

Personstatistik i Danmark

**Et registerbaseret
statistiksystem**

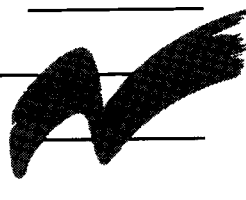
**Statistics on persons in Denmark
A register-based statistical system**

DANMARKS STATISTIK

28 AUG. 2014

BIBLIOTEKET

DANMARKS
STATISTIK



Personstatistik i Danmark Et registerbaseret statistiksystem

Udgivet af Danmarks Statistik

Oktober 1994

ISBN 87-501-0908-1

Pris: 192,00 kr. inkl. 25% moms

Oplag: 600

Danmarks Statistiks trykkeri, København

© Danmarks Statistik 1994

Enhver form for hel eller delvis gengivelse eller mangfoldiggørelse af denne publikation, uden skriftligt samtykke fra Danmarks Statistik, er forbudt efter gældende lov om ophavsret.

Undtaget herfra er citatretten, der giver ret til at citere, med angivelse af denne publikation som kilde, i overensstemmelse med god skik og i det omfang, som betinges af formålet.

Forord

Personstatistik i Danmark - Et registerbaseret statistiksystem er en samlet redegørelse for den danske registerbaserede statistik med hovedvægten lagt på personstatistikken. Der er tale om et sammenhængende system, der gradvis er udviklet siden begyndelsen af 1970'erne.

I bogen gøres der rede for systemets udvikling samt for de tanker, der ligger bag dets nuværende opbygning. Registerlovgivning, formidling af registerstatistik og dokumentation tages op, ligesom problemer i forbindelse med datakvalitet diskuteres. Endvidere findes der i bilag 2 en oversigt over de eksisterende statistikregistre.

En tidligere beskrivelse af registerstatistikken blev udgivet i 1982. Meget er sket siden da, og der har derfor i flere år været et udtalt behov for en opdatering.

Bogen er blevet til på grundlag af et seminar, som blev holdt af Danmarks Statistik i januar 1994. Her deltog godt 30 af de personer i Danmarks Statistik, som har været involveret i udviklingen af den registerbaserede statistik. Desuden medvirkede enkelte eksterne deltagere med særlige erfaringer, først og fremmest i udnyttelsen af registerstatistikken til forskning. Deltagerlisten fremgår af bilag 7. Danmarks Statistik ønsker hermed at takke alle deltagerne for deres bidrag.

Den foreliggende bog er redigeret og sammenskrevet af en mindre redaktionsgruppe bestående af konsulent Finn Spieker (formand), fuldmægtig Marius Ejby Poulsen, fuldmægtig Anna Qvist, kontorchef Carsten Torpe og afdelingschef Lars Thygesen. Marius Ejby Poulsen har fungeret som redaktør.

Danmarks Statistik, august 1994

Hans E. Zeuthen

/Lars Thygesen

Indholdsfortegnelse

1.	Kravene til den officielle statistik	9
2.	Administrative registre	15
2.1	Generelle betragtninger	15
2.2	Forskellige typer af administrative registre	16
2.3	Personregistrering	17
2.4	Ejendomsregistrering	21
2.5	Erhvervsregistrering	22
2.6	Udviklingen i anvendelsen af administrative registre som statistikkilde	23
2.7	Forskellige typer registerdata	24
2.7.1	Identifikationsdata	24
2.7.2	Tidsreferencer	26
2.7.3	Basisdata	27
2.7.4	Specialdata	27
2.7.5	Baggrundsdata	28
2.7.6	Notatdata	28
2.8	Statistikdata	29
2.9	Konklusion	29
3.	Det samlede statistiksystem	31
3.1	Statistiksystemets opbygning	31
3.1.1	Klassifikationsmoduler	33
3.1.2	Statistikregistre	35
3.1.3	Strukturen i det samlede statistiksystem	38
3.2	Samkøring af registre	39
3.2.1	Indholdsanalyse	40
3.2.2	Dannelse af statistikregistre	42
4.	Metode og kvalitet	51
4.1	Administrative data som statistikdata	51
4.1.1	Indhold	51
4.1.2	Identifikationer	53
4.1.3	Tidsreferencer	54
4.1.4	Registret som statistisk model	55
4.1.5	Stabilitet	56
4.1.6	Kvalitet	56
4.1.7	Ændringer i de administrative data	58
4.1.8	Statistikernes mulighed for at påvirke data	60
4.1.9	Dannelse af statistikregistre	61
4.2	Anvendelse af flere kilder	68
4.2.1	Dannelse af en ny variabel	68
4.2.2	Prioritering af oplysninger fra flere kilder	70
4.2.3	Korrektion af oplysninger	72
4.2.4	Anvendelse af kombinerede data over tid	73
4.2.5	Anvendelse af flere kilder som metode	74
4.3	Integreret dataindsamling	75

4.3.1	Lovgrundlaget.....	75
4.3.2	Eksempler på integreret dataindsamling.....	76
4.3.3	Fordele og ulemper ved integreret dataindsamling.....	79
4.3.4	Sammenfatning.....	81
4.4	Imputering.....	81
4.4.1	Imputering af oplysninger, som helt mangler.....	81
4.4.2	Imputering af delvist manglende oplysninger.....	83
4.4.3	Manuel eller automatisk imputering.....	87
4.5	Surveys og registre.....	88
4.5.1	Surveydata suppleret med registerdata.....	89
4.5.2	Registerdata suppleret med surveydata.....	90
4.5.3	Sammenligning af register- og surveydata, kvalitetsanalyser.....	92
4.6	Afslutning.....	96
5.	Integrationsregistre.....	99
5.1	Behovet for integration.....	99
5.2	Horisontal integration.....	100
5.3	Vertikal integration.....	106
6.	Personstatistikens samspil med andre statistikområder.....	121
6.1	Status for enhedsrelationer.....	121
6.1.1	Relationen person/bolig.....	122
6.1.2	Relationen person/arbejdssted.....	122
6.1.3	Relationen arbejdssted/erhvervsarealenhed.....	123
6.1.4	Relationen mellem ejendomme og ejer.....	124
6.2	Anvendelse af fælles data.....	124
6.3	Begrebernes validitet.....	124
6.4	Perspektiv.....	127
7.	Registerlovgivning og datapolitik.....	129
7.1	Registerloven.....	129
7.1.1	Registerforskrifter.....	131
7.2	Datasikkerhedsorganisationen.....	132
7.3	Danmarks Statistiks diskretionspolitik.....	132
7.4	Reaktioner fra statistikbrugere.....	135
7.5	EU-direktiv om personbeskyttelse.....	136
8.	Formidling og beredskab.....	141
8.1	Bøger.....	142
8.2	Ad hoc service.....	142
8.3	Standard-servicesystemer.....	142
8.4	Databanker.....	143
8.5	Lovmodellen.....	143
8.6	Forskningsdatabaser.....	143
8.7	Elektroniske publikationer.....	144
9.	Dokumentation.....	145
9.1	Status.....	145
9.2	Data i centrum.....	145

9.3	Hvorfor dokumentation.....	146
9.4	Hvornår skal dokumentationen skabes	147
9.5	Hvem anvender dokumentationen	147
9.6	Hvad skal dokumenteres.....	148
9.7	Tidsdimensionen	151
9.8	Datakvalitet og informationskvalitet.....	152
9.9	Registrets anvendelse	152
9.10	Specielt om det personstatistiske registersystem	153
9.11	TIMES-2000	153

Bilag:

1.	Litteraturoversigt.....	155
2.	Registre i Danmarks Statistik	161
3.	Milepæle i registerstatistikken siden 1968.....	207
4.	Oversigt over identifikationer	209
5.	Anvendte begreber og forkortelser	215
6.	Administrative kilderegistre anvendt i Danmarks Statistik	219
7.	Deltagerliste for seminar over emnet »Det personstatistiske registersystem«, 10.-13. jan. 1994 på Gentofte Hotel.....	225
	Stikordsregister	227

Kapitel 1. Kravene til den officielle statistik

Hvordan kan man bedst tilrettelægge en officiel statistik baseret på administrative kilder, og er det i det hele taget fornuftigt? En diskussion af disse spørgsmål må tage udgangspunkt i den opgave, statistikken skal løse. Opgaven er meget sammensat, fordi den officielle statistik er en del af et lands *generelle* infrastruktur, og den benyttes til en lang række forskellige formål.

Den officielle statistik skal give en bred og dybtgående beskrivelse af samfundet. Den skal som minimum dække følgende: Befolkningsforhold, sociale forhold, levevilkår, erhvervsforhold, offentlig økonomi og den nationale økonomi¹. Statistikproduktionen skal indrettes således, at den til enhver tid belyser de væsentlige træk i samfundsrelevante forhold.

Brugerne af denne information er i første række *offentligheden*. Statistikken skal give et godt grundlag for den generelle samfundsdebat, som den finder sted i hjemmene, på arbejdspladserne, i de politiske organisationer, osv. Til denne brugergruppe kan man også henregne undervisningen i skoler og på højere læreanstalter. Den vigtigste informationskanal for statistikken til offentligheden er nyhedsmediernes; aviser, radio og TV. Derfor er det vigtigt for Danmarks Statistik at offentliggøre statistikker, der bliver citeret og viderebragt af medierne. Det stiller både krav til indholdet og til den måde, nyhederne bringes på. En del af offentligheden vil dog også være direkte brugere af de officielle statistiske publikationer.

En gruppe af brugere af den officielle statistik, som altid har været anerkendt som uhyre vigtig, er *lovgivningsmagten*. I nogle lande har det endda været sådan, at kun de krav til statistikken, som blev stillet af regering og parlament, blev tillagt betydning. *Lov om Danmarks Statistik*² gør det klart, at dette ikke er tilfældet i Danmark. Men lovgivningen stiller på mange områder krav om meget fyldige og detaljerede statistiske data, som skal gøre det muligt at bedømme, hvilke behov for lovgivning der er, og hvordan påtænkte og vedtagne love vil virke. Disse krav bliver mere og mere udtalte, efterhånden som lovgivningen på det skattemæssige og det sociale område bliver stadig mere kompliceret.

Den officielle statistik bruges også til andet end lovgivningsformål af *den offentlige administration*. Den bruges af staten, kommunerne og

Statistikens formål er at beskrive samfundet...

...til brug for offentligheden...

...lovgivningsmagten

...den offentlige administration...

¹ Jf. *Danmarks Statistik (1991e)*

² Jf. *Økonomiministeriet (1992a)*.

amtskommunerne som basis for planlægning af offentlige aktiviteter, samt til opfølgning og resultatvurdering af disse.

...erhvervslivet...

Et tilsvarende behov findes i *det private erhvervsliv*, som har brug for at få belyst markedsmæssige muligheder, konkurrenceforhold, m.m.

...forskningen...

Endelig er *samfundsforskningen* en vigtig brugergruppe. Den omfatter såvel de sociale videnskaber som grænseområder inden for medicin og naturvidenskaber.

...og internationale organisationer

Alle de nævnte grupper er velkendte på det nationale informationsmarked, men hertil kommer, at også *internationale organisationer* har brug for statistik. De internationale organisationer skal tilfredsstille de samme typer af behov, som allerede er nævnt, idet de skal betjene offentligheden, virksomhederne og andre, som ønsker at kunne foretage sammenligninger af forholdene fra land til land. De internationale behov er omfattende og bliver i disse år i Europa stadig mere påtrængende. EU's forordninger og direktiver på statistikområdet er bindende for medlemslandene, og de krav, der følger heraf, skal under alle omstændigheder opfyldes.

Der er således en række grupper, der har brug for den officielle statistik, og som derfor stiller krav til den. Men der er også andre interesser; dem der skal finansiere statistikken, og dem der skal levere oplysninger til den.

Styrelsen bestemmer indholdet...

I Danmark er Danmarks Statistik den centrale statistikmyndighed³. Institutionen ledes af en uafhængig styrelse, som fastlægger arbejdsplanen inden for de budgetmæssige rammer, der fastlægges i Finansloven. Dette er en prioriteringsopgave, hvor de forskellige brugergruppers behov skal afvejes over for hinanden, idet de konkurrerer om ressourcerne. Juridisk bindende krav til statistikken skal naturligvis forlods opfyldes, herunder de krav, som EU stiller. I prioriteringen skal der også tages hensyn til respondenterne, som normalt må antages at føle det som en byrde, at der skal afgives oplysninger til statistikken. Der ligger en særlig forpligtelse for Danmarks Statistik til at minimere respondentbyrden og til at maksimere effektiviteten.

...og får hjælp fra rådgivende udvalg

Til støtte for prioriteringsprocessen har styrelsen nedsat en række rådgivende udvalg med deltagelse af både brugere af statistikken og oplysningsgivere, hvor hvert udvalg typisk beskæftiger sig med et bestemt statistikområde. Enige indstillinger fra disse udvalg har naturligvis stor vægt, men det er styrelsens opgave at påse, at programmet kan rummes indenfor de økonomiske rammer.

³ Jf. Økonomiministeriet (1992a).

I de fleste lande indsamles data til den officielle statistik helt overvejende ved, at man spørger de enheder, som statistikken omhandler; personer, virksomheder, kommuner. Normalt bruger man mere eller mindre ensartede skemaer, som respondenterne enten selv skal udfylde og indsende, eller som udfyldes af en interviewer. Man kan spørge alle de enheder, der skal indgå i statistikken (totaltælling), eller man kan nøjes med en del af dem, som udvælges i overensstemmelse med forskellige videnskabelige principper, hvormed det sikres, at de repræsenterer hele populationen (stikprøveundersøgelser). Alle disse typer af undersøgelser går under fællesbetegnelsen *surveys*, og der findes en meget omfattende litteratur om, hvordan de kan eller bør tilrettelægges, og om deres fejlkilder og pålidelighed.

**Surveys
repræsenterer
traditionel
dataindsamling**

Det er et fælles træk ved surveys, at de belaster de medvirkende, og de giver i mange lande anledning til stor modvilje. I nogle lande (fx USA) har man haft parlamentariske udvalg, der fik til opgave at stille forslag om, hvordan man kunne nedbringe respondentbyrden. Både som følge af arbejdet med at give oplysninger og på grund af respondenternes frygt for misbrug af data forekommer der altid et vist bortfald i surveys, således at man ikke får svar fra alle de adspurgte. Dette er en af de mest alvorlige fejlkilder ved metoden, idet erfaringen viser, at de, der ikke svarer, altid adskiller sig systematisk fra dem, man opnår svar fra. I næsten alle undersøgelser, hvor man spørger personer, er der fx langt større bortfald blandt arbejdsløse og blandt byboere⁴.

**Problemer ved
surveys**

En anden måde at fremskaffe statistikkens grunddata på er at udnytte allerede indsamlede oplysninger. En meget rig kilde for den officielle statistik er de data, som den offentlige administration indsamler om borgere og virksomheder. Det er ikke noget nyt, at disse oplysninger udnyttes til statistik. Nogle af de ældste statistikker fremkom på denne måde. Det gælder således de tidligste danske statistikker om befolkningsudviklingen, som byggede på kirkebøgerne.

**Registerstatistik er
ikke ny...**

I de sidste 25 år er der imidlertid sket en dramatisk ændring i statistikproduktionen i Danmark, idet en meget stor del af statistikken efterhånden baseres på edb-registre, der føres af den offentlige administration. Dette hænger naturligvis sammen med, at administrationen også har undergået voldsomme ændringer, og der er siden etableringen af Det centrale personregister (CPR) i 1968 opstået en lang række registre, som alle benytter nogle få fælles identifikationer for de enheder, administrationen retter sig imod, dvs. borgere, virksomheder, boliger og ejendomme. På denne måde kan de forskellige registre kombineres og skabe et nuanceret billede af samfundet.

**...men i kraftig
udvikling**

⁴ Jf. fx *Nordisk Statistisk Sekretariat (1981a)*.

Lov om Danmarks Statistik var fremsynet

Når det har været muligt for Danmarks Statistik at udnytte de nye muligheder i så høj grad, skyldes det, at Lov om Danmarks Statistik indeholder en stærk hjemmel, der gør det muligt at indhente oplysninger fra de offentlige myndigheder. Da loven blev udformet i 1966, var forberedelsen af de nye registre begyndt, og lovgiverne så klart de kommende registres muligheder til statistikformål. Dette fremgår af lovens bemærkninger, som omtaler mulighederne for at nedbringe respondentbyrden og endda komme så vidt, at en folketælling i fremtiden måske kunne gennemføres helt uden spørgeskemaer, udelukkende gennem brug af registrene. Lovgiverne har her været meget fremsynede, idet disse tanker først kunne realiseres fuldt ud i 1981.

Stor international interesse

Interessen for den registerbaserede statistik har naturligvis ikke kun gjort sig gældende i Danmark. Den har været livligt diskuteret i det internationale samarbejde siden slutningen af 1960'erne. Blandt de førende lande har fra tidlig tid været Norge⁵ og Sverige, og registerstatistikken har været et fast punkt på nordiske statistikermøder og chefstatistikermøder. I det europæiske samarbejde i EU har der været megen diskussion om spørgsmålet, navnlig i forbindelse med overvejelser om at basere folke- og boligtællinger på registre⁶. Også i USA har interessen været stor⁷.

Diskussionen har afspejlet en blanding af optimistiske forventninger og skepsis⁸. Internationalt har det fremherskende synspunkt været, at der er betydelige begrænsninger ved registerstatistik, hvilket har ført til, at det kun er ganske få lande i verden, der er gået så vidt som Danmark i retning af at basere den officielle statistik på registre.

Indvendinger mod registerstatistikken

De indvendinger, der især har været rejst mod den registerbaserede statistik, og som vil blive diskuteret indgående i denne bog, er:

- *Krænkelser af privatlivets fred.* Registerstatistikken forudsætter, at statistikbureauet opbevarer en stor mængde oplysninger om personer og virksomheder, hvilket skaber en risiko for misbrug, dvs. brug af oplysningerne til andre formål end fremstilling af anonyme statistikker.
- *Folkelig modvilje,* som hænger sammen med ovennævnte punkt. Uanset om der er en reel risiko for misbrug, kan modviljen blive så stærk, at data ikke kan fremskaffes. Dette og det foregående punkt diskuteres i kapitel 7.
- *Datakvalitet.* Der argumenteres for, at registrene ikke indeholder lige præcis de begreber, der er relevante for statistikken, og at data

⁵ En banebrydende artikel er således *Norbotten (1967)*.

⁶ Jf. *Redfern (1987)*.

⁷ Jf. fx *National Center for Health Statistics (1980)* og *US Department of Commerce (1980)*.

⁸ Jf. *Lange (1993)*.

ikke er pålidelige. Grunden er, at statistikerne ikke kan styre indholdet i registrene, som de er vant til fra deres egne undersøgelser (jf. kapitel 4).

I Danmark har disse indvendinger naturligvis også været fremsat, men holdningen har gennemgående været langt mindre pessimistisk. I midten af 1970'erne besluttede Danmarks Statistik således, at det fremover var en strategisk målsætning at udvikle et sammenhængende system for personstatistikken, baseret på data fra administrative registre. Strategien indebar, at man skulle arbejde for at få udbygget de administrative registre med nogle få oplysninger, der var vigtige for statistikken. Et succesmål for strategien var, at folke- og boligtællingen 1981 skulle gennemføres med fuldt informationsindhold og uden brug af spørgeskemaer. Dette mål blev nået. Senere gennemførte også Finland i 1990 en registerbaseret folke- og boligtælling.

Den danske registerstrategi

Den danske strategi har betydet, at man i hvert fald på det personstatistiske område har udviklet et sammenhængende system, hvor rygraden er oplysninger fra registrene. Ikke alle databehov kan imødekommes ved træk fra registre. Fx er der ikke mulighed for at fremskaffe bløde data om holdninger fra registrene, og oplysninger til brug for statistik vedrørende disse må derfor indhentes ved hjælp af surveys, der indgår som en supplerende del af det personstatistiske system.

Det har været afgørende for udviklingen i Danmark, at lovgivningsmagten fra starten har stillet sig meget positivt til, ja ligefrem forudsat en sådan udvikling. De stramme økonomiske vilkår for Danmarks Statistik har yderligere tilskyndet til valget af denne strategi⁹.

⁹ Jf. *Nordisk Statistisk Sekretariat (1981b)*, kap. 4.

Kapitel 2. Administrative registre

I dette kapitel behandles opbygningen og anvendelsen af de administrative registersystemer med særlig vægt på personregistreringen. Endvidere omtales de typer af oplysninger, der indgår i disse registre samt deres egnethed som statistikdata.

2.1 Generelle betragtninger

Betegnelsen register anvendes i mange forskellige betydninger. I bredeste forstand omfatter registerbegrebet enhver logisk afgrænset samling af data, herunder også datasæt beregnet til rent midlertidigt brug. Begrebet register må dog i denne fremstilling defineres mere snævert. Et *register* er en systematisk samling af oplysninger, der kan henføres til enkeltheder på en sådan måde, at opdatering er mulig. *Opdatering* af et register er en behandling af identificerbare oplysninger med det særlige formål at oprette, ajourføre, korrigere eller udbygge registret, således at det kan videreføres som register.

Registerbegrebet

Et register vil som hovedregel indeholde informationer om en bestemt gruppe af enheder (fx personer, firmaer), hvor enhederne er afgrænset af et præcist regelsæt (fx "bopæl i Danmark"), og hvor indholdet ajourføres i takt med de ændringer, enhederne undergår.

Et register må altså omfatte et system til indrapportering af dataændringer. Ajourføringen kan finde sted med større eller mindre forsinkelse i forhold til den hændelse, der forårsager ændringen. I nogle tilfælde kan indrapporteringen ikke ske i direkte tilslutning til hændelsen, således at man i stedet må nøjes med at konstatere de korrekte data med visse tidsintervaller. Det gælder således nogle af de data, der findes i Bygnings- og boligregistret (BBR). Data, der slet ikke kan ajourføres systematisk, betragtes ikke som egentlige registerdata.

Den anførte definition af registre refererer udelukkende til indholdet af de datasamlinger, der er tale om. Registerenes oplysninger kan opbevares på forskellig måde. Et register kan eventuelt opdeles i flere delregistre af praktiske årsager, og der kan anvendes manuelle og/eller maskinelle processer ved vedligeholdelsen, men disse forhold spiller ingen rolle for registerbegrebets indhold.

Det må dog fremhæves, at de mere omfattende administrative registre, som gennem de seneste 10-20 år har været fundamentet for statistikproduktionen i Danmark, alle drives med anvendelse af elektronisk databehandling, og de erfaringer, som skal fremdrages her, vedrører derfor næsten udelukkende edb-registre.

2.2 Forskellige typer af administrative registre

De administrative registre kan grupperes efter deres formål i basisregistre og målrettede registre.

Basisregistre

Ved basisregistre forstås registre, der drives med det formål at ligge til grund for den offentlige administration i almindelighed eller i hvert fald for flere forskellige administrationsgrene. Disse registre har typisk til formål at holde styr på bestanden af enheder, dvs. registrere enhedernes fødsel og død og dermed også hvilke enheder, der til enhver tid er til stede i populationen. Desuden skal de vedligeholde identifikationsoplysningerne, som bruges af de andre administrative registre, og de skal indeholde nogle få *basisdata*, som er af fælles interesse for forvaltningerne.

Basisregistre skal være tilgængelige for flere forvaltninger. Adgangen kan ske på flere måder, fx ved at opdateringer vedrørende de enheder, der er relevante for en forvaltning, automatisk videregives til forvaltningens registre fra basisregistret. I de sidste 10 år er det blevet almindeligt i Danmark, at forvaltningerne har on-line adgang til basisregistre. Det er også almindeligt, at opdateringer til basisregistre kommer fra en lang række forskellige myndigheder, hvorved datakvaliteten kan blive usædvanlig høj. Hertil kommer, at basisregistre ideelt set ikke er påvirket af én bestemt forvaltnings synsvinkel; de tegner så at sige et neutralt billede af enhederne. Basisregistre har derfor også den allerstørste betydning for statistikken.

Som typiske eksempler på basisregistre kan nævnes Det centrale personregister (CPR), Bygnings- og boligregistret (BBR) og Det centrale erhvervsregister¹⁰. En nærmere omtale af registrene findes i bilag 6.

Organisatorisk kan basisregistre være placeret på forskellige måder: Enten hos en "uafhængig" myndighed, som har til opgave at føre registret, eller hos en myndighed, der er den primære bruger. Set fra statistikens synspunkt er det ideelle den uafhængige placering, som kendes fra CPR, hvor det centrale ansvar er placeret i et kontor i Indenrigsministeriet. CPR ajourføres i meget høj grad af kommunerne, og de fleste kommuner har et selvstændigt folkerregister til denne opgave. Administrationen på det lokale niveau vedrørende BBR sker derimod oftest i teknisk forvaltning, som er en stor databrunder.

Målrettede registre

Til forskel fra basisregistre tjener de målrettede registre kun ét eller en velafgrænset gruppe af formål. Registerne føres af den myndighed, som skal bruge oplysningerne, og dataindholdet farves naturligt heraf. De målrettede registre modtager ofte oplysninger om populationen og om basisdata fra et basisregister, men fører selv andre data til.

¹⁰ Også benævnt Erhvervsregistret.

2.3 Personregistrering

Danmark er blandt de lande i verden, som i dag har de mest udbyggede, integrerede og velfungerende administrative edb-systemer. Som nævnt i kapitel 1, startede udviklingen tidligt på personregistreringsområdet, og med udgangspunkt heri foregik der gennem 1970'erne og 1980'erne en bevidst og omfattende opbygning af landsdækkende, administrative systemer.

I løbet af 1800-tallet indførte mange vestlige lande en løbende og systematisk registrering af borgernes forhold omkring bopæl og ægteskabelig stilling i såkaldte personregistre. Sådanne registre blev ikke dengang oprettet i Danmark, men i begyndelsen af 1900-tallet opstod der som følge af væksten i den offentlige administration og stigningen i de kommunale og statslige udgifter, et øget behov for pålidelige oplysninger om borgernes opholdssted. Således viste det sig vanskeligt for kommunerne at inddrive alimentationsbidrag og at opkræve skatter, fordi man ikke i tilstrækkeligt omfang vidste hvor skatteyderne befandt sig. Der var desuden et mere specielt behov for at etablere en effektiv administration af rationeringsordningerne efter 1. verdenskrig. Man blev også opmærksom på, at personregistre kunne tjene andre formål. Således ville de være egnede ved politiets eftersøgning af personer, som grundlag for udskrivning af valglister og som grundlag for befolkningsstatistik.

Folkeregistrene og CPR

I 1924 vedtog Folketinget derfor *Lov om Folkeregistre*, som bestemte, at alle landets kommuner samme år skulle oprette et folkeregister, dvs. et kartotek indeholdende oplysninger om alle personer, der havde bopæl i kommunen, uanset om de måtte være midlertidigt fraværende.

Folkeregistrene skulle indeholde identifikationsoplysninger såsom stilling og navn, fødselsdag og -sted, bopælsadresse, familieforhold og statsborgerskab. Disse grundoplysninger blev indhentet gennem folketællingen i 1924. Kommunerne skulle løbende ajourføre kartotekerne, idet de skulle modtage oplysninger om fødsler, dødsfald, vielser, skilsmisser, m.m. fra forskellige myndigheder, mens borgerne selv blev forpligtet til at indberette flytninger samt ind- og udvandring direkte til folkeregistret.

I den debat, der gik forud for folkeregisterlovens vedtagelse, indgik også forslag om et landsregister, et manuelt drevet CPR i form af et kæmpekartotek, som skulle være en sammensmeltning af de kommunale folkeregistre omfattende hele Danmarks bosiddende befolkning. Dette forslag nød ikke fremme, da man forudså en enorm administrativ byrde.

Folkeregistrene blev i de følgende år gradvist udbygget som hjælpemiddel for den kommunale og den øvrige offentlige administration.

Der blev indført formularsæt til bl.a. skattemyndigheder og sygekasser, således at disse myndigheder kunne føre egne personregistre. Efterhånden anvendtes hulkortteknik ved løsning af periodevise masseopgaver, såsom skatteberegning og udskrivning af valglister. Til løsning af disse opgaver oprettedes kommunalt ejede, regionale hulkortcentraler. Anvendelsen af folkeregistrene var imidlertid i begyndelsen af 1960'erne efterhånden udbygget så kraftigt, at systemet var ved at nå en kapacitetsgrænse, som nødvendiggjorde en strukturel og teknisk ændring.

Det centrale personregister CPR oprettes

Med folkeregistreringsloven af 10. juni 1968¹¹ gennemførtes en afgørende reform, som havde til formål at forenkle og effektivisere den samlede offentlige personregistrering ved anvendelse af edb-teknik. Ved siden af de kommunale folkeregistre, som fortsatte med at fungere, oprettedes Det centrale personregister CPR, et landsdækkende magnetbåndregister over den danske befolkning.

Personnummeret introduceres

En afgørende del af reformen var indførelsen af et fast identifikationsnummer for hver enkelt borger, *personnummeret*. Dette nummer betragtedes som en praktisk nødvendighed for driften af Det centrale personregister. Nummeret skulle tillige indføres overalt i den offentlige administration og således afløse de talrige nummersystemer, som tidligere havde været anvendt af forvaltningsgrenene.

Adressekode også af betydning

I forbindelse med etableringen af CPR blev en række andre standarder og koder fastlagt, fx kommunekoder, sognekoder, stillingskoder og statsborgerskabskoder. Væsentligst var *adressekoden*, som entydigt identificerer den enkelte adresse; en meget central del af CPR-systemet.

Oprettelsen af CPR skyldtes først og fremmest ønsket om at undgå dobbeltregistrering og det ekstra ressourceforbrug, som fulgte heraf. Medvirkende har også været udsigten til en skattereform med indførelsen af kildeskat (gennemført i 1970), som vanskeligt kunne gennemføres uden et meget sikkert system til identifikation af landets borgere.

CPR's delregistre

CPR blev fra starten vedligeholdt via de fælleskommunale edb-centraler gennem løbende indberetninger fra kommunerne, som igen fik indberetninger fra borgerne og en lang række offentlige myndigheder. Der er naturligvis gennem årene gennemført adskillige forbedringer og moderniseringer. I dag sker opdateringen dagligt gennem on-line terminaladgang fra kommunerne. CPR består af flere dele, hvoraf de vigtigste er:

- *Personregistret* med aktuelle og historiske oplysninger om hver person, herunder personnummer, henvisninger til forældre og evt.

¹¹ Jf. *Indenrigsministeriet (1978)*.

børn og ægtefælle, adresse, navn, ægteskabelig stilling, statsborgerskab.

- *Boligregistret*, som for alle adresser (bopæle) rummer oplysninger om boligen (fra BBR) og dens lokalitet.
- *Vejregistret* med oplysninger om alle veje og bynavne i landet. Hver vej (vejstykke) har en vejkode, et navn og oplysning om vejens (eller vejstykkets, angivet ved et husnummer-interval) placering i fx sogn, postdistrikt, politikreds eller kommunalt fastsatte inddelinger såsom skoledistrikt mv.

Alle offentlige myndigheder har adgang til at få oplysninger fra CPR. Dagligt eller ugentligt leveres status- og/eller ændringsudtræk til godt 75 abonnerende myndigheder. Desuden har en lang række myndigheder terminaladgang til at foretage landsdækkende søgning og opslag i hele CPR's personkreds, dvs. personer som bor eller har boet i Danmark efter 1. april 1968.

Anvendelser af CPR

Private virksomheder kan også rekvirere oplysninger fra CPR, forudsat at virksomheden selv, i henhold til *Lov om private registre*¹², kan levere de pågældende personnumre. Det drejer sig i praksis alene om virksomheder i den finansielle sektor, som skal registrere personnummer på kontohaveren for at kunne afgive lovpligtige oplysninger til skattevæsenet om indeståender og renter.

Tidspunktet 1968 for oprettelsen af CPR var overmåde velvalgt, idet CPR stod klar *før* udviklingen på edb-området for alvor var slået igennem i administrationen - altså *før* der var sat fart i opbygningen af store, landsdækkende systemer.

Andre administrative personregistre

CPR's sikre identifikationer og basale personoplysninger blev således fra begyndelsen føderegister for næsten alle andre større, offentlige administrative systemer, og også for mange private systemer, især i den finansielle sektor. I modsætning til næsten alle andre lande undgik Danmark på denne måde, at de forskellige administrative registre "opfandt" hvert sit identifikationssystem, og der var derfor ingen tekniske hindringer for at indføre den nuværende effektive og ressourcebesparende edb-infrastruktur.

Bred anvendelse af personnummeret

Det fælles identifikationssystem er forudsætningen for, at man næsten overalt - skattevæsen, bankvæsen, sociale sager, vægtafgifter, radio- og TV-licenser osv. - baserer administrationen på løbende oplysninger fra CPR om navne, adresser, ægteskabelig stilling mv. kombineret med de specifikke sagsdata på det pågældende område.

¹² Jf. *Justitsministeriet (1992b)*.

Mange offentlige registre

Det er inden for rammerne af denne bog ikke muligt at komme nærmere ind på indholdet i de administrative systemer. Registertilsynet har i dag registreret forskrifter for ca. 2.800 offentlige personregistre, hvoraf langt hovedparten er administrative registre, mens et mindre antal er registre oprettet i statistik- eller forskningsøjemed.

Registrene fordeler sig omtrentligt således:

- 1050 statslige (heraf 56 statistikregistre i Danmarks Statistik)
- 80 fælleskommunale (på Kommunedata I/S)
- 370 enkelt-amtslige
- 1300 enkelt-kommunale

En betydelig del af disse registre omfatter kun en begrænset personkreds og tjener ganske afgrænsede og/eller lokale formål. Hovedinteressen, når det gælder en statistisk anvendelse, samler sig naturligvis om de administrative, landsdækkende statslige og fælleskommunale systemer, som findes på næsten alle områder, herunder

- skatteregistre
- arbejdsmarkedsregistre
- sociale registre
- sundhedsregistre
- sygehusregistre
- registre over uddannelsessøgende
- registre på erhvervsområdet
- ejendoms- og boligregistre

I bilag 6 gives en kort beskrivelse af de væsentligste administrative registre.

De fleste store administrative fællesopgaver løses ved hjælp af landsdækkende, ensartede, centrale systemer samlet på to datacentre, Kommunedata og Datacentralen, hvilket i betydelig grad har lettet de statistiske anvendelsesmuligheder. Der har imidlertid i de seneste år været en tendens til at vælge decentrale kommunale løsninger, fx for oplysninger om daginstitutionsbenyttelse og oplysninger om aktivering af ledige. Dette hindrer ikke i sig selv en statistisk anvendelse, men øger besværet med dataindsamling, dækningsgrad og harmonisering af uensartede datasæt.

2.4 Ejendomsregistrering

Ejendomsregistreringen her i landet kan siges at være startet med oprettelsen af matriklen i 1662, hvis formål var jordbeskatningen. Opmåling af de registrerede arealer og deres grænser blev senere foretaget og indført i registret. Siden begyndelsen af dette århundrede har matriklen mistet sin direkte tilknytning til ejendomsbeskatningen, som så blev foretaget med udgangspunkt i særlige vurderingsregistre (Statens salgs- og vurderingsregister vedrørende fast ejendom, jf. bilag 6, og Det fælleskommunale ejendomsdatasystem).

I 1972 oprettedes et landsdækkende edb-ejendomsregister til varetagelse af ejendomsbeskatning og ejendomsvurdering. Registret føres af kommunerne, og grundlaget for registreringen er oplysninger i matriklen og i tingbøgerne. Foruden matrikulære oplysninger mv. registreres forskellige data af betydning for vurderingen.

