metadata. Et viktig punkt omkring dette er spørsmålet om våre interne systemer kan samhandle effektivt med det som til vanlig brukes mellom næringsliv og staten (i praksis i Altinn). Dels er dette en forutsetning for å få til effektiv gjenbruk mellom etatene, dels er det en forutsetning for å forenkle den elektroniske dataoverføringen fra næringslivet til etatene. Utgangspunktet for et felles metadatasystem har vi i SEMantisk Register for Elektronisk Samhandling (SERES) hos Brønnøysundregistrene.

Der vi legger opp til å hente data elektronisk, enten fra tredjepart (en større samling av data) eller fra det enkelte foretaket, er det SSB som definerer hva vi skal ha. Måten vi kommuniserer det på kan imidlertid bli annerledes enn i skjemasammenheng. For det ene inneholder slike løsninger ofte et større antall dataelementer. For det andre må vi ta høyde for at vi henvender oss til en som skal "finne" de etterspurte dataene i et IKT-system, og som ikke nødvendigvis har fagkompetanse på det vi spør om. Vår bruk av metadata må tenkes nøye gjennom, og tilpasses hva andre etater gjør i tilsvarende tilfeller.

SSB vil i hovedsak bruke SERES i vår kontakt med andre etater og med næringslivet. Dette blir gjort dels ved at våre metadata blir definert i SERES, dels ved at det utvikles kommunikasjon mellom våre interne systemer og SERES.

Vi har lang tradisjon for å arbeide aktivt for å få til brukervennlige og intuitive skjema. Dette er en etablert form for metadata for datainnsamlinger fra næringsliv/forvaltning og personer/husholdninger. Skjemaet kan tenkes å være tilstrekkelig metadata for etablering av elektroniske løsninger også, men i praksis vil det bli forventet minimum en tilpasning/tillegg der en del teknologiske elementer er ivaretatt. Målsettingen må være å kunne definere metadata som kan være grunnlag for å lage både skjema (på papir og elektronisk) og implementasjonsguide for elektroniske løsninger.

4.2. Administrative registre

Et administrativt register etableres og drives for å holde orden på alle enheter som omfattes av en administrativ prosess/prosedyre. Hensikten er å understøtte saksbehandling. Det innebærer bl.a. at hver enkelt enhet i registret må kunne identifiseres, siden saksbehandlingen skal føre til en konklusjon som angår og skal meddeles vedkommende. Datainnholdet i registret må tilpasses regelverket som til enhver tid nyttes i saksbehandlingen.

Grunndataregistrene som er omtalt i kapittel 3 har et bredt formål, der de skal understøtte en rekke registre og mange forskjellige prosesser i forskjellige etater. Slik sett inngår de i et registersystem, på samme måten som deres statistiske varianter gjør det i SSBs registersystem.

Noen av registrene (*fagregistrene*) som ikke er grunndataregistre har også et samarbeid om datainnhold, datainnsamling og vedlikehold. Det siste tiåret har både næringslivet, regjeringene og enkelte etater påpekt et potensial i denne retningen, og det er prosjekter i gang med sikte på å utnytte dette i større grad enn hittil. Datafangststrategien sier at "SSB vil fortsette med arbeidet for samordning av rapportering på tvers av etater". I vedlegg 2 er det tatt inn en skisse av prosjektet EDAG (Elektronisk Dialog med Arbeidsgivere), der tre etater sikter mot felles innsamling og kvalitetssikring av data. I tillegg til egen effektivisering er målet å lette rapporteringsbyrden for arbeidsgiverne. Slike prosjekter kan føre til nye administrative (og statistiske) registre, som erstatter eller forbedrer de som finnes. Samtidig må vi være på vakt i forhold til endringer i innhold og definisjoner i registrene, ettersom arbeid med slike løsninger også vil søke etter mulige tilpasninger/fellesløsninger av datainnhold.

Vi må ha oppmerksomhet omkring den **statistiske kvaliteten** i et administrativt register. Et register som fullt ut fyller sin oppgave i forhold til den administrative bruken behøver likevel ikke være egnet til den konkrete statistiske bruk vi ønsker. Spesielt må vi være oppmerksomme på dekningsproblemer og relevans, i mange tilfeller også integrasjonsproblemer når flere registre er involvert. Hastigheten (tregheten) i meldingsgangen til administrative registre kan i noen tilfeller gjøre det nødvendig med egen datainnhenting til bruk i korttidsstatistikkene selv om de dataene vi bruker (senere) finnes i et register.

4.2.1. Datafangst fra registre

Der SSB mener det finnes et register som kan gi grunnlag for data til statistikkproduksjon, eller kvalitetssikring av andre slike data, inngås det en avtale om overlevering av data fra registreier til SSB. Avtalen vil normalt også inneholde noe om hvordan SSB kan følge med på endringer som skjer eller skal skje i registret. Det er fagseksjonene i SSB som inngår og vedlikeholder disse avtalene.

Overføring av dataene til SSB skjer på forskjellige måter og i forskjellige formater. For basisregistrene overføres endringsdata ofte, slik at våre kopier kan holdes løpende oppdatert. For andre registre overføres vanligvis full kopi av nye versjoner etter hvert som slike er klare.

Tjenesten "formidlingstjeneste" i AltinnII er tenkt å skulle standardisere overføringer av registerdata fra andre etater til SSB. På sikt er det også mulig at utvikling av felles (SERES) eller koordinerte metadatasystem kan lette bruk av registerdata.

4.3. Tredje-parts-data

Tredjepartsdata er data en innsamler (første part) henter om noen (andre part) hos en annen (tredje part) enn den dataene gjelder for. Poenget med å hente inn tredjepartsdata vil i praksis være at tredjeparten har data fra flere av dem innsamleren er interessert i data om. Innsending av data på vegne av én konkret oppgavegiver f.eks. fra regnskapsfører (som tredjepart) skiller seg ikke prinsipielt eller metodemessig fra det å sende data som ansatt hos oppgavegiver. Men med elektroniske ("automatiske") løsninger kan oppgavebyrden ved å sende for alle aktuelle som finnes i det samme datasystemet bli den samme som å sende for en av dem.

Mulighet for innhenting av tredjepartsopplysning innebærer at SSB i en del tilfeller kan redusere den samlede oppgavebyrde ved å henvende seg til et tidligere ledd for å få utlevert individualiserte opplysninger om kundene. Noen eksempler:

- Informasjon om utelivsnæringens omsetning av alkohol hentes hos importørene i stedet for det enkelte serveringssted.
- Informasjon om butikkenes omsetning hentes fra kjedekontoret i stedet for den enkelte butikk
- Informasjon om drosjenæringen hentes fra drosjesentralene i stedet for hos drosjesjåførene.
- Opplysning om den enkeltes bruk av finansielle tjenester hentes hos bank/kortselskap i stedet for den enkelte bruker

Det er avklart at SSB, om det er hensiktsmessig, kan pålegge oppgaveplikt på tredjeparten (se <http://www.byranettet.ssb.no/Artikler/Jus/1099987920.95>).

Vi kunne regne det som en form for tredjepartsdata når foretakene besvarer spørsmål om de enkelte bedrifter i flerbedriftsforetak. Mer reelt blir det om vi hever dette et hakk opp, og ser på konsernet som en tredjepart som kunne gi oss data om flere foretak. Det forutsetter at det finnes et fellesskap der konsernet har

foretakenes data, eller har tilgang til dem der de finnes. Erfaringer så langt viser at det i hovedsak er aggregert informasjon som flyter mellom foretakene i konsern, noe som gjør utnytting vanskelig. Det er kanskje mer å hente ved å bygge på at de i en viss utstrekning har felles programvare.

Kjeder er en samarbeidsform som er utbredt. Dette er også en samarbeidsform som gir gode muligheter for å hente data fra ett sted om mange. Kjedekontorene har som regel mye og detaljerte data om deltakernes virksomhet. Det å effektivisere databehandlingen er en av flere drivkrefter for å etablere kjeder. Kjeder er kanskje mest kjent i butikk- og hotellsammenheng, men vi har tilsvarende i form av "sentraler" innen f.eks. lastebiltransport og drosjeområdene. Vi viser et eksempel for innsamling av data fra drosjevirksomheten i vedlegg 3. Lignende struktur finner vi omkring passasjertransport, der en billett kjøpt av ett selskap gjelder også for transport som utføres av et annet. Avregninger mellom selskapene er som regel basert på elektroniske data, og utføres av en tredjepart (eller en av aktørene).

Mange foretak setter bort administrative oppgaver til andre aktører. Mest brukt er dette innenfor regnskap og lønningsarbeid. Med slike oppgaver må det følge data fra foretakets virksomhet. Siden tjenesteyter normalt har flere kunder, gir dette en "oppsamling" av data om flere foretak hos en tjenesteyter. I praksis vil tjenesteyter være interessert i å ta på seg rapportering, for å utvide sitt oppdrag og øke sin konkurransekraft om den "opprinnelige" oppgaven. En vesentlig del av rapporteringen til Skatteetaten gjøres via slike løsninger. Rapporteringsopplegg med overføring fra system til system, er tatt i bruk i datafangstopplegg til SSB også, men utnytter sjelden at det finnes data for flere oppgavegivere tilgjengelig for oversending i samme rapportering. I våre løsninger hentes data for de som er med i vårt utvalg, ikke for de som er med i avgivers system.

Når vi lager statistikken basert på utvalg (med tilførte variabler fra registre), er det utfordrende å utnytte slike tilleggsdata fra et ukontrollert utvalg. Om vi snur det hele, og bruker registre som basis, og tilfører data fra utvalgsundersøkelser for deretter å imputere verdier i de enhetene som ikke er med i utvalget, kan faktiske data fra flere av enhetene være en styrke for datasettet.

Det er mulig å kjøpe tekniske tjenester uten å sette bort oppgaver; Å kjøpe datakraft og datalagring, men selv stå for behandlingen. Da kan vi også se muligheten til en tredjepartsløsning der data for mange finnes i ett fysisk system. I utgangspunktet har hver kunde sine egne applikasjoner og databaser, og det er derfor ikke gitt at datainnhenting fra slike datasystemer har noen nytte av tredjepartstenkingen. Innenfor enkelte områder kan det likevel finnes fellesskap, f.eks. for kommunesektoren, der "driftssentraler" tilbyr databaser som er likt utformet og felles programvare for drift av kommunal databehandling.

Mange virksomheter bruker internett til markedsføring av sine varer og tjenester. I en del tilfeller tar andre tak i de opplysningene som legges ut, og samler dem til f.eks. sider med prissammenligninger. Slike internetsider representerer en type tredjepartsdata, på samme måten som f.eks. reisebyråer kan gi (tredjeparts)data om priser hos aktører i markedet. Vi kan tenke oss å samle slike data gjennom "automatiske" prisforespørsler både hos den opprinnelige tilbyder og hos tredjeparter. Opplysninger om bolig/eiendom knyttet til salg på Finn.no inngår nå i SSBs datafangst og benyttes bl.a. til kvalitetsheving av boligdata. Da er vi over mot det som omtales som elektroniske spor i kapittel 2.

4.3.1. Datafangst fra tredjeparter

Datafangststrategien sier at "vi skal øke muligheten for automatisert og samordnet innrapportering fra fagsystem". Overføring av data skjer gjennom mer eller mindre avanserte fagsystemløsninger, eller mer enkle filuttrekk med tilhørende overføringsrutine. I begge tilfellene er det en modul/applikasjon i oppgavegivers (tredjepartens) datasystem som finner og tilrettelegger dataene som skal leveres.

Den enkleste formen, der dataene hentes ut, ordnes i en fil og sendes uten funksjonalitet til SSB, omtales som filoverføring. I en mer avansert variant (fagsystemløsning) vil det finnes funksjonalitet i fagsystemet/modulen og/eller i mottakssystemet hos SSB. Så langt er dette bare i bruk når SSB bruker Altinn som mottakssystem.

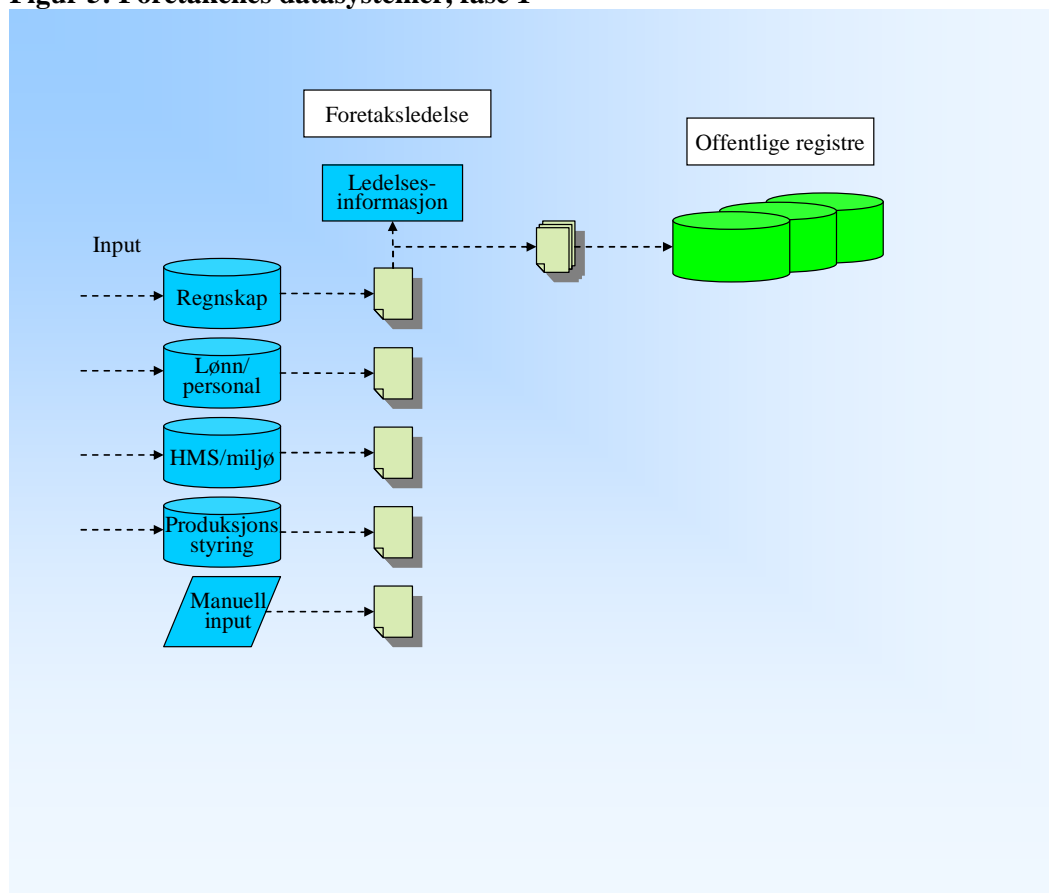
Ytterligere et sitat fra datafangststrategien: ”Vi skal styrke innsatsen overfor programvareleverandører, oppgavegivere og andre sentrale aktører for å øke både muligheten for og kunnskapen om dataleveranser rett fra fagsystem.” Altinn⁷ har implementasjonsguider som rettleider fagsystemleverandørene i arbeidet med å tilpasse systemene til slike leveranser. Gjennom dialog- og samhandlingstjenestene i Altinn2 er det ventet at dette området kan utvikles vesentlig.