Landsdækkende registrering af ejendomme

I 1977 indsamledes i forbindelse med ejendomsvurderingen oplysninger til BBR, som ligeledes føres kommunalt, og som rent teknisk er en udbygning af ejendomsregistret. Formålet med denne udbygning var, at registerkomplekset skulle udgøre et egentligt basisregister for ejendomme, bygninger og boliger. Oplysningerne i BBR vedrører bygningernes og boligernes alder, størrelse, installationer og anvendelse. Registret ajourføres med meddelelser om ændringer, nybygninger og nedrivninger, idet disse oplysninger indhentes i forbindelse med kommunernes byggesagsbehandling.

Forbedrede muligheder med BBR

En stærkt medvirkende begrundelse for oprettelsen af BBR var, at det kunne benyttes til statistiske formål, således at man kunne undgå traditionelle folke- og boligtællinger i fremtiden.

Såvel i CPR som i BBR indgår adresseoplysningen. Ved løbende kommunikation mellem de to basisregistre sikres det, at adresserne er identiske, således at oplysningerne i registrene kan sammenkobles.

Adresseoplysningen forbinder CPR og BBR

I 1980'erne påbegyndtes etableringen af kommunale planregistre over offentlighedsretlige rådighedsindskrænkninger såsom kommune- og lokalplaner samt byfornyelses- og varmeplaner. Planregistret indeholder således oplysninger om den planlagte (og i en række tilfælde tillige den faktiske) arealanvendelse.

Et særligt register er Krydsreferenceregistret, et såkaldt nøgleregister, hvor alle ændringer og ajourføringer af adresser, matrikelnumre, ejendomsnumre, planbetegnelser og tilhørende geokoder (koordinater i form af en geografisk reference) samles. Herved er de forskellige ejendomsregistre bragt til at spille sammen, ligesom samspillet med personregistreringerne er sikret.

Samspil mellem ejendomsregistre

2.5 Erhvervsregistrering

Anvendelse af skatteoplysninger

Opbygningen i 1960'erne af administrative erhvervs- og virksomhedsregistre til brug for bl.a. ATP, toldvæsenet og skattevæsenet medførte indsamling af erhvervsdata, som også Danmarks Statistik kunne benytte til statistiske formål, herunder ajourføring af statistiske erhvervsregistre. De generelle erhvervstællinger, som Danmarks Statistik tidligere havde gennemført med tiårige intervaller, kunne på denne baggrund ophøre.

De administrative registres registrering af erhvervsvirksomhederne og arbejdsgiverne var imidlertid i den periode registeropbygningen fandt sted, kun i mindre omfang koordineret. Der var mange, men ofte små forskelle med hensyn til de forskellige registres enhedsdefinitioner, dataindhold mv. Endvidere anvendte hvert register sit specielle nummersystem til identifikation af enhederne.

Det centrale erhvervsregister oprettes

Med henblik på en forenkling af det offentlige erhvervsregistrering gennemførtes i 1975 en lov om oprettelse af et centralt erhvervsregister. Registret, der administreres af Danmarks Statistik, indeholder oplysninger om selvstændige erhvervsdrivende og arbejdsgivere, som i henhold til anden lov er berettiget eller forpligtet til registrering hos offentlige myndigheder eller institutioner. Registret indeholder endvidere oplysninger om filialer eller afdelinger under de ovenfor nævnte enheder.

Erhvervsregistrets oplysninger om firmaernes navn, adresse og indehavere registreres og ajourføres hovedsageligt ved direkte overførsel af oplysninger fra Told- og Skattestyrelsens registre. Danmarks Statistik tilføjer oplysninger om brancher og særskilt beliggende afdelinger, filialer mv. (arbejdssteder) samt supplerende oplysninger om registerenhederne.

I loven om Erhvervsregistret forudsættes indført en entydig fælles nummerering af registerenhederne til brug for det offentlige erhvervsregistre.

Forbedringer ved sammenlægning af told- og skattevæsen

Mulighederne for indførelse af fælles registreringsregler og enhedsnumre blev forbedret væsentligt, da toldvæsenet og skattevæsenet i 1990 blev samlet til Told- og Skattestyrelsen under Skatteministeriet. Told- og Skattestyrelsen har oprettet et fælles stamregister (SE-registret), der sikrer, at enheder, der er registreringspligtige for moms og kildeskat mv., får samme nummer i de forskellige delsystemer. Der er dog endnu nogle problemer ved opbygningen af stamregistret, men disse forventes at være overvundet omkring årsskiftet 1995/96. Når stamregistret er fuldt udbygget kan det herved sikres, at Erhvervsregistret kan fremtræde med entydigt afgrænsede enheder og et fælles identifikationsnummersystem.

2.6 Udviklingen i anvendelsen af administrative registre som statistikkilde

Selv om folkeregistrene i 1924 blev oprettet med administrative opgaver som det helt klare hovedformål, var man fra begyndelsen klar over, at registrene tillige skulle anvendes til forbedring af befolkningsstatistikken. Folkeregistrene blev derfor pålagt at indsende flyttebeviser for samtlige flyttede personer hvert kvartal til Det Statistiske Departement. Samtidig skulle der sendes en særlig statistikindberetning, som udviste totaltal for antal personer i registret ved kvartalets begyndelse og slutning, hvortil kom en optælling af kvartalets befolkningsbevægelser indenfor kategorierne: fødte, døde, indenlandske til- og fraflytninger samt vandringer til og fra udlandet.

Folkeregistre og befolkningsstatistikken

På grundlag af flyttebeviserne blev det muligt at fremstille en detaljeret statistik over flytningerne mellem kommuner og over ind- og udvandring. Statistikindberetningerne benyttedes til en summarisk opgørelse af "registerbefolkningen og dens bevægelser". Denne opgørelse måtte naturligvis være bundet af indberetningsformen og kunne fx ikke opdeles efter alder og køn. Hovedparten af den detaljerede befolkningsstatistik hvilede fortsat på andre kilder.

Oprettelsen af CPR i 1968 medførte en gennemgribende omlægning af statistikproduktionen. Det blev nu folkeregistrene, der via CPR kom til at danne basis for alle grene af befolkningsstatistikken. Nogle opgørelser henter fortsat oplysninger fra andre kilder, men disse oplysninger opfattes som supplement til data fra CPR.

CPR ændrede grundlaget for befolkningsstatistikken

Indførelsen af den nye produktionsteknik gav i de første år anledning til betydelige vanskeligheder. Allerede i 1969 startede man forsøgsvis en befolkningsstatistik baseret på et CPR-materiale, som skulle beskrive befolkningens sammensætning og alle typer af befolkningsbevægelser. CPR-statistikken for 1969 og 1970 blev imidlertid særdeles mangelfuld, og det var først fra og med 1973, at det samlede statistiksystem kunne betragtes som oprettet.

Vanskelighederne i starten skyldtes ikke alene mangler og fejl i CPR's dataindhold, men også tolkningsproblemer som følge af upræcis og utilstrækkelig kommunikation mellem statistikproducent og registermyndighed.

Det var ikke uden problemer i starten

Det befolkningsstatistiske system har nu gennem mange år fungeret næsten problemfrit, og i denne periode har registerbaseret statistikproduktion over en bred front indtaget en stadig mere dominerende rolle. Som omtalt har Danmarks Statistik nu ansvaret for 56 personorienterede statistikregistre (se bilag 2), som i hovedsagen hviler på løbende dataleverancer fra administrative edb-systemer. Dog baseres enkelte statistikregistre fortsat på hel eller delvis manuel dataindsamling, og

Registeranvendelse dominerer nu, men stadig manuel dataindsamling

nogle af registrene er dannet ved horisontal integration, dvs. ved tværgående anvendelse af eksisterende statistikregistre; et godt eksempel på sidstnævnte er den sammenhængende socialstatistik (jf. kapitel 5), men nævnes kan også visse registre oprettet i forbindelse med forskningsopgaver. En oversigt over hovedpunkter i udviklingen siden 1968 ses i bilag 3.

Det er klart, at denne udvikling kun har været mulig som følge af den tidligere omtalte hurtige udbygning af det offentlige personregistrering, som er enestående i verden, og den lidt senere etablerede centrale registrering af virksomheder, ejendomme mv.

Udvidelsen af registreringsaktiviteten er ofte sket udelukkende med administrative formål, men i visse tilfælde er hensynet til statistikproduktionen indgået med stor vægt i begrundelsen for oprettelse af nye registre eller udvidelse af bestående registres indhold. Det var således af stor betydning for vedtagelsen i 1976 af loven om det nye Bygnings- og boligregister¹³, at registret kunne danne grundlag for en bolig- og beboerstatistik; i 1974 var det nemlig blevet besluttet ikke at afholde den traditionelle folke- og boligælling året efter under den forudsætning, at boligstatistik på lidt længere sigt kunne fremskaffes på registergrundlag¹⁴. Ved opbygningen i 1960'erne af administrative erhvervs- og virksomhedsregistre var de statistiske formål ligeledes med i de forudgående overvejelser.

2.7 Forskellige typer registerdata

Ikke alle registerdata er lige statistikegnede

I det følgende vil det blive illustreret, hvordan data i de administrative registre kan inddeles i nogle kategorier efter det formål, de tjener i registret. De forskellige typer af data er i meget varierende grad egnede som grundlag for statistiske opgørelser.

2.7.1 Identifikationsdata

Identifikationsdata er en forudsætning

En grundlæggende forudsætning for effektiv registerdrift er, at identifikationer er entydige. Dette forudsætter igen, at to betingelser er opfyldt, nemlig

- at enhederne er entydigt defineret
- at identifikationer er præcist og hensigtsmæssigt udformet, samt robuste i forhold til almindeligt forekommende fejlmuligheder

Hvad kravet om entydighed angår, er mulighederne for imødekomme stærkt varierende fra enhedstype til enhedstype. I nogle tilfælde

¹³ Jf. *Boligministeriet (1990)*.

¹⁴ Jf. *Danmarks Statistik (1979b)*.

(personer, biler, skibe, fly) er der efter sagens natur intet - eller næsten intet - problem. I andre tilfælde (ejendomme, erhvervsenheder, bygninger osv.) må der forsøges opstillet operationelle definitioner eller retsregler, der på den ene side tager udgangspunkt i objektivt konstaterbare forhold, og på den anden side bedst muligt modsvarer formålene med registreringen.

Kravet til identifikationernes udformning er af mere teknisk karakter, og de styringsprocesser, som skal sikre mod dobbeltnummerering og mod fejl i numrene, fx checkcifre, er velkendte. Spørgsmålet, om hvorvidt identifikationen skal være informationsløs eller informationsbærende, har også været debatteret i forbindelse med indretningen af de fleste nummersystemer. Det vigtigste identifikationsnummer her i landet, personnummeret, indeholder som bekendt visse informationer, nemlig fødselsdato og køn, men da disse er (næsten) uforanderlige, volder dette ikke større styringsmæssige problemer¹⁵. Tværtimod kan man sige, at netop en fødselsdato er en hensigtsmæssig oplysning at lade indgå i identifikationen, blandt andet for at gøre det let at huske, men det kan desværre ikke praktiseres for alle typer af enheder.

Det samlede registersystem bygger på nogle få generelle identifikationer: Personnummer, adressekode (for boliger mv.), matrikelnummer for ejendomme, samt virksomheds- eller firmanummer (jf. bilag 4). Dette er afgørende for de faktiske integrationsmuligheder. Hver af disse identifikationer må af hensyn til en effektiv drift styres ét sted, og de registre, som varetager denne funktion, får derfor en særlig betydning. Disse styrende, primære identifikationer ligger til grund for de enkelte registers drift og for samspillet imellem disse, men de dækker ikke alle identifikationsbehov. Der anvendes fx en række udadvendte identifikationer, hvoraf navn og adresse for personer selvsagt er vigtige. Oplysning om fødselsregistreringssted er et andet vigtigt element i en personidentifikation, der bl.a. etablerer forbindelsen mellem folkeregistersystemet, herunder CPR, og kirkebøgerne.

Identifikationerne tjener også til at udtrykke enhedsrelationer. For eksempel registreres i CPR relationen "person A er gift med person B" ved, at B's personnummer er anbragt i et særligt felt i person A's CPR-record, og omvendt (det såkaldte ægtefællehenvissningsnummer).

En registreret enhedsrelation kan samtidig være af betydning som sekundær identifikation, som tilfældet er, når ejeren til et firma registreres.

¹⁵ Som det fremgår af kapitel 5, giver det dog nogle problemer, når personnummeret skal rettes.

2.7.2 Tidsreferencer

Tidsangivelser har central betydning

Tiden er en afgørende dimension i vor daglige opfattelse og tolkning af fænomener i verden omkring os. Derfor skal oplysning om tiden indgå som en vigtig del af registrets indhold.

I det følgende peges der på forskellige former for tidsreferencer, som indgår i de administrative registre, idet der skelnes mellem dateringer, der tidsfæster fænomener i virkelighedens verden, og dateringer, der tidsfæster målingen eller registreringen af disse.

Tidsangivelser for fænomenerne selv...

Med hensyn til dateringer, der tidsfæster fænomenerne, kan man sondre mellem dateringer knyttet til selve enhedens eksistens og dateringer, der knytter sig til bestemte variable.

Dateringer, som knytter sig til enhedens eksistens, er de samme som afgrænser enhedens levetid, dvs. fødsels- og død tidspunkt. Disse dateringer vil typisk være centrale i basisregistre, men kan også være vigtige i andre statistiske registre.

Dateringer, som knytter sig til andre variable, må opdeles efter disse variables karakter. Det er her vigtigt at sondre mellem dels variable, hvis værdier er gældende for en periode, *strømvariable*, og dels variable, hvis værdier knytter sig til et tidspunkt, *beholdningsvariable*. Som eksempel på strømvariable kan nævnes en persons bruttoindkomst. Det er vigtigt i et statistisk register, der indeholder strømvariable, at datere den periode, variabelværdien refererer til. Som eksempel på beholdningsvariable (tidspunktvariable) kan nævnes ægteskabelig stilling eller uddannelse, og man kan her sondre mellem to forskellige former for datering, nemlig i *diskret tid* eller i *kontinuert tid*.

Ved datering i diskret tid, kender man kun de pågældende variables værdi på et eller flere givne tidspunkter. Ved datering i kontinuert tid, kender man den periode, inden for hvilken de pågældende variabelværdier gælder. Denne dateringsform giver samtidig en datering af hændelser, hvorved forstås ændringer af variables værdi, der bl.a. er afgørende ved udarbejdelse af egentlig forløbsstatistik.

...og tidsangivelser for målingen af fænomenerne

Endelig kan det som nævnt være vigtigt at angive, hvornår målingen eller registreringen af et bestemt fænomen er foretaget, og der går her en hovedsondring mellem retrospektiv og simultan måling.

Ved en *retrospektiv måling* foretages målingen en kortere eller længere tid efter fænomenets forekomst, og måske foretages der flere retrospektive målinger af det samme forhold. I disse tilfælde er det vigtigt at kunne holde resultatet af de forskellige målinger ude fra hinanden.

Ved en *simultan måling* vil der normalt ikke være det samme behov for en eksplicit datering af målingen, idet denne med god tilnærmelse vil falde sammen med de dateringer, der knytter sig til ændringer af de forskellige variable i registret.

2.7.3 Basisdata

Ved basisdata forstås registeroplysninger, som er beregnet til en mangesidet anvendelse, herunder til fremtidige, måske endnu ukendte formål. Begrebet svarer på dataplanet til det tidligere indførte begreb basisregister (jf. afsnit 2.2). Eksempler på basisdata er hovedparten af CPR's dataindhold og særligt den del, som er egentlige folkeregisterdata, fx bopæl, ægteskabelig stilling og nationalitet. I øvrigt er basisdata karakteriseret ved følgende:

- Lovgivningen benytter ofte basisdata som parametre, fx tildelingskriterier, men det anses som regel ikke for nødvendigt at fastlægge det begrebsmæssige indhold gennem lovgivningen, fordi der er tale om almindeligt anerkendte begreber.
- Datasikkerhedsmæssigt set har de en særstilling, da de efter sagens natur har en betydelig udbredelse. Hertil kommer, at de på grund af deres almene karakter som regel ikke objektivt set er særligt følsomme. For eksempel kan enhver få oplysninger om en række af Erhvervsregistrets data.
- Oplysningerne må begrebsmæssigt set være *objektive* og *neutrale* og mindst mulig præget af forhold indenfor specielle administrative områder.
- Datakvalitet og ajourføringsgrad må være høj, idet manglende nøjagtighed på anvendelsestidspunktet kan give anledning til administrativt besvær og eventuelt få konsekvenser for de registrerede.

Hvis oplysningerne skal opbevares i flere registre, er konsekvensen, at ajourføringsoplysninger må overføres mellem registersystemerne. Hvis de derimod skal opbevares ét og kun ét sted, er konsekvensen, at andre systemer skal hente deres data her, når de skal anvendes. I begge tilfælde vil en form for samkøring være nødvendig.

2.7.4 Specialdata

Ved specialdata forstås oplysninger, som indsamles direkte med henblik på at indgå i bestemte administrative processer eller som er et resultat af sådanne. Som typiske eksempler kan nævnes lønmodtagerfradrag og lejeværdi af egen bolig fra kildeskattesystemets registre samt dagpenge og sygesikring fra det sociale statistikregistersystem.

Basisdata beskriver enhedernes baggrund

Specialdata dannes i den administrative proces

Selv om der er mange grænsetilfælde mellem basisdata og specialdata, er der klare principielle skillelinier i en række henseender:

- Det begrebsmæssige indhold af specialdata vil oftest være fastlagt gennem specielle lovregler og administrative forskrifter.
- Dataindholdet i specialdata er ofte mere følsomt end i basisdata, men også genstand for mindre cirkulation.
- Specialdata må nødvendigvis være defineret i forhold til formålene og underkastet de begrebsmæssige ændringer, som følger af lovændringer mv.
- Datakvalitet og ajourføringsgrad bestemmes ud fra anvendelsesbehovet for de pågældende specialdata.
- Registerfunktionelt må de "fødes" i den pågældende administrationsgrens regi.

2.7.5 Baggrundsdata

Baggrundsdata er supplerende beslutningsgrundlag for administrationen

Foruden de ovenfor omtalte datatyper, som altid er direkte nødvendige for den administrative proces, kan myndighederne have behov for nogle supplerende oplysninger. Når disse oplysninger indsamles systematisk, benævnes de baggrundsdata. Oplysningerne edb-registreres ikke nødvendigvis, men må alligevel siges at udgøre et led i det samlede registersystem.

Hvis administrationens brug af baggrundsoplysninger i konkrete tilfælde er skønsmæssig og relativ sjælden, vil det ikke være nødvendigt systematisk at kontrollere og komplettere oplysningerne i tilknytning til indsamlingen. Det kan tænkes at kræve mindre arbejde blot at rette manglerne op under sagsbehandlingen i de tilfælde, hvor oplysningerne har betydning for afgørelsen af et bestemt spørgsmål. Baggrundsdata anvendes normalt ikke uden for det forvaltningsområde, hvor de er indsamlet som direkte grundlag for administrative processer, og de er ikke umiddelbart velegnede til statistiske formål. Hvis disse oplysninger skal bruges i statistikken, skal de underkastes særlig grundig kontrol.

Som eksempler på sådanne baggrundsdata kan nævnes de specifikationer og oplysninger på de tidligere selvangivelser, som ikke dataregistreredes hos ligningsmyndighederne, og en række af oplysningerne på de tidligere ejendomsvurderingsskemaer.

2.7.6 Notatdata

Notatdata er ikke altid formaliserede

Notatdata er af endnu løsere karakter end baggrundsdata. Det drejer sig om oplysninger, der ikke systematisk indsamles, men som fx er til-

vejebragt under en sagsbehandling. Det kan være hensigtsmæssigt at registrere sådanne oplysninger, hvor de har, eller kan tænkes at få, fremtidig betydning i relation til den pågældende sagsbehandlende enhed. Oplysningerne kan være edb-registrerede og må da opfylde visse formkrav, men de vil oftere have karakter af noteringer på kartoteks-kort eller lignende. Grænsen til egentlige sagsoplysninger er naturligvis ikke skarp.

Som eksempel kan nævnes en række notatfelter i BBR. Den enkelte kommune bestemmer helt på egen hånd, hvad der skal indføres i registret og i hvilken form det skal ske. Tilsvarende har arbejdsformidlingskontorerne registreret en række notatoplysninger i deres interne administrative edb-systemer (AF-MATCH). Oplysningerne vedrører fx jobønsker og kortere uddannelser. Oplysninger af denne karakter vil som regel kun være meningsfyldte for den pågældende sagsbehandlende enhed.

2.8 Statistikdata

En særlig form for data er oplysninger, som er optaget i administrative registre udelukkende med statistiske formål. Denne form for dataindsamling kaldes integreret dataindsamling, og den omtales nærmere i afsnit 4.3. I den administrative proces kan de statistikdata, der indsamles på denne måde, sidestilles med baggrunds- og notatdata i den forstand, at deres kvalitet og fuldstændighed ikke har direkte og afgørende betydning for administrationen. Pålideligheden af denne type oplysninger kan derfor være forringet. Dette er ikke acceptabelt for statistikproduktionen, idet oplysningerne typisk er meget centrale for statistikken, hvilket netop er årsagen til, at de er indført i registrene. Problemet er imidlertid vanskeligt at løse.

Integreret dataindsamling

2.9 Konklusion

Identifikationsdata, basisdata og specialdata er som hovedregel teknisk egnede og ofte også relevante til statistikformål. Det samme gælder i sjældnere tilfælde for baggrundsdata. Bortset fra de former for statistik, som knytter sig snævert til et givet administrationsområde, eller som skal beskrive selve administrationen, er det imidlertid ikke sikkert, at oplysningerne frembringes i de kombinationer, der er relevante til statistiske formål. Samkøring mellem registre giver mulighed for at etablere de kombinationer af data, der er behov for. Som betingelse for samkøring bliver identifikationens kvalitet af afgørende betydning. Kombinationsmulighederne kan udstrækkes til at omfatte traditionelt indsamlede statistikdata, da de generelle identifikationer også anvendes i forbindelse med indsamling af oplysninger gennem postenquete eller interviews.

Kombination af data

Kapitel 3. Det samlede statistiksystem

Dette kapitel omhandler det samlede statistiksystems tilblivelse, fra statistikkens råmateriale via samkøringsprocesser til slutproduktet i form af et system af statistikregistre. Der redegøres for de forhold, der har muliggjort udnyttelsen af de administrative registre til statistikformål, Lov om Danmarks Statistik, samt Det centrale personregister. Det samlede statistiksystem bestående af personregistre, ejendomsregistre og virksomhedsregistre, samt klassifikationsmoduler beskrives.

Efterfølgende beskrives de processer i form af samkørsler, hvorved de administrative registre omdannes til statistikregistre. Dette afsnit er for en stor dels vedkommende en teknisk beskrivelse, som knytter sig til de begreber og analysemetoder, som er vedtagne principper i Danmarks Statistik.

3.1 Statistiksystemets opbygning

Grundoplysningerne i form af udtræk fra de administrative kilderegistre eller andet indsamlet materiale bearbejdes og organiseres i en række statistikregistre og klassifikationsmoduler.

Det enkelte statistikregister er emneafgrænset i den forstand, at det danner grundlag for statistikken inden for et bestemt område. Det betyder, at der opereres med indholdsmæssigt klart afgrænsede registre, hvor der er en tydelig sammenhæng mellem kilden for de centrale oplysninger og det pågældende statistikregister. Modellen er som hovedregel den, at man indhenter de emnespecifikke statistikdata fra en enkelt kilde og til disse føjer en række baggrundsoplysninger, ofte fra andre statistikregistre. Disse sammenstillede oplysninger udgør herefter registret for det pågældende statistikområde.

Denne klare forbindelse mellem grundregister og statistikregister er fortsat fremherskende, men behovet for at anskue tingene i en bredere sammenhæng (horisontal integration, jf. afsnit 5.2) har gjort, at nyere statistikregistre i stigende grad er et produkt af samkørte statistikregistre. Man kan fortsat tale om emneafgrænsede statistikregistre, men emnerne har fået en lidt anden karakter. På tilsvarende måde har interessen for forløbsstatistik bevirket, at forskellige årsversioner af samme register er grundlag for dannelsen af et nyt register. Dette benævnes vertikal integration (jf. afsnit 5.3).

I visse tilfælde bearbejdes grundoplysningerne centralt og opbevares i *klassifikationsmoduler*. Disse moduler er til rådighed i forbindelse med dannelsen af statistikregistre og indgår bl.a. i samkøring med stati-

Et statistikregister kan dannes på grundlag af ét administrativt register...

...eller ved samkøring af mange

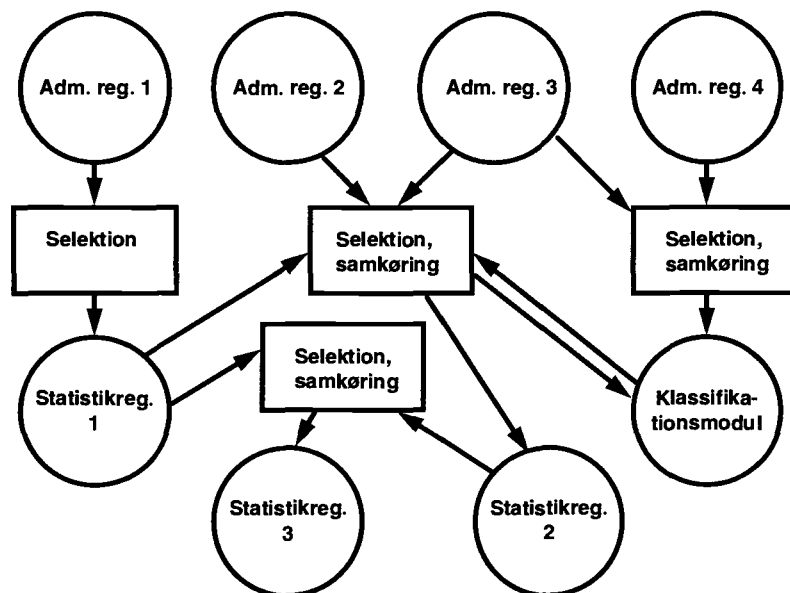
Klassifikationsmodulerne

stikregistre med henblik på dannelse af anonyme datasæt til enkeltstående specialopgaver.

Statistikregistrene og klassifikationsmodulerne udgør samlet Danmarks Statistiks personstatistiske registersystem. Systemet er skematisk illustreret i figur 3.1, hvor statistikregister 1 er den enkle og meget direkte form for dannelse af et statistikregister, og statistikregister 2 har et mere komplekst sammensat kildegrundlag i et delvist integreret forløb med dannelsen af et klassifikationsmodul, mens statistikregister 3 er et integrationsregister baseret på andre statistikregistre. De enkelte elementer kan anvendes isoleret og i sammenhæng. Forbindelsen mellem dem skabes som hovedregel gennem personnummeret og i enkelte tilfælde gennem adressekoden.

Figur 3.1

Personstatistiksystemet



I personstatistiksystemet indgår Ejendomsstatistikregistret som et element. Det indeholder bl.a. en nøgle mellem ejers personnummer og ejendomsidentifikation, således at man via denne nøgle kan trække på oplysninger om fast ejendom til anvendelse i personstatistikken. Endvidere indeholder *Arbejdspladsstatistikregistret* forbindelse til erhvervsstatistikken gennem *arbejdsstedsidentifikationen* (jf. afsnit 4.3). Ved anvendelse af denne identifikation kan Erhvervsregistrets oplysninger inddrages i personstatistikken som et led i klassificeringen af personer efter beskæftigelsesforhold.

3.1.1 Klassifikationsmoduler

Nogle få statistikdata indgår i mange forskellige statistikker og har således karakter af basisdata. Det gælder fx de mest grundlæggende oplysninger vedrørende personers familieforhold, uddannelse og arbejdsmarkedsrelationer. Disse data behandles centralt og indgår i statistiksystemet i form af klassifikationsmoduler. Sædvanligvis dannes der en ny version af klassifikationsmodulerne hvert år. De har karakter af statistikregistre, der har til formål at supplere andre statistikregistre, og statistikopgaver i øvrigt, med visse relevante og generelt anvendte baggrundsoplysninger. Det foregår ved direkte overførsel af moduloplysninger til et statistikregister, eller det kan ske ved behandling af foreliggende data i statistikregistre efter bestemte regler, typisk med anvendelse af omsætningsnøgler. Hvis oplysningerne blot skal samles i et anonymt statistikprodukt, vil det pågældende modul indgå i den samkøringsproces, som skal danne grundlaget for produktet.

Der indgår følgende 6 klassifikationsmoduler i statistiksystemet:

- *Befolkning/datomodulet* afgrænser befolkningen på et givet tidspunkt, og er indbygget i Befolkningsstatistikregistret
- *Områdekodningsmodulet* afgrænser geografiske områder på grundlag af adressekoder.
- *Husstands- og familiemodulet* grupperer personer i familier og husholdninger efter de gældende definitioner.
- *Uddannelsesklassifikationsmodulet (UKM)* anvendes til klassificering af personer efter højest afsluttede almene uddannelse, evt. afsluttet erhvervsuddannelse og evt. igangværende uddannelse.
- *Arbejdsklassifikationsmodulet (AKM)* klassificerer den enkelte person efter væsentligste beskæftigelse (beskæftigelsesstatus, stilling og branche for vigtigste arbejdssted) samt omfanget af beskæftigelse og ledighed i løbet af et kalenderår.
- *Arbejdsstedsmodulet* identificerer de enkelte ansættelsesforhold for lønmodtagere ved personnummer og arbejdsstedsidentifikation, og muliggør derved klassificering af personer efter virksomhedsforhold og virksomhedsklassificering efter personalesammensætning. Endvidere kan arbejdsstedets adressekode sammenholdes med den ansattes bopælsadresse, hvorved der skabes grundlag for pendlingsstatistik.

Seks klassifikationsmoduler

Klassifikationsmodulets funktion som generelt grundlag for hyppigt anvendte oplysninger til forskellige statistikområder indebærer en række fordele. De vigtigste er: højere datakvalitet, større effektivitet og konsistens mellem forskellige opgørelser

Fordele ved klassifikationsmoduler

Statistikens begreber skal være i størst mulig overensstemmelse med den verden, som statistikken beskriver. Kvaliteten af de anvendte statistikoplysninger vurderes i forhold hertil. De oplysninger, som indhentes fra de administrative systemer, kan hver for sig være utilstrækkelige som statistikdata i forhold til konkrete behov. Blandt de foreliggende administrative data udsøges de mest relevante, der kan overføres direkte til klassifikationsmodulet. En del data indgår dog i kombination med andre i dannelsen af nye variable, der anses for at være bedre til at beskrive de egenskaber, der skal belyses. Ofte hidrører de data, der skal kombineres til en ny variabel, fra flere forskellige kilderegistre. Ved dannelsen af kombinerede variable foregår databehandlingen efter et fastlagt sæt regler, der udformes i overensstemmelse med begrebernes definition. Denne proces gør oplysningerne anvendelige på et højt kvalitetsniveau. I vedligeholdelsen af modulet tages der højde for påtvungne ændringer i de administrative grunddata, og de nødvendige undersøgelser gennemføres med henblik på en tilpasning af regelsættet for databehandlingen således, at begrebsdefinitionerne så vidt muligt fastholdes i forhold til tidligere år, naturligvis forudsat at de fortsat er relevante.

Ved oprettelse og vedligeholdelse af modulet gennemføres de procedurer, som er nødvendige for at danne de ønskede baggrundsvariable én gang for alle, hvilket betyder større effektivitet i statistikproduktionen. Den ofte meget komplicerede og ressourcekrævende databehandling, der er forbundet med disse procedurer, skal på denne måde kun gennemføres én gang, i stedet for hver gang data fra klassifikationsmodulerne skal anvendes til en statistik.

Modulerne giver konsistens og letter dokumentationen

Anvendelse af modulernes standardklassifikationer giver sikkerhed for konsistens mellem de forskellige statistiske opgørelser, som klassificeringer af den pågældende art indgår i. Definitioner, tidspunkt for opdatering efter ændringer m.m. vil være det samme. Den opnåede konsistens har meget stor betydning for anvendeligheden af statistikken. Det har endvidere den fordel, at dokumentationen lettes, idet der kan refereres til modulet. Endvidere vil statistikbrugerne i højere grad blive fortrolige med de anvendte begreber.

Fleksibilitet i modulernes klassificeringer

Klassifikationsmodulerne skal være indrettet således, at de giver den bredest mulige dækning med hensyn til klassifikationernes anvendelse. Det betyder, at der skal være fleksibilitet i modulernes klassificeringer, således at relevante tilpasninger til den konkrete opgave kan foretages. På den måde kan det til en vis grad sikres, at brugen af modulerne ikke fører til et stift system, hvor man føler sig tvunget til at bruge klassifikationer, som måske ikke helt passer til formålet. Kravet om fleksibilitet kan forekomme uforeneligt med hensynet til konsistens mellem statistiske opgørelser, men der er en lang række områder, hvor en valgt standardklassifikation både kan og bør benyttes. Hvor dette ikke er

relevant, skal der være mulighed for anvendelse af modulets elementer, så langt det er foreneligt med opgørelsens formål.

Som eksempel kan nævnes brug af Arbejdsklassifikationsmodulet (AKM) i arbejdsstyrkestatistikken. I AKM's standardklassificering afgrænses gruppen af lønmodtagere efter en helårsbetragtning under hensyntagen til, hvad der har været den væsentligste aktivitet i løbet af et år, mens lønmodtagere i arbejdsstyrkestatistikken (statistikregister) afgrænses efter regler, der knytter sig til et givet tidspunkt. Denne forskel i afgrænsningen af lønmodtagere udelukker ikke, at AKM's stillingsklassifikationer kan overføres til og anvendes i arbejdsstyrkestatistikken. Stillingsklassifikationerne er dannet for alle personer, for hvilke grundmaterialet har givet mulighed for det, både beskæftigede og ikke-beskæftigede. Disse oplysninger er derfor til rådighed også for alternativt afgrænsede lønmodtagere.

Eksempel 1: Arbejdsklassifi- kationsmodulet (AKM)

Balancen mellem konsistens og fleksibilitet skal være sådan, at man altid benytter samme standardklassifikation i forskellige statistikker, med mindre der er en saglig grund til at anvende en afvigende klassifikation i en bestemt statistik. I sådanne tilfælde skal det være muligt at gøre rede for forskellene.

3.1.2 Statistikregistre

Kilderne til statistikregistrenes dataindhold kan være administrative registre, klassifikationsmoduler, andre statistikregistre og data indsamlet direkte gennem interviews eller postspørgeskemaer.

Det enkelte statistikregister er som tidligere nævnt afgrænset i forhold til et bestemt emneområde, således at registrets indhold og populationsafgrænsning er foretaget under hensyntagen til, hvad der er relevant inden for netop dette område. Der er til en vis grad sammenfald mellem emneafgrænsningen af de administrative registre og opdelingen af det samlede statistikgrundlag i statistikregistrene, hvilket betyder, at der for mange statistikregistres vedkommende vil være én hovedkilde og én eller flere andre kilder med supplerende oplysninger. De emne- eller sagstypeafgrænsede registre i kommunernes fælles edb-systemer og deres anvendelse, især inden for socialstatistikens forskellige områder, er eksempler på dette.