4.4. Foretakens datasystemer

Som omtalt i kapittel 3 har alle virksomheter behov for og pålegg om å holde rede på en del data om virksomheten. I en god del tilfeller gjøres dette gjennom en ”tredjepartsløsning” (se 4.3.), uten at det i særlig grad vil påvirke hva de velger å ta vare på. I større eller mindre grad vil virksomheten ha egne data tilgjengelig hos seg selv. Tradisjonelt finnes dataene på papir, og da vil eneste rapporteringsform være å overføre opplysninger til et annet papir (skjema) eller et elektronisk medium for oversending. I enkelte tilfeller tilbyr vi å rapportere muntlig pr telefon eller til en intervjuer på besøk.

Etter hvert har lagrede data i virksomhetene i stor grad blitt overført til elektroniske medier for eget bruk og for rapportering.

Figur 5: Foretakens datasystemer, fase 1⁸



⁷ www.altinn.no

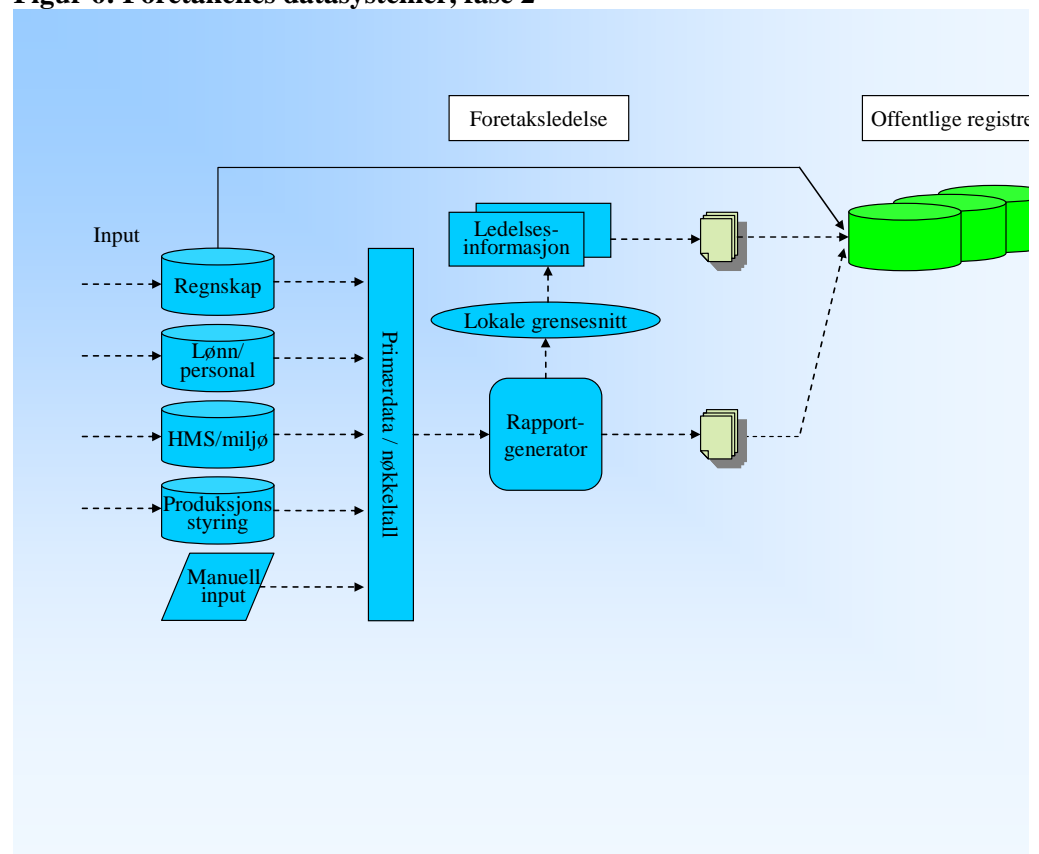
⁸ Ole Jørgen Karud, 2006

Som vist i figur 5 vil dataene være spredt på en rekke forskjellige systemer. Antallet, og innholdet i hver av dem varierer mellom virksomhetene/næringene, og systemene er i større eller mindre grad integrerte. Hvordan data kommer inn i systemene er også forskjellig, fra automatisk oppsamling fra produksjonssystemet til manuell inntasting etter en viss tid/opsamling. Data fra systemene benyttes til ledelsesinformasjon og som underlag for rapportering til offentlige registre, inklusive SSB.

Figur 5 synliggjør at byrden med framhenting av data kan øke vesentlig når en undersøkelse omfatter data fra flere av systemene. Arbeidsmengde og vanskelighetsgrad ved å implementere elektroniske løsninger vil kanskje øke enda mer enn ved de manuelle utfyllingsoppleggene. Slik sett ville det være en fordel om én innsamling kunne være fokusert på ett av systemene framfor én statistikk. Spesielt der samme virksomhet er med i mange undersøkelser bør slike muligheter søkes.⁹

De siste 10-20 årene er det utviklet og tatt i bruk et utall informasjons- og styringssystemer for foretak. Det foregår en utvikling i retning integrering og tilrettelegging med sikte på raskere og bedre ledelsesinformasjon.

Figur 6: Foretakenes datasystemer, fase 2



Dette vil redusere innsatsen som er nødvendig for å legge inn elektroniske overføringer til SSB, forutsatt at vi finner teknologiske og kommunikasjonsmessige standarder for samarbeidet om å lage løsningene. Å lage en "filtuttrekksløsning" med en egen rapportgenerator som skissert i figur 6 vil være enklere enn fra en eller flere av databasene i figur 5. I begge tilfellene er det en forutsetning at vi

⁹ Jf datafangststrategien: SSB skal legge til rette for at samordnet innrapportering fra store foretak i størst mulig grad er tilpasset datakildene og informasjonsrutinene deres.

klarer å involvere programvareleverandører og/eller foretakenes egne IT-ressurser til å utvikle løsningene.

Så langt har insentivene til å utvikle elektroniske løsninger ikke vært sterke nok i forhold til bruken av dem. Dette skyldes i stor grad at det ikke har vært behov for å standardisere informasjonen foretakene har i sine systemer. Unntakene er regnskapene og lønssystemene, der rapporteringspliktene er store, og felles for alle næringer. I vedlegg 4 er det beskrevet en mulig utvikling fra dagens situasjon (fase 1/figur 5 foran) gjennom fase 2 (figur 6) og videre til fase 3 og 4 i en utvikling på området.

Virksomhetenes egne hjemmesider på internett er en del av deres datasystem, og gir oss muligheter til å hente data uten oppgavebyrde. Dette gjøres i enkelte tilfeller for prisinnhenting. Vi regner det som uproblematisk at vi henter data uten virksomhetens egen medvirkning, ettersom det er offentlig tilgjengelige data de selv har lagt ut. Metodemessig er imidlertid slike data ikke helt enkle å utnytte.

Med de registreringer og identifiseringer som gjøres i dag er det mulig å få tilgang til noen individdata om en kundes virksomhet fra en leverandørs system, og motsatt. De aller fleste datainnsamlinger baseres på produsentens eller selgerens data. Vi må være opptatt av å unngå dobbeltregistreringer, og det gjenstår metodearbeid for å kunne utnytte begge sider av dette datatilfanget optimalt.

Ved å se på forholdet mellom kjøper og selger, finner vi underlag for begrepet transaksjonsdata, den enkelte hendelse som er registrert i virksomhetens datasystem. (Det er ikke nødvendigvis en identifiserbar kunde/leverandør.) I SSBs datainnhentinger spør vi som regel etter en aggregering av flere transaksjoner. Med vel utbygde elektroniske systemer kan det bli mer aktuelt å hente inn detaljerte opplysninger, dvs. hver enkelt transaksjon, uten at det gir stor oppgavebyrde.¹⁰ Fordelen ligger i at vi da har opplysninger om to aktører i samme datainnhenting, og data som kan gi stor fleksibilitet i bruk¹¹. Rapporteringsbyrden kan tenkes å bli lavere om vi slipper å kommunisere til dels vanskelige regler for aggregering. Gevinstmulighetene ved transaksjonsdata er størst på kort sikt om vi fokuserer på foretaksinterne transaksjoner. Både teknologiske og metodiske problemstillinger kan med fordel også begrenses til et slikt område i første omgang.

Vi kan fange opplysninger om transaksjoner utenfor deltakernes systemer, og på den måten kan transaksjonsdata være elektroniske spor. Men samtidig er det viktig å påpeke at transaksjonsdata normalt ikke vil være elektroniske spor.

4.4.1. Datainnsamling fra oppgavegiverens datasystemer

Måten dette gjøres på vil i all hovedsak være lik den som er omtalt under 4.3.1. om datafangst fra tredjeparter. Forskjellen er at vi henvender oss til den enkelte oppgavegiver. Investeringen som må gjøres blir den samme, mens gevinsten blir mindre ettersom det bare rapporteres fra én oppgavegiver for hver implementering. Så lenge det kan benyttes felles modul for alle som har samme fagsystem, vil det likevel ofte gi lønnsomhet for oppgavegiver. Selv om vi snakker om uttak av data der en rapportgenerator er tilrettelagt, kan vi møte mangel på kompetanse og/eller vilje i det enkelte foretaket. Det er grunn til å tro at dette vil utvikle seg i positiv retning.

I utredning om hva som kan tenkes levert elektronisk, og hvordan dette i så fall skal ordnes, er det viktig å se på hva som er definert og kan kommuniseres presist. Det kan være best å be om opplysninger på aggregert nivå dersom det ikke finnes noen standard for hva oppgavegiverne har på detaljert nivå. Motsatt kan det være

¹⁰ Dette gjør ikke transaksjonen til en statistisk enhet. Den er en egenskap som knytter seg til to eller flere statistiske enheter, og karakteriserer relasjonen mellom disse.

¹¹ Vi ser personvernproblemet rundt dette.

gunstig å unnlate å gi oppgavegiverne ansvar for aggregeringer der dette kan føre til at noen av detaljene plasseres feil.

4.5. Data som ikke finnes elektronisk

Som nevnt vil data fra personer og husholdninger, og som ikke kan finnes i registre, i praksis falle i denne kategorien.

I noen tilfeller ber vi om vurdering av framtidige utviklinger, men i foretakssammenheng vil det være unntakene at det spørres om opplysninger som foretaket ikke har i sine administrative systemer. (Dette omfatter også papirbaserte systemer.) Når opplysninger etterspørres til statistikkformål (se pkt 1.1.3.), men ikke til andre formål, bør oppgavegiverne varsles så tidlig at de kan samle de nødvendige data underveis i perioden det skal rapporteres for.

I mange tilfeller etterspørres tall som ikke kan hentes ut direkte, men som må beregnes. Begrensningene som ligger i papirskjemaet er ofte årsaken til at det spørres om beregnede (aggregerte) tall. Oppgavebyrden kunne i noen tilfeller vært lavere ved å gi de detaljerte tallene. Det at ”opplysningen ikke finnes elektronisk” må derfor etterprøves i forhold til å supplere med programmerte/maskinelle tilpasninger foretatt enten hos oppgavegiver eller i SSB.

Det bør tas med i betraktningen at det ikke kan forventes at alle oppgavegiverne kan ta i bruk/ benytte mer avanserte automatiske overføringsmetoder. Det vil i overskuelig tid være behov for skjemaløsninger som supplerer de elektroniske overføringsoppleggene der slike tilbys. Ved valg av innsamlingsmetode må vi både vurdere oppgavebyrden og ta hensyn til den. (Se <http://publiser.intranett.ssb.no/Artikler/Jus/1099992246.16>)

4.5.1. Datainnsamling med skjema

Under punktene 4.2. til 4.4. er det omtalt *fangst av elektroniske data*, dvs. data som allerede finnes elektronisk hos en oppgavegiver. I dette punktet (4.5.) er poenget at data ikke finnes elektronisk. Det forhindrer ikke at vi kan snakke om *elektronisk datainnhenting*, med betydningen at dataene er elektroniske når de ankommer SSB. I praksis er de da sendt inn via et elektronisk skjema, eller tastet inn av en intervjuer.

I denne sammenhengen er skjemaet innsamlingsinstrumentet. Poenget er å illustrere forskjellen mellom innhenting av data som allerede finnes elektronisk og data som ikke finnes elektronisk. Et hovedskille innenfor skjemabruk går mellom de selvadministrerte undersøkelsene, dvs. at oppgavegiver får et skjema og selv må stå for alt arbeidet fram til og med innsending, og intervjuundersøkelsene, der oppgavegiver bare har oppgaven å svare, mens intervjueren tar seg av det praktiske rundt det hele. Begge gruppene kan deles videre. Selvadministrerte kan ha papir- og/eller webskjema, mens intervjuene kan foregå ved besøk og/eller over telefon. En god veiledning om skjemaundersøkelser finnes i ”*Spørreskjemametodikk etter kokebokmetoden*”¹². SSBs datafangststrategi slår fast at ”SSBs skjemaløsninger skal forbedres, og kompetansen skal konsentreres om færre teknologier og verktøy.”