Den tilsyneladende klare forbindelse mellem hovedkilde og statistikregister betyder dog ikke, at de administrative oplysninger kan anvendes umiddelbart. Der kan for eksempel være problemer med at tidsfæste oplysninger. I den administrative praksis kan tidsreferencen være bestemt af administrative handlinger, som ligger senere end tidspunktet for den hændelse eller observation, som skal indgå i statistikken. Endvidere kan begrebsdefinitioner i det administrative system afvige

**Administrative data
er ikke altid
problemfri**

fra de statistisk relevante begreber på en måde, som umiddelbart vil forringe statistikkens kvalitet, i visse tilfælde til det uacceptable, hvis de administrative oplysninger bruges i deres rå form. Endelig forekommer utilstrækkelig ajourføring og deraf følgende fejl og mangler i kildeoplysningerne. Det ses i administrative systemer med blandet automatisk og manuel indberetning, eller hvor fejlagtige angivelser ikke har administrative konsekvenser.

Indkomstskatte-systemet

Eksempel på sidstnævnte er indkomstskattesystemet, hvor det hidtil har været uden administrativ betydning, hvordan en A-indkomst er specificeret. Skattekategorien vil give samme resultat.

Inddragelse af andre kilder kan være nødvendig

De nævnte forhold har ført til, at der i forbindelse med oprettelse og ajourføring af et statistikregister ofte er behov for at trække på flere kilder, for at sikre en korrekt tidsafgrænsning, opnå de mest relevante begrebsdefinitioner, samt, så vidt det er muligt, kompensere for fejl og mangler.

Spørgeskemaer

Direkte indsamling af statistikoplysninger foregår i forbindelse med undersøgelser af repræsentative udsnit af befolkningen. Det sker enten ved postenqueter, hvor svarpersonerne selv udfylder et spørgeskema modtaget med posten og tilbagesender det til Danmarks Statistik, eller det sker ved computerunderstøttede telefoninterviews. Som eksempler på interviewundersøgelser kan nævnes Arbejdsstyrkeundersøgelsen, Forbrugsundersøgelsen og Omnibusundersøgelserne. Postenqueter benyttes i dag ikke indenfor personstatistikken, men er fremherskende som grundlag for erhvervsstatistikken. Ved tilrettelæggelsen af denne type statistik søger man at udnytte oplysninger, der allerede foreligger i diverse statistikregistre. Der indgår naturligvis i den forbindelse en vurdering af registeroplysningernes kvalitet, idet formålet med den direkte henvendelse kan være at kompensere for utilstrækkelige registeroplysninger. En særlig rolle har statistikregistre i forbindelse med estimeringsprocedurer, der har til formål at rette op på skævhed forårsaget af bortfald i interviewundersøgelser.

Integrationsregistre

Behovet for at undersøge årsagssammenhænge mellem forskellige fænomener har ført til, at der oprettes registre med et tværgående indhold. Den emneorienterede opbygning af registersystemet har betydet, at nye samkøringer skal foretages, hvis der opstår konkrete behov for at se emner i sammenhæng og dermed datasammenstillinger ud over, hvad der foreligger i det enkelte register. Opgaver af denne karakter er blevet og bliver fortsat gennemført på grundlag af anonyme datasæt baseret på samkøringer, men mulighederne for at analysere forløb, som systemet nu giver mulighed for, har skabt stor interesse især blandt forskere, og det har ført til, at der dannes såkaldte *integrationsregistre*. Register af denne karakter er dels et led i det almindelige statistikberedskab, og dels en konsekvens af specielle databehov i forbindelse med konkrete serviceopgaver.

En særlig form for grunddata i det statistiske registersystem er *indleverede datasæt*. I forbindelse med forskningsprojekter gennemført uden for Danmarks Statistik kan der være indsamlet data, som ønskes samkørt med oplysninger i Danmarks Statistiks registre for at opnå en bredere belysning af et konkret emne. Hvis personnumre indgår i primærmaterialet er en sådan samkøring mulig, og forskningsmaterialet kan på denne måde tilføres supplerende oplysninger eller klassifikationer. Det samkørte produkt er dog ikke et led i registersystemet, da det vil være anonymiseret. Det kan derved kun bruges isoleret til analyser i forbindelse med det konkrete forskningsprojekt i Danmarks Statistik eller til tabelkørsler.

**Indleverede
datasæt, en
mulighed for
forskere**

Strukturen med en række emneopdelte eller formålsbestemte statistikregistre kan diskuteres. Den umiddelbare konsekvens af dette er en ret betydelig redundans i systemet. Den samme oplysning om en given enhed vil forekomme i flere og, for nogle enkelte oplysningers vedkommende, i en del registre. Hensigtsmæssigheden af dette må ansues ud fra forskellige synsvinkler. Det teknisk/økonomiske synspunkt tilsiger umiddelbart, at en given oplysning kun skal være registreret ét sted. Pladskravene ville derved blive reduceret, og dokumentation og vedligeholdelse ville blive forenklet. Beredskabet i forhold til anvendelsen er lidt vanskeligere at vurdere, da det vil afhænge af den valgte tekniske løsning og adgangsbestemmelserne. Netop sidstnævnte er et væsentligt led i argumentationen for det nuværende system. Med opdelingen i en række registre fastlægger man for hvert af disse, hvad det må indeholde, hvad det må bruges til, og hvem der må bruge det. Specielt ved fastlæggelsen af adgangen til et register sikres det, at kredsen er lille, og at samme person ikke har adgang til mange registre. Samkøring af forskellige statistikregistre kan kun finde sted efter en formel beslutning og skal følge bestemte regler. Dette gør det gennemskueligt, hvad der foregår, når de mange oplysninger anvendes i statistisk sammenhæng, og det betyder, at den utryghed, der kunne forekomme, reduceres. Endvidere er beredskabet højt, idet det enkelte register umiddelbart er klar til brug inden for det område det dækker. Alt taget i betragtning er det Danmarks Statistiks vurdering, at den emneorienterede opdeling af registrene er hensigtsmæssig og nødvendig at opretholde. Hensynet til databeskyttelse er selvfølgelig også afgørende for denne vurdering.

**Samme oplysninger
findes i flere registre**

I bilag 2 er de enkelte statistikregistre anført med en kort omtale af deres indhold.

3.1.3 Strukturen i det samlede statistiksystem

Alt kan kombineres, men saglige krav skal opfyldes

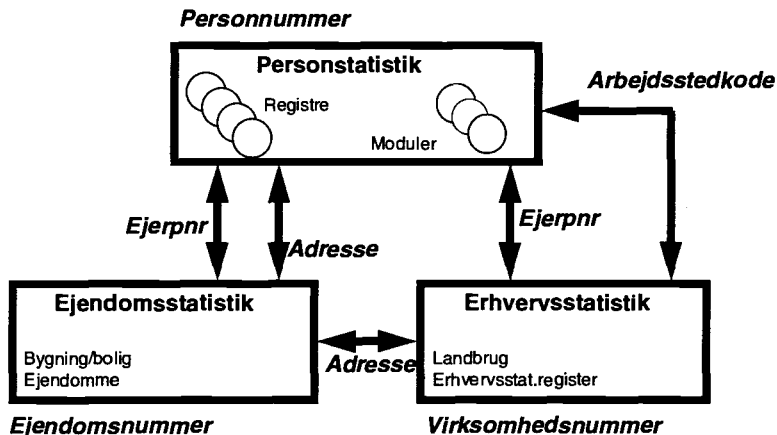
Det grundlæggende element i statistiksystemet er personnummeret, som er den fælles identifikation af statistikenheder i personstatistikregistrene. Ved anvendelse af dette nummer kan i princippet alle oplysninger kombineres. I praksis sker det ikke, da de foretagne samkøringer af oplysninger kun omfatter data, som er relevante i forhold til konkrete statistikopgaver. Der sættes grænser for, hvad der må kombineres, idet kombinationer af oplysninger skal være sagligt begrundede, og det må ikke være muligt at få viden om enkeltpersoner via statistikken.

Bolig- og erhvervsoplysninger kan knyttes til personoplysninger

Som tidligere omtalt omfatter systemet også nøgler til ejendoms-/boligenheder samt arbejdssteder via adressekode og arbejdsstedskode. Der er derfor mulighed for at overføre bolig- og erhvervsstatistiske oplysninger til personstatistikken, og der kan overføres personstatistiske oplysninger til boligstatistikken og erhvervsstatistikken. I figur 3.2 er det sammenhængende statistiksystem skitseret med angivelse af de identifikationer, som knytter systemets elementer sammen.

Figur 3.2

Det samlede statistiksystem



Adressekoder forbinder personstatistik og ejendomsstatistik samt ejendomsstatistik og erhvervsstatistik. Mellem ejendomsstatistikken og personstatistikken er der en personnummersammenhæng for så vidt angår ejere af fast ejendom. Tilsvarende gælder mellem erhvervsstatistikken og personstatistikken for virksomhedsindehavere. Endelig er der en sammenhæng mellem lønmodtagere og virksomheder gennem arbejdsstedskoden, således at oplysninger om ansatte personer kan forbindes med oplysninger om ansættelsessted.

3.2 Samkøring af registre

Ved en samkøring af registre forstås den proces eller metode, hvorved informationer fra forskellige statistikregistre om en bestemt statistisk *enhed* knyttes sammen, eller hvorved informationer om forskellige *enheder*, der indgår i bestemte relationer, knyttes sammen. Samkøring kaldes også integration.

Formålet med denne sammenknytning af informationer er at opnå ny viden om de *enheder* man er interesseret i. Ved samkøringen sker der en bearbejdelse af den eksisterende information, hvorved værdien af informationen forøges.

Samkøring af statistikregistre har en overordentlig høj nytteværdi som alternativ til traditionel dataindsamling. For at en samkøring skal fungere tilfredsstillende, kræves det imidlertid, at identifikationen af personer og andre enheder i udtrækkene fra de pågældende kilderegistre er entydige og veldefinerede. Er der mangler i denne henseende, kan det medføre store omkostninger at få oprettet fejl.

Ved samkøring vil der også vise sig fejl og uoverensstemmelser i andre oplysninger, og der må tages stilling til, i hvilket omfang sådanne fejl skal føre til rettelser.

Ved beslutninger vedrørende opretning af fejl er det vigtigt at gøre sig målsætningen med samkøringen klar, dvs. hvad det er samkøringen skal hjælpe med at belyse. En fejlretning ved samkøring vil primært tage sigte på at finde de kombinationsfejl, der opstår i forbindelse med den aktuelle samkøring. Det er ikke givet, at man derved får editeret data således, at det automatisk egner sig til en hvilken som helst anden *mulig* samkøring.

En samkøring eller integration kan optræde i flere forskellige fysisk-logiske fremtrædelsesformer; det er nemlig ikke nødvendigvis givet, at der ved en samkøring overhovedet sker en permanent fysisk samkøring. Der kan skelnes mellem fire forskellige typer:

- *Logisk samkøring*. Herved forstås det forhold, at det ved en logisk analyse (jf. afsnit 3.2.1) fastslås, at det er muligt at integrere to registre. Resultatet af den logiske samkøring er en enhedsgraf, hvor relationerne mellem de logiske enheder fastlægges.
- *Ad hoc-samkøring*. Til dannelse af et konkret sæt af tabeller med oplysninger fra flere registre, foretages en éngangssamkøring af de pågældende registre. En variation af dette er de tilfælde, hvor der foreligger et fast beredskab med henblik på dannelse af bestemte typer af tabeller fra et givet sæt af registre. I begge tilfælde foreligger den fysiske integration kun mens databehandlingen foregår.

Fejl i identifikationer

Fejlretning må tilpasses målsætningen med samkøringen

Fysisk integration

Statistikservicesystemerne KÅS (Kommune-Års-Service) og Markedsstatistik er eksempler herpå.

- *Permanent samkøring.* En gang for alle foretages en samkøring, hvor alle relevante oplysninger for de pågældende enheder samkøres til ét permanent register. Sygehusbenyttelsesstatistikregistret er et eksempel på dette.
- *Dannelse af en database.* Hvis de relevante registre, der ønskes samkørt, er dannet efter et relationsdatabaseskema, kan dette forhold overflødiggøre yderligere fysisk samkøring, idet alene den logiske samkøring er nødvendig (jf. ovenfor). Ved en logisk samkøring fastlægges de nødvendige relationer mellem enheder på tværs af registre og dermed de relevante nøgler, der siden kan anvendes ved forespørgsler i databasen¹⁶. Naturligvis kræver det, at registre ligger lagret i formater, der er tilgængelige for det pågældende spørgesprog. En variation af denne integrationsform udgør de tilfælde, hvor registre er beskrevet i et fælles meta-informationssystem, hvor de er defineret med henblik på logisk samkøring, og hvor der findes oplysninger om den fysiske lagring.

Hvilken type samkøring, der bør foretrækkes, afhænger i det konkrete tilfælde af målsætningen med integrationen, mængden af data, ønsker om beredskab, ressourcer, maskiner, software m.m. Ad hoc- og permanente samkøringer er de mest anvendte i dag, men det må forventes at dannelse af databaser vil blive mere udbredt, efterhånden som omkostningerne forbundet med de fysiske lagringsmedier bliver mindre og databaseværktøjerne mere effektive.

3.2.1 Indholdsanalyse

Nedenfor gennemgås nogle vigtige begreber vedrørende indholdsanalyse¹⁷.

Figur 3.3

Illustration af en enhed

Enhed	Egenskaber
Person	PNR køn fødselsdato civilstand

¹⁶ Forespørgsler kan eventuelt foretages i SQL (Structured Query Language).

¹⁷ Beskrivelsen er udformet i overensstemmelse med de begreber, der er anvendt i Danmarks Statistiks Planlægningshåndbog, jf. *Danmarks Statistik (1990)*.

Enheder er de personer eller genstande, som man vil fremstille statistik om, dvs. tællingsenhederne. En enhed anføres grafisk ved et rektangel. Se figur 3.3.

Enheder

Egenskaber eller *variable* er det, man beskriver *enheden* med, fx alder eller størrelse. Der skelnes mellem for det første de egenskaber, der identificerer den enkelte enhed, *nøglen* eller *identifikationen* (fx personnummer), og for det andet de egenskaber, der i øvrigt beskriver enheden, de *beskrivende egenskaber*. Egenskaber kan være *kvalitative*, som køn (mand, kvinde), eller *kvantitative*, som indkomst (200.234 kr.) og alder (29 år).

Enheder har egenskaber

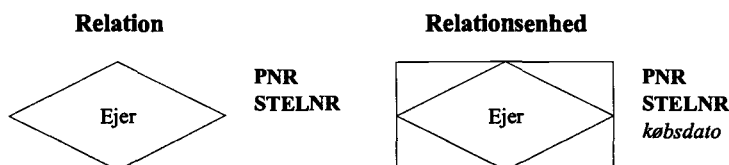
Nøglen er på grafen anført med fed tekst (fx PNR¹⁸) og de beskrivende egenskaber er kursiv. Hvis man ønsker at sammenholde oplysninger om en enhed fra flere forskellige kilder, er det *nøglen*, man bruger til sammenkoblingen. En samling af enheder med bestemte egenskaber betegnes en *enhedsgruppe* eller *population*. Hvis man ønsker at undersøge forbindelserne mellem forskellige enheder, det vil sige relationer, er det *nøglerne* fra de indgående enheder, man bruger.

En *relation* karakteriserer et forhold mellem to eller flere *enheder*, fx "en person bor i en bolig" eller "en person ejer en cykel". En relation identificeres ved kombinationen af de indgående enheders *nøgler*, fx "person: PNR ejer cykel defineret ved STELNR". En relation anføres grafisk ved en rombe, som vist til venstre i figur 3.4. Identifikationen for relationen består af identifikationerne for de enheder, der indgår i den, her PNR og STELNR.

Relationer mellem enheder

Illustration af relationer og relationsenheder

Figur 3.4



En relation kan - ud over *nøglerne* - også indeholde specifik information om relationens karakter. I eksemplet ovenfor til højre indeholder relationen oplysning om købsdato for cyklen. Hvis en relation indeholder oplysninger, ud over de indgående *nøgler*, betegnes den som *relationsenhed* og anføres grafisk som et rektangel med en indskreven rombe.

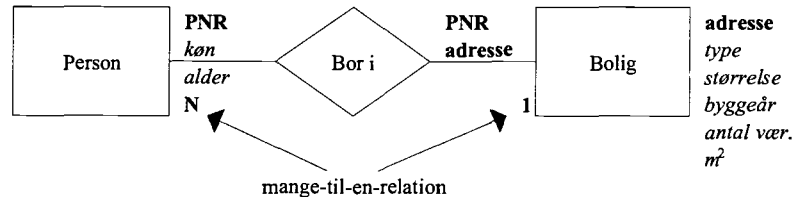
¹⁸ Forkortelse for personnummer.

Enhedsgrafen

Den logiske sammenhæng mellem *enheder*, *egenskaber* og *relationer* er afbildet i figur 3.5.

Figur 3.5

Illustration af en enhedsgraf



Relationerne mellem enhederne opdeles ofte i typerne "en til en", "en til mange" og "mange til mange". Herved angives, hvordan enhederne antalsmæssigt står i forhold til hinanden. Én person kan højst have én ægtefælle, medens én person kan eje flere cykler. Mange gymnasieskoler udbyder mange gymnasiale retninger. Angivelse af én- og mange-relationer sker ved betegnelserne 1 og N.

Figuren, der viser sammenhængen mellem personer og boliger, læses således: Personer bor i boliger; der kan bo *mange* personer i *én* bolig, og *én* person kan kun have *én* bopæl¹⁹; de konkrete personer, der bor i konkrete boliger, bestemmes af PNR og adressekode.

3.2.2 Dannelse af statistikregistre

Næsten alle personrelaterede *statistikregistre* i Danmarks Statistik er som tidligere nævnt produceret på grundlag af *administrative registre* i statslig eller kommunal regi, evt. suppleret med manuelt indsamlede oplysninger fra andre kilder (fx statistik over levendefødte: CPR og fødselsanmeldelser).

Fra kilderegistre...

De administrative kilderegistre kan være egentlige dagligdags brugerregistre, som er dynamiske registre i den forstand, at de løbende modtager *transaktioner* (opdateringer) af informationer på baggrund af *hændelser* vedrørende de *enheder*, der findes i registret, således at registret til en hver tid afspejler den administrative virkelighed. Kilderegistret kan dog også være et centralt integreret register, der periodisk modtager oplysninger fra de egentlige brugsregistre. For eksempel er Landspatientregistret et register i Sundhedsstyrelsen, der løbende modtager udtræk fra de enkelte sygehuses patientregistre.

¹⁹ Efter danske regler kan man kun have bopæl i én bolig, selv om man måtte råde over flere boliger, som man fra tid til anden opholder sig i.

Fra kilderegistret modtager Danmarks Statistik udtræk til statistisk bearbejdelse, og dette udtræk vil ofte (bl.a. af praktiske grunde) kun afspejle kilderegistrets tilstand på et bestemt tidspunkt eller for en bestemt periode. Danmarks Statistik modtager typisk et udtræk én eller flere gange om året fra det administrative kilderegister, der indeholder de oplysninger, som er relevante. Udtrækket afbilder derved udelukkende kilderegistret på det pågældende tidspunkt eller for den pågældende periode. Der kan opridses tre hovedtyper af udtræk:

- udtræk fra *status for populationen pr. en given dato* (fx personer med social pension og pensionens størrelse pr. januar måned)
- udtræk fra *status for population pr. en given dato, suppleret med oplysninger om udviklingen i centrale egenskaber i en periode* (fx Bistandslovsstatistik i året 1992)
- udtræk af *transaktioner for alle de relevante enheder i en periode* (fx sygehusbenyttelse)

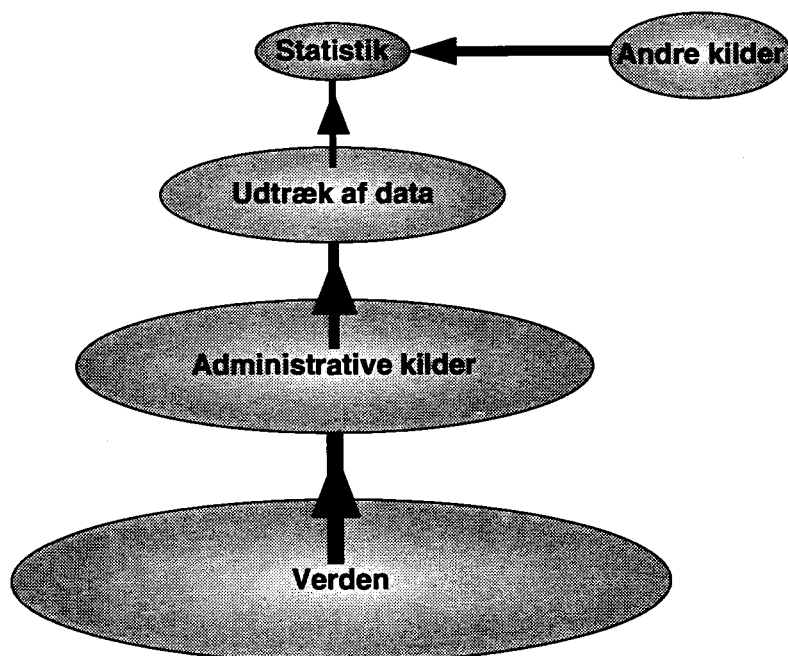
Udtrækket bliver editeret og ofte sammenholdt med andre oplysninger og registre, for til sidst at blive til det færdige *statistikregister*. Denne fase af processen er omfattende.

...modtages udtræk...

...der omdannes til statistikregistre

Fra virkelighedens fænomener til statistik

Figur 3.6



Statistikregistret skal kunne beskrive en del af "verden". Som man kan se af figur 3.6, er der lang vej fra "verden" til statistikregistret, og problemerne er mange.

Tidsproblemer

Ofte er det et problem, at man kun har oplysninger for givne perioder eller tidspunkter. Opgaven kan være

- at beskrive forløb ud fra to statusudtræk via tilgang og afgang eller
- at sammenligne statusdata fra et tidspunkt med forløbsdata over en periode.

Eksempel 2: Oplysninger om sygehusbenyttelse

Sygehusbenyttelsesstatistikregistret indeholder oplysninger om de personer, der i et kalenderår udskrives fra somatiske heldøgnsafdelinger. Kilderegistret er Sundhedsstyrelsens Landspatientregister, der modtager oplysninger fra de enkelte sygehuse. Registret indeholder ikke oplysninger fra psykiatriske afdelinger. Oplysninger fra skadestuer og ambulatorier indberettes endnu kun delvist.

Fra Landspatientregistret foretages et årligt udtræk. Oplysningerne vedrører de personer, der i årets løb er udskrevet fra en somatisk heldøgnsafdeling, mens udskrivninger fra skadestuer og ambulatorier endnu ikke tages med, da oplysningerne først fra 1995 vil være fuldt indberettet. Enhederne er: *personer* og *afdelinger*; mellem en person og en afdeling eksisterer relationen: *udskrivning* (og naturligvis indlæggelse). Se figur 3.7.

Enhedsgraphen i figur 3.7 viser følgende enheder og relationer:

Enheder:

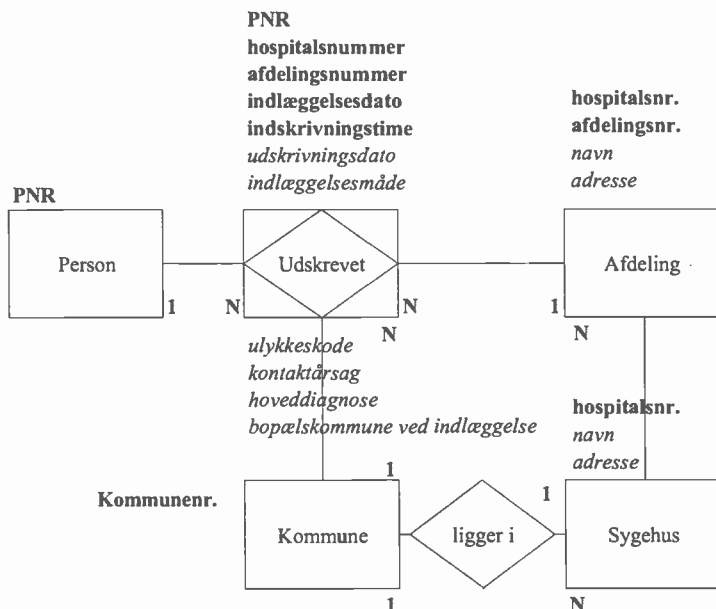
- Personer (ca. 700.000)
- Sygehuse (ca. 90)
- Afdelinger (ca. 500)
- Udskrivninger (ca. 1 mio.)

Relationer:

- Et sygehus kan have mange afdelinger
- Personer udskrives fra én afdeling
- En person kan være udskrevet/indlagt flere gange og på forskellige afdelinger
- En afdeling kan have mange personer udskrevet/indlagt

Enhedsgraf for udtræk fra Landspatientregistret

Figur 3.7



Befolkningsstatistikregistret består af et statusregister over befolkningen, samt et bevægelsesregister. Oplysningerne i statusregistret baseres på udtræk fra CPR, og indeholder oplysninger om de faktisk tilstedeværende personer i Danmark pr. 1. januar i et givet år. Fra udtrækket kan der på grundlag af personernes egenskaber dannes familier og husstande. Figur 3.8 viser de væsentlige enheder og relationer for befolkningsstatus.

Eksempel 3: Befolkningsstatistikens statusregister

Enhedsgrafen i figur 3.8 viser følgende enheder og relationer:

Enheder (1992):

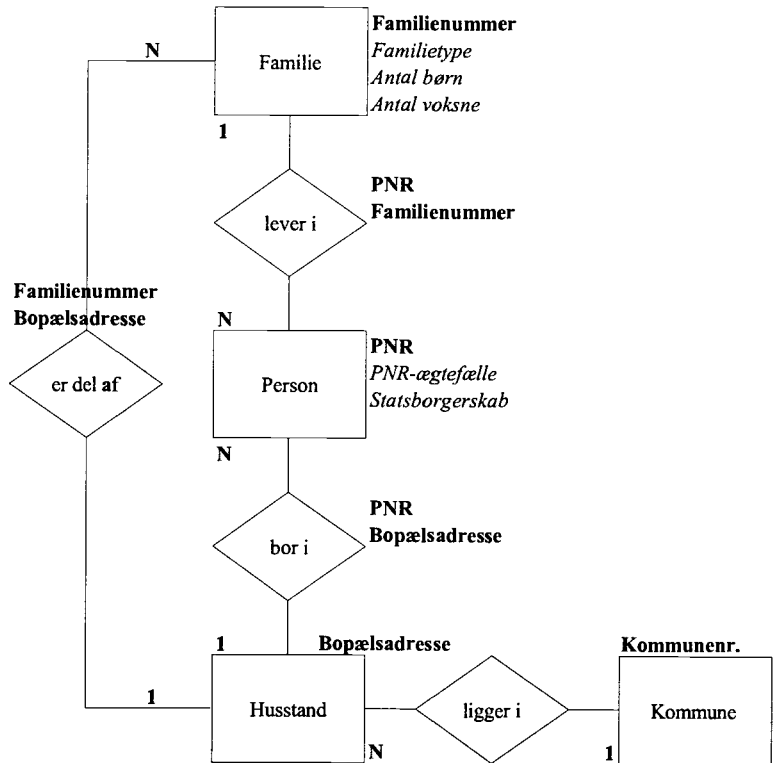
- Personer (ca. 5.150.000)
- Familier (ca. 2.800.000)
- Husstande (2.300.000)

Relationer:

- Én familie kan bestå af flere personer
- Én husstand kan bestå af flere personer
- Én husstand kan bestå af flere familier
- Én husstand har én adresse i én kommune

Figur 3.8

Befolkningen pr. 1. januar



Sygehusbenyttelsesstatistikregistret dannes ved samkøring

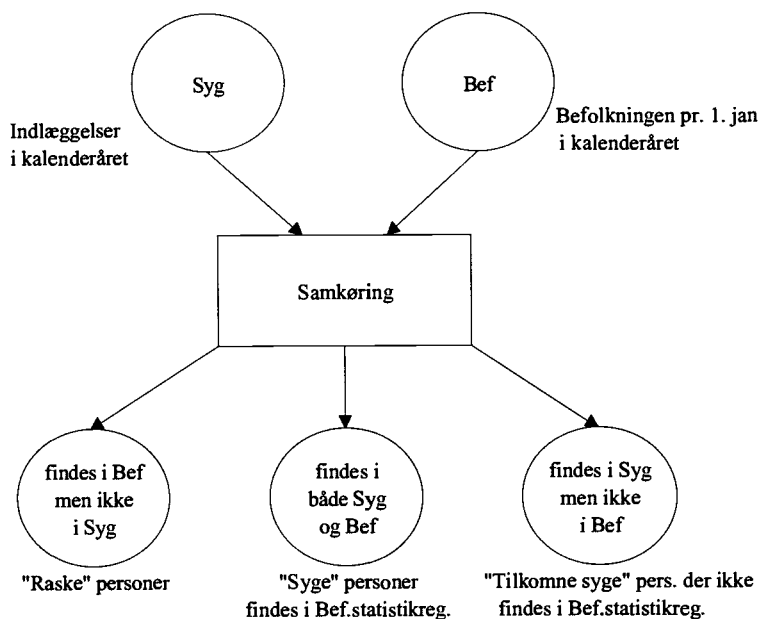
Et eksempel på samkøring er dannelse af Sygehusbenyttelsesstatistikregistret, hvor et udtræk fra Landspatientregistret om personers udskrivninger fra somatiske sengeafsnit (se figur 3.7) samkøres med en række baggrundsregistre i Danmarks Statistik, bl.a. Befolkningsstatistikregistret (statusregistret, se figur 3.8), Bygnings- og boligstatistikregistret, Uddannelsesklassifikationsmodulet mfl. Ved samkøringen integreres oplysninger om personers sygehusbenyttelse med deres familiemæssige, boligmæssige, uddannelsesmæssige forhold osv., og man kan på denne måde sammenligne levevilkår og sygelighed for forskellige grupper i den danske befolkning i den pågældende periode.

Figur 3.9 viser et overordnet systemdiagram for samkøring mellem udtræk fra Landspatientregistret og Befolkningsstatistikregistret. Cirklerne i systemdiagrammet repræsenterer datasæt. Udtræk vedrørende indlæggelser samkøres med udtræk fra Befolkningsstatistikregistret. Der dannes tre datasæt:

- *Raske.* Personer i Befolkningsstatistikregistret pr. 1. januar, der ikke findes i udtrækket fra Landspatientregistret, dvs. personer, der ikke har været udskrevet fra et sygehus i årets løb.

- *Syge*. Fællesmængden, dvs. personer, der findes i begge registre. De er udskrevet fra en sygehusafdeling én eller flere gange i løbet af året.
- *Tilkomne syge*. Personer i sygehusudtrækket, der ikke findes i befolkningen pr. 1. januar. Denne gruppe består af personer, der er indvandret eller født siden 1. januar. Det kan oplyses, at i 1990-versionen var der ca. 64.400 "sygehuspersoner", der ikke havde bopæl i Danmark ved starten af året. Langt hovedparten, nemlig 61.700, var født i årets løb.

Systemdiagram for samkøring mellem sygehusindlæggelser i kalenderåret og befolkningsudtræk pr. 1. januar **Figur 3.9**



Resultatet af samkøringen kan indholdsanalytisk beskrives på flere måder. Man kan for eksempel vælge at formulere resultatet som subtyper af persongrupper. Tilsammen udgør de tre grupper hele populationen. For hver subtype kan der dannes en enhedsgraf ved at kombinere de to enhedsgrafer i figur 3.7 og 3.8.

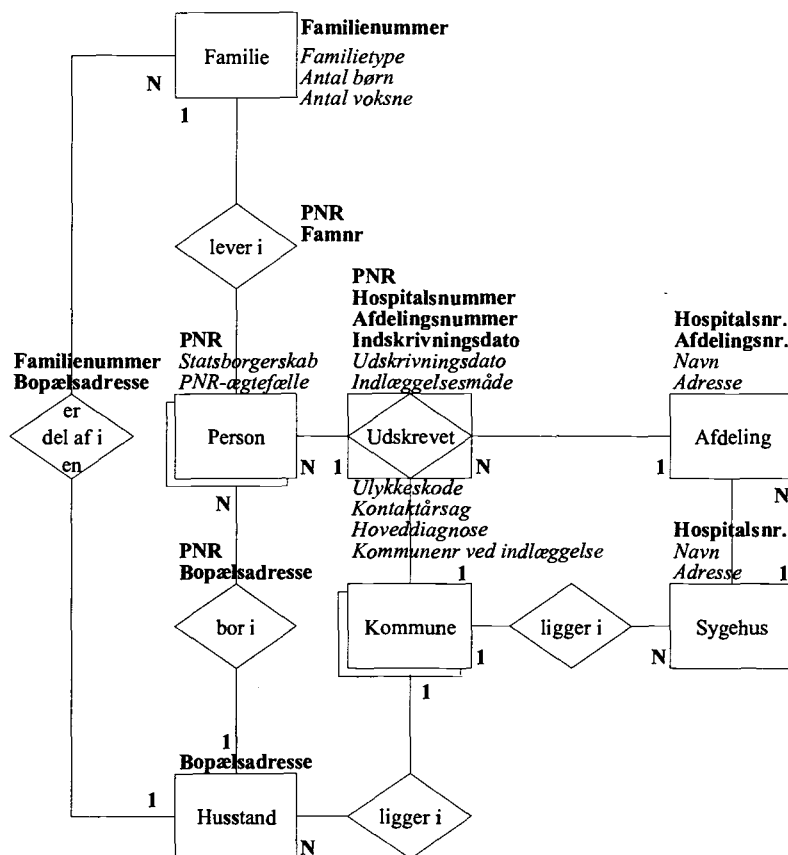
Den mest fuldstændige enhedsgraf findes for subtypen "syge" fundet i Befolkningsstatistikregistret (jf. figur 3.10), som er resultatet af den logiske samkøring og viser hvilke relationer, der findes mellem enhederne i de to registre.

De to andre subtyper består hver af én basisenhedsgraf, som den fremgår af figur 3.7 og 3.8. Man kan kombinere de tre subtypegrafer i ét totalbillede.

Som det fremgår af enhedsgrafen i figur 3.10, får man ved integrationen dannet forbindelseslinier mellem de hidtil adskilte enheder via de fælles nøgler. Herved bliver de to registres indhold til en fælles informationsmængde, og helt nye og spændende sammenhænge kan afdækkes. For eksempel kan man undersøge sammenhængen mellem sygehusindlæggelser og familie- og husstandsmønstre via relationen fra personenhed til hhv. husstandsenhed og familieenhed.

Figur 3.10

Enhedsgraf for integration mellem Sygehusbenyttelsesstatistikregistret og befolkningsstatusregistret



Supplering af data

Dannelse af Sygehusbenyttelsesstatistikregistret er et eksempel på en samkøring, der har som et væsentligt formål at supplere et register med data fra et andet. I Sygehusbenyttelsesstatistikregistret er udtrækket fra

Landspatientregistret - ud over Befolkningsstatistikregistret - integreret med Arbejdsklassifikationsmodulet, Bygnings- og boligstatistikregistret, Det medicinske fødsels- og dødsfaldsstatistikregister, Uddannelsesklassifikationsmodulet, Sygesikringsstatistikregistret og Registret over indkomsterstøttende ydelser. Denne omfattende integration af oplysninger giver mulighed for at belyse et meget stort antal spørgsmål.