4.6. Én innsamling – flere kilder og innsamlings-instrumenter

Å sette sammen et datasett til statistikkbruk kan omfatte å forholde seg til alle typer kilder som er omtalt. Ofte kan data fra administrative registre (enten de allerede er i SSB eller hentes spesielt for formålet) ha full dekning også til statistikkformål, men det er ikke nødvendigvis slik. For eksempel er statistikk over befolkningens høyeste utdanning basert på register som oppdateres med data fra

¹² Gustav Haraldsen

undervisningsstedene, men registeret mangler data for de som innvandrer til Norge. Dette tilføres ved periodiske (hvert 10. år) skjemaundersøkelser blant innvandrere, eller ved utdanning tatt i Norge etter innvandring. Tilsvarende er det mange undersøkelser som bruker data fra registre som basis, og fyller på med egen datainnsamling.

Data som ikke finnes i register må samles inn fra de enkelte oppgavegiverne eller fra tredjeparter. Ved innsamling av data fra personer og husholdninger er skjema hovedalternativet, ettersom dataene ikke i særlig grad kan finnes andre steder og/eller elektronisk. Bruken av tredjeparter og elektroniske løsninger for å overføre elektroniske data kommer inn i undersøkelser der næringslivet og/eller offentlig forvaltning er oppgavegivere.

I noen få tilfeller, der det er et lite antall oppgavegivere, kan en tenke seg å tilby bare en variant av kildebruk eller innsamlingsmåte. I de fleste undersøkelsene er det nødvendig å tilby flere, og la oppgavegiver velge mellom disse. Valget vil avgjøres av om oppgavegiver har data hos tredjepart, har data elektronisk i sitt system, har kunnskap og vilje til å nytte elektroniske løsninger osv.

Om vi ser bort fra administrative registre, er løsningene vi tilbyr slik at oppgavegiver velger en kilde/ett instrument for hele datasettet han skal levere. Det kan lages løsninger der noen av dataene sendes inn fra en tredjepart eller fra eget datasystem, og noe fylles ut i et web- skjema. Sett fra oppgavegivers side kan det fremstå som at noe blir "forhåndsutfyllt" fra datasystemet, og resten må fylles ut manuelt. Hvor mye som hentes fra datasystemet kan variere mellom oppgavegiverne.

De elektroniske innsamlingsinstrumentene kan sies å være i en tidlig fase, og åpner ikke godt nok for enkelte av de nevnte alternativene. For eksempel er det ikke funksjonalitet i IDUN til å kunne ta i mot data fra oppgavegivers datasystem. Skjemautvikling i Altinn har så langt vært opplevd som tungvint, i tillegg til at kompetansen i SSB har vært begrenset. Med Altinn II har vi langt bedre muligheter til å utvikle gode og framtidsrettede elektroniske løsninger for datafangst.

I vedlegg 6 er det tatt inn en oversikt over datafangsten til lønnsstatistikk. I den ligger det valgmuligheter for oppgavegiver. Papirskjema eller en skjemaportal kan brukes for manuell overføring, mens levering fra fagsystemer i dette tilfellet kan gjøres gjennom grensesnitt mot Altinn eller ved uttak/overføring av definerte filer ved hjelp av f eks disketter. Den siste varianten forsøkes bevisst faset ut.

Det er nærliggende å bruke begrepet mixed-mode om slike datainnsamlinger. I de fleste tilfellene til nå er begrepet brukt om blant annet parallell bruk av papirskjema og webskjema, eller undersøkelser som gjennomføres dels som intervjueradministrert og dels som selvadministrerte. Slik vi beskriver det her omfatter mixingen en større variasjon i både kilder, oppgavegivere og instrumenter.

5. Primært og sekundært i en statistikkdata-sammenheng

Gjennom satsing på bruk av administrative registre har SSB lenge hatt oppmerksomheten på bruken av sekundære kilder. Utvikling av elektroniske løsninger for datafangst fra tredjeparter har vist at uttrykket har et videre innhold. I noen tilfeller er uttrykket brukt om elektronisk overføring heller enn sekundær kilde, og da kan det skapes uklarhet om hva som ligger i uttrykket. Også spørsmålet om primære og sekundære data og om primær og sekundær bruk av data er en del av dette.

5.1. Ved innhenting av data

Primære data er data som blir registrert fra en kilde i produksjonsløpet. Sekundære data er de avledninger og beregninger som gjøres ut fra disse. Her er det åpenbart gråsoner og følgelig vansker med å trekke opp et helt entydig skille ut fra denne definisjonen.

Primær bruk av data gjøres der hvor den tyngste begrunnelsen for å ta vare på dataene ligger. Når dataene kan nyttes i andre sammenhenger, får vi en sekundær bruk. Det finnes flere begrunnelser som sammen utgjør underlaget for å ta vare på data. Et datasett bør dermed kunne ha flere primære bruksområder. Betegnelsen gjenbruk synes å kunne få samme betydning som sekundær bruk.

Når vi henter data fra den opprinnelige registreringen bruker vi den *primære kilden*, mens data som er overført til andre, og som vi henter derfra, blir data fra en sekundær kilde. Heller ikke dette er noen enkel avgrensing. Her regnes alle tredjeparter som sekundære kilder, men tar høyde for at f.eks. regnskapsdata ofte blir registrert direkte hos en regnskapsfører.

Med disse definisjonene kan vi finne både primære og sekundære data hos både primære og sekundære kilder. SSB er primær bruker av data dersom de er registrert for å sendes til oss, mens vi er sekundær bruker av data som i utgangspunktet er registrert til andre behov. Vi vil i praksis alltid være sekundær bruker når det er registerdata vi henter inn.

Om dataene, kildene eller bruken er primær eller sekundær avgjøres ikke av om overføringen skjer på papir, webskjema eller ved å overføre elektroniske data.

SSBs strategi er å kunne gjøre *sekundær bruk (gjenbruk) av primære data fra sekundære kilder*.

Sekundære kilder er i hovedsak kilder der vi får data om mer enn en oppgavegiver fra ett sted. Primære data betyr at vi ikke ber oppgavegiver om å gjøre beregninger for oss. Sekundær bruk innebærer at oppgavegiver ikke trenger ta vare på data eksplisitt for å rapportere til oss. Alle tre poengene går på reduksjon av oppgavebyrden.

5.2. Bruk av innsamlede data

Individdata og datasett som hentes inn til statistisk bruk blir kontrollert og eventuelt justert før de er klare til bruk. Hvordan og hvor omfattende slike bearbeidinger legges opp varierer. Datakilde og innsamlingsmetode er viktige faktorer rundt valg av bearbeidingsopplegg. SSB har en egen håndbok om revisjon av data: *Datarevisjon. Kontroll, gransking og retting av data. Anbefalt praksis*.¹³

5.2.1. Statistikkssystemer krever en felles ramme for dataene

SSB har ambisjoner om å vise et sammenhengende bilde av det norske samfunnet¹⁴. Det betyr at statistikk skal være konsistent gjennom forskjellige områder. For å få til det brukes de fire basisregistrene (se figur 3) som fundament (populasjonsstyring) og tilførsel av grunndata, og de aller fleste datasett som hentes inn, enten det er som registre eller med skjema på total- eller utvagsbasis forholder seg til disse basisregistrene (se figur 1).

Det er interessant å sammenlikne det som her trekkes opp med Svein Nordbottens tanker fra 1960-tallet, formulert i et notat¹⁵ fra 1960. *Han definerte der et*

¹³ Statistisk sentralbyrås håndbøker, nr 84

¹⁴ Se strategi 2007-

¹⁵ Nordbotten: Elektronmaskinene og statistikkens framtidige utforming. Invitert foredrag for 7. Nordiske Statistikermøte i Helsinki, 1960. Se også SSB Artikler nr 34, 1969: Odd Aukrust og Svein Nordbotten: Dataregistrering, dataarkiver og samfunnsforskning.

"arkivstatistisk system" som "en statistikkproduksjon hvor innsamling av grunnmateriale foregår kontinuerlig og uavhengig av de tradisjonelle statistiske fagområder og tellingstidspunkter, og hvor bearbeidingen av det arkiverte materiale foretas etter hvert som behovene oppstår og uavhengig av innsamlingen." I senere terminologi har 'registerbasert statistikk' blitt brukt som et uttrykk som, om det ikke er sammenfallende, vel dekker essensen i Nordbottens ide som opprinnelig formulert.¹⁶

Samspillet mellom- og via basisregistre gir muligheter til å bruke data i flere sammenhenger. Data samlet inn og tilrettelagt for flere formål allerede ved innsamlingen gir *flerbruk*, mens data som i ettertid tas i bruk i andre sammenhenger enn opprinnelig planlagt gir *gjenbruk*. Vi betegner vanligvis bruk av data fra administrative registre som gjenbruk, ettersom de er samlet inn for administrativ bruk. Dette gjelder selv om den statistiske bruken også var kjent og planlagt før innsamlingen.

Behovet for å kunne gjøre forskjellige bearbeidinger avhengig av formålet viser behovet for å ta vare på rådataene, dataene slik de ankommer SSB fra registreier eller oppgavegiver.

¹⁶ Sitatet tatt fra SSBs intranett, presentasjon av Nordbotten i forbindelse med hans 80-årsdag.

Referanser

Haraldsen, G; *Spørreskjemametodikk – etter kokebokmetoden*, Gyldendal, 1999.

Karud, O. J; *Datafangst til prestasjonsmålinger, benchmarking og analyser i norsk bygge- og anleggsnæring*. Doktorgradsavhandling ved NTNU, mai 2006.

Wallgren, A og Wallgren, B; *Registerbased Statistics, Administrative Data for Statistical Purposes*, Wiley, 2007.

Lov av 16. juni 1984 nr 54: *Lov om offisiell statistikk og Statistisk sentralbyrå*, <http://www.ssb.no/omssb/statlov/>

Statistisk sentralbyrå:

- *Strategi 2007* - <http://www.ssb.no/omssb/strategi/strategi2007.pdf>
- *Strategi for datafangst*, <http://www.ssb.no/omssb/strategi/strategi-datafangst.pdf>
- *Datarevisjon – Kontroll, gransking og retting av data*. SSBs håndbøker nr 84.
- *Virksomhetsmodell for Statistisk sentralbyrå*. Notater nr 2008/47.
- *Noen sentrale begreper knyttet til metadata – til bruk i SSBs felles metadata-systemer*. Hustoft og Sæbø, Interne dokumenter nr 2006/2.

Vedlegg 1

Registerforum - mandat

1. Bakgrunn

Skattedirektoratet, Statistisk Sentralbyrå, Kommunenes sentralforbund og Statens kartverk, har over flere år samarbeidet om å gjennomføre Boligadresseprosjektet. I den forbindelse er det gjennomført flere tiltak for å samordne og kvalitetsheve informasjonen i Folkeregisteret og i GAB. Erfaringene viser at det er stor nytteverdi i et strukturert samarbeid mellom de store registereierne, og med SSB.

Ideen om å opprette et registerforum er hentet fra Finland, der et tilsvarende forum ("Registerpoolen i Finland") har eksistert siden 1997. Forumet er knyttet opp mot den finske regjeringens eGovernment-program og utnevnes av det finske "Inrikesministeriet".

Registerforum ønsker å knytte sin virksomhet tett opp mot moderniseringsarbeidet i Norge.

2. Formål

Registerforum har som formål å bidra til bedre og mer effektiv registerforvaltning i Norge.

Forumet skal være et bidrag til å øke registrenes samfunnsnytte gjennom et tettere samarbeid mellom registereierne i den praktiske hverdag. Dette skal oppnås gjennom erfaringsutveksling, kompetansebygging, idéutvikling og felles innsats for å sikre en samordnet modernisering i takt med den teknologiske utvikling og brukernes behov.

Registerforum skal bidra til økt kunnskap om basisregistrenes innhold og funksjon i samfunnet.

3. Sammensetning, varighet, mm

Registerforum har følgende sammensetning:

- Brønnøysundregistrene
- Rikstrygdeverket
- Skatteetaten
- Statens kartverk (leder Registerforum 2005 og 2006)
- Statistisk sentralbyrå

Moderniseringsdepartementet kan delta som observatør i forumet.

Ledelsen av Registerforum går på omgang mellom de etater som har opprettet forumet. Etatene leder Registerforum i to år om gangen og representerer i denne perioden forumet utad. Den etat som leder Registerforum har ansvar for sekretariatet.

Registerforum avgjør selv om man skal invitere nye medlemmer inn i forumet.

4. Oppgaver

Registerforum har følgende hovedoppgaver:

- Bidra til økt (register)samarbeid mellom de ansvarlige etater
- Bidra til gjensidig informasjonsutveksling
- Understøtte aktivitet i andre fora etatene deltar i (innenfor registerområdet)
- Samordne interesser i forbindelse med:
 - Offentlige høringer
 - Prinsipper for prising og tilgjengeliggjøring av offentlige data
 - Lov- og regelverk
 - Taushetsplikt og personvern
- Ta initiativ til regelendringer for å økt samfunnsnytte
- Øke samfunnsnyttene av registervirksomheten i etatene ved å bidra til
 - økt bruk av informasjonen i basisregistrene
 - økt kvalitet i alle basisregistrene
 - standardisering
 - metadatadefinering
 - at data bare blir etablert én gang
 - økt gjenbruk av informasjon
 - kompetanseutvikling i medlemmenes etater
- Gjennomføre informasjonsvirksomhet for å øke kunnskapen om basisregistrenes innhold og funksjon i samfunnet

Registerforum skal én gang pr. år utforme en rapport om virksomheten. Rapporten oversendes de aktuelle "eierdepartementer".

5. Saksbehandling

Registerforum kan på selvstendig initiativ behandle og ved enighet fremme saker overfor berørte parter, organisasjoner eller departementer.

Etatene melder inn aktuelle saker til sekretariatet som setter opp saksliste til møtene. Sekretariatet sørger for å avtale nødvendige saksforberedelser med den aktuelle etat. Så langt som mulig skal alle saker være vel forberedt. Der det er aktuelt skal det også foreligge et forslag til konklusjon.

Det legges opp til 2 formelle møter i året. Det forutsettes at det praktiske samarbeidet mellom etatene fra sak til sak, skjer utenom forumets møter.

Det utarbeides skriftlige referater fra møtene, med angivelse av saksdokumenter, ev. konklusjoner og ansvar for videre oppfølging.