Andre formål med samkøringer kan være kontrol af validitet og afgrænsning af en population.

Ved at samkøre to registre, om hvilke det vides, at visse oplysninger i det ene skal genfindes eller kunne beregnes i det andet, opnås mulighed for at checke validiteten af de pågældende oplysninger. I ovennævnte eksempel vedrørende dannelse af Sygehusbenyttelsesstatistikregistret ved man fx, at sygehusbenytterne skal findes i Befolkningsstatistikregistret, og ved samkøringen kontrolleres deres identifikation. Dette gælder dog ikke for personer, der er født eller indvandret i årets løb, hvis identifikation kontrolleres via Befolkningsstatistikregistrets bevægelsesregister.

Kontrol af validitet

Ved samkøring af to registre kan der foretages en afgrænsning af en population. Ved hjælp af kombinerede egenskaber fra de to registre vil nye populationstyper kunne afgrænses, jf. ovenstående eksempel, hvor befolkningen inddrages i to grupper; sygehusbrugere og andre. Dette giver muligheder for at belyse forskelle mellem grupper.

Afgrænsning af population

Samkøring af registre udgør en metode af fundamental betydning for udnyttelsen af administrative registre til statistiske formål. I planlægningsfasen er det vigtigt at få rede på de implicerede enheder og begreber og formålet med hele processen. Til at opnå dette kan indholdsanalysens metode og begreber benyttes.

Sammenfatning

Kapitel 4. Metode og kvalitet

I dette kapitel diskuteres det, hvilke krav der skal være opfyldt, for at man med held kan basere den officielle, samfundsbelysende statistik på administrative registre, hvilke typer af data, der må anses for helt centrale, og hvordan forskellige datatyper kan indgå i statistikgrundlaget. Desuden behandles nogle praktiske problemer i forbindelse med tilrettelæggelsen af en registerbaseret statistik, og hvordan de kan løses. Endvidere omtales alternative løsningsmuligheder, når de ønskede oplysninger ikke umiddelbart foreligger i de administrative registre.

4.1 Administrative data som statistikdata

De krav, der må stilles til datagrundlaget, dvs. til de administrative registre, har at gøre dels med datas indholdsmæssige dækning, dels med kvalitative forhold. Det er umuligt at opstille helt præcise krav, som under alle omstændigheder skal opfyldes, før det er frugtbart at producere statistik på grundlag af administrative registre. Det er da også klart, at selv isolerede, enkeltstående administrative registre kan anvendes og bliver anvendt over alt i verden som statistikgrundlag, og nogle af de ældste statistikker i Danmark har været baseret på optegnelser som fx kirkebøgerne.

I det følgende opstilles nogle krav, som skal være opfyldt, for at det skal være muligt at skabe et *sammenhængende, dækkende personstatistisk system*, som i hovedsagen er baseret på registrene, dvs. et system af samme karakter som det danske.

4.1.1 Indhold

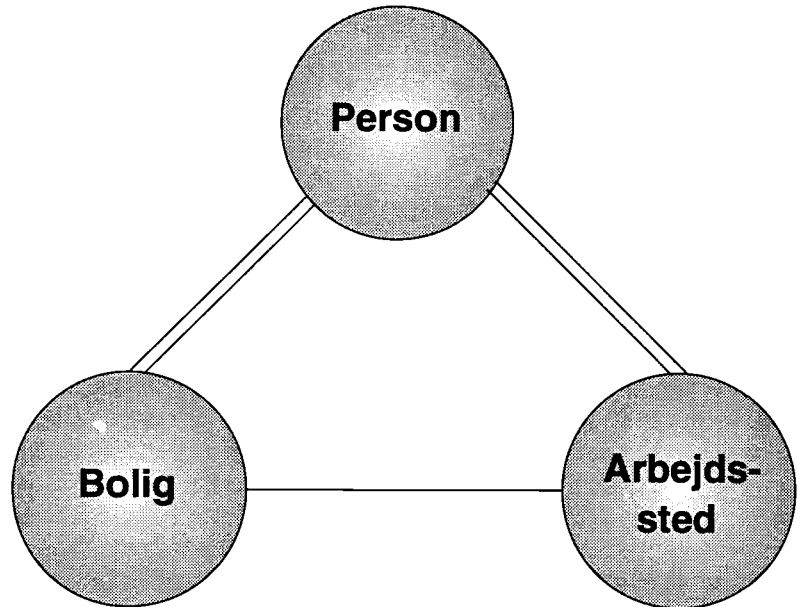
Først og fremmest skal registrene indeholde data, der dækker de vigtigste emner i den samfundsbelysende statistik.

Derudover er der mange forskellige enheder, som det er ønskeligt at kunne belyse i statistikken, men der er tre centrale enheder, der er væsentlige for statistikkens strukturering: Personer, virksomheder og boliger. I personstatistikken er personerne selvfølgelig i centrum, men også de to andre enheder har stor betydning inden for dette statistikområde. Det fremgår fx af, at en traditionel folke- og boligtælling også knytter an til de boliger, hvor personerne bor, og de virksomheder (arbejdssteder), hvor de arbejder. En stor del af statistikbrugerne, både i den offentlige og den private sektor, har brug for netop disse relationer.

Enheder

Figur 4.1

Identifikationsrelationer



Der er naturligvis mange andre enheder, det er vigtigt at have registreringer om. Nogle af disse enheder er *hændelser*, som vedrører personerne, fx de klassiske demografiske hændelser (fødsler, dødsfald osv.) eller sociale begivenheder af forskellig slags. Blandt andre enheder af betydning kan nævnes biler.

Præcis definition af enheder

For den statistiske anvendelse er det vigtigt, at enhederne er veldefinerede. Dette er ikke noget problem for "naturlige" enheder som personer eller biler, hvor definitionen og afgrænsningen over for andre enheder så at sige giver sig selv. Også begrebet bolig er nogenlunde let at håndtere, selv om der er nogle grænsetilfælde, hvor eksistensen eller afgrænsningen af en bolig kun kan fastslås efter nærmere fastlagte regler. Nogle enheder fremkommer som kombinationer af "naturlige" enheder, fx familier og husstande. Problemer i den forbindelse er omtalt i eksempel 7.

Det bliver hurtigt mere problematisk, når man ser på andre typer enheder, og her er erhvervsenheder et godt eksempel. En virksomhed eller et arbejdssted er et abstrakt begreb. Eksistens og afgrænsning kan kun konstateres efter meget nøje fastlagte regler. Uklarhed og uenighed om disse definitioner har gennem årtier i høj grad vanskeliggjort udnyttelsen af registre i erhvervsstatistikken og har i øvrigt også ødelagt sammenhæng og konsistens i statistikken.

Registrene skal gerne indeholde oplysning om alle enheder i landet af de nævnte typer, dvs. de skal være totaldækkende. En ufuldstændig dækning vil ofte være systematisk, fx således at visse kommuner ikke indgår, hvilket betyder, at statistik på et sådant grundlag umiddelbart vil indeholde skævheder, som det kan være vanskeligt at kompensere for.

**Registrene skal
gerne være
totaldækkende**

En af de store styrker ved den registerbaserede statistik i forhold til den traditionelle, spørgeskema-baserede statistik er netop, at man ofte har mulighed for at få totaldækning, eller er meget tæt på dette. I spørgeskemaundersøgelser må man dels som regel nøjes med en repræsentativ stikprøve, dels acceptere et uundgåeligt bortfald, der oftest er skævt, således at de, der indgår i undersøgelsen, adskiller sig systematisk fra dem, der ikke indgår. Bortfaldet har ofte en betydelig størrelse, typisk fra 15 til 50 pct., når det drejer sig om frivillige undersøgelser blandt privatpersoner. I nogle undersøgelser i udlandet ser man endog bortfald på op til 80 pct., hvilket i de fleste sammenhænge gør resultaterne helt værdiløse²⁰.

**Totaldækning er
værdifuld...**

Det er en vigtig forudsætning for en dækkende statistik på registergrundlag, at dataindholdet i de administrative registre er bredt og omfatter mange variable om de relevante enheder. Mangler fører til supplerende indsamling af oplysninger efter traditionelle metoder eller til begrænsninger i statistikens indhold.

...ligesom bredde

4.1.2 Identifikationer

Identifikationer spiller en meget stor rolle, både for den administrative registerdrift (jf. kapitel 2) og for den statistiske udnyttelse. For statistikken er det især vigtigt, at fælles identifikationer anvendt i flere registre kan bruges til at sammenknytte oplysninger fra de forskellige kilder om én enhed. Da de administrative registre ikke er dannet til statistiske formål, og derfor oftest ikke umiddelbart indeholder de ønskede datakombinationer, er det et vigtigt led i dannelsen af næsten enhver registerbaseret statistik, at data fra forskellige registre samkøres ved hjælp af fælles nøgler.

Den vigtigste nøgle i denne forbindelse er naturligvis *personnummeret*. *Beliggenhedsadressen* og *arbejdsstedsnummeret* er også af afgørende betydning for at fastholde relationerne mellem personer, boliger og arbejdssteder, jf. figur 4.1 ovenfor.

²⁰ Jf. fx *Van Bastelaer (1993)*.

4.1.3 Tidsreferencer

Tidsdimensionen spiller en ganske særlig rolle i den samfundsbelysende statistik, og inden for alle emneområder er det nødvendigt, at man kan foretage sammenligninger over tid. I statistiske publikationer eller databanker er tidsdimensionen således stort set altid med. Derfor er det helt nødvendigt for den statistiske anvendelighed, at pålidelige oplysninger om forskellige datoer findes i registrene.

Hændelsesdatoer er nødvendige...

Først og fremmest skal datoen for ændringer eller hændelser indgå. Nogle af de vigtigste er datoerne for enhedernes "fødsel" og "død", men også tiden for andre ændringer vedrørende enhederne har betydning. Der tænkes her på det reale tidspunkt, hvor en hændelse har fundet sted, der ændrer ved et datum, fx dato for en flytning eller dato for et brancheskift for en virksomhed.

...men ikke altid mulige at fastsætte

Det kan i praksis være overordentlig vanskeligt at få fastslået den reale hændelsesdato. Med branche for en virksomhed er der tale om en variabel, som så at sige ændres gradvist, efterhånden som virksomhedens aktiviteter skifter, og det er måske ikke muligt at fastslå, hvornår en ny aktivitet er den dominerende. Men også for mange andre variable, som i princippet kan observeres på ethvert tidspunkt, og hvor en ændring derfor kan tidsfæstes ganske nøjagtigt, er der i praksis problemer med registreringen. Det kan skyldes, at det ikke nødvendigvis er interessant for den administrative brug af data, hvornår en dataværdi er trådt i kraft, blot man på ethvert tidspunkt har adgang til den mest aktuelle værdi. I andre tilfælde kan problemerne skyldes, at det ikke er muligt eller praktisk gennemførligt at få registreret hændelserne i real tid, men at man i stedet må nøjes med at få kendskab til dataændringer ved tilbagevendende forespørgsler, fx en gang om året i forbindelse med en selvangivelse.

Hvis der ikke er mulighed for at få registreret nøjagtige datoer, er tilnærmelser selvfølgelig anvendelige og langt bedre end ingenting.

Registreringsdatoen er næsten lige så vigtig

I tillæg til hændelsesdatoer eller ændringsdatoer er der brug for *registreringsdatoer*, dvs. angivelse af hvornår den pågældende dataværdi er blevet optaget i registret. Dette har afgørende betydning, når registret skal opfattes som en *statistisk model* af den virkelighed, som statistikken skal beskrive, jf. afsnit 4.1.4.

Det ideelle er derfor, at enhver oplysning i det administrative register er ledsaget af to datoer: En *ikrafttrædelses- eller hændelsesdato* og en *registreringsdato*. Virkelighedens registre adskiller sig ofte stærkt fra dette ideal.

Foreløbige og endelige tal

Forsinkelser i opdateringen af de administrative registre fører i mange tilfælde til, at man må publicere flere udgaver af tallene: *Foreløbige*

tal, der er nødvendige af hensyn til behovet for aktuelle tal, men som mangler præcision på grund af manglende opdateringer, og *endelige tal*, som er mere fuldkomne, men som først fremkommer på et senere tidspunkt. Flere mellemversioner kan forekomme. Eksistensen af flere versioner af tallene kan imidlertid forårsage stor forvirring og ulempe hos brugerne og bør af den grund undgås, hvis det er muligt.

4.1.4 Registret som statistisk model

Når et registersystem skal bruges som grundlag for statistik, betyder det, at registret skal kunne opfattes som en model af den gruppe enheder, statistikken skal beskrive.

I nogle tilfælde må det yderligere forudsættes, at de dataændringer, der indføres i registret, kan opfattes som en model af hændelser, der beskrives i statistikken. Det gælder fx befolkningsstatistikken, hvor opgørelser af hændelser i befolkningen (flytninger, vielser etc.) skal kunne udledes af dataændringer i CPR. Registerne skal derfor være indrettet således, at opgørelser af registrets bestand og af bevægelserne heri tidsmæssigt kan afgrænses i overensstemmelse med de tidspunkter, hvor hændelserne faktisk har fundet sted.

Til bevægelsesopgørelser kræves det, at det er muligt i registret at skelne virkelige hændelser fra korrektioner og mere teknisk betonedede dataændringer. Indholdet af dette sidste krav kan bedst belyses gennem nogle problemer, som rent faktisk har gjort sig gældende ved befolkningsstatistikens anvendelse af CPR, selvom netop CPR må siges at være et registersystem, der på et meget højt niveau imødekommer statistikproduktionens behov.

Opgørelserne over indenlandske flytninger tager udgangspunkt i oplysninger om de personer, som i en given periode får ændret adresseangivelse i CPR. I denne gruppe indgår de personer, som faktisk er flyttet i perioden, men der indgår tillige en række persongrupper, som ikke har interesse for statistikken, nemlig bl.a.:

- personer, der er omfattet af grænsereguleringer mellem kommunerne eller kommunesammenlægninger
- personer, hvis adresse ændres som følge af en vejnavneændring eller en omnummerering eller omkodning af vejen
- korrektion af forkerte adressekoder

Navnlig adresseændringer som følge af vejnavneændring m.m. har haft et betydeligt omfang, som frem til slutningen af 1970'erne blev anslået til ca. 1/4 mio. personer årligt. Selv om kommunerne efter reglerne skulle give indberetningerne af disse adresseændringer en særlig markering, skete det kun i knap halvdelen af tilfældene, og af denne grund

Registret som en model af enheder...

...og en model af hændelser

Nødvendigt at kunne skelne mellem korrektioner og hændelser

Eksempel 4: Korrektioner og hændelser i CPR

var det ikke muligt at fremstille en brugbar generel statistik over samtlige indenlandske flytninger, men kun over den del af flytningerne, der passerede kommunegrænser²¹. Disse problemer blev stort set løst ved en teknisk omlægning af CPR, som fandt sted i 1978.

De egentlige korrektioner af forkert indberettede data frembyder imidlertid et problem for statistikken, selv om man er i stand til at skelne dem fra de virkelige hændelser. Korrektioner må nemlig i mange tilfælde betragtes som rettelser til tidligere indberettede hændelser: En adressekorrektion kan være en rettelse til en indberetning om flytning. En korrektion kan også annullere en tidligere indberettet hændelse, hvilket i CPR-systemet navnlig finder sted, når en person, efter at have meldt udvandring til folkeregistret, ombestemmer sig og bliver i landet.

Korrektioner til hændelser er vanskelige at behandle, fordi de i realiteten kan blive indberettet med vilkårlig lang forsinkelse. I befolkningsstatistikken har man valgt at holde dem helt uden for opgørelserne af bevægelserne, men denne fremgangsmåde er kun brugbar, fordi korrektionernes antal er ringe, sammenlignet med hændelsernes antal.

4.1.5 Stabilitet

Stabilitet over tid er nødvendigt

Det er en vigtig egenskab ved statistikken, at den kan beskrive tidsforløb, dvs. hvordan en bestemt størrelse udvikler sig fra måned til måned, fra år til år. Derfor er det også meget vigtigt, at begreberne i de administrative registre ligger fast over længere perioder. I modsat fald opstår der ofte store problemer med at opnå sammenlignelige tal fra den ene periode til den næste. I nogle tilfælde, men langt fra alle, har man dog mulighed for at korrigere mere eller mindre præcist for ændringerne. Der er også eksempler på, at man har måttet opgave at offentliggøre en statistik i en periode, fordi det har været umuligt at korrigere for konsekvenserne af ændringer i registrene, jf. afsnit 4.1.7.

4.1.6 Kvalitet

De kvalitetskrav den statistiske anvendelse stiller, falder i nogen grad sammen med de krav, som også må være opfyldt af hensyn til registrenes primære formål.

²¹ Adresseændring fra én kommune til en anden vil normalt ikke kunne opstå som et resultat af en omkodning af vejen. I de relativt sjældne tilfælde, hvor en kommunegrænse ændres, eller hvor fx to kommuner lægges sammen, vil statistikerne normalt have kendskab til ændringen på forhånd og kan derfor tage højde for den i statistikken.

Det første kvalitetskrav man må stille til grunddata er, at de skal være *relevante* for statistikken. Med dette menes, at data skal handle om de begreber, der ønskes belyst i statistikken. Det er væsentlig lettere at kontrollere relevansen i traditionel statistik, som baseres på data, der er specielt indsamlet til formålet. Når statistikken skal baseres på data, der er indsamlet til andet formål, må man som oftest lade sig nøje med det, som er nødvendigt eller ønskeligt til dette formål.

I gunstige tilfælde kan statistikerne dog have held til at påvirke indholdet i registrene, jf. afsnit 4.3.

Der er også mulighed for at lade forskellige datakilder supplere hinanden gennem samkøring, således at man danner den ønskede oplysning ved at sammenholde en række oplysninger fra forskellige registre, der mere eller mindre perfekt belyser forholdet. Anvendelsen af flere datakilder behandles nærmere i afsnit 4.2.

I forbindelse med diskussionen om relevans må man være opmærksom på risikoen for, at et statistisk system, der udelukkende eller i meget høj grad er baseret på offentlige myndigheders registre, kan komme til at give et verdensbillede, der er præget af myndighedernes briller. Begreber eller persongrupper, som ikke eksisterer i skattevæsenets eller socialforvaltningens referenceramme, kommer måske slet ikke med i statistikken. Det kunne fx være hjemløse personer.

For at imødegå denne risiko er det vigtigt, at der eksisterer nogle basisregistre som fx CPR, som har til opgave at registrere alle enheder uden at tænke på noget specifikt administrativt formål. Endvidere er det en sikkerhed, at mange registreringer fra forskellige myndigheder kan samkøres, så man kan få en statistik, der ikke er bundet til en enkelt administrativ synsvinkel. Alligevel er det vigtigt at være opmærksom på, at registrene ikke kan eller skal give os hele samfundsbilledet, og at der er behov for at indsamle data specielt til at belyse de sider, som ikke dækkes af registrene. Disse ekstra data kan indsamles gennem interview- eller spørgeskemaundersøgelser, og de bør ideelt indeholde identifikationer, som gør det muligt at sammenholde dem med registerdata. Dette omtales mere detaljeret i afsnit 4.5.

Det næste krav som stilles til statistikkens grunddata, drejer sig om pålidelighed. Det gælder uanset om data kommer fra spørgeskemaer eller registre. Der må være en høj grad af sikkerhed for, at de registrede data belyser de faktiske forhold. Hvis det fremgår, at en person er maler, og han faktisk er rigsstatistiker, risikerer man at drage fejlagtige konklusioner på baggrund af den færdige statistik.

Det kan ganske vist indvendes, at så længe fejlene ikke er systematiske, vil de ikke resultere i skæve randfordelinger, da modsatrettede fejl

Grundoplysningerne skal være relevante

Risiko for et skævt verdensbillede...

...mindskes ved eksistensen af basisregistre

Grundoplysningerne skal være pålidelige

Systematiske og ikke-systematiske fejl

har en tendens til at ophæve hinanden. Men når man, som det ofte er nødvendigt, ser på *statistiske sammenhænge* mellem flere variable, bliver det et problem med sådanne fejl, idet de kan forstyrre bedømmelsen af sammenhængene.

Hvis fejlene systematisk går i en bestemt retning, bliver problemerne umiddelbart forstærket, men systematikken kan medvirke til, at det bliver lettere at korrigere for dem.

Grundoplysningerne skal registreres med passende præcision

Endelig er det vigtigt, at data registreres med en præcision, der passer til statistikkenes behov, dvs. at den skala, der benyttes, er tilstrækkelig fin. Som eksempel herpå kan nævnes, at specifikationsgraden af stillingsoplysninger skal være temmelig fin, hvis den skal kunne bruges til epidemiologiske undersøgelser, hvor sygdomsrisiko skal ses i forhold til arbejdsspecifikke påvirkninger.

Kravet til præcisionen gælder også selvom oplysninger ikke skal anvendes i statistiske opgørelser på det registrerede niveau. I mange tilfælde foretages grupperinger på grundlag af variable, der er afledt med udgangspunkt i elementaroplysningerne. Definitionen og tilpasningen af sådanne variable til det statistiske behov er meget afhængig af præcisionen i grundmaterialet.

4.1.7 Ændringer i de administrative data

Som omtalt kan det give anledning til store problemer for statistikken, når der sker ændringer i dataindholdet i de administrative registre, som følge af lov- eller regelændringer. Dels kan det være vanskeligt eller umuligt at bedømme den tidsmæssige udvikling i en bestemt størrelse, hvis der pludselig bruges andre definitioner i grundlaget, dels kan der opstå problemer med at fastslå, hvilke ændringer i dataværdier, der skal opfattes som udtryk for egentlige hændelser, og hvilke, der kun afspejler nye begreber eller definitioner.

Dataændringer som følge af lovændringer

Ændringerne i de administrative registre kan skyldes ændringer i den gældende lovgivning på området og konsekvenserne for statistikken afhænger af, hvilken type statistik der er tale om. Hvis der netop er tale om en *lovbelysende statistik*, som skal vise hvordan loven administreres i forhold til fx borgere, må statistikken naturligvis blot følge med og adoptere den nye lovgivnings begreber. Statistikken om kontanthjælp må således til enhver tid afspejle de regler, der gælder på området, og det kan så være vanskeligt at bedømme fx adfærdsmæssige ændringer.

Hvis det derimod drejer sig om en mere *generel statistik*, kan det ikke accepteres, at statistikkenes begreber fra før og efter lovændringen ikke kan sammenlignes. Den generelle statistik forsøger at belyse nogle

begreber, der ikke er fastlagt i lovgivningen, fx arbejdsløshed, som er defineret som begreb i en international konvention²². Hvis statistikken bygger på udbetaling af forskellige ydelser (fx dagpenge), og reglerne for disse ændres, kan det være vanskeligt eller umuligt at kompensere for det i statistikken. Man må så i det mindste forsøge at opstille skøn over, hvor meget ændringerne har betydet, så en vis sammenkædning af tidsserierne muliggøres.

Et eksempel herpå er den konjunkturbelysende statistik over beskæftigelsesudviklingen baseret på arbejdsgivernes samlede ATP-indbetalinger²³. Det væsentligste formål med denne statistik er at beskrive korttidsudviklingen, og det er derfor afgørende at eliminere eller reducere betydningen af regelændringer på målingen af beskæftigelsen.

I 1977 blev ATP-medlemskredsen udvidet bl.a. med værnepligtige og 16-17 årige lønmodtagere. For at kunne vurdere beskæftigelsesudviklingen fra kvartalerne i 1977 til 1978, blev det nødvendigt at korrigere tallene; grundlaget for denne korrektion var resultater fra arbejdsstyrkeundersøgelserne, en stikprøvebaseret postskemaundersøgelse.

Et andet og mere subtilt eksempel på betydningen af regelændringer i denne statistik er den lovmæssige sammenkobling af ATP-indbetaling og kompensation for indbetalte arbejdsmarkedsbidrag²⁴ i 1988. Der syntes i perioden indtil 1988 at være en vis tilbøjelighed hos nogle arbejdsgivere til at udskyde en større del af den samlede årlige indbetaling til 4. kvartal, end der relativt skulle forventes. Denne udskydelse havde ingen konsekvenser i form af bøder, renter eller lignende, og det var derfor et fordelagtigt lån. Med indførelse af kompensation for indbetalte arbejdsmarkedsbidrag til arbejdsgiverne på basis af indbetalte ATP-bidrag blev det en fordel for arbejdsgiverne at indbetale rettidigt. Der skete derfor en ændring af indbetalingsmønstret, som gjorde det problematisk at vurdere beskæftigelsesudviklingen på det hidtidige grundlag. På basis af en række undersøgelser og vurderinger blev det derfor nødvendigt at korrigere tallene på kvartalsbasis.

I nogle få tilfælde har det ikke været muligt at vurdere betydningen af ændringerne tilstrækkeligt præcist, og det har, bl.a. i forbindelse med omfattende regelændringer i 1993 og 1994, været nødvendigt midlertidigt at indstille beskæftigelsesstatistikken på grundlag af ATP-indbetalingerne.

Der kan naturligvis også være tale om ændringer i andre administrative regler end love, hvilket kan give anledning til ganske lignende problemer. Eksempelvis har det i ATP-statistikken i flere kvartaler været

Eksempel 5: ATP-statistikken

Andre regelændringer

²² Jf. ILO (1982).

²³ Bidrag til arbejdsmarkedets tillægspensionsordning.

²⁴ En række arbejdsgiverbidrag samledes i 1988 til ét bidrag, AMBI.

nødvendigt at tage forbehold og foretage korrektioner som følge af antallet af *lønudbetalingstidspunkter*. ATP-beløb bliver i de fleste tilfælde beregnet i forbindelse med lønudbetalinger, og antallet af lønudbetalingstidspunkter i et kvartal kan for ugelønnede variere mellem 12 og 14. Et andet eksempel på andre ændringer end lovmæssige med betydning for ATP-indbetalingerne er overenskomstmæssige ændringer. Fx betød ferieforlængelsen fra 4 til 5 uger i 1981-82 en nedsættelse af det årlige ATP-bidrag for ugelønnede, men ikke for månedslønnede. Da det i forbindelse med ATP-indbetalingerne ikke er muligt at skelne mellem den ene og den anden type lønmodtagere, betød denne ændring alt andet lige en "nedgang" i beskæftigelsen. Lignende forhold har gjort sig gældende ved arbejdstidsnedsættelser kombineret med ændringer i timegrænserne for ATP- indbetaling.

Systemomlægninger Endelig kendes der eksempler på, at de administrative registre ændrer sig, uden at det skyldes ændringer i de grundlæggende regler. Typisk drejer det sig om ændringer foretaget af tekniske grunde eller for at opnå rationaliseringsgevinster. Det kan fx være problematisk, hvis man beslutter, at et bestemt datum ikke længere er nødvendigt.

Systemomlægninger kan også have en mere lokal karakter, som når der fx ændres på en kommunegrænse eller på husnummereringen på en vej, hvorved et vist antal personer får en ny adresse, uden at der er tale om nogen flytning. Her vil det være nyttigt, hvis ændringerne i det mindste kan identificeres som korrektioner.

Konsekvenser for statistikken Det er klart, at de omtalte ændringer er alvorlige for statistikken, hvis de gennemføres uden tilstrækkelig hensyntagen til denne. Det er derfor vigtigt, at statistikerne er med i forberedelsen af ændringerne, så man i det mindste kan blive opmærksom på konsekvenserne af disse, inden de foretages.

4.1.8 Statistikernes mulighed for at påvirke data

Ændrede muligheder ved overgangen til registerbaseret statistik

Lige siden den registerbaserede statistiks opblomstring fra 1970 har det været anerkendt som et stort problem, at man ikke havde den kontrol over indholdet i grunddata, som man var vant til, eller troede man havde, i den skemabaserede statistik. Man kan ikke være sikker på, at registrene dækker lige præcis de enheder, der er relevante, eller at data er defineret i overensstemmelse med brugernes behov. Yderligere problemer kommer som nævnt til, når registrenes indhold ændrer sig over tid, og paradoksalt nok kan der opstå et problem, hvis registrene bliver mere pålidelige, for også i det tilfælde får man et databrud.

Det er naturligvis ønskeligt, at statistikerne har en vis indflydelse på dataindholdet, men samtidig må det erindres, at registrene føres med ganske bestemte administrative formål for øje, og "registerførerne" har

det som deres opgave at opfylde disse formål mest effektivt. De kan derfor ikke lytte alt for meget efter ønsker, der "kun" tjener statistiske formål.

I Danmark er det imidlertid bestemt i *Lov om Danmarks Statistik*, §1²⁵, at den, som planlægger at oprette et register, skal drøfte planerne med Danmarks Statistik, for at registret kan blive mest muligt egnet som grundlag for statistikken. Der er da også et ganske godt og tæt samarbejde mellem Danmarks Statistik og registerførerne på mange områder. Danmarks Statistik sidder som medlem af brugerkomitéer for både statslige og fælleskommunale systemer, og planer om registrenes udvikling diskuteres normalt også i de rådgivende udvalg, som Danmarks Statistiks styrelse har nedsat på mange statistikområder. Herigennem bør det i hvert fald kunne sikres, at man kender statistikken ønsker, før man foretager ændringer, og at man ved opbygning af nye systemer får drøftet, om der kan opnås statistiske fordele gennem mindre justeringer.

Man må imidlertid ikke forvente, at statistikerne kan stille og få opfyldt meget betydelige krav om ekstra data, andre definitioner eller lignende. Registerførerne må se nøje på deres egen effektivitet og ressourceindsats. Statistikerne må derfor være meget beskedne og kun fremføre større ønsker, når der er en meget stor samfundsmæssig gevinst at hente ved en tilpasning af registrene. Der kendes kun nogle få eksempler på, at ekstra data indsamles gennem registrene udelukkende af hensyn til statistikken, jf. afsnit 4.3 om integreret dataindsamling.

Beskedenhed er en dyd

4.1.9 Dannelse af statistikregistre

Inden de administrative grunddata kan benyttes til statistik, skal de underkastes en statistisk databehandling, som skal gøre dem velegnede til formålet. Som resultat af denne proces indordnes data i nogle *statistikregistre*, som er datasamlinger med identifikationsnøgler (altså registre), hvor populationen og dataindholdet i hvert register er afstemt, så det passer med ét statistikområdes behov, fx befolkningsstatistikens.

Data i et statistikregister kan komme fra et eller flere administrative registre. Alle registerdata, som kan yde et væsentligt bidrag til at belyse statistikken hjørne af virkeligheden, bør inddrages. Dannelsen af statistikregistret indebærer derfor ofte samkøring af grunddata på enhedsniveau. På basis af de forskellige data skal man forsøge at finde de bedste estimatorer for statistikken variable.

Tidsreferencerne skal bearbejdes, så man får den bedst mulige statistiske model af enhedsbestanden og dens data på de ønskede tids-

²⁵ Jf. Økonomiministeriet (1992a).

punkter eller i de ønskede perioder. De administrative registre følger normalt ikke en sådan model, idet et af formålene ofte er at give de senest tilgængelige data på ethvert område. Det betyder, at nogle data er helt nye, mens andre er noget ældre.

Kontrol af registeroplysningerne

En af de vigtigste og mest ressourcekrævende processer i registerdannelsen er kontrol af de indgående grunddata og rettelse af eventuelle fejl, også kaldet editering. Statistikregistret skal ideelt set være "renset" og konsistent, så der ikke findes indbyrdes modstrid mellem forskellige data. Usandsynlige oplysninger skal undersøges. I denne proces er det ikke usædvanligt, at der afsløres systemfejl i de administrative registre eller i de udtræk, dataleverandøren foretager inden Danmarks Statistik modtager dem.

De formelle krav er ikke opfyldt

Fejlene kan bestå i, at oplysningerne ikke opfylder de formelle krav for den pågældende oplysningstype. I sådanne tilfælde, som ofte optræder gruppevis, skal det afgøres, om den pågældende værdi skal ændres, så oplysningen falder inden for de tilladte intervaller, eller om de pågældende oplysninger ikke kan indgå i statistikgrundlaget, da en korrektion ikke vil kunne foretages på en meningsfyldt måde.

Enkeltoplysninger rettes

Enkeltstående fejl kan i visse tilfælde konstateres og rettes ved direkte opslag i kilderegistrene. Danmarks Statistik har fx adgang til opslag i CPR og Det centrale erhvervsregister.

Da der her er tale om udnyttelse af administrative oplysninger, der ikke er indhentet med det direkte formål at foretage statistiske opgørelser, må der efter registerlovens bestemmelser kun i begrænset omfang rettes henvendelse til den administrative myndighed om den pågældende fejl, da dette vil kunne få indflydelse på den administrative afgørelse for enkeltpersonen.

Samarbejde om rettelse af fejltyper i basisregistrene

Modsat bør man imidlertid påpege, at det er afgørende at eventuelle grundlæggende fejltyper i de administrative grundregistre, hvorfra der udtrækkes data til statistiske formål, bliver rettet. Dette kan kun ske ved en indgående kontakt med de registeransvarlige myndigheder. Der er mange eksempler på, at statistikerne har opdaget fejltyper, som senere er blevet rettet i basisregistrene, til glæde både for basisregisteret og den endelige statistik.

Eksempel 6: Fejl i Personnummeret

En af de alvorligste fejl der kan forekomme, er fejl i identifikationen. Der findes som nævnt særlige sikkerhedsforanstaltninger, som modvirker, at der registreres fejlagtige identifikationer. Men det forekommer, at enhederne tildeles fejlagtige identifikationer, når de skal registreres i basisregistrene. Således har det vist sig, at der i få tilfælde registreres forkerte personnumre. Når dette kan forekomme, skyldes det, at personnummeret ikke overholder det ideale krav om, at identifikationer ikke må indeholde information. Derfor er det nødvendigt at

rette i et personnummer, hvis det viser sig, at en person har fået registreret forkert fødselsdato eller køn som en del af personnummeret.

I mange år anså man dette problem for marginalt eller glemte det helt, men i forbindelse med epidemiologiske forløbsanalyser viste det sig efterhånden, at det blev et problem, man var nødt til at gøre noget ved. I *Lynge & Thygesen (1990)* beskrives en analyse af sammenhængen mellem cancerrisiko og erhvervsdeltagelse, hvor man i en årrække følger en kohorte af personer, som var levende den 9. november 1970, og konstaterer senere diagnosticerede kræfttilfælde fra Cancerregisteret. Når personnummeret ændrer sig, mislykkes koblingen til cancerregistreringerne, og det vil synes som om personernes risiko for cancer er nul. Blandt 270.000 cancertilfælde i 1970-80 konstaterede og rettede man denne fejl i 286 tilfælde. Senere har man foretaget lignende rettelser i forbindelse med analyserne af sammenhænge mellem dødelighed og erhverv, hvor fejlene ellers ville resultere i, at nogle personer syntes at leve evigt.

I det næste eksempel beskrives det, hvordan man af de eksisterende enheder i et register kan aflede nye enheder.

Danmarks Statistik modtager hvert år pr. 1. januar et statusudtræk fra CPR. Det består af en record for hver person, der den 1. januar er tilmeldt CPR, og indholdet beskriver den aktuelle tilstand for personen. Udtrækket underkastes en statistisk bearbejdning, der omfatter kontrol og tidsafgrænsning, og som resulterer i en ny version af Befolkningsstatistikregistrets statusdel. På grundlag heraf fremstilles statistik over befolkningen fordelt på køn, alder, bopælskommune, statsborgerskab og ægteskabelig stilling. Disse variable ligger i udtrækket for hver person klar til brug, eller de udledes af personnummeret.