6. Informasjon

Sekretariatet sørger for informasjonsflyt gjennom møteinnkallinger, sakspapirer og møtereferat. Det opprettes en egen internettside for informasjon til interesserte ut over forumets egne medlemmer. Hva slags informasjon som skal legges ut, drøftes nærmere mellom partene.

Vedtatt på møte 9. februar 2005

Vedlegg 2

EDAG – Elektronisk dialog med arbeidsgivere

Utklipp fra mandatet og utkast til forstudierapport.

Hovedmålet er å komme med forslag til en ny innrapporteringsordning for arbeidsgivere på området ansettelse, lønn og skattetrekk, og som dekker flere etaters databehov. Hovedmålet for ny ordning er nedfelt i mandatet (kapittel 2.1):

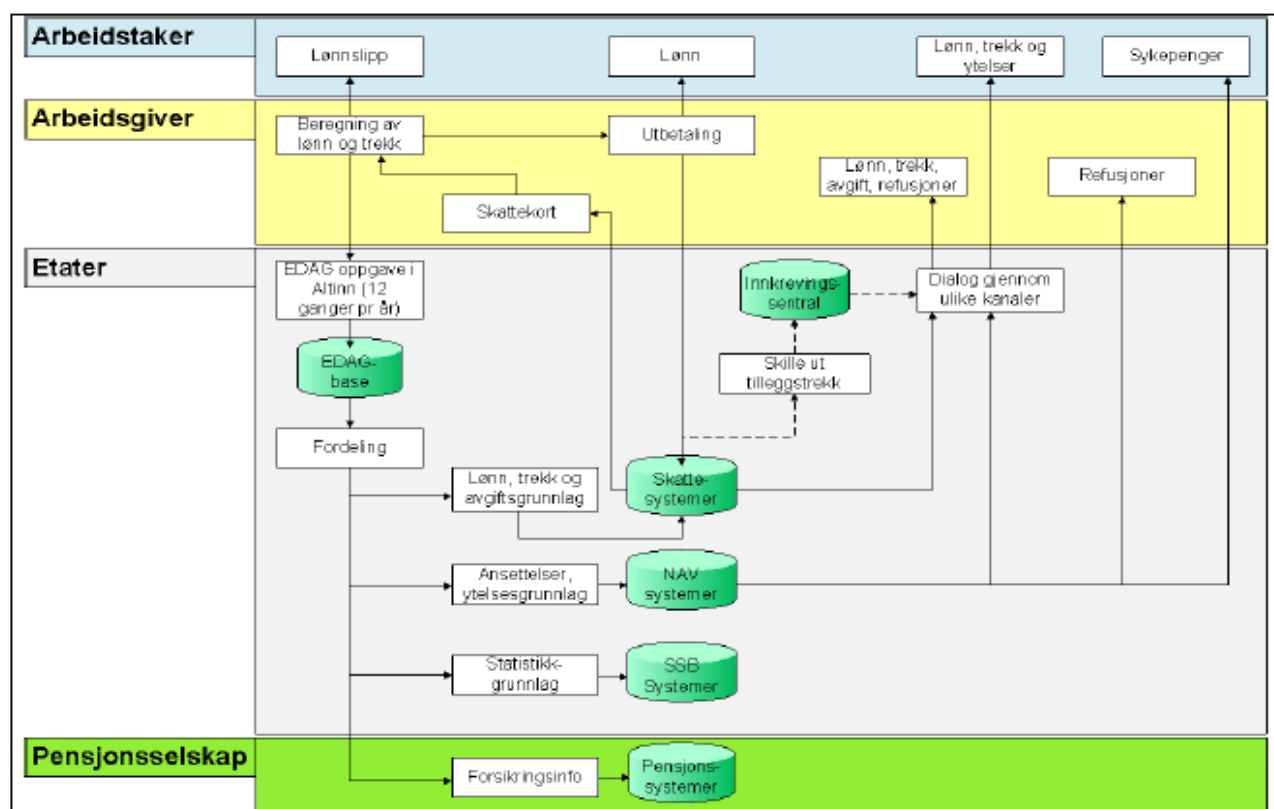
”En ny ordning for kommunikasjon med arbeidsgivere skal bidra til at det samlet sett skal bli enklere for arbeidsgiver å ivareta sin opplysningsplikt. Samtidig skal ordningen legge grunnlag for bedre oppgaveløsning i forvaltningen, inkludert bedre tjenester for arbeidsgivere. Arbeidsgiver skal kommunisere opplysning om samme forhold om en lønnstaker én gang og til ett sted.

Lønnstakere skal få forbedret innsyn i data rapportert av arbeidsgiver og skal kunne dokumentere gjennomført forskuddstrekk, uten å måtte oppbevare lønnsslipp. Lønnstaker vil slippe å måtte dokumentere opplysninger om arbeidsforhold i sitt møte med forvaltningen. En ny felles ordning skal ivareta personvern hensynene og skal sikre den enkelte lønnstaker fullt innsyn i de data som rapporteres om en selv.

Forvaltningen skal få tilgang til data som er bedre ajourholdt og som holder bedre kvalitet enn i dag, og dermed få grunnlag for mer korrekt og mindre ressurskrevde behandling. Eksempler på dette er at skatteetaten vil kunne få et bedre og mer ajourført datagrunnlag for arbeidsgiverkontroll og innfordring, arbeids- og velferdsetaten et bedre og mer ajourført grunnlag for inntektsavhengige ytelser og Statistisk sentralbyrå vil kunne produsere bedre og ferskere statistikker ut fra et mer ajourført datagrunnlag”.

Denne forstudien skisserer en ny ordning for kommunikasjon med arbeidsgivere. Hovedelementet i forslaget er at arbeidsgiver gir inntektsopplysninger på personnivå i forbindelse med avlønning. Skissen til løsning må ytterligere utdypes og detaljeres i et forprosjekt.

Bedre måloppnåelse - skisse til ny felles ordning



Figur 2 - Skisse til ny felles ordning

Kort forklaring til de enkelte elementene i skissen:

- **Lønnsberegningen** beregner brutto lønn, trekk og netto lønn. Beregningen gir grunnlag for lønsslipp, samt betaling til lønnstaker og trekk til myndighetene.
- **EDAG-oppgaven**, en utvidet lønns- og trekkoppgave, produseres normalt maskinelt fra arbeidsgivers lønns- og personalsystem og leveres via Altinn.
- Uavhengig av innleveringskanal lagres innsendte oppgaver i **EDAG-basen**. EDAG-basen blir således en felles database der all innsendt informasjon er lagret. Hver etat får tilgang til opplysninger i henhold til hjemmel.
- Oppgaven angir hvilken lønn som er utbetalt til lønnstaker, og grunnlaget for innbetaling av trekk og avgift til myndighetene.
- Ordningen gir mulighet for at arbeidsgiver kan angi hvilket selskap det er avtalt pensjonsforsikring med, og at angitt forsikringsinformasjon vil kunne formidles videre til forsikringsselskapet.
- EDAG-oppgaven gir faktagrunnlag for videre dialog mellom etatene og arbeidsgiver/lønnstaker. Dette kan gjelde både direkte personkontakt og ulike former for "selvbetjeningstjenester" gjennom ulike kanaler.
- EDAG innebærer at dagens rutiner for utleggstrekk kan bortfalle, og at alle eventuelle tilleggstrekk i lønn kan innkreves gjennom skattekortet.
- Opplysningene fra arbeidsgiver kan legges til grunn for tilståelse og beregning av ytelser fra NAV. Omfanget av typer opplysninger som oversendes fra arbeidsgiver vil påvirke fremtidige muligheter.

Vedlegg 3

Forkortet versjon av innlegg i:

Seminar on Registers in Statistics - methodology and quality 21 - 23 May, 2007 Helsinki

How Statistical Data Collection could be adjusted to an industry specific organization.

*Yngve Bergstrøm
Statistics Norway
yngve.bergstrom@ssb.no*

Summary

Public administrative registers and Industry Specific Administrative Systems supplemented with survey data can be the basis for statistical data collection. A case from Transport statistics shows that a mix-mod survey design can ease the response burden for businesses and increase the frequency in the data collection. A tailor made data collection requires that the data collector has very good information about the organisation of the whole industry that is going to be described in the statistics. This information helps us to design a survey that is efficient for both the enterprises that has to answer the surveys and the data collector.

For many activities the enterprise need a public permission to perform the activity, one example is taxi operation. These permissions are registered in an administrative public register. On the other hand has there been developed industry specific electronic logistic/production systems. For taxi operation are these administrative systems administrated from regional taxi central agency offices. These systems have production information concerning each car operating within the taxi central region. The IT-system register turnover, number of kilometres with and without passenger etc. This paper discusses a new data collecting system from these Industry specific systems, and how the information is combined with the governmental register information. A main challenge has been to have control of the statistical population as the industry specific register does not cover the production for the whole industry. Another important issue is to combine the register information from both the public and industry specific register with survey information.

Chapter 1. Introduction

This paper describes the data collecting system for taxi operation statistics in Norway. The new system combines register information with electronic data collection from the regional taxi central agency offices. The survey is starting this year with a quarterly data collection from the taxi central agencies, and will be completed with a yearly survey to the rest of the population by the end of the year.

Traditionally has the data been collected yearly by a paper survey. A taxi company is a small enterprise with 1-2 employees. The driver owns the car and is registered as an enterprise. These small companies do not have any administrative resources. Answering questionnaires usually also involve their bookkeepers, who often also are small specialized companies. Every new questionnaire that is sent to them implies a risen response burden for these small companies.

A way to ease the burden is to develop a data collecting system that uses information that is already electronically registered in one or another form among the businesses data systems or in public registers.

Chapter 2. The statistics objectives and need of data

Chapter 2.1 Objective

Statistics Norway is publishing a yearly statistics that gives some key information on the structure of the industry. The National Taxi organisation also publishes some statistics covering the members of the organisation, number of members and their turnover. There is a need to build better information on fluctuations of the transport market. This information is asked for by both the Bank of Norway (national bank) and the National Account.

The statistics for taxi operation services is for the moment not very good.

Chapter 2.2 Definition of the data need

The data that is needed for the statistics has been defined by the subject matter division in cooperation with researchers and The Norwegian Taxi association. It is also important to define the statistical population. In this project is the statistical population defined as all enterprises that have permission to operate taxi. The population is collected from the register of taxi licensees. The register contains all licensees. It is a link to the responsible enterprise by the organization number and/or the district in which the permission is valid. The register of vehicles gives information of the cars that are registered as taxi cars, showing i.e. the number of taxi cars. This register also has information of the owner of the car, but as many taxi cars are leased by the operator it usually is the organisation number of the leasing company (owner of the car).

The production is measured by the number of hours a taxi car is in operation, the length in km that a car is going with or without passengers and the turn-over.

Chapter 2.3 Where do we find the information needed?

There is no single register information that gives information about the transport production within the industry. The total turn-over is possible to find for each enterprise that has a taxi license from the value added tax system. Transport of persons has its own tax rate – 8 %, but there is no information if the turn-over is from taxi operation or other business within the registered enterprise.

The information is known by the enterprises, and the traditional way to find this kind of information is to ask. To find out if it was possible to organize the data collection differently we had to look at how the taxi operation industry is organized in Norway.

The contact with the national taxi organization informed us that most of the taxi centrals around in the country stored basic information in their logistics system. When a customer phones for a taxi, the requirement is stored in the system. Both which car/license that has produced the service and the price of the service is registered. They also store the distance the car have gone with and without passengers.

Many of the centrals use the same software.

Chapter 3 Organization of the taxi operation industry – NACE 60.22

Every one who transports people for cash/payment in Norway must have a license from a competent authority – in Norway is this The County. The holder of the license could be a person or a company. If it is a company, the daily leader must have the required qualifications defined by law. The license gives the holder the right and duty to operate taxi services within a region – license district. This is usually the same as the county. A taxi car can have at the most 8 passengers. In some occasions can the authority give a license with the right to transport up to 16 passengers in the same vehicle -

minibuses. All licensees are before approval tested for provision of needs. The license also states if the holder has to be a member of a taxi central or if the taxi could be operating from the place of living. All approved licensees are registered in a public administrative register with licence number, link to organisation number of the holder and the name of the holder.

Taxi operation services in Norway consist today of 6500 enterprises. Totally has the Department of Transportation given approval to 8000 licensees. These enterprises operate 6900 cars and about 1000 minibuses. It is estimated that the industry produce about 12000 man-labour per year. Figure 1.

1100 of these enterprises are operating locally from the place of living to the owner. 5400 enterprises are coordinated thru 227 Taxi agency offices¹⁷ – taxi centrals in 200 license districts. Attached to these centrals are often different kinds of common purchase companies. The companies could for example buy gasoline, give service to the car or buy/own cars collective for the members of the taxi central. On the other hand have large public users of taxi operation build up coordinating offices for buying patient transport for hospitals etc.

There are 5 different software vendors that has installed and/or developed the logistics system used by these centrals.

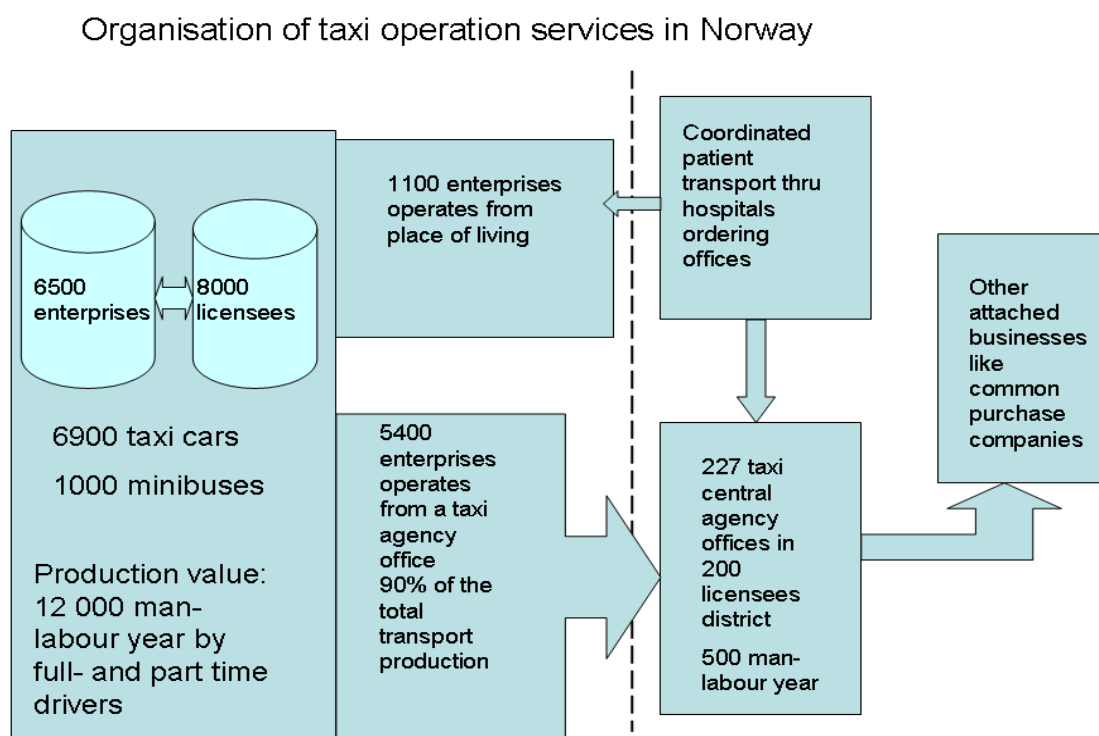


Figure 1. Organisation of taxi operation services in Norway¹⁸

Chapter 4. Definition of the statistical population

The statistical population is defined as all enterprises that are classified within activity class NACE 60.22 Taxi operation. First there has been a need to see if this population corresponds to the taxi licence register. A link between the licence register and the Central Register of enterprises and

¹⁷ NACE code 63.211 Central agencies for goods and transportation procurement in the Central Register of Enterprises and Establishments

¹⁸ Kilde: Jørn Bertelsen, Norges Taxiforbund

Establishment gives information of all enterprises with a taxi license. The license register also states if a holder of the license has to be a member of a taxi central. However, there are no links in the license register directly to a Taxi central and CRE. An old overview is available from the Taxi association. It has therefore been necessary to develop routines to update these links by asking the taxi centrals which taxi enterprises that are connected to the central. This updating has in the future to be done at least on yearly basis, but in some time we hope the public licence register will be complemented with the information. The information is essential; as it helps us to know for which taxi enterprises a single taxi central can report production data.

Another challenge is that Taxi centrals can operate other services outside the statistical population, for example operating in NACE class 60.21 (other scheduled passenger transport - transport of school children) or 60.23 other land passenger transport on contract, i.e. touring cars. The data extraction from the IT-systems in the taxi centrals therefore has to define exactly all units they are reporting for.

Figure 2 illustrates these main challenges in defining the population. The survey is designed as quarterly reporting from the centrals for all their production, and a yearly reporting for the rest, i.e. taxi operation from the place of living to the holder of a license. The size of the yearly survey has not yet been decided. As soon as there is a full updating of the enterprises that operates from a taxi central could the remaining enterprises be analysed thru information in the VAT-register. There are indications that these enterprises are small, and we hope that the yearly survey could be easily stratified by region. This work is going on for the moment and has to be finished by the end of the year. The data collecting from the taxi centrals is starting from the 1. Quarter of 2007, and the yearly survey is going to be conducted by the end of this year.

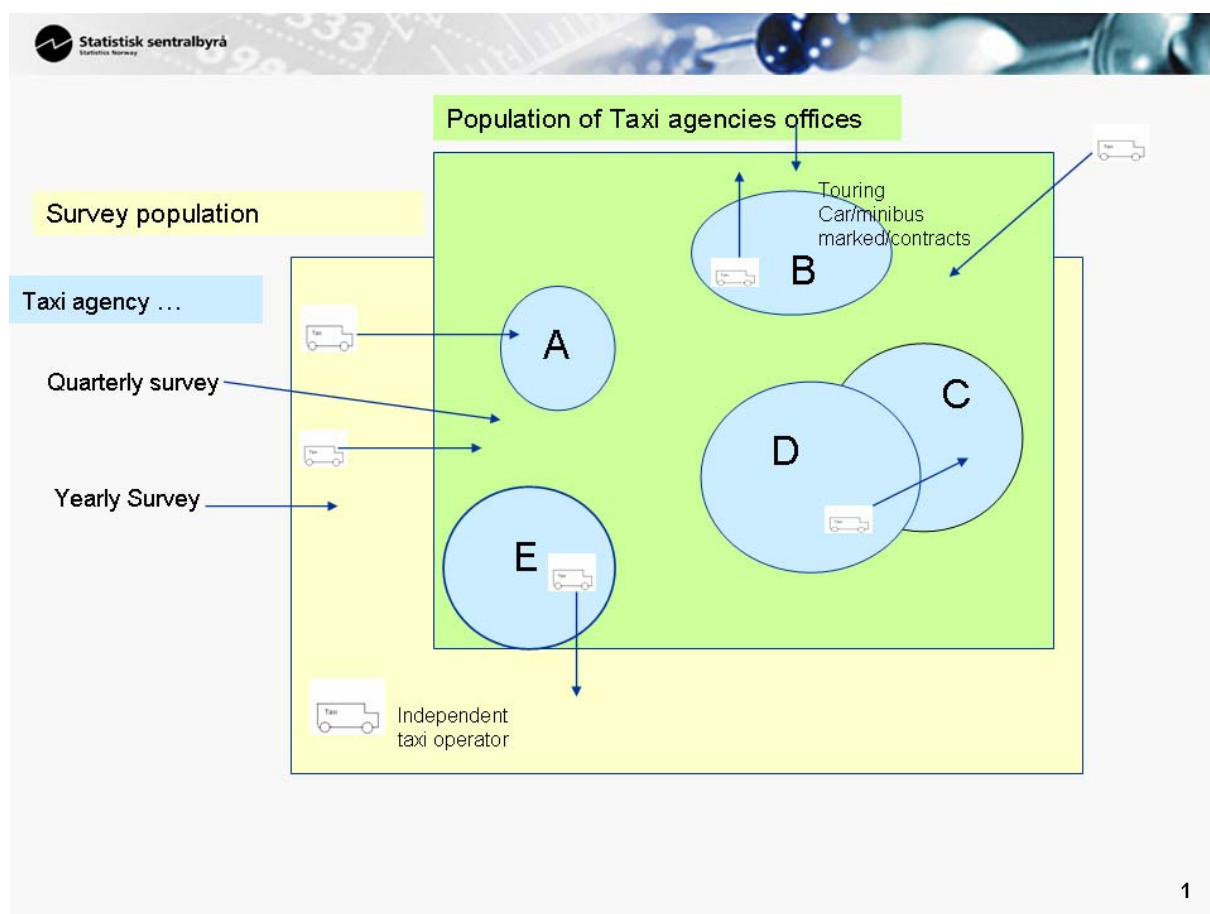


Figure 2. The survey population and the population covered by the taxi centrals

Chapter 5 Suggested organization of the data collection

The data collecting system has been built up by using the Altinn-portal. Altinn is a common Internet portal for public reporting in Norway. A short description is given in appendix 2.¹⁹

Figure 3 illustrates the organization of the data collecting system.

The register of taxi operation licensees and the CRE are linked to identify the statistical population. It is the organization number from the CRE that is the link key. Phase 1, 2 and 3 in the figure

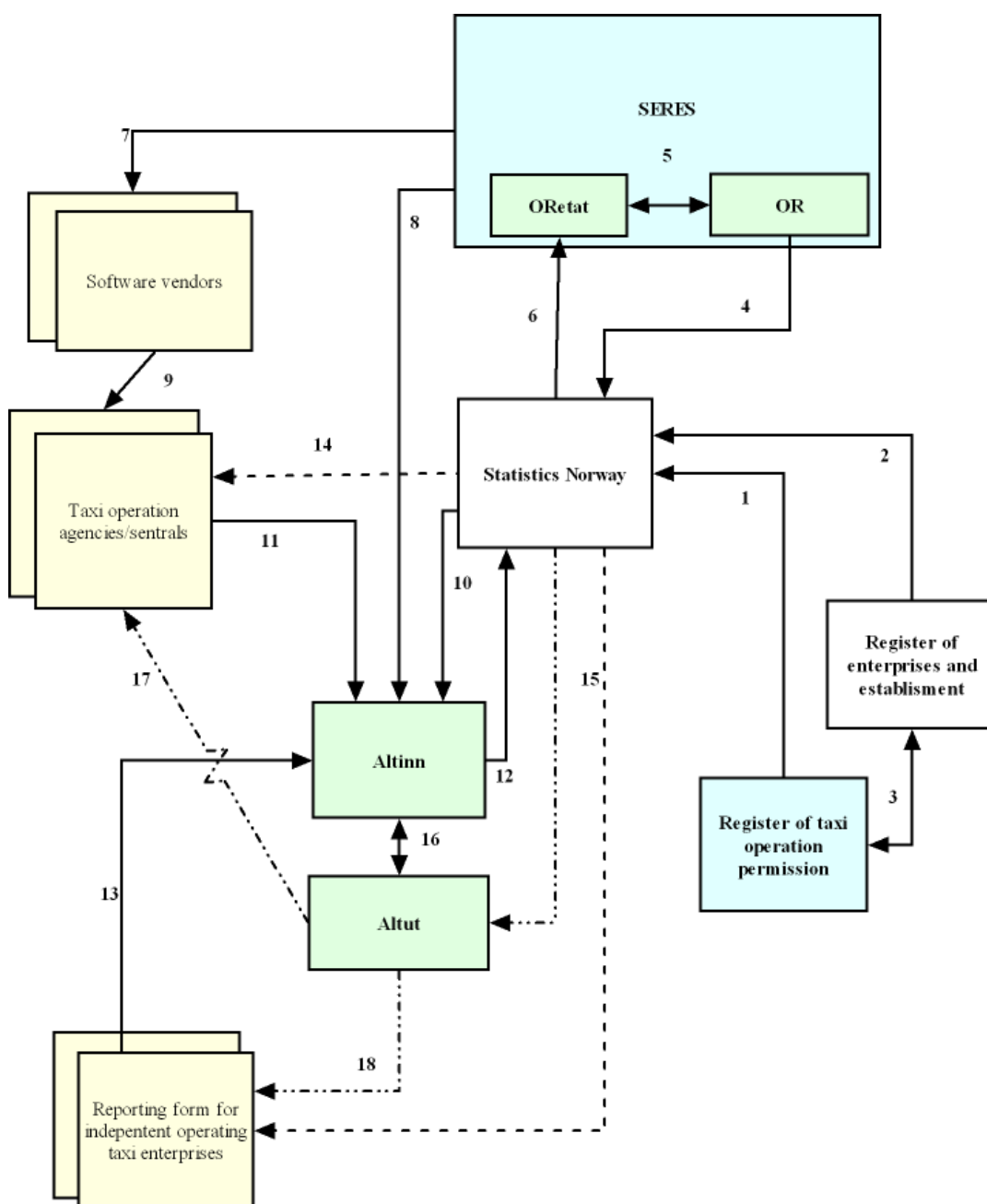


Figure 3. Organization of the Data collection

¹⁹ Se also www.Altinn.no

The questionnaire has been implemented in the Altinn-portal. As the enterprises are obliged to answer the questionnaire, the data must be registered in *The Register of Reporting Obligations* – OR. Appendix 3.

When a form is implemented in Altinn is the metadata that is needed to define the form registered in the metadata base SERES (see 5.1.). From the database is it possible to create a XML-specification of the data request and the XML-specification is published thru the Altinn-portal. All software vendors can download the XML-specification for public forms.

The whole specification could be downloaded from:

http://w2.brreg.no/oppgaveregisteret/spesifikasjon_skjema.jsp?skjemanr=1220.

Chapter 5.1 **SERES**

The Norwegian SEMantics Register for Electronic Services (SERES²⁰) is an IT system and information model repository for promotion of semantic interoperability in the public sector. It is based on the idea of establishing a coherent information model described in an easily comprehensible and commonly accepted and used modelling language, namely UML. The system is implemented as an Internet-based multi-user, multi-layer Java-application (J2EE). Users from any public institution may have access to SERES.

Only one of the UML diagram types is used, the class diagram. The datatypes are called semantic types and are organised as a taxonomy starting with the set of Core Datatypes from UN/CEFACT. The information model is composed of a set of domain models that may have different levels of harmonisation with other domain models and varying adherence to an agreed set of modelling guidelines. These guidelines require that information classes are directly or indirectly related to a harmonised core information model and that all data types are semantic types.

Document models are extracted from the information model and are combined with format elements to form message specifications. Each semantic type is directly or indirectly connected to some format element. A message specification reflects the structure and semantics of the corresponding document model and employed format element set. Dublin Core-type meta-information is used for administration and tracking of individual model elements. The resulting specifications are published on a web site available for all.

Chapter 5.2 **Software vendors**

Software vendors can now download the specification and implement it in the logistics system used by the Taxi central agencies (activity 7 and 9 in figure 3). The centrals that do not have direct reporting possibility can report by filling in an electronic form in Altinn. The same form can also be used by independent taxi operators that is not participating in a Central, i.e. those who reports once a year.

Chapter 5.3 **Reporting of data**

The taxi centrals receive a letter from Statistics Norway that give some information of the survey and log on information to Altinn. To the quarterly survey, there has only been developed an electronic form. In appendix 2 there is a copy of the electronic form used in Altinn. For the yearly survey we see a need of a paper survey as there are some enterprises that do not have Internet connection.

²⁰ SERES Short Description by Even Thorbergsen, Brønnøysund Register Centre, Norway May 2006

Altut is a new possibility in the electronic portal. This service gives Statistics Norway possibility to send electronic information to a specific enterprise. The connection is also secure.

Altinn is used by many different public authorities in Norway. It means that the users get familiar with the user-interface and know how to use the system. Another advantage with Altinn is that the users of Altinn can either fill in the forms directly in the Internet portal or they can use their own IT systems to transfer data, for example salary and accounting systems or a year-end accounting package. The companies' own IT systems transfer pre-filled forms to the portal through a simple interface where subsequently the forms can be completed and signed in the portal. During the time the software vendors develop these routines can a respondent fill in the Altinn-form manually by using an on-line Internet connection with Altinn.