Udtrækket viderebearbejdes i husstands- og familiemodulet, som er et af Danmarks Statistiks seks klassifikationsmoduler. Resultatet er en husstandsfil, en familiefil og en personfil, alle med husstands- og familieoplysninger. På grundlag af disse filer udarbejdes den årlige husstands- og familiestatistik. Men de dannede filer benyttes også inden for en række andre statistikområder som baggrunds- og klassifikationsoplysninger. Anvendelsen er særlig udbredt inden for socialstatistikken, men også i indkomststatistikken, Forbrugsundersøgelsen, Lovmodellen mfl. benyttes husstands- og familiemodulets klassifikationsvariable.

Dannelsen af familieoplysninger på registergrundlag foregår efter nogle regler, der er knyttet til familiedefinitionen. Den familiedefinition, der for tiden benyttes, er indført pr. 1. januar 1991. Et vigtigt formål med denne definition er at få en bedre belysning af de ikke-gifte par end tidligere.

Eksempel 7:
Familiestatistik på grundlag af udtræk fra CPR.

Familieafgrænsningen foregår på grundlag af følgende variable fra CPR-udtrækket: *bopælsadresse, køn, alder, ægteskabelig stilling, henvisningsnummer* til seneste ægtefælle eller registreret partner, samt *henvisningsnummer* til forældre.

Følgende grundregler gælder for familieafgrænsningen:

- En husstand omfatter alle personer på en adresse. Den består af en eller flere familier, og en familie består af en eller flere personer.
- Alle personer er enten børn, enlige eller danner sammen med en anden person en af de fire partyper, familiedefinitionen opererer med: ægtepar, registrerede partnerskaber, samlevende par og samboende par (partyperne beskrives nedenfor).
- Enlige og personer i par kan have børn, der indgår i familien.
- En familie består af én eller to generationer. I to-generationsfamilier består yngste generation af hjemmeboende børn under 18 år.
- Kun personer, der ikke selv har børn, aldrig har indgået ægteskab, og ikke indgår i et par, regnes med til forældrenes familie, hvis alderskravet er opfyldt.

Definitionen indebærer, at hjemmeboende børn, der er fyldt 18 år udgør deres egne familier, og at søskende (uanset alder), der ikke bor sammen med deres forældre, ikke regnes til samme familie, ligesom gamle personer, der bor hos deres voksne børn, ikke indgår i disses familier.

Denne familiedefinition må på mange punkter betragtes som resultatet af afvejninger af forskellige forhold mod hinanden. En af de mest betydende ændringer i forhold til den tidligere familiedefinition var, at man sænkede aldersgrænsen for hjemmeboende børn fra 26 til 18 år.

Aldersgrænsen for hjemmeboende børn på 18 år svarer til myndighedsalderen og ophør af forældrenes forsørgelsespligt, hvilket har haft betydning for valget af denne grænse. Der opereres dog også med en familietype, kaldet D-familien, for hvilken der ingen aldersgrænse er for børn hørende med til forældrenes familier (men det kræves, at de skal være ugifte og ikke have børn). Der fremstilles ingen statistik på denne familietype. Den benyttes til klassifikation af husstande, idet antallet af D-familier i en husstand indgår i koden for husstandstype. Desuden udgør D-familieoplysningerne et beredskab i husstands- og familiemodulet.

Den anden store ændring i forhold til tidligere var dannelsen af en ny partype, samboende par, bestående af ikke-gifte par uden fællesbørn. Samboende par er defineret som par boende på samme adresse, bestå-

ende af en mand og en kvinde, der ikke er gift med hinanden, og som ikke har børn sammen.

Der opereres herefter med to slags ikke-gifte par:

- *Samlevende par*, som har fælles børn, og som derfor med stor sikkerhed kan erkendes som reelle samlivsforhold.
- *Samboende par*, som ikke har fælles børn, og hvor man således kun kan gætte på, at der er tale om et egentligt samlivsforhold.

Mht. definitionen af samboende par er det i praksis nødvendigt at stille et krav om, at der ikke må være mere end to voksne personer på adressen. Endvidere bør der ikke indgå søskendepar i denne gruppe. Ældre søskendepar vil det dog ikke være muligt at sortere væk. De kan kun identificeres på fælles forældrehenvisningsnumre, og disse er kun undtagelsesvis til stede for personer født i begyndelsen af 1950'erne eller tidligere (jf. nedenfor om datakvalitet). Unge ned til 16 år danner samboende par, hvis de øvrige betingelser er opfyldt.

For samboende par kræves det endvidere, at aldersforskellen er mindre end 15 år. Som grundlag for beslutningen om denne grænse indgik omnibusundersøgelsernes²⁶ resultater og samkøring med oplysninger om reelt enlige i følge Børnetilskudsstatistikregistret. I Omnibusundersøgelserne blev interviewpersonerne bl.a. spurgt om, hvorvidt de var "papirløst samlevende". Svarene kunne sammenholdes med befolkningsstatistikkens oplysninger vedrørende husstands- og familieforhold. For husstande bestående af to enlige af modsat køn med under 15 års aldersforskel betragter langt den overvejende del sig som "papirløst samlevende", mens der blandt dem med større aldersforskel var en mindre andel af disse par. Denne gruppe indeholder fx unge loggerende hos ældre personer, og par bestående af en person og en af dennes forældre. Den sidstnævnte type kan ikke i tilfredsstillende grad frasorteres ved hjælp af forældrehenvisningsnumre, da disse henvisninger som nævnt mangler for mange voksne. Femtenårsgrænsen er arbitrært fastsat og har været meget diskuteret. Der er selvfølgelig tilfælde, hvor et reelt samboende par har en større aldersforskel. Men betydningen af dette er forholdsvis ringe. Det kan nævnes, at under 2 pct. af ægteparrene og de ikke-gifte par med fællesbørn har en forskel på 15 år eller mere.

Ved omlægningen af familiestatistikken indførte man en fjerde partype, de registrerede partnerskaber. Loven om registrerede partnerskaber trådte i kraft den 1. oktober 1989. Den giver mulighed for, at to personer af samme køn kan indgå registreret partnerskab med det resultat, at

²⁶ Omnibusundersøgelserne er regelmæssigt udførte interviewundersøgelser af et repræsentativt udsnit af befolkningen.

de juridisk stilles som et ægtepar²⁷. Denne partype forekommer i lighed med de øvrige både med og uden børn - i disse tilfælde er der naturligvis alene tale om særbørn. Udskillelsen af de registrerede partnerskaber volder ingen problemer, idet der i CPR er indført tre nye koder for ægteskabelig stilling for henholdsvis personer der er i registreret partnerskab, personer der har fået opløst partnerskabet, og personer der har overlevet deres partner (de øvrige koder gør det muligt at skelne mellem ugifte, gifte (inkl. separerede), fraskilte og personer i enkestand). Talmæssigt betyder de registrerede partnerskaber ikke meget²⁸, men da registrerede partnerskaber juridisk set er som ægteskaber, anses det for vigtigt at få dem med som familietype. Derimod er det forbundet med alt for store problemer og af ringe interesse at forsøge at få tal på ikke-registrerede parforhold mellem personer af samme køn.

Ikke-gifte par med fællesbørn betegnes som samlevende par. Det skal ikke nødvendigvis være hjemmeboende børn. Det er tilstrækkeligt, at parret på et tidspunkt²⁹ har boet sammen med hjemmeboende børn (uanset disses alder), og ikke har boet hver for sig ved et senere årsskifte. Derved opstår den umiddelbart lidt uforståelige gruppe af samlevende uden børn, dvs. uden hjemmeboende børn under 18 år.

Med den anvendte metode til dannelse af familier vil der i gruppen af samboende par indgå nogle personer, der betragter sig som enlige, ligesom der er personer, der regnes for enlige, selvom de betragter sig som par. Det sidste sker antagelig mest i tilfælde, hvor der bor andre voksne på adressen. Men alt i alt er der grund til at tro, at den nye familiedefinition giver en rimelig beskrivelse af befolkningens pardannelser.

Officiel ægteskabelig stilling og konkret familiesituation har ikke så tæt en relation som før i tiden. I statistikkens klassifikationer indgår den ægteskabelige stilling som betingelse på følgende punkter: Personer i ægtepar og i registrerede partnerskaber skal have ægteskabelig stilling som henholdsvis gift eller registreret partner, og børn, hjemmeboende såvel som ikke-hjemmeboende, skal være ugifte. Alle øvrige kombinationer af ægteskabelig stilling og familiestatus kan forekomme; fx har kun 95 pct. af de gifte personer familiestatus som gift.

De grundlæggende data fra CPR, som familiestatistikken bygger på, har generelt set meget høj kvalitet. De omfatter adresse, køn, alder og ægteskabelig stilling, samt ægtefællehenvisningsnummer.

²⁷ Kun adoptionsloven gælder ikke for registrerede partnerskaber.

²⁸ Pr. 1. januar 1994 var det samlede antal personer med en af de tre nye civilstandskoder på 2.809 personer. Antallet af familier af typen registrerede partnerskaber var 1.078.

²⁹ Tidspunktet skal være den 1. januar 1990 eller ved et senere nytår.

Adressen afgrænser husstanden og dermed også familien, og adresseoplysningerne i CPR er ret præcise. Da de for langt den overvejende del af befolkningen også er udtryk for, hvor de virkelig bor, er adresseoplysningerne velegnede som grundlæggende kriterium ved afgrænsningen af familier. I de antageligt meget få tilfælde, hvor en person ikke er tilmeldt CPR på den korrekte adresse, bliver familietyperne forkerte på både CPR-adressen og på den virkelige adresse. Det er ikke muligt at korrigere for en sådan fejl. Det må også antages, at i en del af de tilfælde, hvor en person står tilmeldt en adresse, der ikke svarer til den faktiske bopæl, sker det bevidst for at opnå bestemte fordele enten med hensyn til skat eller sociale tilskud. I sådanne tilfælde vil det ofte være vanskeligt at få den korrekte oplysning ved interview.

Køn, alder og ægteskabelig stilling beskriver den enkelte person og indgår som kriterier ved fastlæggelsen af familieforhold. Oplysningerne fra personnummeret (køn og alder) kan med meget få undtagelser anses for fuldstændig korrekte, og oplysningerne om ægteskabelig stilling kommer tæt på denne tilstand.

Henvisningsnumre til seneste ægtefælle eller registrerede partner benyttes til afgørelsen af, om to personer på samme adresse udgør et ægtepar (eller et registreret par). Denne oplysning er af meget høj kvalitet, bl.a. på grund af betydningen for skatteligningen og som følge af, at der foregår en direkte indberetning af denne oplysning til CPR fra de registrerende myndigheder.

Henvisningsnumre til forældre benyttes til identifikation af familierelationer. Disse numre henviser til de juridiske forældre, hvilket vil sige, at man ikke kan skelne biologiske forældre fra adoptivforældre. Desværre er hyppigheden af disse numres tilstedeværelse faldende jo længere tilbage i tiden man går, hvilket har betydning for muligheden for at adskille søskendepar fra samboende par. Personer under 18 år har meget fin dækning med hensyn til henvisninger til forældre; kun 0,15 pct. af dem mangler begge henvisningsnumre. Opdelingen af børn i hjemmeboende og ikke-hjemmeboende foregår derfor på et ret sikkert grundlag. Personer født i 1960 mangler for 7 pct.'s vedkommende begge forældrehenvisninger, og for 1950-årgangen drejer det sig om 96 pct. Disse henvisningsnumre slettes ikke længere i CPR, når de én gang er indført. Derfor forbedres befolkningens dækning i denne henseende gradvist år for år. De forbedrede forældrehenvisningsnumre vil i fremtiden bl.a. give mulighed for opgørelse af børnefamilier, der bor sammen med bedsteforældregenerationen.

Husstands- og familiemodulet dannes pr. 1. januar hvert år. I princippet kan man på grundlag af et statusudtræk fra CPR for et hvilket som helst tidspunkt danne modulets delregistre. Dette gøres i begrænset omfang som led i specielle opgaver.

4.2 Anvendelse af flere kilder

Når dataindsamling til statistik gennemføres med anvendelse af traditionelle indsamlingsmetoder, søges spørgsmål og vejledning til disse formuleret på en sådan måde, at man får oplysninger, som er i overensstemmelse med de definitioner, som anvendes inden for det pågældende statistikområde. Man har så at sige mulighed for at styre indsamlingen, således at man får netop de oplysninger, der er brug for.

Hvis de ønskede oplysninger ikke findes umiddelbart i et register....

Anvendelsen af administrative registre som datagrundlag for statistik giver ikke den samme styringsmulighed. Man kan vælge blandt foreliggende oplysninger og udsøge dem, som passer bedst til det konkrete formål. Det vil imidlertid ofte være sådan, at selv de bedst egnede oplysninger ikke umiddelbart dækker behovet i tilstrækkelig grad. Årsagen kan være afvigende definitioner, begrænset dækningsgrad, fejl eller manglende opdatering. I den situation kan det være en mulighed at trække på oplysninger fra flere administrative kilder.

...så kan samtidig anvendelse af flere kilder være løsningen

Når behovet for en konkret oplysning til én eller flere statistikopgaver ikke umiddelbart kan dækkes gennem en direkte anvendelse af foreliggende oplysninger, kan problemet i visse tilfælde løses gennem dannelsen af såkaldt *afledte variable*. En sådan variabel dannes ved at sammenkæde forskellige oplysninger, som tilsammen giver den ønskede information, om ikke helt præcist så dog med god tilnærmelse.

Ved dannelsen af en afledt variabel vil man som udgangspunkt opstille en definition af det begreb, som ønskes anvendt i statistikken. Herefter vil man søge efter administrative oplysninger, evt. oplysninger der allerede anvendes til statistik, som kan bidrage til at opfylde de definitoriske krav. Man kan skelne mellem tre typer af variabeldannelse, *dannelse af en ny variabel, prioritering af oplysninger for en given variabel og tilpasning eller korrektion af given information*.

4.2.1 Dannelse af en ny variabel

Datakombinationer giver den ønskede oplysning

Dannelse af nye variable kan komme på tale, hvis den nødvendige information foreligger i tilstrækkeligt omfang, uden dog at være indeholdt i et enkelt administrativt datum. Flere oplysninger, indsamlet og behandlet efter administrative regler, kan hver for sig bidrage til at give en samlet beskrivelse af enkeltindivider i forhold til det ønskede statistikbegreb. Man kan sammenligne det med et puslespil, hvor hver brik er en del af det samlede billede uden dog isoleret set at kunne give et indtryk af det færdige billedes udseende.

Eksempel 8: Beskæftigelsesstatus

Som eksempel på en sådan ny variabel kan nævnes *beskæftigelsesstatuskoden*, som dannes i Danmarks Statistiks Arbejdsklassifikationsmodul. Den klassificerer personer efter den væsentligste status i løbet af et kalenderår, og den indeholder grupperne *selvstændigt erhvervsdrivende, medhjælpende ægtefæller, lønmodtagere, pensionister og andre ude af erhverv*. Oplysninger som denne er klart af statistisk karakter, og man vil ikke finde den i samme form i de administrative registre. Her indgår oplysninger som fx modtaget indkomst specificeret efter art og evt. regnskabspligt fra indkomstskattesystemets registre og evt. ejerforhold til erhvervsvirksomhed fra Det centrale erhvervsregister. Sådanne oplysninger er egnede til at indgå i en samlet vurdering af, hvad der for en given person må anses for værende den dominerende status i løbet af et bestemt år. De udgør en række indikatorer, som vægtes gennem et sæt af regler, der fører frem til en klassifikation efter beskæftigelsesstatus.

I de fleste tilfælde er denne klassificering ganske ukompliceret. Alle indikatorer vil pege på samme gruppe. Der kan fx være tale om, at der som eneste klassificerende indkomst er modtaget løn, og den pågældende er ikke indehaver af en erhvervsvirksomhed. Denne person vil blive klassificeret som lønmodtager, idet den modtagne løn dog skal være af en vis størrelse. Hvis dette ikke er tilfældet, klassificeres personen som værende ude af erhverv.

Mere kompliceret bliver det, når der er indikation for flere muligheder. Det kan være en virksomhedsindehaver, der har modtaget løn og pension i det omhandlede år. Her skal der foretages en afvejning af informationerne. Ideelt set burde den tidsmæssige udstrækning af de enkelte forhold være afgørende for vægtningen, men noget sådant er ikke muligt. Som den bedste tilnærmelse til det ideelle kan vægtningen ske efter størrelsen af de relevante indkomstarter, overskud af virksomhed, løn og pension. Det praktiseres da også i et vist omfang, men det kan give alvorlige fejltolkninger at vurdere omfanget af en selvstændig erhvervsdrivendes arbejdsindsats på grundlag af virksomhedsresultatet i et enkelt år, idet man kan få underskud trods en betydelig arbejdsindsats. Derfor tillægges oplysningen om ejerforhold til en virksomhed fra Det centrale erhvervsregister meget stor vægt ved afgrænsningen af selvstændige. Det er udformet på den måde, at en virksomhedsindehaver klassificeres som selvstændig, medmindre en eventuel løn er større end overskud af virksomhed, og ATP-bidraget indikerer fuld beskæftigelse hele året som lønmodtager, eller overskud af virksomhed er nul eller mindre, og den pågældende har modtaget pension. De to undtagelser fører til klassificering som henholdsvis lønmodtager og pensionist.

Ved dannelsen af en ny variabel er det væsentligt, at de nødvendige komponenter foreligger for alle personer på et til formålet højt kvalitetsniveau. I det anførte eksempel gælder det, at oplysninger om ejer-

Oplysningernes dækningsgrad og kvalitet

forhold til virksomhed kun findes for personer, der efter reglerne herom skal være optaget i Det centrale erhvervsregister. Det er arbejdsgivere og momspligtige. En mindre gruppe erhvervsdrivende falder uden for denne kreds, fx inden for persontransportområdet og visse sundhedsydelser. Her anvendes oplysninger om regnskabspligt samtidig med, at forekomsten af virksomhedsover- eller underskud tillægges større betydning. Afgrænsningen af lønindkomst i forhold til andre indkomstarter, som beskattes på samme måde som løn, er lidt præget af, at den ikke er af væsentlig betydning for skattemyndighedernes ligningsarbejde, hvorfor der er en lille usikkerhed knyttet til denne. Det må dog konkluderes, at trods de nævnte mangler, er resultatet i god overensstemmelse med den opstillede definition.

4.2.2 Prioritering af oplysninger fra flere kilder

Varierende dækningsgrad...

Der er situationer, hvor de ønskede oplysninger findes, men foreligger i forskellige administrative registre med varierende dækningsgrad og/eller kvalitet.

Den varierende dækningsgrad hænger sammen med, at visse registre er oprettet med henblik på varetægelse af bestemte administrative opgaver vedrørende en mindre afgrænset del af befolkningen. Eksempelvis omfatter et lønsystem typisk en veldefineret gruppe af lønmodtagere.

...og svingende kvalitet

Oplysningernes kvalitet afhænger af deres betydning for det administrative formål, de skal tjene. Endvidere kan kvaliteten afhænge noget af, hvem der har indberettet den pågældende oplysning til registret. Oftest er det en administrativ myndighed eller en arbejdsgiver, men i enkelte tilfælde, som fx stillingsbetegnelser, kan det være personen selv, hvilket kan indebære en kvalitetsforringelse af statistikgrundlaget i form af begrebsmæssigt afvigende oplysninger, egentlige fejl eller mangler, og en kontrol af oplysningens indhold kan være vanskelig at gennemføre.

Eksempel 9: Stillingsoplysninger

Anvendelsen af stillingsoplysninger som grundlag for klassifikation af lønmodtagere efter fag er et eksempel på prioritering af oplysninger for en given variabel. Stillingsoplysninger er ikke et fuldt tilfredsstillende udgangspunkt for en sådan klassifikation, men er de bedst egnede, og de giver også acceptable resultater.

Det generelle grundlag for stillingsoplysninger er CPR. Dette register indeholder oplysninger om alle personer med bopæl i Danmark, og blandt disse oplysninger indgår eventuel stilling. Oplysningerne er afgivet af personerne selv enten via selvangivelsen (tidligere ordning) eller gennem det lokale folkeregister (nuværende ordning). Totalt dækkende oplysninger om stilling skulle derfor kunne hentes fra CPR. Interessen i befolkningen og hos administrative myndigheder for an-

vendelse af stillingsbetegnelser er med tiden blevet mere og mere begrænset, hvilket har betydet, at ændringer ikke er blevet indberettet i tilstrækkeligt omfang. Resultatet er derfor, at en stillingsbetegnelse mangler i mange tilfælde, hvor den burde være der, og at en del betegnelser ikke er i overensstemmelse med de aktuelle forhold. Hertil kommer fortolkningsproblemer og intetsigende titler som en konsekvens af, at den enkelte person inden for vide rammer selv kan vælge sin stillingsbetegnelse.

Manglerne i forbindelse med CPR-stillingsoplysningerne, som var et erkendt problem på et tidligt tidspunkt, har ført til bestræbelser i retning af at basere fagklassifikationen på oplysninger fra andre kilder. Der findes imidlertid ikke sådanne kilder, som dækker hele den relevante del af befolkningen. Derfor har Danmarks Statistik været henvist til at benytte flere kilder, der hver især kun dækker en del af befolkningen. Arbejdsgivernes lønadministration har i første række påkaldt sig interesse. Her indgår arbejdsgiveroplysninger om den enkelte ansattes stilling til brug for lønberegning og personaleadministration i almindelighed.

Oplysninger fra de offentlige lønsystemer anvendes i første række til lønstatistik samt til identifikation og karakteristik af arbejdsstedet for offentligt ansatte, men de indeholder også oplysninger, der umiddelbart kan indgå som stillingsklassifikationsgrundlag for denne gruppe lønmodtagere. På det private område har stillingsoplysninger ikke været tilgængelige på samme måde før 1992, men det er ændret, således at private arbejdsgivers indberetning af individualoplysninger til lønstatistikken omfatter de ansattes personnummer. De særdeles velguede stillingsoplysninger, der foreligger i denne sammenhæng, kan derfor nu bruges sammen med andre oplysninger.

Stillingsoplysninger fra de offentlige lønsystemer og indberetninger fra private arbejdsgivere udgør herefter tilsammen det vigtigste grundlag for fagklassificering. CPR-oplysningerne udgør et alternativt grundlag, hvor brugbare oplysninger ikke foreligger fra disse kilder. Endvidere indgår eventuel medlemskab af en arbejdsløshedskasse, som i nødstilfælde kan være grundlag for en fagklassificering på et groft klassifikationsniveau, samt oplysninger om eventuel igangværende uddannelse, der indebærer aflønning og dermed placering i arbejdsstyrken som lønmodtager, da lønindkomst af en vis størrelse medfører klassificering som sådan.

4.2.3 Korrektion af oplysninger

Korrektion af registeroplysninger foretages om muligt, hvor fejl eller mangler forekommer. De vil i et vist omfang kunne gennemføres med udgangspunkt i forventede sammenhænge mellem oplysninger fra samme kilde, men det kan være nødvendigt at bruge oplysninger fra andre kilder.

Eksempel 10: Afgrensning af efterløn og pensioner

Afgrensningen af visse beløbsarter i indkomststatistikken er eksempler på dette. Efterløn er markeret med en særlig kode i Told- og Skattemyndighedens Oplysningssedelregister (COR), men koden er ikke anført i alle tilfælde, og det forekommer også, at beløbet er indberettet i et forkert felt. Til støtte for en endelig afgrensning af efterløn anvendes derfor oplysninger om, hvem der ifølge Det centrale register for arbejdsmarkedsstatistik (CRAM) er efterlønsmodtager. En markering af dette indgår sammen med COR-oplysningerne i den endelige fastlæggelse af, hvad der er efterløn, og hvem der har modtaget det.

Et andet eksempel vedrører udbetalt pension, hvor hovedgrundlaget også er COR. Registret indeholder en dækkende specifikation af de *socialle pensioner*, men dækningen af *tjenestemandspension* er meget ufuldkommen, og udbetalte pensioner i henhold til *private ordninger* indgår i en restsum af uspecificerede ydelser. Behovet for oplysninger om pensionsforhold er stort, så det tilstræbes at korrigere grundoplysningerne fra COR ved at trække på supplerende oplysninger om udbetaleren fra Det centrale erhvervsregister. COR indeholder en identifikation af udbetaleren i form af et SE-nummer. De relevante numre i forhold til udbetaling af pension udsøges på den måde, at numre, der har været anvendt mindst 100 gange ved udbetaling af ydelser til personer over 70 år, udtrækkes, hvis det udbetalte beløb er registreret i det relevante felt i COR. Med denne betingelse er der sikkerhed for, at de udsøgte SE-numre anvendes ved udbetaling af ydelser til pensionister, og med en bestemt placering af ydelsen i COR er der stor sandsynlighed for, at det udbetalte beløb er pension. Ved samkøring med uddrag fra Det centrale erhvervsregister på grundlag af SE-nummeret klassificeres udbetalerne i grupperne offentlig sektor, finansiel sektor og andet. Ydelser udbetalt over SE-nummeret i den første gruppe betragtes som tjenestemandspension. Anden gruppe omfatter privattegnede pensionsordninger, herunder pensionskasser, og den tredje gruppe indeholder andre pensionslignende ydelser. I denne endelige afgrensning tages der ikke hensyn til modtagerens alder, kun det anvendte SE-nummer og beløbets registreringsfelt i COR er afgørende.

Ved afgrensningen af efterløn opnås ret præcise resultater, medens pensionsafgrænsningen fører til resultater, der i varierende grad er til-

nærmet den ønskede information. Tjenestemandspension afgrænset som beskrevet er i god overensstemmelse med de faktiske forhold, private pensionsordninger må antages at være dækket i rimelig grad, medens andre pensionslignende ydelser ikke skal betragtes som pensionsydelser alene. Specielt i den sidste kategori trækker metoden en række andre ydelser med sig, som ikke er pension.

Korrektion på grundlag af supplerende oplysninger fra andre kilder kan føre frem til brugbare oplysninger på områder, hvor det foreliggende materiale er fejlbehæftet eller på anden måde mangelfuldt, men der kan ligge en fare i at gennemføre systematiske korrektioner, hvis man ikke er opmærksom på de begrænsninger, som kan være forbundet med den anvendte metode.

4.2.4 Anvendelse af kombinerede data over tid

Anvendelse af kombinerede data vil ofte kræve et ret komplekst regelsæt for behandlingen af oplysningerne. Det bliver på den måde vanskeligt at gennemskue, hvad der egentlig sker, og det kan føre til, at der ikke vil være tilstrækkelig opmærksomhed omkring betydningen af de enkelte elementer. En isoleret set forholdsvis ubetydelig ændring i grundoplysningerne kan få meget uheldige konsekvenser i forbindelse med behandlingen af de kombinerede oplysninger. Der kan være tale om regelændringer, ændringer i den administrative praksis eller blot om en ændret kodning af en bestemt oplysning. Sådanne ændringer kan slå så voldsomt igennem, at virkningerne umiddelbart er synlige, men der kan også, specielt når behandlingen er betinget af bestemte værdier for flere variable, forekomme mindre iøjnefaldende virkninger, som medfører, at man umærkeligt fjerner sig fra den begrebsdefinition, som man tilstræber og tror er gældende.

Det er vigtigt, at regelsæt for behandlingen er beskrevet meget nøje og med en præcis angivelse af, hvilke oplysninger der indgår, hvordan de indgår, og hvilke betingelser der er knyttet til deres anvendelse. Endvidere skal man så vidt muligt holde øje med eventuelle ændringer i oplysningernes indhold, således at reglerne kan vedligeholdes i forhold til dette.

Det kan være svært på forhånd at sikre sig kendskab til alle relevante ændringer, så det er væsentligt at teste behandlingen i forløbet, fx gennem kontrol af, at de forskellige betingelser indvirker på slutresultatet i samme omfang som tidligere. Afvigelse fra dette er ikke i sig selv udtryk for, at noget er ændret, men det viser, hvor der kan være behov for en nøjere undersøgelse.

Brugen af afledte variable dannet på grundlag af oplysninger fra forskellige kilder vil i særlig grad være sårbar over for administrative

Komplekse regelsæt er følsomme over for administrative ændringer...

...derfor nøje beskrivelse og opmærksomhed mod ændringer

ændringer. Anvendelsen af flere kilder vil øge risikoen for, at sådanne ændringer af betydning for statistikformålet forekommer. Ændringerne kan være kendt, men det kan være umuligt at tage højde for dem, fx hvis en oplysning udgår. Her kan man være påtvunget en tilpasning af den pågældende begrebsdefinition for at bringe den i overensstemmelse med de foreliggende muligheder. Det vil betyde, at man fjerner sig fra den hidtidige definition, og får et såkaldt databrud set i en tidsammenhæng.

Muligheder for forbedringer

Administrative ændringer kan også have positiv karakter i den forstand, at nye oplysninger dukker op eller andre forbedres med hensyn til at imødekomme de statistiske behov. Ændringer i retning af det bedre vil naturligvis også give databrud, men gevinsten ved det forbedrede grundlag skulle gerne mere end opveje ulempen ved databrudet.

4.2.5 Anvendelse af flere kilder som metode

Risiko for utilsigtede afvigelser fra idealdefinitioner

Brugen af oplysninger fra flere kilder kan være løsningen, når den ønskede oplysning ikke umiddelbart foreligger i en til statistikformålet anvendelig form. Man må imidlertid gøre sig klart, at selv om man som udgangspunkt søger at fastlægge en definition ud fra de statistiske behov, så vil de til rådighed værende oplysninger og den behandling, man har mulighed for at give dem, meget let føre til afvigelser fra denne definition. Ved udvælgelsen af oplysninger og under fastlæggelsen af reglerne for behandlingen af disse, bør man være meget opmærksom på, at disse afvigelser ikke bliver for voldsomme. Det kunne føre til, at man i statistikken opgør noget andet end det man tror.

Forvirring hos statistikbrugere bør undgås

De faktiske definitioner, der er knyttet til de afledte begreber vil ofte være noget sindrige, idet de vil afspejle det anvendte regelsæt. Det kan skabe unødigt forvirring hos statistikbrugerne, hvorfor de i præsentationen fremtræder i en mere forenklet form, sådan som det også sker i forbindelse med traditionel dataindsamling. Når oplysninger fra flere kilder anvendes, kan man få et mere nuanceret billede af bestemte forhold for en given person. Eksempelvis kan der foreligge oplysninger om indtjent løn fra COR samtidig med, at den pågældende har været aktiv som virksomhedsindehaver ifølge Erhvervsregistret. Regelsættet skal fastlægge, hvorledes man skal vægte oplysningerne, når en sådan person klassificeres efter beskæftigelsesstatus som lønmodtager eller selvstændig erhvervsdrivende.

Disse forskellige oplysninger foreligger også i interviewsituationen, men de er kun kendt af interviewpersonen, så her overlades det til denne at vælge klassificering og dermed vægtning af elementaroplysningerne. Dette valg kan være ret subjektivt, så det er umuligt at angive en præcis afgrænsning af de enkelte grupper i en klassifikation.

Man bliver klassificeret som lønmodtager, hvis man selv har angivet dette som svar på et spørgsmål.

Ved præsentationen af den registerbaserede beskæftigelsesstatus defineres den som væsentligste aktivitet i løbet af et givet kalenderår. De nærmere detaljer kan naturligvis stilles til rådighed, hvis særlige undersøgelser kræver en analyse af afgrænsningen mellem de forskellige grupper.

4.3 Integreret dataindsamling

Når en administrativ myndighed indsamler oplysninger for en anden myndighed samtidig med indsamling af oplysninger til eget brug, kaldes indsamlingsformen *integreret dataindsamling*.

Definition

Dette indebærer, at en administrativ myndighed, fra virksomheder, institutioner mv., dels indsamler oplysninger, som den selv skal benytte til løsning af sine egne opgaver, dels indsamler oplysninger, som den ikke selv har behov for, men som skal benyttes af en anden myndighed, fx Danmarks Statistik. De to datasæt indhentes i én og samme integrerede arbejdsgang.

Formålet med den integrerede dataindsamling er at begrænse omkostningerne i forbindelse med dataindsamlingen. De samlede samfundsmæssige omkostninger reduceres, idet det er langt billigere at indsamle oplysninger i en samlet proces end i flere processer. Dette gælder især, hvor kildeoplysningerne ikke findes i edb-registre, og hvor der derfor skal iværksættes ressourcekrævende dataindsamling via skema eller interview.

Formål

Hertil kommer, at den integrerede dataindsamling bidrager til, at de indsamlede oplysninger er sammenhængende og dækker de samme enheder, mens dette ikke nødvendigvis ville være tilfældet, hvis indsamlingen blev foretaget af to institutioner i adskilte processer.

4.3.1 Lovgrundlaget

Lov om Danmarks Statistik indeholder bestemmelser om, hvilke data der kan indsamles af institutionen, men giver ingen retningslinier for hvilke indsamlingsmetoder der skal anvendes. Det er således i loven fastsat, hvilke oplysninger eller hvilke typer af oplysninger, der kan indsamles fra erhvervsvirksomheder, institutioner mv., men der er ingen bestemmelser om, hvorvidt oplysningerne skal indsamles ved hjælp af spørgeskema, ved direkte videregivelse af oplysninger, som institutionerne eller virksomhederne er i besiddelse af, ved maskinel dataoverførsel til Danmarks Statistik, om oplysningerne indsamles af

Lov om Danmarks Statistik

en anden myndighed ved integreret dataindsamling, eller eventuelt ved en helt femte metode.

Lov om offentlige myndigheders registre

*Lov om offentlige myndigheders registre*³⁰ indeholder bestemmelser om behandlingen af personoplysninger på elektronisk medium. Loven skal sikre, at det offentliges oprettelse og brug af edb-registre med personoplysninger sker på en sådan måde, at den enkelte borgers retsbeskyttelse og integritet ikke krænkes samtidig med, at administrationen kan foregå effektivt. Efter loven må registre, der føres for en statslig myndighed, kun oprettes efter godkendelse af en minister. Inden et register tages i brug, skal der være fastsat forskrifter for registrets opbygning og drift. For registre, der føres af en kommunal myndighed, skal beslutning om oprettelse og om forskrifternes udformning, som hovedregel træffes af vedkommende kommunalbestyrelse. Registerforskrifterne skal godkendes af Registertilsynet.

Hvem må bruge oplysningerne?

Kapitel 3 i lov om offentlige myndigheders registre indeholder bestemmelser om registrering og opbevaring af oplysninger. I §9 stk. 1 står der således: "Der må kun registreres oplysninger, der klart er af betydning for varetagelsen af vedkommende myndigheds opgaver. Der må endvidere registreres oplysninger, der klart er af betydning for varetagelsen af en anden myndigheds opgaver, såfremt registret indrettes således, at oplysningerne kun kan benyttes af den anden myndighed". Det vil typisk være tilfældet, hvor en myndighed i forvejen indhenter en række oplysninger fra borgerne (eller virksomhederne om fx de ansatte) til brug for egen sagsbehandling, medens den anden myndighed skal anvende en oplysning, der knytter sig snævert hertil.