Chapter 6. *Further development and challenges*

An important task that so far not has been decided is the design of the yearly survey. There is also a need to build up knowledge concerning the development of the response burden for the enterprises. In the project has there so far been developed new samplings, survey and production routines. We assume that this will lighten the response burden, but we do not know. Statistics Norway has developed a set of questionnaires to measure the perceived response burden²¹. We hope that it will be possible to start collecting this information from the first yearly survey. An analyse of this information with the same information from taxi central agencies will over some time give information and knowledge if the response burden actually has dropped.

Some years ago, value-added-tax was introduced in taxi services, with a tax rate of 8 %. It would be valuable to use the turn-over data in revision of the collected data, and in updating of the statistical population. This information could also be used in the design of the yearly sample, i.e. give background information before the sample size is decided.

The work has so far also shown that it will be a challenge to always have good control and updating routines of the statistical population. As statistical data is collected from a few centrals Statistics Norway always has to know which enterprises are part of a taxi central, and which are not.

References:

²¹Developing Methods for assessing perceived response burden, by Hedlin, Haraldsen, Dale and Jones

SERES Short Description by Even Thorbergsen, Brønnøysund Register Centre, Norway May 2006

Jørn Bertelsen, Norges Taxiforbund

Runar Kjellin, drosjelovgivning gjennom tidene, Norges Taxi forbund

Kollektivtransporten har et statistikkproblem, Eivind Frøysadal, Transportøkonomisk Institutt

Markedsorientert statistikk for kollektivtransport, Norheim og Frøysadal, Transportøkonomisk Institutt

<http://www.transportloyve.no> – Information about licensees in taxi transport (only in Norwegian)

Vedlegg 4

Nye metoder for prestasjonsmåling i byggenæringen

Doktoravhandling av Ole Jørgen Karud:

Avhandlingen

Avhandlingen har tittelen "Datafangst til prestasjonsmålinger, benchmarking og analyser i norsk bygge- og anleggsnæring", og ble levert i mai 2006.

Nye datafangst- og rapporteringsmetoder

Karud foreslår konsepter for lokal og sentral datafangst av nøkkeltall og prestasjonsindikatorer i BA-næringen, basert på bruk av ny teknologi og nye metoder for **automatisert datafangst og rapportering**. Videre foreslår han standard utvekslingsformater mellom bedriftslokale prosjektdatabaser og nasjonale registre, slik at **også offentlig rapportering samtidig kan forenkles**. Det er en målsetting å redusere bedriftenes oppgavebyrde overfor det offentlige. Samtidig er det vesentlig å kunne tilby næringens aktører raskere og sikrere beslutningsgrunnlag gjennom effektiv tilgang til oppdatert informasjon i ulike offentlige registre.

Avhandlingen diskuterer og beskriver også et konsept for en ny sentral nasjonal database for prestasjonsmålinger, benchmarking og analyser i BA-næringen. Denne vil imøtekomme næringens behov for felles benchmarkingsgrunnlag og vil kunne ha stor nytteverdi for næringens aktører, bransjeorganisasjoner og myndigheter.

Utdrag fra avhandlingen²²:

4 Utviklingstrinn for datafangst og rapportering

Et mål om automatisert datafangst for lokal og sentral rapportering vil naturlig nok skje gjennom flere trinn. Nedenfor presenteres 4 utviklingstrinn for en slik prosessmodell. Figurene er satt opp for å markere utviklingen uten å være for detaljerte. I praksis kan det være kombinasjoner av disse utviklingstrinnene.

Følgende forenklinger er her foretatt:

- Figurene viser intern og ekstern rapportering pr foretak²³ samlet. Prosjekter og eventuelle avdelinger og distriktskontorer er derfor ikke vist.
- Det er her visualisert at alle foretakene har ERP²⁴-systemer / fagapplikasjoner med egne databaser for: regnskap, lønn / personal, HMS / miljø og produksjonsstyring. I enkelte bedrifter er disse slått sammen til større og færre systemer eller de kan også bestå av langt flere systemer, uten at dette er viktig i denne sammenheng.

²² Det kan være mindre avvik mellom det som er gjengitt her og det som ble publisert.

²³ Betegnelsen *foretak* er her valgt i stedet for konsern eller bedrift, fordi dette betegner juridisk enhet i henhold til Enhetsregisteret og brukes i dag av Statistisk sentralbyrå ved deres skjemarapportering.

²⁴ ERP = Enterprise Resource Planning, samlebetegnelse for foretakenes administrative datasystemer.

- I noen grad trengs rapportering til ledelsen som ikke kommer fra ERP-systemene. Dette er illustrert med en manuell inputprosedyre. Data herfra antas i denne sammenheng å bli behandlet og lagret på linje med data fra ERP-systemene. Samlet kalles dette subsystemer.
- Rapportering til foretaksledelsen er illustrert via et LIS²⁵-system, i prinsippet kan dette være manuelt og papirbasert, automatisk og databasert eller kombinasjoner av disse.
- Offentlige registre er visualisert som 3 databaser. Dette er registre hos Statistisk sentralbyrå og eksempelvis Brønnøysundregistrene, GAB-registeret, Byggsøk-databasen, bransjeregistre etc.

I figurene er det brukt følgende strek- og fargekoder:

-----I hovedsak manuelt initiert rapportering

_____I hovedsak automatisk rapportering

 Lokalt

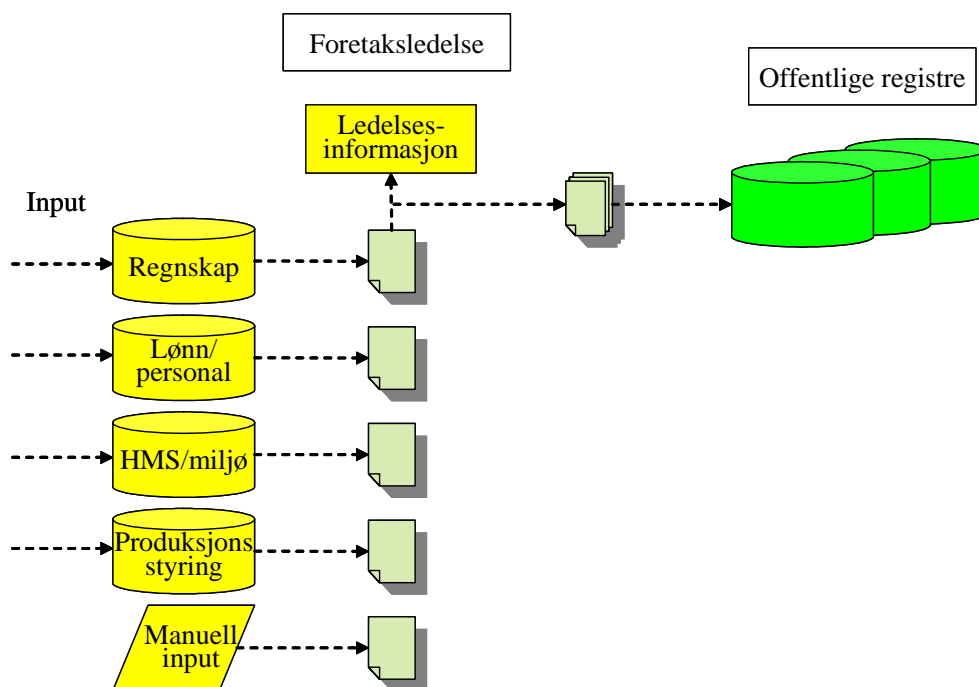
 Sentralt

 Offentlig

Trinn 1 – dagens situasjon

Første trinn illustrerer dagens situasjon for de aller fleste foretak og rapporteringsforhold, se figur 1. Input og output i ERP-systemene er i stor grad manuelt initiert og skjer direkte i hvert subsystem uten at disse er koblet sammen i vesentlig grad. Rapportering til ledelsen skjer ved at det lages rapporter fra hvert system. Det er liten eller ingen strategisk samlet nøkkeltallsrapportering. Rapportering til det offentlige skjer manuelt i hvert enkelt tilfelle og er for det meste papirbasert. Det er da heller ikke krav til grensesnitt mellom subsystemene.

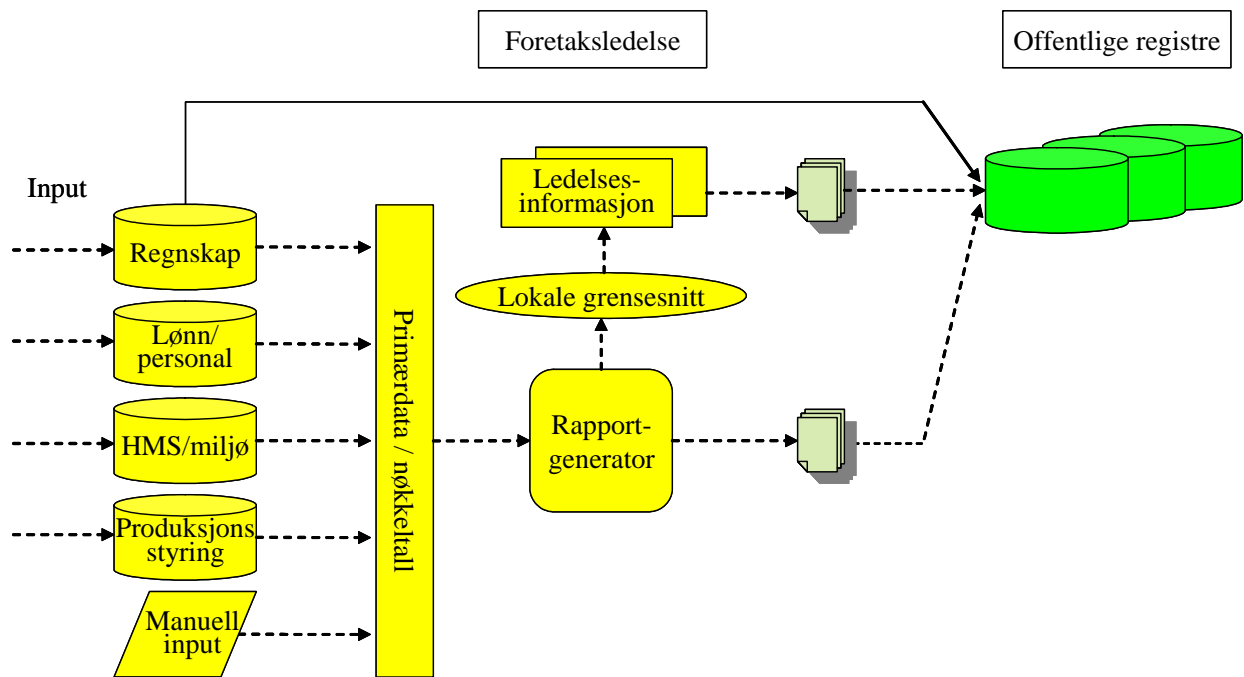
²⁵ LIS = Ledelsesinformasjonssystem, systemer som støtter beslutningsprosessen hos toppledelsen.



Figur 1 Prosessmodell datafangst og rapportering, utviklingstrinn 1 - dagens situasjon.

Trinn 2 – lokal integrasjon

Figur 2 viser trinn 2. Input til ERP-systemene er fortsatt manuelt initiert og direkte til hvert enkelt-system, men subsystemene integreres til en viss grad. Nå er også strategiske nøkkeltall og prestasjons-indikatorer definert slik at viktige primærdata fra subsystemene kan trekkes ut og kombineres. Rapporteringen internt skjer ved at primærdata og nøkkeltall går fra subsystemene og via en rapport-generator som skreddersyr informasjonen til ledelsen. Rapportgeneratoren illustrerer her en system- og rapporteringsmessig overbygning for subsystemene. Ledelsesinformasjonen får dermed en ny dimensjon og kan gi øket verdi for flere nivåer og bruksområder i foretaket. Dette fokuset på nøkkel-tallsrapportering kan greit kombineres med systemer for balansert målstyring. Internt er grensesnittet mellom rapportgeneratoren og LIS-systemet proprietært og tilpasset foretakets egne behov.



Figur 2 Prosessmodell datafangst og rapportering, utviklingstrinn 2 - lokal integrasjon.

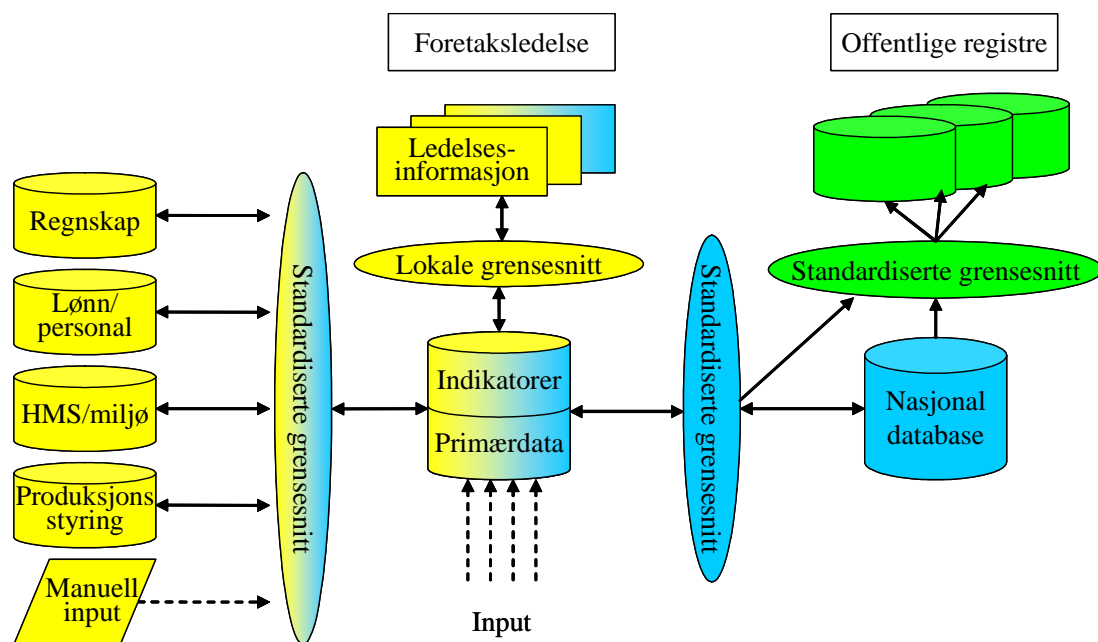
Rapportering til offentlige registre skjer fortsatt ved manuelle rapporter, men rapportgeneratoren kan alternativt bidra til at dette forenkles i betydelig grad. Et annen mulighet er at de mest sentrale ERP-systemene åpnes og tilpasses for manuelt initierte, men ellers automatiske uttrekk til enkelte offentlige registre. Et verktøy for innlevering av offentlige skjemaer er web-portalen AltInn (se delkapittel 7.5). En del fagapplikasjoner har allerede tilpasset grensesnitt for uttrekk til AltInn.

De foretak som har arbeidet spesielt med intern nøkkeltallsrapportering har en prosessmodell som antagelig ligner mest på trinn 2.

Trinn 3 – lokal og sentral database

I trinn 3 er integrasjonen mellom de lokale subsystemene ytterligere utbygd, se figur 3. Nå er det også etablert en intern database for primærdata, nøkkeltall og indikatorer med automatiske uttrekk fra subsystemene. Input kan skje direkte i denne lokale databasen uten at man må forholde seg til de underliggende ERP-systemene direkte. Det bidrar til at data kan registreres én gang og på ett sted. Databasen utveksler kontinuerlig data med ERP-systemene.

Samtidig er det i trinn 3 etablert en nasjonal database for prestasjonsmåling, benchmarking og analyse. Denne nasjonale databasen er bygd opp av omforente og standardiserte primærdata, nøkkeltall og prestasjonsindikatorer for effektiv benchmarking og statistikk på bransje- og nasjonalt nivå. Dette nasjonale nivået stiller også krav til at grensesnittet og utvekslingsformatet overfor den lokale databasen er standardisert. Figuren illustrerer også at dette kan være en toveis kommunikasjon.



Figur 3 Prosessmodell datafangst og rapportering, utviklingstrinn 3 - lokal og sentral database.

Det er en klar fordel at de lokale prioriterte primærdatabaser og nøkkeltall samspiller med de som er valgt på nasjonalt nivå og vise versa. Det forenkler kommunikasjonen og forbedrer statistikk- og analyse-grunnlaget. Det betyr at datautvekslingen mellom ERP-systemene og den lokale databasen også til en viss grad må følge standardiserte grensesnitt.

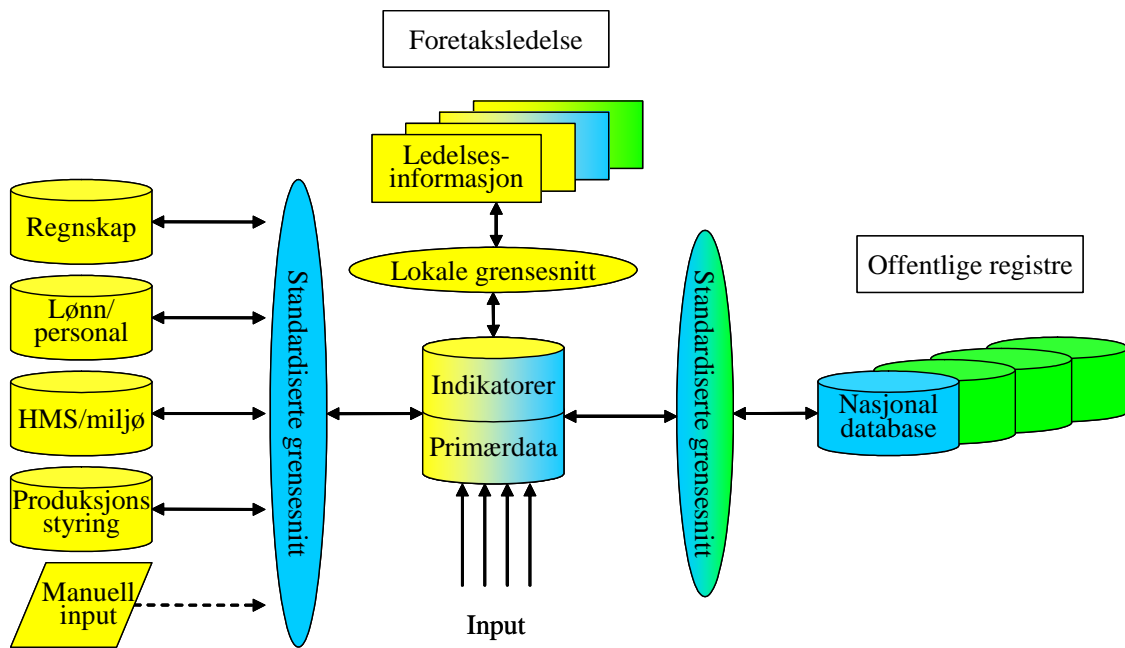
Grensesnittet mellom den lokale databasen og LIS-systemet kan fortsatt være lokalt tilpasset.

Ledelsesinformasjonssystemet er her bygd ut ytterligere med automatisert rapportering gjennom tilpassede dashbord på Intranettet. Figuren viser at LIS-systemet nå også inneholder nasjonal bransje-informasjon som er hentet tilbake fra den nasjonale databasen.

På dette utviklingstrinnet bør rapportering til offentlige registre kunne foretas direkte og automatisk i kombinasjon fra den lokale databasen og fra den nasjonale databasen. Det blir de offentlige registrene som da styrer grensesnittene for denne datakommunikasjonen. Denne rapporteringsformen vil utgjøre en betydelig forenkling både for foretakenes rapportering og det offentliges datafangst. Avgiverpunktet for offentlig rapportering flyttes og offentlig statistikk tilbys et større mangfold av informasjon inklusive prosjektrelaterte data for gjensidig nytte.

Trinn 4 – integrasjon med offentlige registre

Selv om trinn 3 kan stå som en målsetting for en prosessmodell for lokal og sentral datafangst, er modellen ytterligere videreutviklet i trinn 4, se figur 4.



Figur 4 Prosessmodell datafangst og rapportering, utviklingstrinn 4 - integrasjon med offentlige registre.

Her tenkes datafangsten til ERP-systemene via den lokale databasen ytterligere automatisert ved at manuell registrering erstattes av datafangstsystemer hvis mulig.

Hovedforskjellen i forhold til trinn 3 er at den nasjonale databasen her er den del av de offentlige registrene. Rapportering og uttrekk av bransjestatistikk til og fra den nasjonale databasen skjer da via de samme kanaler som de offentlige registrene. For foretaksledelsen betyr det konseptuelt at åpen informasjon fra flere offentlige registre lettere kan bli del av det lokale beslutningsgrunnlaget. For den nasjonale databasen betyr det at deler av datafangsten kan bli underlagt offentlig rapporteringsplikt.

De interne grensesnittene må nå i enda større grad være standardiserte og offentlige krav til grensesnitt og utvekslingsformater vil prege kommunikasjonen fra den lokale databasen. Dette vil igjen bety at ERP-systemene, kanskje også LIS-systemene, vil standardiseres, internasjonaleses og det vil gå i retning av færre og større systemer på markedet.

Vedlegg 5.

Utdrag fra en bearbeidet versjon av et innlegg²⁶;

Use of Administrative Data in Official Statistics – Past, Present and Future – With Special Reference to the Nordic Countries

Svein Nordbotten, Professor Emeritus, University of Bergen, Norway

4. Future use of administrative data

4.1 Use of administrative data in the next decade

So far, most of the administrative data acquired by NSIs for producing statistics have been data collected by central, government agencies. But there are an increasing number of public as well as private organizations, e.g. special public services, banks and insurance companies, collecting data based on the official identification system to perform their respective tasks. A central question will be if these data are wanted by and can be made available to the NSIs for producing official statistics. The answers will depend on political decisions about social benefits of statistical information and risk of misuse of sensitive data.

The Nordic NSIs are in a unique position to utilize administrative data from new sources since their official identification systems are extensively used and the NSIs have already authorized access to produce statistics from multiple administrative data sources using the official identification systems. Other organizations than the NSIs will probably request similar access to administrative data, but the probability that they will be trusted and provided with similar authorization is small due to confidentiality concerns and the NSIs' high performance. The crucial dilemma is whether information embedded in the data collected for administrative purposes should be wasted because of the potential risk for misuse or made available to NSIs for producing official statistics to increase the understanding of the complexity of our societies and contribute to higher efficiency. Extending access to new administrative data by trusted NSIs seems to represent a smaller risk for misuse than authorizing one or more other organizations to access new administrative data.

Which are then the existing data sources and which are the statistical products of interest? A number of data sources already exist which have in common that the objects they observe are identified within the official identification system and the attribute values observed are electronically recorded. These data are frequently referred to as *electronic tracks*, *electronic footprints*, etc.

A few examples of such data are:

- Data from public electronic services
- Credit card transactions data
- Tracking data by RFID technology
- Traveling data from electronic ticket records
- International migration data from electronic passports

A number of public services are currently offered on the Internet, for example registration of new addresses, income declarations, social benefits, medical services, etc. They all have in common that they save the users for time visiting the service organizations in person and the respective administrative organizations for unnecessary use of resources. Statistics on who are the users of these services will provide information on the use of the services, the time and cost savings compared with traditional service deliveries, the possibilities for service improvements, and so on.

²⁶ This chapter is based on a key note speech at the 2008 Conference of the International Association of Official Statistics, October 14, 2008 in Shanghai.

An increasing amount of commodities and services are paid by means of credit cards. Since credit cards in the Nordic countries frequently are connected to accounts identified by the owner's official identifiers, the data recorded in connection with purchases and services paid by credit cards can, when linked to other data about the users, provide a basis for more detailed statistics about private consumption and expenditures, e.g. which type of people are making the purchases in which type of shops, and at which time, how far away from their registered home address, and changes from previous periods for groups of objects.

In the last 5-10 years, a technology known as *Radio Frequency Identification (RFID)* by which object identification is stored in small, wireless chips embedded in objects that can be read by readers near the object has been expanding [Wikipedia 2009a; Miles, Sarma and Williams 2008]. The RFID chips can either be *passive*, i.e. obtain their operating energy wireless from the reader/environment, or *active* with their own energy sources. This technology is being increasingly used in many fields, for instance as a substitute for bar codes because the RFID does not require a near-by aimed optical reading device and because of low prices (about US\$ 0.10 per chip). The RFID can be used for direct identification of commodities in a basket by a reader at the cash-register, and the amount to be paid instantly computed. An electronic time-stamped record with specification of the commodities in the basket with their prices will exist, and can, if the payment is done by credit card, be associated with customer's identification. These records could give the statistician fast access to precise data about the transactions and provide the economist with the statistics she needs for dynamic consumption analysis.

Toll roads have become common also in the Nordic countries. Application of the RFID technology is used to identify cars passing toll stations if the owners subscribe to an automatic payment system. From a small device on the front screen, the RFID reader at a toll gate forward the passing car's identification to a central system in which it is time-stamped and the toll fare is subtracted from the car owner's account. In case of a private owner, the account will be associated with her personal identification number as well as the car's license number. Company cars are linked to the organizational identification. The records are saved for a certain time in case of complaints. If transferred to the NSI, the traffic data can be linked to data about the owner. These integrated statistical records can be important increments to statistics on use of cars, traffic patterns, etc.

Electronic travel tickets are frequently purchased over the Internet charging a bank account. In many cases, the bank accounts are associated with the official identification system. Access by the NSI to the electronic ticket transactions can provide a basis for statistics on air, land and sea traffic, travel patterns and migration improving the present situation considerable, particularly if access to international immigration records is also granted.

4.2 Technological trends

Technological trends point toward development of small wireless transmitting and receiving electronic devices which can be attached to or embedded in artifacts. These devices can obtain required operating power from their environment, be programmed remotely and receive instructions and transmit data interactively to communicating devices. The communication technology will vary from short range radio technology as the RFID, use of the mobile phone GSM nets, to GPS and low-orbit satellite communication [Wikipedia 2009a; IRIDIUM 2009].

A few examples indicate types of applications. Owners of cars and pleasure boats are now offered small boxes based on GPS, GSM and satellite technology that can be embedded in the car/boat and make it possible to track the whereabouts of the object. GPS is already an almost standard device for navigation in cars and boats. It is likely that in a few years the tracking box will be offered as a standard feature in new cars and boats with the option to subscribe to a tracking service storing tracking data. The system will at any time be able to provide the subscriber with electronic tracks for the object and its recent movements. Insurance companies may in the future offer reduced premiums to vehicle owners subscribing to such service. Special insurance rates depending on where and when the vehicle is used may also be offered if the owner grants the insurance company access to the tracks of his vehicle. The tracking data will most probably be connected to official identifications of the vehicle and owner, and can become another new data resource for statistics on use of cars and boat as well as traveling statistics.

Corresponding devices exist for humans. In some countries, criminals can serve their prison sentences at home with restricted mobility. The condition is that the person serving is willing to wear an electronic *chain* around his leg. The chain contains a GPS unit tracking the person's movements supplemented by a wireless transmitter reporting to a center controlling that he abides to the restrictions.

In hospitals, patients in intensive care can be monitored even when they are allowed to move out of bed by means of devices they wear and which communicate with a hospital control center by means of a wireless local net. Out-patients needing frequent control can be equipped with and carry similar types of devices communicating by means of GSM relieving the patient from staying in a hospital and saving the hospital from spending unnecessary resources. Tracking devices are also offered to help locate old people having problems with losing their way back home.

We have become accustomed to the idea that we always carry our finger prints with us as identification. Biometric passports for entry into a number of countries are recently introduced [EU Council Secretariat 2007]. To serve their purpose, the read identifications for people passing a border will be stored for some time. Advanced coding/decoding algorithms are used to reduce the risk for identity thefts.

Applications are not limited to devices installed or worn. Unique serial numbers have been *embedded* as part of the cars for many years. *Implantable Medical Devices, IMDs*, are, as the name indicates, *implanted* in human patients in connection with serious illnesses. The pace makers for heart function control are a well-known example. The IMDs, which are programmable, allow hospitals remotely to extend care to their out-patients [Fu 2009]. *VeriChip* has already got their first chip for implantation approved in USA [Wikipedia 2009b].