4.3.2 Eksempler på integreret dataindsamling

I en række tilfælde vil Danmarks Statistik have behov for både de oplysninger der indsamles for og udelukkende er til brug for Danmarks Statistik, og som derfor ikke må benyttes af den indsamlede myndighed, og de oplysninger, som den administrative myndighed indsamler til eget brug. I disse tilfælde vil det samlede sæt af oplysninger indgå i de data, som leveres til Danmarks Statistik, og indgå i de relevante statistikregistre i Danmarks Statistik. Dette gælder i det følgende eksempel.

Eksempel 11: Arbejdspladsprojektet

Arbejdspladsprojektet har til formål at give en fortegnelse over alle arbejdssteder og fordele alle beskæftigelsesforhold på de korrekte arbejdssteder. Oplysningerne til arbejdspladsprojektet indhentes ved integreret dataindsamling via de oplysningssedler, som arbejdsgiverne hvert år skal aflevere til Told- og Skattestyrelsen for hver enkelt ansat. Oplysningssedlerne indeholder oplysninger om løn, pension, honorarer, uddelinger fra fonde mv. for den enkelte ansatte eller modtager af

³⁰ Jf. *Justitsministeriet (1992a)*.

offentlige ydelser. Denne er identificeret ved personnummeret, mens arbejdsgiveren er identificeret ved SE-nummeret, som gennem arbejdspladsprojektet suppleres med en kode for ansættelsessted, hvis den pågældende arbejdsgiver har flere arbejdssteder. Oplysningssedlen rummer desuden oplysninger om ansættelsesperiode, ATP-bidrag og A-skat, samt enkelte andre ydelser, hvor der ikke er indeholdt A-skat.

Arbejdspladsprojektet gennemføres én gang årligt i et samarbejde mellem Told- og Skattestyrelsen og Danmarks Statistik. Told- og Skattestyrelsen udsender i november/december de rå oplysningssedler samt vejledninger til virksomhederne om årets indberetninger. I denne udsendelse medsendes arbejdsstedsfortegnelser til virksomheder med flere arbejdssteder. Arbejdsstedsfortegnelserne er udarbejdet af Danmarks Statistik. De indeholder en fortegnelse over de arbejdssteder, som Danmarks Statistik har registreret for den pågældende virksomhed, således som det fremgår af Danmarks Statistiks Erhvervsregister. Fortegnelserne indeholder navn, adresse og branche (branchenavn og kode) for hvert enkelt arbejdssted. Endelig indeholder arbejdsstedsfortegnelserne den kode for arbejdsstedet, som skal påføres oplysningssedlen for hver enkelt ansat. Arbejdsgiveren skal ajourføre arbejdsstedsfortegnelsen med hensyn til oplysningerne om det enkelte arbejdssted, hvis der måtte være fejl og mangler. Arbejdsgiveren skal endvidere komplettere arbejdsstedsfortegnelsen med nye og nedlagte arbejdssteder, således at fortegnelsen, efter eventuelle korrektioner, indeholder en komplet liste over alle arbejdssteder, som har været aktive i årets løb.

Lønoplysningerne indhentes af Told- og Skattestyrelsen til brug for deres indkrævning og kontrolfunktion, og til brug for ligningsarbejdet, som foregår i et samspil med de kommunale ligningsmyndigheder.

Arbejdsstedskoderne indhentes til brug for Danmarks Statistik. Desuden indhentes oplysninger om den ansattes ansættelsesperiode i det pågældende år udelukkende til Danmarks Statistiks brug.

Danmarks Statistik modtager en kopi af samtlige oplysninger på oplysningssedlen, dels de oplysninger, der er indsamlet af Told- og Skattestyrelsen alene til Danmarks Statistiks brug, og dels de oplysninger, der indsamles til Told- og Skattestyrelsens egen administrative brug.

For Danmarks Statistik er oplysningerne en helhed. Skulle Danmarks Statistik selv indsamle oplysningerne, ville de administrative oplysninger være lige så betydningsfulde som oplysningerne om arbejdssted og ansættelsesperiode. Den tolkning, man kan give de oplysninger, som indhentes alene til Danmarks Statistik, afhænger af værdierne for de øvrige oplysninger.

For den offentlige sektors vedkommende findes oplysningen om arbejdssted mv. i stort omfang allerede registreret i de statslige og kommunale edb-lønanvisningssystemer. I arbejdspladsprojektet benytter man derfor oplysninger, som hentes direkte fra disse systemer, som dog forinden må kompletteres og systematiseres. Dette er en omfattende proces.

Arbejdspladsprojektets oplysninger anvendes af Danmarks Statistik til statistiske opgørelser. Oplysningerne danner i første omgang grundlaget for individoplysninger i det personstatistiske system, der dækker arbejdsmarkedsoplysninger for en årsperiode samt for en tidspunktsopgørelse (ultimo november). De indgår således i det personstatistiske systems klassifikationsgrundlag.

Oplysningerne indeholder imidlertid også det afgørende bindeled mellem de personstatistiske opgørelser og erhvervsopgørelserne, idet SE-nummeret, med tilhørende arbejdsstedskode, giver oplysninger om personernes tilknytning til konkrete virksomheder og arbejdssteder. Dette giver et grundlag for opdatering af Erhvervsregistrets enheder, og personsummationerne opdaterer Erhvervsregistrets beskæftigelsesoplysninger på arbejdsstedsniveau.

Oplysningerne på personniveau giver således mulighed for erhvervsklassificeringer på individniveau, og erhvervsoplysningerne giver gennem de tilknyttede personnumre mulighed for at knytte personvariable fra det personstatistiske registersystem til erhvervsenhederne.

Den registerbaserede arbejdsstyrkestatistik (RAS) inddrager arbejdspladsoplysningerne. RAS giver en beskrivelse af befolkningens tilknytning til arbejdsmarkedet på et givet tidspunkt i løbet af året, nemlig slutningen af november, svarende til oplysningerne om ansættelsesperiode fra oplysningseddelen. Datagrundlaget udgøres af arbejdspladsprojektets grundoplysninger samt et uddrag fra en række administrative registre og statistikregistre, hvis oplysninger er bearbejdet med henblik på at belyse arbejdsmarkedstilknytningen i den sidste uge af november det pågældende år, for alle personer med bopæl i Danmark den 1. januar i det efterfølgende år.

Materialet fra arbejdspladsprojektet og oplysningseddelerne anvendes også til udarbejdelse af erhvervsbeskæftigelsesstatistikken, som har til formål at belyse beskæftigelsen set fra erhvervsenhedernes side. Arbejdsstederne klassificeres efter bl.a. deres beskæftigelse i slutningen af november. Erhvervsbeskæftigelsesstatistikken efter den her beskrevne metode blev udarbejdet første gang for året 1990. Der anvendes samme definitioner i erhvervsbeskæftigelsen som i RAS, og de to opgørelser er koordinerede. Der er opbygget et basisregister, hvorfra der foretages udtræk til både den registerbaserede arbejdsstyrkestatistik og erhvervsbeskæftigelsesstatistikken. I sidstnævnte opgøres både

den primære og den sekundære beskæftigelse, dvs. en person kan optages med to beskæftigelsesforhold.

Man kan udtrykke det på den måde, at den registerbaserede arbejdsstyrkestatistik opgør arbejdsmarkedstilknytningen og beskæftigelsen set fra personsiden, mens erhvervsbeskæftigelsen opgør beskæftigelsen set fra erhvervsiden.

Som et andet eksempel på statistikdata indsamlet ved integreret dataindsamling, er oplysninger om personers stilling, jf. afsnit 4.2.

Stillingskoden blev efter ønske fra Danmarks Statistik indføjet i CPR og i skatteregistrene fra omkring 1970. Stillingsbetegnelsen blev anført af borgerne på selvangivelsen, og de kommunale skattemyndigheder skulle sørge for, at stillingsændringer blev registreret. Resultatet af dette var ikke alt for opmuntrende, bl.a. fordi nogle kommuner helt eller delvist undlod at registrere de oplysninger, borgerne havde indsendt. Danmarks Statistik har i flere omgange iværksat oplysningskampagner overfor kommunerne for at få dem til at opfylde deres forpligtelser, hvilket har ført til en vis forbedring.

Eksempel 12: Stillingskoder

4.3.3 Fordele og ulemper ved integreret dataindsamling

Fordelene ved den integrerede dataindsamling kan opsummeres som følger:

Mange fordele...

- Der spares omkostninger i forbindelse med dataindsamlingen.
- Det sikres, at de indsamlede oplysninger er sammenhængende.
- Virksomhederne ulejlighes kun en enkelt gang.

Rækkefølgen ovenfor er ikke helt tilfældig. Omkostningerne ved den integrerede dataindsamling ligger væsentligt lavere end de omkostninger, der skulle have været afholdt, såfremt de supplerende statistiske oplysninger skulle have været indhentet af Danmarks Statistik ved henvendelse til virksomhederne. Der er tale om omkostninger af meget betydelig størrelse, så store, at det formentlig i praksis ville have været økonomisk ufremkommeligt at gennemføre en sådan supplerende dataindsamling. Arbejdspladsprojektets oplysninger, som her er benyttet som eksempel på integreret dataindsamling, er et uundværligt led i det personstatistiske system. De danner bindeledet mellem person- og erhvervsstatistiske opgørelser og er desuden et betydningsfuldt grundlag for den generelle erhvervsstatistik.

...ikke mindst økonomiske

Det forhold, at statistiske oplysninger kan indgå i sammenhæng med administrative data medfører, at de administrative data bliver statistisk berigede og kvalificerede. Oplysningerne bliver sammenhængende, og de kan derfor bedre dække de statistiske behov. Den integrerede data-

Sammenhængende data beriger og kvalificerer

indsamling sikrer således også, at nogle af de traditionelle svagheder ved brugen af administrative data til statistiske formål bliver reduceret. Det fremføres ofte, at de administrative data er vanskelige at udnytte som grundlag for statistiske opgørelser, fordi de er underlagt ændringer i lovgivning og ændringer i administrative rutiner og procedurer. Denne ulempe kan således til en vis grad reduceres ved udformning af de statistiske data, der supplerer de administrative data i den integrerede dataindsamling.

Reduceret respondentbyrde

Respondentbyrden for virksomhederne reduceres. Selv om den samlede datamængde er næsten den samme, hvad enten man sender forespørgsel til virksomhederne i en eller flere omgange, er ulejligheden for virksomheden betydeligt mindre ved en enkelt koordineret henvendelse.

Nogle ulemper...

Der er dog også visse ulemper ved integreret dataindsamling:

- De oplysninger, der indsamles til brug for Danmarks Statistik, kan være fremmede for den administrative myndighed.
- Kontrollen med den korrekte indberetning af oplysningerne til Danmarks Statistik kan blive nedprioriteret af den indsamlede administrative myndighed.

...men de kan reduceres ved samarbejde

Det kan volde problemer at indsamle oplysninger for en anden myndighed, da det administrative personale ikke er bekendte med de oplysninger, der indsamles for en anden myndighed. Instrukser og vejledninger kan afhjælpe problemet, men fortolkningen af tvivlstilfælde kan være vanskelig at håndtere for den indsamlede myndighed. I arbejdspladsprojektets tilfælde kanaliseres nogle af disse problemer til Danmarks Statistik, idet en del af oplysningssedlens oplysninger om arbejdssteder afleveres direkte til Danmarks Statistik.

For den indsamlede administrative myndighed er det naturligt, at man prioriterer egne oplysninger højest. Dem er man bekendte med, og de er desuden af afgørende betydning for hele institutionens arbejde, mens man ikke har direkte interesse i oplysningerne til Danmarks Statistik. Hvis de data, der indsamles til eget brug, ikke opfylder de formelle og legale krav, må man kontakte virksomheden og bede om nye og korrekte oplysninger. Fejlsøgningsrutinerne er indarbejdede og kendte, mens dette ikke gælder for de øvrige oplysninger. I arbejdspladsprojektet har man derfor været ude for, at kontrollen med statistikdataene hos skattemyndighederne ikke i alle tilfælde er tilstrækkelig effektiv. Det er derfor af afgørende betydning, at så mange procesdetaljer som muligt aftales mellem den administrative myndighed og Danmarks Statistik for at sikre, at indberettede fejl opdages og rettes så hurtigt og effektivt som muligt.

4.3.4 Sammenfatning

Fordelene ved den integrerede dataindsamling overstiger langt ulemperne. Denne konklusion er ikke overraskende da dataindsamlingsmetoden ellers næppe ville have været valgt. Afgørende for succesen er imidlertid, at der etableres et godt samarbejde mellem den administrative myndighed og Danmarks Statistik omkring indsamlingen af oplysningerne. Det kan sikre en god kvalitet og en effektiv udnyttelse af den administrative myndigheds professionalisme i indsamling af oplysninger hos virksomheder eller borgere. Kombineres det med professionalisme hos Danmarks Statistik, er den integrerede dataindsamling en effektiv metode til at opnå data af høj kvalitet.

4.4 Imputering

I statistikproduktionen støder man hyppigt på det problem, at ikke alle ønskede og relevante data er oplyst, eller at der ikke er data for alle enheder i statistikken. Et problem man kender fra statistik produceret på basis af administrative registre, såvel som når dataindsamlingen foregår via surveys.

Datamangler

Manglerne kan forekomme i mange forskellige grader. Lige fra det tilfælde, hvor en relevant variabel slet ikke er belyst i registret, til det tilfælde, hvor det drejer sig om mangel af en enkelt ud af flere elementaroplysninger, som skal stykkes sammen for at fastlægge værdien af en bestemt variabel for en enhed.

De manglende oplysninger kan i nogle tilfælde indhentes ved at kontakte datadonorerne igen, men selvom dette er muligt, vil det ofte kræve mange ressourcer og belaste datadonorerne yderligere. Derfor foretrækker man ofte at imputere de manglende oplysninger, dvs. at generere de pågældende data ved hjælp af nærmere fastsatte regler, der knytter sig til andre kendte størrelser³¹.

Behov for imputering

I denne fremstilling benyttes følgende brede definition på *imputering*: Erstatning af en fejlagtig eller manglende værdi med en anden værdi uden fornyet kontakt med dataleverandøren. Dette indebærer, at en meget betydelig del af fejlproportionerne i Danmarks Statistik foregår ved imputering.

Definition

4.4.1 Imputering af oplysninger, som helt mangler

Mangler der information om en variabel i statistikgrundlaget, kan den eventuelt indhentes ved *samkøring* med andre registre eller ved at ud-

Forskellige metoder

³¹ Jf. Danmarks Statistik (1982).

vide indberetningerne til registret således, som det er beskrevet i afsnit 4.2 og 4.3.

I nogle tilfælde vil der kun være mulighed for at indhente oplysningerne ved en *selvstændig dataindsamling*, og det er ofte meget bekosteligt at indhente data for hele registerpopulationen. Dette kan afbødes ved, såfremt det er muligt og forsvarligt, at indhente oplysningerne for et udsnit af populationen, hvorefter der efterfølgende opregnes til hele statistikpopulationen. Nedenfor vises et eksempel, hvor denne fremgangsmåde har været anvendt, og hvor imputeringen således sker i forbindelse med opregningen.

Eksempel 13: Årsag til ledighed

Arbejdsløshedsstatistikken bygger på Det centrale register for arbejdsmarkedsstatistik (CRAM), som modtager indberetninger fra kommuner og arbejdsløshedskasser om alle personer, der i en periode har været berørt af ledighed, og som får udbetalt arbejdsløshedsdagpenge eller kontanthjælp som ledige. På et tidspunkt ønskede man at supplere statistikken med oplysninger om årsagerne til ledighed - altså årsagen til, at personen påbegynder en ledighedsperiode og dermed optræder i registret. Disse oplysninger kunne ikke indhentes ved samkøring med andre registre, og en udvidelse af indberetningerne til registret ville ikke kunne sættes i værk så billigt og hurtigt, som det var ønsket. Endvidere var der generelt stærk modstand mod en sådan udvidelse blandt arbejdsløshedskasserne, som står for indtastningen.

Man har derfor valgt at indhente oplysningerne om forekomsten af ledighed og årsagerne hertil ved interviews, men kun for et udsnit af befolkningen. For hver interviewperiode (et kvartal) sammenholder man interviewpersonerne med Arbejdsløshedsstatistikregistret dækkende samme periode og finder de interviewpersoner, der optræder i registret. Man opregner derefter svarene angående ledighedsårsager for disse interviewpersoner til den samlede registerpopulation.

Kommentar til eksemplet

Metoden er velegnet, hvis man ønsker at udbygge oplysningerne fra et register, der som CRAM er et omfattende register med høj datakvalitet men uden den store variabelrigdom.

Kombination af registerdata med data indsamlet gennem surveys er endvidere velegnet, hvis man ønsker at udbygge statistikgrundlaget med oplysninger af mere kvalitativ karakter, som ikke altid egner sig til at blive indhentet fra administrative registre.

Indsamling af data ved interviewmetoden er ofte meget ressourcekrævende. I forbindelse med den beskrevne undersøgelse af ledighedsårsager holdes omkostningerne dog nede, idet det ikke er nødvendigt at opbygge en selvstændig interviewundersøgelse. Det var muligt at

koble sig på den løbende dataindsamling, der finder sted ved den interviewbaserede arbejdskraftundersøgelse.

4.4.2 Imputering af delvist manglende oplysninger

I visse tilfælde er de relevante oplysninger kun delvist belyst i daggrundlaget. Oplysningerne kan mangle helt for enkelte enheder, eller de kan være af så ringe kvalitet, at de er uanvendelige til statistikbrug.

Generelle overvejelser

I en statistik, der bygger på data indsamlet ved surveys, vil man kunne tage hensyn til dette forhold ved en opregning, idet man antager, at bortfaldet (de delvist manglende enheder) har samme fordeling, som man finder for de kendte enheder eller for en bestemt del af dem.

Det er dog oftest således, at statistikregistre dannes ved samkøring mellem flere registre i sammenhængende systemer, og desuden danner de ofte grundlag for belysning af flere forskellige statistikområder. Man skal derfor foretage nye opregninger hver gang registret indgår i en (ny) specialstatistik.

Da dette ikke er hensigtsmæssigt, foretrækker man at generere en værdi for den manglende variabel gennem anvendelse af et sæt regler knyttet til kendte størrelser. De pågældende værdier kan være forkerte i forhold til den omhandlede enhed, men vil i den aggregerede statistik fremtræde med en fordeling, som med stor tilnærmelse giver en korrekt belysning af variabelen³².

Fastlæggelsen af reglerne afhænger af manglernes karakter og af de oplysninger, man har til rådighed, og der findes en lang række imputeringsteknikker, der dog kan grupperes i 3 hovedtyper³³.

Tre metoder

- imputering ved brug af afledt information
- imputering ved brug af modeller
- imputering ved at overføre data fra andre enheder

Imputering ved brug af afledt information repræsenterer de tilfælde, hvor den manglende information udledes deduktivt, dvs. korrekt eller med stor sikkerhed fra anden information. Dette sker fx i den registerbaserede arbejdsstyrkestatistik, hvor man bl.a. tæller personer, der er beskæftiget på en bestemt dato (sidste arbejdsdag i november). I nogle tilfælde mangler oplysningen om ansættelsesperiode, og dermed om ansættelsen var gældende i november, men er der indbetalt et bidrag til ATP svarende til et årsbidrag, udleder man heraf, at personen har væ-

³² Jf. *Danmarks Statistik (1982)*.

³³ En nærmere beskrivelse af metoderne findes bl.a. i *Australian Bureau of Statistics (1993)* og *Ferguson (1993)*.

ret beskæftiget i november. Hvis ATP-beløbet er mindre end et fuldt årsbidrag afgøres det ved lodtrækning, om personen skal betragtes som beskæftiget pr. novemberdatoen, hvor sandsynligheden for at blive udtrukket som beskæftiget, er proportional med ATP-bidragets størrelse.

Oftest indgår der et skøn i imputeringen. På grundlag af andre oplysninger om enheden, fx uddannelse og branche, vil man kunne imputere eksempelvis en arbejdsstilling. For en løbende statistik vil oplysningen om variabelværdien for samme person fra en tidligere opgørelse eventuelt kunne anvendes.

Eksempel 14:
Arbejdsstederne i
beskæftigelses-
statistikken

Eksemplet er hentet fra den registerbaserede arbejdsstyrkestatistik. Alle beskæftigede personer knyttes her til et arbejdssted, og har arbejdsgiveren flere lokale arbejdssteder, skal hver enkelt ansat placeres på den filial, hvor han eller hun arbejder. Personerne kan så tildeles en række variable vedrørende deres erhvervsforhold, fx branche. Oplysningen om arbejdssted findes i langt de fleste tilfælde, men der er tilfælde, hvor der mangler oplysning om hvilken filial personen arbejder i, og hvor det ikke har været muligt at fastlægge dette i den manuelle behandling. Ved imputering sker der så en placering på de lokale arbejdssteder af de ufordelte personer.

Imputeringen foretages i to trin. Først undersøges det, om det er muligt at henføre personen til samme arbejdssted som sidste år. Dette kræver dog, at arbejdsstedet var oplyst for personen året før, at arbejdsstedet stadig eksisterer, at personen er ansat hos samme arbejdsgiver i år, og at personen ikke har ændret bopæl (over større afstande).

Når imputering til sidste års arbejdssted ikke kan foretages, sker der i næste trin en imputering til nærmeste arbejdssted i forhold til bopælen. Har flere arbejdssteder samme afstand til bopælen, vælges det største arbejdssted (målt på lønsum).

Til fastlæggelse af nærmeste arbejdssted anvendes en såkaldt *afstandsmatrice*, der indeholder oplysninger om alle afstande mellem kommunecentre. Har arbejdsgiveren flere arbejdssteder i samme kommune, kan kun ét indgå i imputeringen, og det største arbejdssted målt på lønsum udvælges. Har man ingen oplysninger om arbejdsstedernes størrelse (fx når ingen af de ansatte hos arbejdsgiveren er fordelt ud på filialerne), udvælges det arbejdssted i kommunen, der har den laveste identifikationskode (arbejdsstedskode), altså et helt vilkårligt valg.

Resultaterne af denne maskinelle imputering til nærmeste arbejdssted kontrolleres herefter manuelt. Har et arbejdssted ved imputeringen fået tildelt et større antal ansættelsesforhold, undersøges det, om imputeringen er acceptabel, eller om der skal ske korrektioner. Her kan sup-

plerende oplysninger fra andre kilder inddrages. Specielt er en manuel efterkontrol nødvendig i de tilfælde, hvor arbejdsgiveren har flere arbejdssteder i en kommune, eksempelvis en detailhandelskæde. Dette skyldes, at den maskinelle imputering henfører alle personer med manglende filialoplysninger, som bor i eller nær den pågældende kommune, til det ene arbejdssted, der er udtaget til at indgå i imputeringen.

Korrektioner foretages ved at fastlægge, hvilken andel af ansættelsesforholdene, der skal flyttes til de forskellige arbejdssteder. Udvalgelsen af hvilke personer, der skal tilknyttes hvilke arbejdssteder, sker tilfældigt.

Ovenstående eksempel beskriver et relativt kompliceret imputeringsforløb, der inddrager en lang række variable og beregninger. Inden imputeringsmetoden vælges, bør den valgte metode testes. I eksemplet beskrevet ovenfor blev metoden således først testet på et fuldstændigt oplyst materiale. Man udtog en gruppe enheder med fuldt oplyste og valide arbejdssteds-koder, undertrykte derpå oplysningen om arbejdssteds-kode og gennemførte en imputeringsproces, hvorefter resultatet heraf kunne sammenlignes med de korrekte resultater. Metoden blev justeret, til den tilnærmelsesvis gav de samme resultater som ved brug af korrekte data.

Evaluering af metoden

Man bør løbende være opmærksom på om der sker ændringer for de (typer af) enheder, der henvises til imputering. Ligeledes skal man løbende følge udviklingen og resultaterne af de foretagne imputeringer, for at sikre sig de mest hensigtsmæssige regler til fastlæggelse af de manglende værdier.

Da de anvendte regler bygger på tidligere erfaringer, kan imputering få en konserverende virkning på statistikken. Den skønnede værdi, der tildeles ved imputeringen, kan være forkert i forhold til den pågældende enhed, og dette kan være et problem, når enheden indgår i flerdimensionale opgørelser. Man bør derfor altid markere, når en værdi er imputeret.

Imputering ved brug af modeller går ud på at anvende viden om gennemsnitsværdien eller medianværdien for den variabel, der skal imputeres for, eller at inddrage viden om variabelens værdi i forhold til værdien for andre variable, som det sker ved ratio- og regressionsimputering.

Er der fx behov for at imputere et lønbeløb, kan en gammel værdi for enheden fra året før blive fremskrevet med den forudsætning, at udviklingen i det seneste år for enheden har været den samme, som blev fundet for de fuldt oplyste enheder af samme type.

Tænkt eksempel

Mens man ved brugen af modeller ved tildeling af imputeringsværdier anvender viden om, hvorledes andre lignende enheder opfører sig, går *imputering ved at overføre data fra andre enheder* ud på at imputere en konkret værdi fra en af de andre lignende enheder. To hyppigt anvendte metoder i denne forbindelse er *cold-deck* og *hot-deck* imputering.

Cold-deck og hot-deck

Ved cold-deck imputering anvender man data fra en anden tælling - der vil ofte være tale om at benytte en tilsvarende, tidligere gennemført tælling - mens man ved hot-deck imputering anvender data fra den aktuelle tælling.

Som centrale elementer i metoderne indgår en donorenhed og en modtager- eller værtsenhed³⁴. I eksemplet nedenfor er anvendt såkaldt hierarkisk hot-deck imputering, hvor man gentagne gange sorterer og forsøger at matche donorer og modtagere på et faldende antal variable, indtil man opnår match.

Eksempel 15: Arbejdskraftundersøgelsen

I arbejdskraftundersøgelserne indhentes for interviewpersonerne bl.a. en lang række oplysninger om beskæftigelsesforhold, jobsøgning og deltagelse i uddannelsesaktiviteter. Interviewdata indhentes for enkeltpersoner, men man ønsker også målvariablene belyst for hele familien.

I Befolkningsstatistikregistret finder man de øvrige personer i familien. Disse personer skal tildeles værdier på målvariablene fra de allerede indhentede interviewdata.

Dette sker ved, at der først samkøres med flere registre, således at man får hver enkelt af interviewpersonens familiemedlemmer beskrevet med køn, alder, bopæl, familietype, uddannelse, arbejdsløshedsberøring m.m. Blandt de gennemførte interviews afgrænser man derefter de interviews, der er gennemført med personer, der matcher den type af personer, der er i familien, hvorefter man for hvert familiemedlem tilfældigt udvælger en passende interviewperson. Herefter kan interviewpersonens familiemedlemmer få tildelt værdier på målvariablene, ud fra hvordan de matchende donorpersoner har svaret.

Lykkes det ikke at finde en donorenhed i første omgang, reduceres antallet af variable successivt, indtil match opnås.

Generelt om metoden i eksemplet

I processen indgår en værtsfil A, som indeholder information om variabelsættene (X,Y) og en donorfil B, som indeholder variabelsættene (X,Z). Man ønsker en samlet belysning af (X,Y,Z) eller (Y,Z) for de enheder, der indgår i værtsfilen. Variabelsættet X, der forbinder enhederne i de 2 filer, kan fx bestå af demografiske variable.

³⁴ En nærmere beskrivelse af forskellige teknikker til udvælgelse af den mest egnede donorværdi findes bl.a. i *Australian Bureau of Statistics (1993)*.

Der er her tale om en metode, hvor man med et begrænset ressourceforbrug får suppleret en statistik med manglende oplysninger, idet man undgår den dyre proces med en selvstændig dataindsamling.

Evaluering

Endvidere er metoden velegnet i forbindelse med registerstatistik, da man her arbejder med datasæt, som er tilstrækkeligt store til at opnå et sikkert match. Anvendes metoden ved mindre datasæt, kan man få problemer med at finde en donorenhed, der helt svarer til variabelsættet X . Og jo mere specificerende X er, jo hyppigere vil man komme i den situation, at man må svække på matchningskravene for at kunne finde en donorenhed. Eventuelt vil matchningsgraden derfor være forskellig for de enkelte sæt af (Y, Z) i datasættet.

Når metoden anvendes, skal man være opmærksom på, at flere forhold kan få betydning for det opnåede resultat. Det skal fx fastlægges, om Z -værdierne skal tildeles med eller uden tilbagelægning, dvs. om samme donorenhed må anvendes flere gange. Metoden bygger på antagelsen om uafhængighed mellem datasættene Y og Z , en forudsætning, man skal teste inden anvendelsen. Et problem for et sikkert match kan være, at de identificerende variabelsættet X (køn, alder, mfl.) har høj korrelation med nogle af variablene i Z , men ikke med dem alle³⁵.

4.4.3 Manuel eller automatisk imputering

Udvælgelsen af en imputeringsværdi kan ske manuelt eller automatisk. Ved manuel behandling er det en person med kendskab til de sammenhænge, der ligger i data, der foretager valget af den imputerede værdi. Ved en automatisk imputering sker valget af værdier maskinelt efter nærmere fastlagte regler.

Ved brug af automatiske såkaldte dataediteringsprogrammer, der omfatter automatiske kodnings- og fejlsøgningssystemer, kan også imputeringsregler være tilkøbet, således at behandlingen af uoplyste variable, fx skoleuddannelse, sker integreret med en automatisk kodning og fejlsøgning. I de tilfælde, hvor skoleuddannelsen mangler, men en senere erhvervsuddannelse er oplyst, kunne afsluttet skoleuddannelse, efter nærmere fastlagte regler, eventuelt genereres i en del tilfælde.

Imputering som led i editering

Der er adskillige fordele ved maskinel imputering, da metoden

³⁵ En nærmere gennemgang og kritik af metoder findes i *Rodgers (1984)* og *Singh, et.al. (1993)*.

- er *billig*, forudsat at antallet af uoplyste enheder har et vist omfang, da der er et vist arbejde forbundet med at udvikle imputeringsprogrammer
- sikrer at alle enheder bliver *behandlet ens*, dvs. efter de samme imputeringsregler
- sikrer at alle har (eller nemt kan få) *kendskab til de regler, der anvendes*

...men manuel kontrol er vigtig

Det fremgik dog også af eksempel 14, at man ikke fuldstændigt bør forlade sig på automatisk dataeditering eller imputering, idet man ved den manuelle datarevision får kontrolleret de regler, imputeringen bygger på.

4.5 Surveys og registre

De to metoder

Hovedkilden for de oplysninger, der anvendes i personstatistikken, er som nævnt administrative registre. Som det ligeledes er fremgået er oplysningerne i nogle tilfælde utilstrækkelige eller af manglende aktualitet, og det kan derfor være nødvendigt at indsamle dem direkte hos de omhandlede personer eller et udsnit af disse, dvs. gennemføre en survey.

Der har tidligere været en tendens til at betragte statistik byggende på administrative registre og statistik byggende på surveys som to helt adskilte verdener. Denne indgangsvinkel har ikke været frugtbar; de to former for dataindsamling har ganske vist hver sine styrker og svagheder, men pointen er, at metoderne netop derfor kan spille sammen og berige hinanden.

Fordele og ulemper ved surveymetoden

Fordelen ved surveydata er især, at man direkte får mulighed for at styre dataindsamlingen, idet man kan stille spørgsmål om netop de emner, man ønsker belyst. Hertil kommer, at surveydata - især hvis moderne teknik i form af CAPI, CATI mv. anvendes - på nogle områder vil kunne bearbejdes og offentliggøres meget hurtigt i forhold til registerdata.

Ulempen ved surveybaseret statistik er, at det er en ret omkostningskrævende dataindsamling, at man uvægerligt støder ind i stikprøveskævheder og frafaldsproblemer, samt at de interviewede husstande, personer mv. måske ikke altid er i stand til eller ønsker at give korrekt svar.

Fordele og ulemper ved registermetoden

Fordelen ved registerbaseret statistik er især, at den (her i landet) er stort set totaldækkende, at dataene fra de enkelte administrative kilderegistre gennemgående er konsistente og af høj kvalitet, samt at indsamlings- og bearbejdningsomkostningerne er relativt små. Registerdata rummer desuden meget store potentialer, idet der kan foretages

sammenkobling mellem forskellige registre ud fra veldefinerede nøgler (personnummer, adressekode osv.).

Ulempen ved registerstatistik er bl.a., at man i vid udstrækning er bundet af de definitioner og den administrative praksis, der anvendes af de registeransvarlige myndigheder. Endvidere kommer statistikproducenten forholdsvis langt væk fra dataindsamlingen, hvilket kan medføre mangelfuld viden om det præcise dataindhold, om datakvaliteten m.m.

Begge typer af statistik har altså helt klare fortrin. Det forekommer derfor oplagt, at man ved at kombinere metoderne vil kunne opnå et godt samspil og dermed en forbedret statistik.

Muligheder for samspil

I det følgende ses der nærmere på tre former for samspil:

- surveydata suppleres med registerdata
- registerdata suppleres med surveydata
- sammenligning af register- og surveydata med henblik på kvalitetsanalyser

Den første form for samspil har med stort held været anvendt i mange sammenhænge, mens de to sidste er forholdsvis uopdyrkede områder i Danmark.

4.5.1 Surveydata suppleret med registerdata

I Forbrugsundersøgelsen fra 1987 samt i den nye, løbende forbrugsundersøgelse, som blev iværksat i 1993, anvendes registerdata i vidt omfang som supplement til surveydata indhentet ved besøgsinterviews hos husstande. De registerdata, der anvendes, er hovedsagelig indkomstdata, samt oplysninger om erhverv, uddannelse og bolig.

Anvendelsen af registerdata er teknisk set simpel, idet Forbrugsundersøgelsen indeholder personnumre og adressekoder for alle de personer og husstande, der indgår i undersøgelsen. Det er derfor enkelt at foretage en sammenkobling med eksisterende registeroplysninger og herved overføre registerbaserede oplysninger til Forbrugsundersøgelsen.

Eksempel 16: Forbrugsundersøgelsen

Når dette sker i et betydeligt omfang, er årsagen især, at det sparer tid i selve interviewsituationen, og tid er i den sammenhæng en direkte omkostningsfaktor og kvalitetsparameter. Endvidere får man på denne måde i en del tilfælde mere præcise oplysninger - fx kan oplysninger om modtagen løn normalt fås bedre fra registre end ved at spørge personerne selv. På den anden side er det klart, at man inden for en del emneområder ikke kan nøjes med at benytte registerdata, men er nødt

til at stille supplerende interviewspørgsmål, enten fordi registerdata ikke findes, eller fordi de ikke kan specificeres i den fornødne grad.

I forbrugsundersøgelserne er det fx fundet vigtigt at opgøre husstandenes samlede indkomster og andre økonomiske tilgange. De store beløb i form af lønindkomst, offentlige pensioner og virksomhedsoverskud mv. hentes via registre, men en del indkomster kan ikke findes i de administrative registre. I interviewet med husstandene indhentes derfor oplysninger om mere perifere indkomstarter, herunder bl.a. støtte fra familie, forsikringsudbetalinger, gevinster og indkomst ved sort arbejde. Disse indkomstarter tegner sig for alle husstande under ét for relativt beskedne beløb, men for enkelte husstande kan de være af afgørende betydning som forklaringsvariabel i relation til forbrug, levestandard mv.

Metoden er dog ikke problemfri. For det første er det vanskeligt at sikre en helt konsistent begrebsafgrænsning. De indkomstarter, som er dækket af Indkomststatistikregistret, må ikke indgå i de beløb, som indhentes gennem interview, da der herved opstår dobbeltregning. Det er i visse tilfælde ganske kompliceret at sikre sig mod denne fejltipe, som kan skyldes upræcise interviewspørgsmål og/eller manglende viden hos respondenterne. Specielt har dette problem været følt omkring forskellige former for pensionsudbetalinger, hvor det er ganske svært for husstandene præcist at redegøre for, hvilken type udbetaling, der faktisk er tale om, og herunder om den er underkastet almindelig indkomstbeskatning (og derfor indgår i registrene).