Will implantation in humans of chips emitting personal identifications also be widely offered in the future? With an identity chip implanted, your driver's license, passport, bank account, etc. can in principle be checked when passing a reader with the required connection to the respective administrative system. In addition, the geographic position of the reader with a time-stamp can be stored with the rest of the data.

Implanting a chip in the human body could eliminate the need for carrying identity papers. As indicated by Figure 5, it would for example be possible that personal identities and those of the commodities carried in shopping basket being read automatically and the customer's account charged when passing an exit reader in the supermarket. With implanted chips, immigration officers at the border could be substituted by a portal screening your identity, your driver's license could be checked even if you had left it at home, and the patient's medical history be automatically retrieved when a patient is brought through the entrance of the hospital emergency room.

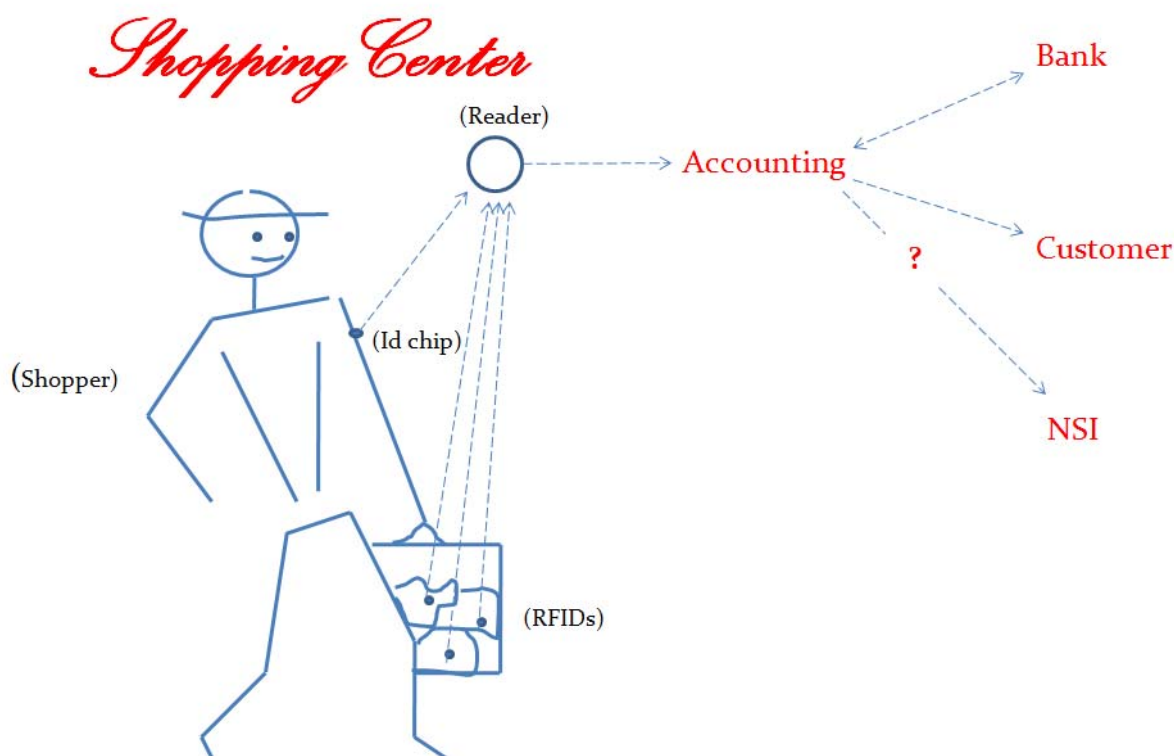


Figure 5: A future shopper passing the cash reader

These applications have few things in common: they generate huge masses of data about individuals over time, they will probably be based on the official identification system and most of the data read will be considered sensitive. We may not like it and oppose the applications, but the technology exists and increased use will most likely continue. Because of its efficiency, experience indicates that the technology will be used and misused. But access to an analysis of such data can also revolutionize the conditions for certain aspects of social, economic and behavioral analysis and research, and provide trade and industry with data for more efficient operations. Obviously, access to the data will be considered very desirable, and the idea of National Data Centers is likely to surface again followed by new intensive privacy debates [Albrecht and McIntyre (2005)].

We shall refer to data discussed above as *Electronically Observed Data*, *EOD*, in contrast to the traditional data that are usually recorded with an active participation of the objects.

4.3 Implications for official statistics²⁷

Which implication may the increasing mass of micro data collected have for official statistics? What are the limits of the official statistics? Should an NSI refrain from using this growing stream of EOD in producing official statistics, or should it take advantage of the data to be able to provide the public with more information about the society? What would be an acceptable solution for balancing between the benefits of using these data to improve our knowledge of the society and the serious risk for misuse of the data? Is it likely that commercial interests will also see the value of and be interested in exploiting the data for commercial use? Should the NSIs engage actively in the development of a system that may reduce the risk for misuse to an acceptable level, and prepare themselves for the responsibility of being caretakers of EOD? Is it possible to design a legal framework regulating the use of sensitive data in a satisfactory manner, and can such regulations be effectively enforced? These are only a few of many questions official statisticians will have to consider in preparing their strategies for the future.

Two main future alternative scenarios can then be envisaged, one in which the NSI is not interested in nor entitled to access and use the EOD of administrative organizations, and a second scenario in which the NSI is permitted by a Statistical Law to utilize administrative EOD similar to the present access to administrative data. The 2 alternatives are outlined in Figures 6 and 7.

²⁷ The discussion in this section is limited to data stored with direct or indirect reference to the official identification system. In addition to data satisfying this requirement, there will be other data collected and stored for special purposes which may also be of interest to NSIs.

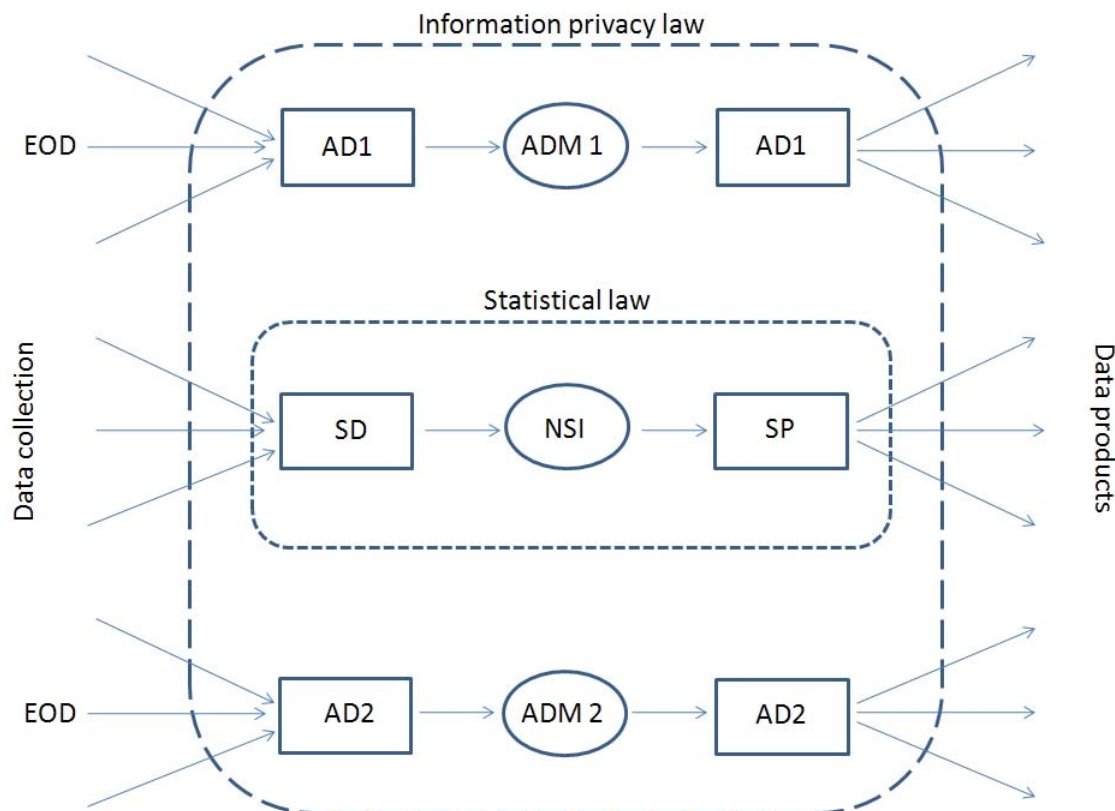


Figure 6: Organization without NSI engagement in the new administrative data collection

Figure 6 indicates that according to the first scenario, administrative organizations, ADM 1 and ADM 2, will record EOD about objects for their administrative purpose, and probably as far as the Privacy law permits, also make the data available for third party users. If the NSIs decide not engage in the use of EOD, official statisticians cannot be blamed for using administrative data considered too private or sensitive to be used for statistics. To collect statistical data comparable to EOD from administrative sources would be an expensive alternative for the NSIs if official statistics for the corresponding events would be required.

If the NSI decides that the use of the EOD collected for administrative sources is a wanted extension of the current use of administrative data for preparation of statistical products, and is authorized by a Statistical Law to access and use administrative EOD as indicated in Figure 7, the NSI will, because of the official identification system, be able to integrate the EOD from different sources with its own data and provide better and more comprehensive statistical products without the costs of repeated observation of the objects.

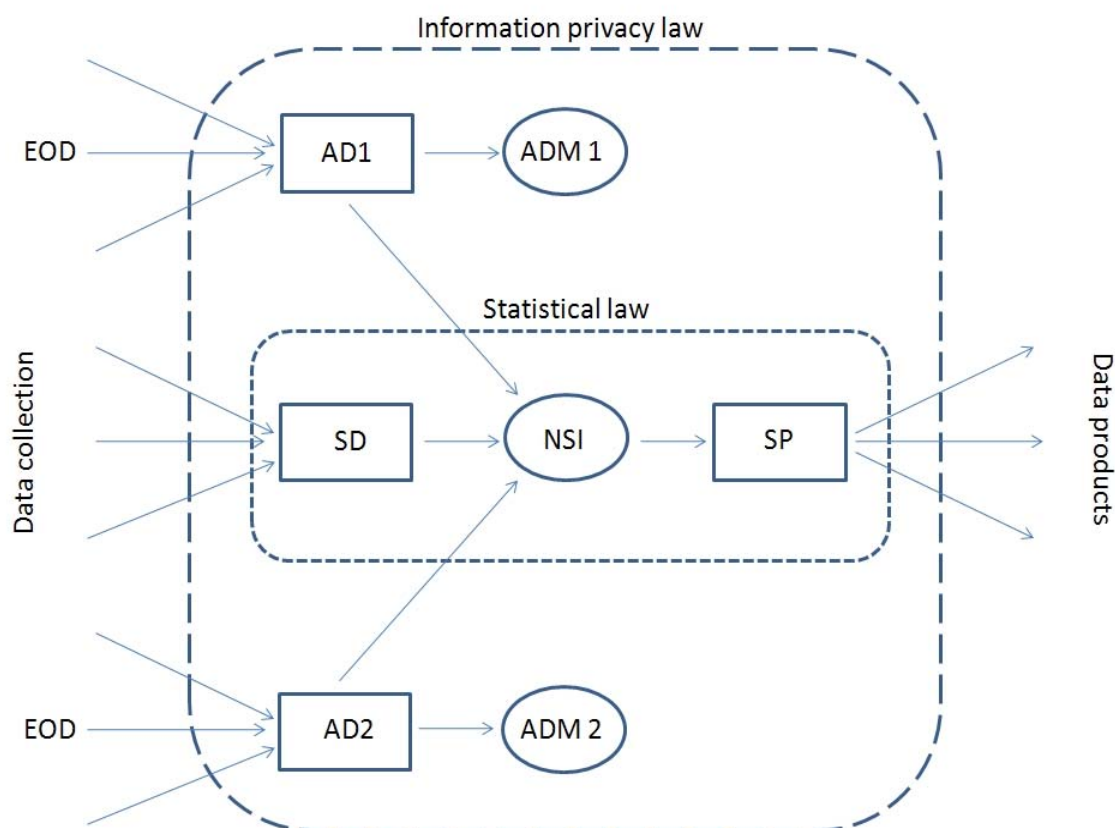


Figure 7: Organization with NSI engaged in application of the new administrative data collection

In support for engagement of NSIs as caretakers of EOD warehouses, the main arguments are that the EOD are too valuable to be wasted, the NSIs' are well established with reputation for safe handling of confidential data, and the NSIs have statistical and technical competence to obtain more useful information from the data than any other organization. The second scenario would require that present statistical and privacy legislations are extended and adjusted to regulate the future situation for safe handling of EOD. The statistical legislation should give an exclusive right to the NSI for accessing, organizing and processing any data collected by other operators using the official identification system directly or indirectly. The NSI should be charged with responsibilities for guarding the data against any access which can lead to misuse as specified by the legislation. The requirements would generate new methodological and technical challenges for the NSIs. In addition to the improved statistical methods required, technical methods for safeguarding the data against unauthorized use are the most important.

Since the National Data Center proposal in the 1960s, requests from researchers for access to micro data sets for research have been increasing. Also an increasing number of business users want access to micro data sets for applying their own analytical methods to solve their particular problems, or presenting statistics for new users. Several approaches to by-pass the confidentiality problem have been evaluated and tested. Among the approaches tried are letting NSI staff members carry out the necessary processing according to the researchers' prescriptions, hiring the researchers as unpaid NSI staff members and as such subjected to the statistical laws, permitting the researchers to work with the micro data within closed and regulated processing systems, generating synthetic micro data set versions with the same distributional characteristics as the real data set, etc.

A challenge for the keepers of the EOD warehouse would be to develop practical and safe ways to avoid that data and statistics for which they are responsible could be available and used in conflict with the privacy and statistical legislations. The general public must be informed and convinced that the developed solutions are effective and secure. It is important to be prepared for the future, and have a strategy for how to be prepared for the development. Some possible future technological trends of significance, some problems and possible solutions have been indicated. It remains to analyze these in detail, draw the necessary conclusions and wait for necessary political decisions.

To be prepared, the NSIs should put scenarios for possible future use of administrative EOD in production of official statistics on their agendas for discussion including requirements to an adequate legislative framework, and development of methods needed to take care and safe use of the EOD.

Vedlegg 6.***Datafangst lønnsstatistikk***