Hertil kommer periodiseringsproblemet, som opstår fordi skattemyndighederne ikke altid anvender samme periodisering, som husstande eller personer finder naturlig.

Fordele

Som det fremgår af eksemplet, er der i praksis forskellige vanskeligheder. Metoden med at supplere interview med registeroplysninger er dog i eksemplet klart mere effektiv, end at hente alle data ved interviews, og der vurderes også samlet at være kvalitative fordele. I 1987-forbrugsundersøgelsen blev 93 pct. af de samlede indkomster hentet via registerdata, og man skulle derfor kun indhente oplysninger om de resterende 7 pct. gennem interview³⁶.

4.5.2 Registerdata suppleret med surveydata

Metoden er ikke udforsket i praksis

I Danmarks Statistik er der kun få konkrete eksempler på, at registerdata suppleres med surveydata. Et eksempel herpå er dog omtalt i afsnit 4.4.1. Det er sandsynligt, at der i fremtiden vil blive gjort langt større anvendelse af denne metode.

³⁶ Jf. *Danmarks Statistik (1992d)*.

Inddragelsen af supplerende surveyoplysninger kan ske på 2 måder:

Principper

- På *makroniveau*, ved at supplere med summariske surveydata, fx i form af at bringe et tillæg med tabeldata fra surveys ved offentliggørelsen.
- På *mikroniveau*, ved at udnytte surveydata til at overføre/imputere supplerende oplysninger eller egenskaber for de enkelte enheder i registrene.

Et eksempel til belysning af problemstillingen er sorte indkomster, som pr. definition ikke indgår i de officielt registrerede indkomster, og dermed ikke i indkomststatistikken. Man ved ikke præcis, hvor stort et beløb disse indkomster tegner sig for, eller hvordan de er fordelt. Samtidig er der ikke megen tvivl om, at disse ikke-registrerede indkomster for en del husstande har et ret betydeligt omfang. Hvordan kunne man evt. forbedre statistikken på dette punkt?

Fra interviewundersøgelser (fx Rockwool-projektets undersøgelser³⁷) har man en beskrivelse af det sorte arbejde og dets fordeling blandt forskellige typer af husstande. Beskrivelsen er nok langt fra dækkende, men er alligevel bedre end ingenting. Det kunne derfor være naturligt at inddrage disse supplerende oplysninger om de sorte indkomster mv. i Danmarks Statistiks generelle indkomststatistik.

Eksempel 17: Sorte indkomster

På makroniveau vil dette kunne ske ved at tillægge et skøn over de sorte indkomster i alle eller udvalgte tabeller, idet skønnene vil skulle foretages for alle de kategorier af husstande og personer, som indgår i tabellerne. Det bemærkes, at disse skøn kan være ret komplicerede at foretage i det omfang, tabelkategorierne ikke svarer til de kategorier, der anvendes i de supplerende surveys.

For at komme ud over dette problem, kunne man i stedet imputere på mikroniveau. Dette ville give den fordel, at man ved den efterfølgende anvendelse af materialet ikke var bundet til på forhånd valgte grupperinger af husstande og personer mv., men kunne bruge materialet fleksibelt ved fx at være i stand til at kunne ændre gruppering i indkomstintervaller m.m.

Et sådant projekt vil således bestå i, ud fra informationer i de pågældende surveys, at importere skøn over sorte indkomster til enkeltindivider i registrene. Hertil må anvendes specielle metoder - fx en form for statistisk match³⁸. Proceduren kunne fx være følgende:

³⁷ Jf. fx Smidt, et.al. (1989) og Mogensen (1990).

³⁸ Statistisk match er en metode, der stort set svarer til cold-deck imputering (jf. afsnit 4.4.2).

Survey materialet opdeles i et større antal relevante strata. Inden for hvert stratum opstilles en fordelingsfunktion over forekomsten af og størrelsen af de sorte indkomster. Registerpopulationen opdeles i tilsvarende definerede strata. På tilfældig vis tilregnes nu et antal individer inden for hvert stratum sorte indkomster på en sådan måde, at fordelingsfunktionen efter denne import er den samme i registerpopulationen som i surveypopulationen.

Resultatet heraf vil være, at man stadig på makroniveau vil have de bedst mulige tal for de sorte indkomster. Samtidig vil omgrupperinger af husstande og personer kunne ske fleksibelt, idet resultaternes kvalitet dog vil være helt afhængig af, hvor heldig man har været ved den gennemførte stratificering, dvs. hvor entydig og stabil fordelingen af de sorte indkomster kan bestemmes ud fra de tilgængelige baggrundsplysninger. I det omfang, stratificeringen er stærkt korreleret med forekomsten af de sorte indkomster, vil resultatet blive godt.

Evaluering af eksemplet

Man kan med en vis ret hævde, at de "datamanipulationer", som er beskrevet ovenfor, ikke kan give egentlig ny erkendelse, men blot er udtryk for en særlig databearbejdning, som åbner for en mere fleksibel og hensigtsmæssig præsenteringsform.

Det er klart, at opgørelsen af fx det sorte arbejde byggende på surveydata aldrig vil kunne blive bedre end surveymaterialet i sig selv tillader. Man skal derfor være ganske forsigtig ved anvendelsen af sådanne syntetiske totaltal. Hvor detaljeret, man på faglig forsvarlig vis kan tillade sig at offentliggøre sådanne syntetiske tal, må afhænge af en vurdering af kvaliteten af den pågældende survey samt en vurdering af, hvor præcis modellen eller fordelingsfunktionen for de sorte indkomster er.

Der er som nævnt ikke megen erfaring i Danmarks Statistik på dette område, og der er uden tvivl behov for et betydeligt udviklingsarbejde, både teoretisk og praktisk.

4.5.3 Sammenligning af register- og surveydata, kvalitetsanalyser

Mulighederne

De oplysninger, man indsamler i en survey, kan dels ses som en statistik i sig selv og dels som en udbygning af registerstatistikken. Beskæftigelsesstatistikken i Danmarks Statistik er et eksempel på dette. Statistikken omfatter en *registerbaseret* arbejdsstyrkestatistik (RAS) og en *surveybaseret* arbejdsstyrkeundersøgelse (TASU). De to dele er hver for sig grundlag for belysning af arbejdsmarkedsforhold i Danmark, men surveyoplysningerne kan også ses som supplerende information, der kan integreres i RAS med henblik på en udbygning af visse områder med nye eller omdefinerede variable. De sammenfaldende

begreber i de to typer af oplysninger kan samtidig danne grundlag for en sammenligning af dataindholdet, hvilket er temaet i dette afsnit.

Ved at benytte registerdata, som er knyttet til de til enhver tid gældende regler, vil der, når reglerne ændres, være en risiko for, at de anvendte begreber ikke er i overensstemmelse med de oprindelige definitioner, eller at betydningen af valgte kriterier kan vise sig at være anderledes end forventet. Derfor er der behov for en analyse af registerstatistikbegrebernes faktiske indhold for at sikre, at man belyser det tilsigtede. På surveysiden formuleres en række spørgsmål, som tilsammen repræsenterer en form for "idealdefinition" af de begreber, man tilsigter at belyse. Man kan ikke altid forvente, at de indsamlede og behandlede oplysninger er i overensstemmelse med denne idealdefinition, men en sammenstilling af oplysningerne herfra med tilsvarende oplysninger fra registerstatistikken giver alligevel mulighed for at undersøge, om der er utilsigtede virkninger af den fremgangsmåde, man anvender i behandlingen af data fra de to typer af kilder.

I efteråret 1993 påbegyndtes i Danmarks Statistik et nyt projekt - Datakvalitetsprojektet. Den første opgave gik ud på at sammenligne data fra arbejdstyrkeundersøgelsen og den registerbaserede arbejdsstyrkestatistik og et eksempel hentet herfra kan benyttes til at illustrere mulighederne.

Der er i høj grad brug for sådanne kvalitetsstudier, og de er desværre kun blevet gennemført i meget begrænset omfang i Danmark³⁹. Man må være opmærksom på, at sådanne kvalitetsanalyser kan påvise forskelle mellem kilders resultater, men de kan ikke påvise, hvad der er "sandheden" eller hvad der er det mest korrekte.

Analysen kan enten foretages på *summarisk plan*, fx ved at sammenligne de to statistikkers resultater mht. arbejdsløshedens størrelse og sammensætning, beskæftigelsesstrukturen mv.; eller analysen kan foregå helt detaljeret, dvs. på *enhedsniveau*, hvor oplysningerne om de enkelte personer i hvert af de to datasæt sammenlignes. Det er sidstnævnte indgangsvinkel, som er benyttet i det følgende eksempel.

I undersøgelsen indgik på *surveysiden* data fra den telefoninterview-baserede Arbejdsstyrkeundersøgelse fra 1991 (*TASU-91*)⁴⁰ og på registersiden Den registerbaserede arbejdsstyrkestatistik fra 1991 (*RAS-91*)⁴¹ samt bearbejdede oplysninger fra Det centrale oplysningsseddelregister og Arbejdspladsstatistikregistret (CON-91).

Behovet for kvalitetsanalyser

Datakvalitetsprojektet

Analyse på summarisk niveau og på enhedsniveau

Eksempel 18: RAS-91 og TASU-91

³⁹ I Sverige er der gennemført en del undersøgelser, jf. *Sandgren (1993)*. I Finland er en meget omfattende analyse dokumenteret i *Korp (1989)*.

⁴⁰ Jf. *Danmarks Statistik (1992a)*.

⁴¹ Jf. *Danmarks Statistik (1993a)*.

Uanset hvilken fremgangsmåde man vælger ved sammenligning af surveys og registre, med henblik på evaluering, bør følgende forudsætninger være opfyldt:

- *Tids-konsistens*: De oplysninger der sammenstilles skal tidsmæssigt stemme overens.
- *Begrebs-konsistens*: De begreber eller variable der sammenlignes skal begrebsmæssigt stemme overens.

Med hensyn til forudsætningen om tidsmæssig konsistens, er denne ikke opfyldt hvad angår oplysninger fra TASU-91 og RAS-91. Referenceperioden i TASU-91 er fra 11. marts til 28. april 1991, mens RAS-91 refererer til beskæftigelsesforhold i sidste uge af november 1991. Dette skaber umiddelbart et problem, idet uoverensstemmelser mellem oplysninger blot kan skyldes, at personer har skiftet status i den mellemliggende periode.

For at tage højde for ovenstående problem er oplysninger fra CON-91 inddraget i analysen, med det formål at agere tidsmæssigt "bindeled" mellem de to opgørelser. CON-91 indeholder bearbejdede oplysninger om alle de ansættelsesforhold, der er registreret via de oplysningssedler, som alle arbejdsgivere hvert år indsender til Told- og Skattestyrelsen. På baggrund af disse er det til en vis grad muligt at tidsfæste ansættelsesforhold, og derved kan fx personer med ansættelsesforhold, der strækker sig over perioden fra 11. marts til sidste uge i november 1991, afgrænses.

Det er dog forbundet med visse problemer at anvende disse oplysninger. Det er flere gange vist, at specielt oplysningerne om ansættelsesperiode på oplysningssedlerne ikke stemmer helt overens med virkeligheden⁴², men det er fundet nødvendigt at anvende oplysningerne, for at kunne skabe en form for tidskonsistens. Resultaterne der præsenteres i det følgende, skal derfor ses i lyset af disse problemer.

Det begreb som man sammenligner i denne analyse, nemlig branche, er principielt det samme i de to kilder. De forsøger altså at måle samme forhold, men målingsprocessen er væsentlig forskellig. Som udgangspunkt er der væsentlig forskel i på den ene side at anvende et formaliseret begrebs- og klassifikationsapparat, som det gøres i det meste registerstatistik, og på den anden side at "begrebsdefinere" og klassificere ud fra svar i en survey. Dataindholdet i de to datasæt om det samme forhold (fx personers branchetilhørsforhold) er påvirket af de metoder, der anvendes til at klassificere, af troværdigheden i de interviewede personers svar samt af kodningens kvalitet.

⁴² Jf. *Danmarks Statistik (1991d)* og *Nielsen (1987)*.

I eksemplet nedenfor forsøges det at vurdere, i hvor stort omfang brancheoplysningerne i RAS-91 og TASU-91, for det man kan kalde de *stabile personer*, stemmer overens. Med stabile personer menes personer, der indgår i såvel TASU-91 som RAS-91, og som har ét ansættelsesforhold (i følge CON-91), der strækker sig over minimum den periode, der ligger mellem disse (10. marts - 1. december 1991)⁴³.

8.707 personer opfylder stabilitetskravet. For disse personer er branchetilhørsforholdet ifølge TASU-91 og RAS-91 sammenlignet. I TASU-91 drejer det sig om spørgsmålet: "Hvilken slags virksomhed er De ansat i/arbejder De i?" og i RAS-91 om formålsbranchekoden for arbejdsstedet, som stammer fra Det centrale erhvervsregister⁴⁴. Begge oplysninger er kodet med den samme 5-cifrede brancheklassifikation (DSE-77)⁴⁵, som i øvrigt bygger på FN's klassifikation (ISIC-68)⁴⁶.

Resultatet af sammenligningen på den 5-cifrede branchekode er, at for 5.936 personer (68,2 pct.) stemmer branchekoden overens og for 2.771 personer (31,8 pct.) er der tale om uoverensstemmende branchekoder. Sidstnævnte gruppe er så stor, at den bør analyseres nærmere.

I branchegrupperingen refererer 1. ciffer til hovedbrancheområde, hvorfor man kunne opstille den hypotese, at en del af forskellen i den 5-cifrede branchekode blot er et udtryk for, at det kan være vanskeligt at placere personer på et så detaljeret niveau, mens en placering på 1. ciffer skulle passe. Derfor er 1. ciffer i de to branchevariable krydset for de 2.771 personer.

Ved at sammenligne på branchekodens 1. ciffer reduceres uoverensstemmelsen til 11,6 pct., hvilket til dels støtter ovenstående hypotese.

Der kan være flere årsager til den generelle uoverensstemmelse:

1. Manuel kodning af svar i TASU.
2. Registeroplysningernes kvalitet.
3. Kvaliteten af brancheoplysningerne i Det centrale erhvervsregister.
4. De interviewedes "viden" om deres branchetilhørsforhold.
5. Andelen af interviews, hvor interviewpersonen ikke selv svarer.
6. Andre årsager.

⁴³ Det skal bemærkes, at der ikke er sat som betingelse, at personerne skal være beskæftigede i følge RAS-91.

⁴⁴ Det skal bemærkes, at i de tilfælde hvor personen indgik i TASU-90, er der mulighed for at markere, hvorvidt den branche, der blev oplyst her, stadig gælder.

⁴⁵ Danmarks Statistiks erhvervsgrupperingskode af 1. april 1977.

⁴⁶ International Standard Industrial Classification of All Economic Activities, 1968-versionen.

Det er vanskeligt at estimere, hvor meget de forskellige faktorer betyder. Hvad angår pkt. 1. foregår der i TASU en kodning af interviewpersonernes svar. Kodningen går til dels ud på at omsætte en branchetekst til en defineret branchekode. En mangelfuld branchetekst kan således give anledning til en forkert branchekode. Desuden skal man være opmærksom på, at det er forskellige personer, der foretager kodningen, og der vil erfaringsmæssigt opstå uoverensstemmelser på baggrund af dette. Ligeledes kan der være tale om enkelte fejlkodninger.

Registeroplysningernes kvalitet mht. periodeangivelser er allerede nævnt. Det bør også nævnes, at en stor enquete i 1993 blandt virksomheder med henblik på at rette op på brancheoplysningerne i Det centrale erhvervsregister viste, at der var fejl i ca. 20 pct. af branchekoderne. Det skal dog bemærkes, at der for hovedpartens vedkommende var tale om små virksomheder.

Med hensyn til pkt. 5 bliver det i TASU registreret, hvem i husstanden, der er blevet interviewet. En overrepræsentation af interviews, hvor andre end interviewpersonen selv er interviewet, kan tænkes at være en indikator for usikkerheden i de anførte svar. Ud af de 2.771 personer (interviews) viser det sig, at i 2.196 tilfælde er det interviewpersonen selv der er interviewet, i 482 tilfælde er det interviewpersonens ægtefælle/samlever og i 93 tilfælde er der tale om andre personer. En lignende fordeling gør sig gældende for de 5.936 personer for hvem branchekoderne stemmer overens. Der synes således ikke at være nogen indikation af, at de uoverensstemmende branchekoder skyldes, at interviewpersonen ikke selv er interviewet.

Kommentar til eksemplet

Eksemplet viser et enkelt resultat fra de første analyser under Datakvalitetsprojektet, som omfatter sammenligninger vedrørende branche-, stillings- og ledighedsoplysninger. Hvad angår brancheoplysningerne fremgik det, at der var tale om ikke uvæsentlige uoverensstemmelser.

Der er formentlig mange forklaringer på dette, men det er vanskeligt at udpege de mest betydende. Dog har én forklaringsfaktor været fremhævet, nemlig betydningen af kvaliteten i de oplysninger om ansættelsesforhold - specielt periodeangivelser - der hvert år indsendes til Told- og Skattestyrelsen via oplysningssedlerne.

4.6 Afslutning

Brug af registre ikke uden problemer...

Som det er fremgået af det foregående, er der en del problemer forbundet med at basere statistikken på administrative registre. Det kan undertiden være vanskeligt at få de relevante oplysninger til statistikken, fordi statistikerne ikke har kontrollen over data, og i den forbin-

delse kan man risikere at få et skævt verdensbillede. Data kan være fejlbehæftede, hvilket der dog undertiden kan rådes bod på ved omhyggelig fejlsøgning og editering. Tidsreferencerne kan være mangelfulde, og korrektioner kan være vanskelige eller umulige at skelne fra oplysninger om virkelige hændelser. Endelig kan det give store kontinuitetsproblemer, når registrene eller reglerne bag dem ændres.

På den anden side har registerstatistikken store fordele, hvilket de danske erfaringer har vist. Det gælder ikke mindst de økonomiske fordele, som opnås, når oplysningerne allerede er indsamlet til andet formål, og samtidig kan man nedbringe befolkningens indberetningsbyrde. Statistikkens hurtighed kan i nogle tilfælde forbedres væsentligt. Kvalitetsmæssigt er registrenes totale populationsdækning en stor gevinst, og data er ofte mere pålidelige i registrene. Endelig skal det nævnes, at registerdata giver helt enestående muligheder for at gennemføre forløbsundersøgelser.

I nogle tilfælde er det nødvendigt at sammenstykke oplysninger fra flere kilder for at kunne danne de inddelinger, som skal anvendes i statistikken. Metoden benyttes især i den registerbaserede arbejdsstyrkestatistik, hvor der trækkes på mange kilder for at klassificere personer efter beskæftigelses- og erhvervsforhold.

Fremgangsmåden er ofte ganske kompliceret at håndtere i praksis. Det kan være vanskeligt at opnå konsistens, bl.a. fordi opdateringstidspunkter, definitioner, afgrænsninger og prioriteringsregler kan variere i de forskellige kilderegistre. Der er dog opnået gode resultater, og metoderne udvikles og forbedres løbende.

På enkelte områder, bl.a. i Arbejdspladsprojektet, gør Danmarks Statistik brug af integreret dataindsamling, hvilket betyder, at de statistiske oplysninger til Danmarks Statistik indhentes af en anden myndighed samtidig med, at myndigheden indsamler oplysninger til eget brug. Dette har vist sig at indebære store fordele både for indberetter og producent, men forudsætningen for et godt resultat er et tæt samarbejde mellem statistikproducenten og den myndighed, som varetager dataindsamlingen.

Mangelfulde eller manglende data kompletteres i reglen ved imputering, dvs. uden at kontakte dataleverandøren. En udstrakt anvendelse af imputering er nærmest en forudsætning for at realisere en registerbaseret statistikproduktion af den kvalitet og i den skala, som forekommer i Danmark.

Der benyttes flere former for imputering, og nogle varianter bygger på raffinerede modeller og antagelser. Der er her behov for et fortsat metodearbejde.

...men også store fordele

Brug af flere kilder...

...er kompliceret, men har givet gode resultater

Integreret dataindsamling

Imputering er et fundament i fejlopretning

Flere varianter

Surveys kombineret med registre

Det samme gælder for de metoder, der benyttes til at kombinere data fra surveys og registre. Hovedkilden er registerdata, men der kan naturligvis opregnes fordele og ulemper ved hver af de to indsamlingsformer. Der kan opnås store økonomiske fordele ved at kombinere data fra de to typer af kilder. I praksis er der mange eksempler på, at surveys suppleres med registerdata, hvorimod det omvendte indtil videre er et forholdsvis uopdyrket område i Danmark.

Kvalitetsanalyser

Kombination af survey- og registerdata benyttes desuden i forbindelse med kvalitetsanalyser. I Danmarks Statistik er iværksat et såkaldt datakvalitetsprojekt, hvor der i første omgang er foretaget sammenligninger på personniveau af beskæftigelsesoplysninger indhentet hhv. i RAS og TASU. Væsentlige forskelle kan konstateres, men fortolkningen af de foreløbige resultater er vanskelig, og der er behov for et fortsat udviklingsarbejde.

Kapitel 5. Integrationsregistre

I dette kapitel omtales dannelsen af integrationsregistre, dvs. statistikregistre der er kombinationer af flere andre statistikregistre. Først omtales behovet for integration, idet der sondres mellem horisontal og vertikal integration. Disse integrationsformer belyses gennem eksempler.

5.1 Behovet for integration

Som tidligere omtalt er muligheden for at opbygge et samlet, statistisk system baseret på administrative registre stærkt afhængig af mulighederne for samkøring eller *integration* af flere registre eller registerveritioner. Herved forstås sammenkædning af information på enhedsniveau ved hjælp af en nøgle. De statistiske grundregistre, som er omtalt i bilag 2, er i de fleste tilfælde dannet ved hjælp af integration af administrative registre.

Hvert statistikregister er indrettet til at belyse én statistikgren, dvs. til at belyse ét emneområde, og registrene er adskilt fra hinanden, bl.a. for at undgå unødige ophobninger af store mængder af oplysninger om de enkelte personer. Af datasikkerhedsmæssige grunde er der skodder mellem de forskellige registre, idet det kræver særlig autorisation at få adgang til et register. Registerne er endvidere typisk organiseret sådan, at hvert register findes i et antal adskilte versioner, der hver for sig tilstræber at beskrive emnet på et bestemt tidspunkt, fx 1. januar i år x, eller i en bestemt periode, typisk et år.

Når der er et særligt behov for at kombinere data fra forskellige statistikgrene eller fra forskellige perioder til brug for en bestemt statistik eller til en aktuel opgave, kan man imidlertid integrere forskellige statistikregistre eller flere versioner af et statistikregister. I sådanne tilfælde taler man om *integrationsregistre*. De statistikregistre, der indgår i det integrerede register, er allerede blevet bearbejdet og verificeret hver for sig, så indholdet så godt som muligt modsvarer begreberne inden for det statistikområde, registret er skabt til. Erfaringen viser imidlertid, at der altid dukker en lang række nye problemer op, når statistikregistre bliver integreret.

Efterhånden som det samlede statistiksystem er blevet udbygget og de mest grundlæggende behov dermed er blevet opfyldt i den løbende statistikproduktion, synes der at komme en stigende interesse for at gå på tværs af statistikområderne. Statistikbrugere, specielt samfundsforskere, ønsker at belyse sammenhænge mellem fænomener, som traditionelt er henregnet til to eller flere emneområder, og de ønsker at ana-

Statistikregistre er adskilte...

...men kan bringes sammen efter behov

Behov for at gå på tværs og på langs

lysere tidsforløb for de enkelte grupper af individer, hvor den løbende statistik kun kan belyse tværsnit.

Hvis man ikke havde muligheden for integration, ville den eneste mulighed for at tilfredsstille sådanne behov være at udforme skræddersyede spørgeskemaundersøgelser, hvor hver respondent skulle svare på spørgsmål om alle de forhold, man ønskede at sammenstille oplysninger om. Det ville i øvrigt ofte være umuligt at opnå blot nogenlunde anvendelige svar på sådanne spørgeskemaer, dels på grund af stort bortfald, dels på grund af respondenternes manglende evne til at huske eller gengive lange tidsforløb. Det er desuden klart, at sådanne undersøgelser ville blive meget omkostningskrævende. Derfor er integration af statistikregistre en meget attraktiv kilde til ny viden.

Horisontal integration

I nogle tilfælde drejer det sig om at integrere informationer fra to (eller flere) forskellige registre med henblik på at danne en foreningsmængde af *egenskaber* eller kombinationer af mængder af *enheder* fra de to registre. Dette kaldes *horisontal integration*. Ordet *horisontal* betyder, at man så at sige bevæger sig vandret mht. tidsdimensionen, idet data typisk vil vedrøre ét bestemt tidspunkt eller én periode.

Vertikal integration

Når et register i to eller flere tidsmæssigt forskudte versioner sammenholdes på individniveau med det formål at konstatere tilgang og afgang af *enheder* eller ændringer i enhedernes *egenskaber*, taler man om *vertikal integration*.

5.2 Horisontal integration

Horisontal integration kan have forskellige formål:

- at supplere egenskaber fra ét register med nye egenskaber fra et andet
- at kontrollere validiteten af oplysninger i et register mod oplysninger i et andet
- at afgrænse en population

Ofte vil der være tale om flere af disse formål på en gang.

Supplering af egenskaber

Supplering af egenskaber fra ét register med nye egenskaber fra et andet kan betyde, at en variabel overføres til "hovedregistret". Der kan også være tale om, at værdisættet i en variabel i hovedregistret udvides eller ændres ved hjælp af de supplerende oplysninger. I begge tilfælde er det helt afgørende, at man kan bringe oplysninger sammen om eksakt samme enhed i de to kilder, hvilket igen beror på, at enhederne er entydigt definerede, og at identifikationerne er korrekte. Dette skulle normalt ikke være noget problem, når det som her drejer sig om at samkøre to konsoliderede statistiske registre.

En væsentlig faktor, der dog kan vanskeliggøre integration, er forskelle i tidskomponenten i de to kilderegistre. Det kan fx skyldes, at to registre har væsentligt forskellige opgørelsestidspunkter, eller at det ene register er et statusregister med opgørelse pr. en given dato og det andet et register, der indeholder forløbsdata for de samme enheder over en periode, hvilket er en meget almindelig situation.

Et andet formål med horisontal integration er at kontrollere data i et register ved at sammenholde dem med data i et andet register. Det skal bemærkes, at det ofte vil være umuligt at afgøre, hvilken af de to kilder, der er den mest korrekte, med mindre man kan supplere med meget intensive kvalitetsstudier. Eksempler på denne type integration findes omtalt i afsnit 4.5, hvor statistikregistre baseret på administrative kilder sammenholdes med registre baseret på spørgeskemaer.

Kontrol af validiteten

Endelig kan formålet med integration være at afgrænse en eller flere populationer. Det kan være foreningsmængden af enheder i de to registre, hvilket kan være interessant, hvis man har to registre, som ikke er komplette. Det kan også være fællesmængden eller komplementærmængden, dvs. mængden af enheder, som er i det ene, men ikke i det andet register.

Afgrænsning af en population

Den sammenhængende socialstatistik er et skoleeksempel på horisontal integration, hvor en lang række statistikregistre forenes i et kompliceret system. Kompleksiteten forøges af, at statistikken skal beskrive virkningerne af en lovgivning, som er meget indviklet, og som ændrer sig ganske meget fra år til år. Eksemplet indeholder også vertikal integration, idet man beskriver tidsforløb for de enkelte modtagere af sociale ydelser.

Eksempel 19: Sammenhængende socialstatistik

Statistikens baggrund

Danmarks statistik har fra starten af 1980'erne etableret en række personorienterede statistikregistre på det sociale område, som hver for sig belyser én type overførselsindkomster eller et lovområde (se bilag 2). Statistikkerne er baseret på oplysninger i administrative registre, der føres af kommunerne (fælleskommunale afregningssystemer). Der var imidlertid behov for at få en sammenhængende statistisk belysning af de forskellige ordninger. Forskellige undersøgelser fra midten af 1980'erne påviste, at der syntes at være tale om en ophobning af langtidsofrettede inden for visse dele af det sociale system, samtidig med at de enkelte sociale delsystemer - i kraft af forskellige lovgivninger - i ringe grad var i stand til at koordinere de sociale opgaver.

I 1983 iværksattes et forsøgsprojekt med tværgående opgørelser over modtagere af indkomsterstøttede ydelser, og de følgende år gennem-

førtes flere forskningsprojekter, der inkluderede såvel horisontale som vertikale registersamkøringer.

Den sammenhængende socialstatistik blev optaget på Danmarks Statistiks arbejdsplan for 1987, og første offentliggørelse omfattede oplysninger om perioden 1984-89.

Statistikens afgrænsning og formål

Den sammenhængende socialstatistik er et *personorienteret register* over modtagere af *indkomsterstattendes ydelser*. En indkomsterstattendes ydelse har til formål at sikre personens eller familiens forsørgelsesgrundlag, når der optræder et indkomstbortfald som følge af ledighed, sygdom, invaliditet, alderdom eller en anden social begivenhed.

Statistiksystemet har til formål at belyse personens/familiens kontakt med de indkomsterstattendes ordninger, målt på:

- ordningens art (lovområde)
- ydelsens art
- omfang; varighed (målt i dage) og størrelse (målt i kroner)
- bevægelser mellem de enkelte ordninger/ydelser

Desuden beskrives personens og familiens sociale baggrund⁴⁷.

Ordningerne

Statistikken omfatter følgende indkomsterstattendes ordninger, som overvejende er offentligt finansierede:

- dagpenge ved arbejdsløshed
- dagpenge ved sygdom og barsel
- aktiverings- og uddannelsesordninger
- bistandslovens forsørgelsesordninger (kontanthjælp eller revalidering)
- sociale pensionsordninger
- efterløn
- tjenestemandspension

Det er et mål at udvide statistikken med private pensionsordninger samt aktiverings- og orlovsordninger.

⁴⁷ Jf. Danmarks Statistik (1994c).

Der er i nogle tilfælde tale om ordninger, som principielt ikke udelukker hinanden. Fx kan modtagere af tjenestemandspension samtidig oppebære social pension, og personer over 67 år kan være i beskæftigelse og ved sygdom modtage dagpenge, samtidig med at de modtager reduceret folkepension. Modsat er der også nogle ordninger, der principielt udelukker hinanden.

Der er tale om en udbredt vandring mellem de enkelte ordninger, og inden for den enkelte ordening en vandring mellem forskellige specifikke ydelser. Visse vandringer opstår som en logisk følge af lovgivningen, andre som følge af ændringer i ydelsesmodtagerens sociale situation, og andre igen som en følge af administrative forhold.

Registermodel

De fleste af de nævnte indkomsterstattendes ordninger findes som nævnt allerede beskrevet i etablerede, personorienterede statistikregistre, indsamlet med henblik på en belysning af de specifikke lovområder og ydelsesordninger. Det er imidlertid ikke muligt at addere de enkelte statistikker til et samlet billede. Løsningen på dette problem er integration.

Det integrerede register indeholder ensartede og sammenlignelige statistiske oplysninger om alle ordningerne. Den centrale enhed er personen, med personnummeret som nøgle.

Ud over den statistiske belysning af det enkelte lovområde, som man allerede har, kan det integrerede register belyse den totale population (hvor mange modtager en eller flere indkomsterstattendes ydelser?), forskellige fællesmængder (hvor mange modtager samtidig to eller flere bestemte ydelser?) samt vandringen mellem de enkelte områder, brutto såvel som netto. Det integrerede register kan desuden måle resourcemængden (varighed og udbetaling).

En grundlæggende sondring i den sammenhængende socialstatistik går mellem *midlertidige* og *varige* ordninger. De *midlertidige* ordninger omfatter kontanthjælp, revalidering, syge-/barselsdagpenge og arbejdsløshedsdagpenge, mens *varige* ordninger omfatter sociale pensioner, efterløn og tjenestemandspension.

Registerdannelsen

De eksisterende statistikregistre forbehandles og integreres i et sammenhængende registersystem. Der findes i Danmarks Statistik emnespecifikke statistikregistre, vedrørende de indkomsterstattendes ordninger, der er nævnt ovenfor, med undtagelse af aktiveringsordningerne og de private pensionsordninger. Hvad angår de private pensionsord-

ninger er det på grundlag af oplysninger i Indkomststatistikregistret muligt at afgrænse modtagerne, men analyser har vist, at det er forbundet med stor usikkerhed at bringe data på en form, så de kan indgå med samme detaljeringsgrad som de øvrige indkomsterstøttende ordninger.

Danmarks Statistik har hidtil heller ikke rådet over et statistikregister vedrørende aktiveringsordninger, men det er under etablering og vil være dækkende fra 1. januar 1994.

Opnåelse af en *fælles måleenhed for ydelseernes varighed* forudsætter en del omregninger, da de indgående registre på dette punkt i særlig grad er uensartede: For arbejdsløse er oplyst en *ledighedsgrad pr. uge*, for sygedagpenge kendes de faktiske *dage* (datoer), for kontanthjælp og sociale pensioner er der tale om *hele kalendermåneder*, og for efterløn og tjenestemandspension er informationerne umiddelbart indskrænket til *hele året*. Ved en række omformninger og opretninger, som delvis baseres på beslutningsregler ud fra sandsynlighedsprincipper, og hvor oplysninger fra andre statistikregistre anvendes, bringes alle varighedsdata på formen: *antal dage pr. måned*. Summen af antal dage for en person med en bestemt ydelse kan højst udgøre 30 dage pr. måned (360 dage/år). Summen af varighederne for en person kan derimod være betydeligt længere, fordi man kan modtage flere ydelser på én gang.

Oplysningerne vedrørende *ledige* hentes fra Arbejdsløshedsstatistikregistret, som omfatter alle personer i arbejdsstyrken, der i årets løb har været registreret ledig i én eller flere perioder. Der skelnes mellem forsikrede ledige, der modtager dagpenge, og ikke-forsikrede ledige, der under ledighedsperioden modtager kontanthjælp. Ved samkøring med Indkomststatistikregistret overføres oplysninger om udbetaling af *arbejdsløshedsdagpenge*, idet beløbsdata ikke indgår i Arbejdsløshedsstatistikregistret.

Oplysningerne om *syge- og barseldagpenge* bygger på Sygedagpengestatistikregistret, som omfatter alle, der som følge af indtægtstab ved uarbejdsdygtighed på grund af sygdom, graviditet, fødsel eller adaption har modtaget ydelser via dagpengesystemet. En del af dagpengeperioden finansieres af arbejdsgiveren, mens resten finansieres af staten. Statistikken omfatter alene den statsfinansierede del. Arbejdsgiverperiodens længde - den første del af fraværsperioden - har været fastlagt forskelligt for offentligt og privat ansatte, og periodens længde har varieret. Det betyder, at den persongruppe, der er omfattet af statistikken de enkelte år ændres, hvilket forringer tallenes sammenlignelighed over tid.

*Kontanthjælpsstatistikregistret*⁴⁸ omfatter personer, der har modtaget ydelser efter bistandslovens kapitel 9 og 10. Enheden er familien. Som selvstændige familier regnes ægtepar, ugifte i alderen 18 år og derover samt ikke hjemmeboende børn i alderen 15-17 år. Det familiebaserede ydelsessystem betyder, at sager vedrørende ægtepar rent administrativt er registreret under én af ægtefællerne (den såkaldte hovedperson). Der er således ikke oplysninger om, hvem i familien, de forskellige ydelser er beregnet for. I den sammenhængende socialstatistik er begge ægtefæller regnet som modtagere med samme varighed og ydelsesart, og de udbetalte ydelser er fordelt ligeligt mellem ægtefællerne.

Oplysninger om *pensioner* stammer fra Pensionsstatistikregistret, der rummer oplysninger om *sociale pensioner* (førtidspension og folkepension) samt oplysninger om modtagere af *efterløn* og *tjenestemandspensioner*. Registret rummer oplysninger om alle personer, der i løbet af året har modtaget én eller flere pensionsydelse. Grundoplysningerne om sociale pensioner stammer fra det fælleskommunale pensionsystem og er meget detaljerede. Derimod er oplysningerne om modtagere af efterløn og tjenestemandspension mere summariske, og data stammer fra andre kilder.

Som det fremgår beskriver den sammenhængende socialstatistik en meget kompleks virkelighed, hvilket gør det vanskeligt at etablere nogen mekanisk fejlsøgningsprocedure. En særlig kompliceret opgave består i at indfange alle betydningsfulde ændringer i lovgivning, personkreds og data inden for de enkelte indkomsterstättende ordninger, så statistikens begreber kan blive opdateret i overensstemmelse hermed.

Et andet eksempel på horisontal integration er folke- og boligtællingen fra 1981.

Den sidste spørgeskemabaserede folke- og boligtælling i Danmark blev gennemført med tællingsdato d. 9. november 1970.

De administrative, landsdækkende registre, der var opbygget i 1970'erne, udnyttedes i 1976 til gennemførelse af den første registerbaserede folketælling i Danmark. 1976-tællingen indeholdt ikke boligoplysninger, da Bygnings- og boligregistret først blev etableret efter denne tælling.

1976-tællingens oplysninger blev dannet ved sammenkobling af Befolkningsstatistikregistrets personoplysninger med oplysninger fra statistikregistre baseret på kildeskattesystemet og Erhvervsregistret.

Eksempel 20:
Folke- og
boligtælling
1981

⁴⁸ Delregister i Bistandslovsstatistikregistret.

Nøglerne, der gjorde sammenkoblingen mulig, var personnummeret og CIR-nummeret⁴⁹ som identifikation for erhvervsenhederne.

Folke- og boligtællingen 1981 var den første fulde folke- og boligtælling i verden, der udelukkende byggede på oplysninger fra administrative registre. Dens indhold svarede stort set til, hvad der indsamledes på skemaerne i 1970-tællingen.

Ved 1981-tællingen sammenkoblede op mod en halv snes grundlæggende statistikregistre og -moduler. De vigtigste registre i denne sammenhæng var:

- Befolkningsstatistikregistret
- Bygnings- og boligstatistikregistret
- Arbejdsklassifikationsmodulet
- Beskæftigelsesstatistikregistret
- Uddannelsesklassifikationsmodulet
- Områdeklassifikationsmodulet (bruges til afgrænsning af byer)
- Husstands- og familiemodulet

Sammenkoblingsnøglerne i 1981-tællingen var personnummeret, bopælsadressen og CIR-nummeret, suppleret med en eventuel arbejdsstedskode. Alle oplysningerne samledes i Folke- og boligtællingsregistret.

Efter publiceringen af 1981-tællingen var der på næsten alle de områder, der traditionelt dækkes af en folke- og boligtælling, etableret årlige strukturbelysende statistikker. Man fik så at sige en folke- og bolig-tælling hvert år, og detaljerede statistikker stilledes til rådighed for brugerne, især gennem on-line databanker. Behovet for folke- og bolig-tællinger hvert femte år var således til en vis grad faldet bort, og de traditionelle, omfattende tabelværker produceres da heller ikke længere. Af hensyn til internationale sammenligninger produceres dog særlige statistiksamlinger til EU og FN til de officielle folke- og bolig-tællingsterminer hvert tiende år.

5.3 Vertikal integration

Definition

Ved *vertikal integration* forstås som nævnt den proces, hvorved et register i to eller flere tidsmæssigt forskudte versioner sammenholdes på personniveau for at konstatere tilgang og afgang af *enheder* eller ændringer i enhedernes *egenskaber*. På denne måde kan man belyse for-

⁴⁹ Identifikationen for den juridiske erhvervsenhed (CIR står for Det centrale indeholdelsesregister). Nummeret har i 1982 fået navneforandring til SE-nummer, jf. bilag 4, og omtales som sådan andre steder i bogen.

løb eller årsagssammenhænge over tid. Det er næsten altid således, at der anvendes flere registre, hvorfor den vertikale og den horisontale integration er snævert knyttet sammen.

Ved starten af registerstatistikken i begyndelsen af 1970'erne var det naturligvis afgørende at få dannet det enkelte årsregister med henblik på statistikproduktionen for det pågældende år, og der blev ikke gennemført mange store forløbsundersøgelser⁵⁰. De var af eksperimentel karakter, og handlede i høj grad om at udforske mulighederne og belyse de problemer, som er forbundet med sådanne projekter. Udviklingen er som bekendt gået meget stærkt siden, og fra slutningen af 1970'erne begyndte man i større omfang at tage den vertikale integration i brug i en række projekter. Banebrydende var en analyse af sammenhængen mellem dødelighed og erhverv i perioden 1970-75⁵¹. I dag er det muligt at opbygge meget omfattende longitudinelle registre, der uden alt for store vanskeligheder kan udnyttes effektivt.

Udviklingen

De tekniske muligheder har støttet interessen især blandt forskere for at udvikle såkaldte multivariate modeller, som tillader inddragelsen af mange variable (horisontalt såvel som vertikalt) i analyserne. Alle blot nogenlunde omfattende statistikpakker, der i dag tilbydes (SAS, SPSS, BMDP mv.), har disse muligheder indbygget. I den kontakt, Danmarks Statistik har med forskere, er der næsten altid et ønske om datasammenstillinger til anvendelse i longitudinelle, multivariate analyser.

Nedenfor belyses de problemer, der opstår ved vertikal integration.

Problemerne

Årsregistre, dvs. de enkelte statistikregistre, lever så at sige deres eget liv. Det er en betydningsfuld opgave at gøre det enkelte årsregister så perfekt som muligt, og gerne bedre end sidste år. Et eksempel herpå er udviklingen af Arbejdsklassifikationsmodulet (AKM), der fra sin spæde start i 1976 har været under stadig udbygning. Den uoplyste restgruppe "lønmottagere uden nærmere angivelse", som altså ikke kan beskrives nærmere på grund af manglende oplysninger, er bevidst søgt reduceret gennem indhentning af data fra flere kilder, hvilket har ført til en kvalitetsforbedring. Det giver imidlertid anledning til et kontinuitetsproblem, når man ønsker at belyse den tidsmæssige udvikling fx i størrelsen af en specifik gruppe lønmottagere. AKM er naturligvis ikke det eneste eksempel på sådanne kontinuitetsproblemer, og for nogle af de øvrige eksemplers vedkommende er der sågar tale om forringelser i et registers kvalitet fra år til år.

Ændringer i datakvaliteten ødelægger kontinuiteten

⁵⁰ Der gennemførtes dog på stikprøvebasis en del undersøgelser i det såkaldte Minipopulationsregister, hvor bl.a. beskæftigelsesundersøgelserne, baseret på spørgeskema, blev samkørt med henblik på analyser af bruttoforskydninger på arbejdsmarkedet.

⁵¹ Jf. *Lyng (1979)*.

Egentlige databrud

Det for al statistik klassiske problem med egentlige databrud giver naturligvis særlig store problemer i den vertikale registerintegration. Klassifikationer, opgørelsesmetoder og det administrative datagrundlag ændres over tid. Et eksempel, som snart bliver aktuelt, er indførelsen af nye stillings- og brancheklassifikationer. Danmarks Statistik må naturligvis følge med den internationale udvikling af fælles klassifikationer, hvilket betød, at man i 1993 indførte en ny stillingsklassifikation (DISCO), der er i overensstemmelse med FN's ISCO (1988 version)⁵² samt en klassifikation af brancher (Dansk Branchekode 1993)⁵³, der bygger på EU's klassifikation. Netop stillings- og branchekoder er vigtige led i nogle af de fundamentale definitioner, der redegøres for i eksempel 22 nedenfor. Sådanne ændringer medfører en stor arbejdsindsats for at bevare de integrerede registres værdi.

De longitudinelle registres datamængder har tilbøjelighed til at blive meget omfattende. Det er ofte vanskeligt (specielt for forskere) at forestille sig, at der kan undværes noget fra grundregistre i den longitudinelle version. Synspunktet er logisk nok, idet man i mange situationer først ved om en variabel er væsentlig, når den så at sige er afprøvet i praksis. Danmarks Statistik har imidlertid gjort sig den erfaring, at variabelvalget *skal* være (stærkt) begrænset af hensyn til overskueligheden og *bør* dikteres af den teori, der er fremherskende på det statistik- eller forskningsområde, der er tale om. Man kunne i princippet tænke sig, at alle Danmarks Statistiks registre med personidentifikation blev sat sammen til ét stort register over tid. Rent bortset fra, at dette ville være helt uacceptabelt af hensyn til databeskyttelsen, ville det også i praksis vise sig helt umuligt at arbejde med.

Endelig skal kravet til dokumentation af registret præciseres. Dokumentation er altid vigtig, men når data indeholder forløb over tid, stilles der særligt store krav. Ikke blot skal det enkelte års dokumentation samles sammen fra mange, ofte meget forskelligt opbyggede dokumentationssamlinger, men databrud og kodeændringer skal nøje beskrives. Ved dannelsen af to longitudinale forskningsdatabaser, IDA og Fertilitetsdatabasen, har dokumentationssystemet TIMES (jf. kap. 9) været en betydelig hjælp. I det efterfølgende belyses problemstillingerne nærmere ved tre eksempler; Fertilitetsdatabasen, IDA og Det integrerede elevregister.

⁵² Jf. ILO (1990).

⁵³ Jf. Danmarks Statistik (1992c).

Eksempel 21: Fertilitetsdatabasen

Fertilitetsdatabasen er etableret i Danmarks Statistik i årene 1990-1992 med støtte fra Statens Samfundsvidenskabelige Forskningsråd. Den har til formål at danne grundlag for analyser af fertiliteten i forskellige befolkningsgrupper (socioøkonomiske grupper mv.). Til dette formål er forløbsbeskrivelser for de enkelte personer helt nødvendige, idet man skal vide, hvornår de eventuelt har fået børn, og hvordan deres uddannelseshistorie, erhvervshistorie mv. er forløbet.

Ved etableringen af databasen er dannelsen af populationen og specielt fastlæggelsen af relationen mellem forældre og børn foregået på en for Fertilitetsdatabasen specifik måde. Der blev i denne forbindelse foretaget en bearbejdning af de henvisninger til forældre, der eksisterede i Befolkningsstatistikregisteret.

Der blev tidligt i planlægningen besluttet et princip om at foretage en grundig kontrol og de nødvendige korrektioner i forbindelse med etableringen af populationerne, herunder forældrerelationen, mens oplysningerne i de "ydre" registre, dvs. registrene vedrørende fx uddannelse og arbejdsmarkedstilknytning, blev taget for givet. Disse registre er således anvendt uden kontrol, idet man bl.a. har afstået fra et konsistenscheck mellem forskellige årgange af samme register eller mellem forskellige registre, dækkende det samme kalenderår.

Problemer relateret til vertikal integration er derfor (indtil videre) kun identificeret i forbindelse med etableringen af populationen i Fertilitetsdatabasen.

Fertilitetsdatabasens opbygning og indhold

Fertilitetsdatabasen indeholder demografiske og sociale oplysninger om samtlige kvinder og mænd i den fertile alder (defineret som 12-49 år for kvinder og 12-64 år for mænd), der har haft bopæl i Danmark i løbet af 1980'erne. Desuden er der oplysninger om, hvor mange børn disse kvinder og mænd har fået, samt hvornår. Antalsmæssigt omfatter de årlige populationer ca. 1,3 mio. kvinder og 1,8 mio. mænd. Det antal levendefødte børn, de enkelte voksne har fået på et givet tidspunkt og dermed ved en given alder, er blevet optalt ud fra henvisningerne fra børn til forældre i Befolkningsstatistikregistret.

De demografiske oplysninger om de voksne er indhentet fra årligt færdigdannede registre og omfatter bl.a. køn, alder, ægteskabelig stilling, bopælskommune og statsborgerskab. *De sociale oplysninger*, som også er indhentet årligt, karakteriserer kvinders og mænds uddannelse, erhvervsarbejde og socioøkonomiske placering, indkomst, familie- og boligforhold samt omfanget af eventuel modtagelse af sociale ydelser. Mens de demografiske oplysninger om kvinder og mænd findes for

hver 1. januar i årene 1980-89, findes de sociale oplysninger for perioden 1981-88⁵⁴.

Der er således tale om en omfattende samling af data, både med hensyn til antal omfattede personer og datamængden om hver person. Efter gennemførelsen af den første analyse blev der i 1993 påbegyndt en opdatering til og med 1992. Herefter er det hensigten, at Fertilitetsdatabasen skal opdateres årligt.

Etableringen af en database som Fertilitetsdatabasen indebærer store muligheder for så vidt angår udnyttelsen, men giver også en del problemer undervejs i etableringen.

Populationen

Populationen i Fertilitetsdatabasen består af en række årlige bestandsopgørelser (pr. 1. januar) af kvinder og mænd i den fertile alder i Danmark, uanset om de har børn eller ej. Den tidligere publicerede statistik er således baseret på en række tværsnitopgørelser og -beregninger⁵⁵.

For hver årlig population er der påført en række oplysninger, fx om vedkommende person genfindes næste år, udvandrer eller dør. Indvandrere er indført i populationen pr. 1. januar i det år de indvandrer, da de har mulighed for at føde et barn i landet det pågældende år.

Karakteristika ved kvinder og mænd (de voksne) i Fertilitetsdatabasen kan således beskrives hvert kalenderår i 1980'erne. For børnenes vedkommende er derimod kun medtaget såkaldte stamoplysninger vedrørende karakteristika ved selve barnet og om forældrene på barnets fødselstidspunkt.

Basis for fastlæggelse af forældrerelationen har været de forældrehenvisninger, der findes om børnene i Befolkningsstatistikregistret. Børn, som kun har haft henvisning til en mor og/eller far, der ikke indgår i en af de årlige populationer, indgår ikke i Fertilitetsdatabasen.

Opgørelser ud fra Befolkningsstatistikregistret pr. 1. januar 1989 viste, at andelen af de enkelte fødselsårgange, der havde henvisning til mor eller far, steg kraftigt for årgangene fra 1950'erne. Fra omkring årgang 1960-62 har mere end 90 pct. en henvisning til mor og/eller far, og blandt børn født i perioden 1973-89 har 99,4 pct. en henvisning.

Blandt personer, der er født uden for Danmark fra 1960 og frem, mangler omkring 75 pct. en henvisning til en mor, mens næsten alle

⁵⁴ Jf. *Danmarks Statistik (1993b)*.

⁵⁵ Jf. *Knudsen (1993a)*.

dem, der er født i Danmark (fra 1960 og frem), har en henvisning. De personer, der især mangler forældrehenvisning, er altså i stor udstrækning personer, der er indvandret som voksne.

Vertikal integration i Fertilitetsdatabasen

Ved etableringen af Fertilitetsdatabasen opstod der to typer af problemer. Den samme persons personnummer kunne være forandret over tid, og der var uforudset uoverensstemmelse mellem dokumentation og dataindhold i de ældste årgange af Befolkningsstatistikregisteret.

Under etableringen af populationen var det et problem, at der fra år til år kunne være foretaget rettelse af personnummeret for nogle få personer. Det er dog kun ændringer af børnenes personnumre, der er konstateret og bearbejdet. Når det har været så vigtigt at kontrollere børnenes personnumre, skyldes det, at antallet af børn, hver voksen har fået, den såkaldte *paritet*, er en væsentlig variabel til analyse af både fertilitetsmønstre og udvikling.

Fremgangsmåden var følgende: For hver kvinde i den samlede population, dvs. enhver kvinde, der blot forekom på ét tidspunkt i perioden, blev der dannet en oplysning for hver gang, hun var registreret som mor til et barn. For hver kvinde blev der dernæst kontrolleret for inkonsistens i børnenes fødselsdatoer. Hvert tilfælde, hvor en kvinde havde to enkeltfødte børn, der var født med mindre end 5 måneders interval, blev kontrolleret.

En stor del af de korrektioner, der herefter blev foretaget, skyldtes at der var to oplysninger om det samme barn, hvoraf den ene var tidligere og mangelfuld i forhold til den seneste. Andre fejl skyldtes korrektioner af indvandrede børns fødselsdatoer, som kunne finde sted op til flere år efter indvandringen og omfatte samtlige en kvindes børn. En tredje type var de rettelser, der fandt sted blandt adoptivbørn, der ikke ønskede at beholde et personnummer, der kunne vise, at der var tale om en person født i udlandet⁵⁶.

Da man principielt ikke kan indlægge en kontrol for afstanden mellem en mands børn, er rettelserne udelukkende gennemført i forbindelse med kvinderne, hvorefter de rettede oplysninger om barnet er overført til mandepopulationerne.

De ældste registerversioner, der blev anvendt til at etablere populationerne, var Det medicinske fødsels- og dødsfaldsregister fra 1973 og

⁵⁶ I de første år (1968-80) blev der anvendt en særlig serie personnumre for personer, der skulle have tildelt et nummer på et senere tidspunkt end ved fødslen, dvs. alle førstegangs-indvandrede. Dette blev ændret, da der opstod offentlig kritik af, at en tidligere indvandring på denne måde blev signaleret gennem nummeret.

Befolkningsstatistikregistret fra 1979. Det førstnævnte register var veldokumenteret, mens der var problemer i dokumentationen for Befolkningsstatistikregistrets første årgange. Desuden viste det sig, at CPR-systemet, som dette register er baseret på, indeholdt mange fejl i de første år - fejl, som først blev åbenlyse under etableringen af populationen til fertilitetsundersøgelsen.

En af de fejl der forekom i et vist antal tilfælde var ombytning af moderens og faderens personnummer (henvisning) i recorden for barnet⁵⁷. Det var disse henvisninger, der skulle bruges til optælling af pariteten. Selv om det er en simpel fejl, skal man - for at finde den - på forhånd være opmærksom på netop denne mulighed. For Fertilitetsdatabasen betød opdagelsen af dette, at en række påtænkte maskinelle rutiner ikke kunne gennemføres som først planlagt.

Konsekvenser af de foretagne korrektioner

En af de store, umiddelbare fordele ved etableringen af en database som Fertilitetsdatabasen er, at den beskrevne opretning af populationerne sikrer en bedre kvalitet i materialet og betyder, at fremtidige brugere ikke behøver bekymre sig om at rette op på inkonsistens.

Konsistensen i oplysningerne fra de ydre registre er der som nævnt ikke gjort noget ud af. Da Fertilitetsdatabasen endnu er relativt nyetableret, har den kun været brugt til forholdsvis få analyser, hvor eventuelt manglende konsistens over tid i de forskellige sociale oplysninger ikke har spillet nogen væsentlig rolle og ikke har kunnet identificeres i de pågældende udtræk.

I det næste eksempel beskrives forskningsdatabasen IDA, der ligesom Fertilitetsdatabasen bl.a. er udviklet for at tilfredsstille nogle behov blandt samfundsvidenskabelige forskere.

Eksempel 22: IDA - databasen

I Danmarks Statistik er der i perioden 1988-90 blevet opbygget en database på grundlag af registeroplysninger til brug for arbejdsmarkedsforskningen. Denne samling af data kaldes for Integreret Database for Arbejdsmarkedsforskning eller IDA. Personer, ansættelser og virksomheder kan her følges i Danmark fra 1980 og frem. Formålet med oprettelsen af databasen er at kunne foretage analyser af personers mobilitet mellem forskellige virksomheder og virksomheders livsforløb⁵⁸.

Enheder og variable i databasen

Oplysningerne i databasen omfatter samtlige personer i befolkningen og samtlige ansættelser samt alle virksomheder med lønnet beskæfti-

⁵⁷ Som "børn" i denne forbindelse opfattes alle personer født efter 1941.

⁵⁸ Jf. *Danmarks Statistik (1991d)*.

gelse. Databasen omfatter således pr. år ca. 5 mio. personer, ca. 2,5 mio. ansættelser og ca. 200.000 virksomheder (arbejdssteder). Oplysningerne vedrører fortrinsvis situationen i slutningen af november måned i de enkelte år. Enhederne kan karakteriseres ved hjælp af godt 300 variable.

Kildematerialet til IDA-databasen består af oplysninger fra en række statistikregistre. Nedenfor gives nogle eksempler på typer af variable, der findes i databasen:

<i>Personer</i>	køn, alder, ægteskabelig stilling, uddannelse, beskæftigelse, arbejdsløshed, indkomst
<i>Ansættelser</i>	stilling, timeløn, anciennitet, af- og tilgang
<i>Arbejdssteder</i>	branche, antal ansatte og lønniveau, identitet over tid

For en oversigt over nogle af de konkrete anvendelser af IDA-databasen henvises til fx Arbejdsministeriet (1993).

Konstruktion af en virksomhedsdatabase: Virksomheders identitet over tid

Et grundlæggende problem, det har været nødvendigt at løse i forbindelse med opbygningen af databasen, er at afgøre, hvornår en virksomhed er den samme over tid (fra et år til det næste). På personsiden er det teknisk set uden problemer at følge personerne over tid ved at anvende personnummeret som nøgle⁵⁹. For virksomhederne findes på tilsvarende måde et nummersystem af hensyn til bl.a. skatte- og momsbetaling, men virksomhederne kan skifte numre, uden at der er tale om reelle ændringer. Hvis man baserer en opgørelse over af- og tilgang af virksomheder imellem to år på disse numre, vil man overvurdere antallet af virksomhedsnedlæggelser og -oprettelser.

At en virksomhed stadig er den "samme" fra det ene år til det næste indebærer, at visse kendetegn, som i den konkrete sammenhæng anses for centrale, stort set er uforandrede fra det ene år til det næste. En række forskellige kendetegn eller egenskaber kunne tænkes at indgå ved fastlæggelsen af, om en virksomhed stadig eksisterer i det følgende år. Problemet er derfor at vælge, hvad der er afgørende for, at der er tale om samme virksomhed.

Problemet kan belyses ved at tage udgangspunkt i, hvornår der foreligger en ny virksomhed. Man kan vælge at lægge vægt på, om en stor del af de ansatte er nye. En anden mulighed kunne være at tage udgangspunkt i ejeren. En tredje mulighed kunne være, at man var inter-

⁵⁹ Bortset fra de få tilfælde, hvor personnummeret bliver korrigeret, jf. forrige eksempel.

esseret i virksomhedernes beliggenhed, og herunder om der inden for et givet område var kommet nye virksomheder, der evt. kunne være tilflyttet fra et andet område. Fastlæggelsen af regler for, hvornår man vil sige, at der foreligger en ny virksomhed, afhænger således i høj grad af den sammenhæng, som denne oplysning skal anvendes i⁶⁰.

Problemet med at fastlægge, hvad der definerer en (identisk) virksomhed, har også betydning uden for den statistiske verden. Fx skal virksomheder ved afskedigelser af et større antal ansatte give meddelelse til de lokale arbejdsmarkedsnævn. Problemet har her bl.a. været, hvad der skal forstås ved "større" afskedigelser i de tilfælde, hvor indskrænkningen skal ske på et lokalt arbejdssted, der indgår i et større firma⁶¹. Tilsvarende kan der opstå fortolkningsproblemer i forbindelse med at afgøre, om overenskomster er gældende ved salg af virksomheder, hvor fagforeningerne ønsker, at overenskomster ikke skal kunne ophøre på grund af fx proforma salg.

I forbindelse med konstruktionen af IDA-databasen har hensigten som nævnt bl.a. været at kunne belyse arbejdspladsmobiliteten og virksomhedernes livsforløb, og fastlæggelsen af kriterierne for, om en virksomhed er den samme, afspejler dette. Begge formål har medført, at en virksomhed bør kunne skifte ejer og flytte, uden at der dermed er tale om en ny virksomhed.

Ved definitionen af identiske arbejdssteder i IDA er hovedvægten blevet lagt på graden af gengangere blandt de ansatte på det enkelte arbejdssted fra det ene år til det næste⁶².

På det operationelle plan indgår følgende oplysninger fra to på hinanden følgende år ved fastlæggelsen af virksomhedernes identitet over tid:

- ejerskab
- branche
- beliggenhed
- arbejdsstyrke

Det endelige regelsæt vedrørende virksomhedernes identitet over tid er udformet som en trinvis procedure⁶³.

⁶⁰ Jf. Baldwin, Dupuy & Penner (1992).

⁶¹ Jf. Arbejdsministeriet (1977).

⁶² I Canada har man anvendt samme grundlæggende ide, jf. Baldwin, Dupuy & Penner (1992).

⁶³ Jf. Danmarks Statistik (1991d).

I forbindelse med den operationelle fastlæggelse af identiteten over tid opstod muligheden for at klassificere ændringer på forskellig måde. Ved at anvende personsammenfald blandt de ansatte i to på hinanden følgende år viste det sig, at en del nye virksomheder var kendetegnet ved, at en større andel af de ansatte kom fra samme arbejdssted året før. Denne type oprettelse er blevet klassificeret som *oprettet ved udskilning* fra et arbejdssted, hvis en række betingelser er opfyldt. Tilsvarende gør sig gældende for nedlagte virksomheder, der kan være *nedlagt via opsugning* i et andet arbejdssted. Disse to kategorier gør det muligt at foretage analyser, hvor der kan tages hensyn til, om ændringer af denne type skal medtages eller ej. Dette er bl.a. af interesse i forbindelse med opgørelser af jobskabelse, som vil blive omtalt nedenfor.

Analysen over tid

Opgørelserne i IDA-databasen er baseret på oplysninger om situationen i slutningen af november måned i de enkelte år for såvel personer⁶⁴ som virksomheder⁶⁵. Ændringer kan derfor opgøres ved at sammenligne oplysningerne for to tidspunkter. Der er dermed ikke tale om en kontinuert registrering, men om sammenligning af to statusopgørelser.

At personer og virksomheder følges over tid giver mulighed for at foretage både prospektive og retrospektive analyser. Dvs. at der kan gives svar på spørgsmål eller problemstillinger, der typisk indebærer, at man ser "fremad": Hvor stor en andel af ufaglærte har været "opadgående socialt mobile"?⁶⁶ Tilsvarende kan spørgsmål, der fx vedrører tilgangen af en bestemt gruppe personer - hvor man ser "tilbage" - besvares: Hvilken erhvervskarriere har de personer bag sig, der deltager i voksenuddannelseskurser i 1989? Endvidere kan der foretages opgørelser over en række år op til de 10 år, som databasen på nuværende tidspunkt dækker.

"Job-creation and -destruction"

Et område der har påkaldt sig stor interesse er opgørelser af det, man med begreber på engelsk kalder *job-creation* og *job-destruction*⁶⁷. Når man sammenligner beskæftigelsesniveauet i et land fra et år til det næste, er der normalt kun tale om relativt beskedne ændringer. Som supplement hertil er man begyndt at se på omfanget af nye jobs i nyetablerede virksomheder og i nettovæksten af jobs på virksomhedsni-

⁶⁴ Jf. *Det økonomiske Råd (1992)* og *Leth-Sørensen (1993)*.

⁶⁵ Jf. *Vejrup-Hansen (1993)*.

⁶⁶ Jf. *Rohwer & Leth-Sørensen (1993)*.

⁶⁷ Jf. *Leonard (1987)*; *OECD (1987,1993)*; *Finansministeriet (1992)*; *Ministeriet for Erhvervspolitisk Samordning (1993)*.

veau i allerede eksisterende virksomheder. Tilsvarende for virksomheder der forsvinder eller indskrænker. Det viser sig typisk, at de beskudne ændringer på makro-planet dækker over en meget betydelig jobomsætning på virksomhedsniveau opgjort på denne måde. Det skal bemærkes, at opgørelsen af jobomsætningen på denne måde er en minimumsbetragtning over den årlige jobomsætning. Da IDA-databasen netop indeholder årlige oplysninger om af- og tilgang af virksomheder og om antallet af beskæftigede, har det været nærliggende at anvende data herfra til opgørelse af jobskabelse og -nedlæggelse.

Arbejdsstederne kan opdeles i tre overordnede kategorier eller mængder:

- tilgang af arbejdssteder
- fortsat eksisterende arbejdssteder
- afgang af arbejdssteder

Et særligt problem knytter sig til, at arbejdsstederne kan afgang via *opsugning* eller tilgå via *udskilning* fra andre arbejdssteder. For det bevarede arbejdssted, der indgår i "opsugning" eller "udskilning", vil der på tilsvarende måde kunne ske relativt store ændringer i beskæftigelsen. Samtidig findes der i IDA-databasen oplysninger om, hvorvidt en opsugning/udskilning sker inden for samme firma. Hvis dette er tilfældet, har man i denne sammenhæng valgt ikke at betragte ændringerne som nye eller forsvundne jobs, men kun *nettovirkningen* af opsugningen/udskilningen for de berørte arbejdssteder medregnes.

Et tredje eksempel på vertikal integration, af en lidt anden karakter, er Det integrerede elevregister.

Eksempel 23: Det integrerede elevregister

Det integrerede elevregister⁶⁸ er et kumuleret register, hvor alle *uddannelsesforhold* indsamlet fra 1973/74 og fremefter er organiseret på en sådan måde, at man kan følge den enkelte persons *uddannelseskarriere*, som den udvikler sig over tid.

Et uddannelsesforhold er en *persons* deltagelse i en *uddannelse* på en *institution*. En uddannelse betragtes i første række som en aktivitet, der er tidsmæssigt afgrænset med et *påbegyndelsestidspunkt* og et *afgangstidspunkt*.

Uddannelsesaktiviteten resulterer, hvis den gennemføres med succes, i en *kompetence*, der med andre ord er aktivitetens resultat.

Den uddannelsesmodel der bruges, indeholder en kode for uddannelsen, hvortil der kan knyttes en eller flere koder for kompetencer, idet

⁶⁸ Uddannelsesstatistikregistret.

der kan være mulighed for at standse uddannelsen på forskellige uddannelsesniveauer (fx cand. phil. eller cand. mag.), eller der kan være tale om kvalitativt forskellige kompetencer (retninger, fx kemiingeniør, bygningsingeniør) på samme uddannelsesniveau.

De registrerede uddannelsesforhold består således af følgende data (identifikationer i kursiv):

- personnummer
- uddannelse
- påbegyndelsestidspunkt
- afgangstidspunkt
- afsluttende uddannelse (kompetence)
- institution (indberettende)

Datafangst

Institutionerne indberetter årligt til Danmarks Statistik, og alle elever/studerende, der har været på den enkelte institution i tællingsperioden, dvs. det forløbne år forud for tællingstidspunktet den 1. oktober, skal indberettes.

Institutionerne er fordelt på fire områder:

- skoler (folke-/gymnasieskoler)
- erhvervsskoler mv.
- videregående uddannelsesinstitutioner
- frie skoler (ungdoms-, landbrugs-, søfartsskoler mv.)

Der findes en række indberetningsmedier og kanaler; skema, diskette, magnetbånd, m.m. Der kan indberettes direkte fra den enkelte institution eller via et for flere institutioner fælles elevadministrationssystem, fx systemet ESAS, som dækker alle erhvervsskoler (handels- og tekniske skoler).

Dataintegration

Når årets indberetninger er i hus, skal data indlægges i Det *integrerede elevregister*. Dette register har for tiden registreret ca. 7 mio. uddannelsesforhold, hvoraf de fleste ligger som afsluttede uddannelsesforhold, der udgør den uforanderlige historiske del af registret. En mindre del af registreringerne udgøres af *de igangværende uddannelsesforhold*. For disse gælder, at man skal have at vide, om de stadig er i gang, eller om de er blevet afsluttet og i givet fald hvordan.

For at afklare dette samkøres indberetningerne med den aktive del af Det integrerede elevregister, og de enkelte uddannelsesforhold ajourføres. En del af registreringerne i de enkelte årsregistre vil ikke matche, nemlig de der vedrører nytilgangen af elever/studerende. Disse indlægges derfor efterfølgende separat i Det integrerede elevregister.

Billedligt kan man sammenligne Det integrerede elevregister med et træ, hvor der hvert år tillægges en ny årring udad til i form af tilgang og afsættes én indad til i form af afsluttede uddannelsesforhold.

Tidligere var indlæggelsen en meget tidskrævende proces, idet man behandlede alle tilfælde af manglende match manuelt. I dag er indlæggelsen en ren maskinel proces med mulighed for manuel overvågning af de maskinelle processer.

Der dannes en foreløbig version af registret på basis af indberetninger, der er "aflusede" men ikke fejlsøgt til bunds. Den *foreløbige* version er klar ca. 8 måneder og den *endelige* ca. 12 måneder efter tællingstidspunktet.

Uddannelseskarriver

I Det integrerede elevregister er personernes uddannelsesforhold lagt i kronologisk orden, dvs. de kan læses som personlige *uddannelseskarriver, der udfolder sig i tid*. Da registret dækker (næsten) alle uddannelser, giver dette register enestående muligheder for at give en tilstandsbeskrivelse i uddannelsessystemet på et hvilket som helst tidspunkt samt at belyse de samlede bevægelser i en given periode.

Det er naturligvis ønskeligt, at registret dækker alle offentligt anerkendte uddannelser. For øjeblikket er dækningen god bortset fra voksenuddannelserne. En udbygning er i gang på dette område.

En person kan i princippet kun registreres som værende i gang med én uddannelse ad gangen. Dette hænger sammen med, at registret (på nær visse undtagelser) kun indeholder fuldtidsuddannelser. En person kan derfor pr. definition ikke være i gang med to samtidige uddannelser. I praksis viser det sig at være svært umiddelbart at leve op til dette krav af to grunde:

- Der findes konkurrerende indberetninger, hvor forskellige institutioner gør krav på at have den samme person som studerende.
- Visse uddannelser er bygget op af elementer, som indberettes hver for sig og som kan forløbe parallelt (fx to overbygningsuddannelser på Roskilde Universitetscenter).

I det første tilfælde er der overvejende tale om, at indberetternes registre ikke er ajourført frem til indberetningstidspunktet, dvs. en afmelding ikke er kommet på plads. Der er her tale om det, man kunne kalde "ulovligt" tidsmæssigt overlap mellem to uddannelsesforhold. Overlapsproblemer af denne type fjernes gennem tidsmæssige justeringer/sletninger af records efter visse regler.

I det andet tilfælde er der derimod tale om et helt igennem lovligt tidsmæssigt overlap, som man ikke uden videre kan fjerne. Der er med andre ord ikke tale om en fejl, men det har dog den ubehagelige konsekvens, at princippet om entydighed i uddannelseskarrererne brydes. Det vanskeliggør fx opgørelser over antal studerende på et givet tidspunkt, idet der er et dobbelttællingsproblem.

Et andet forhold, som vanskeliggør statistikken, er institutionsskift. Da indberetningerne kommer fra institutionerne, vil institutionsskift betyde, at den første institution må indberette en afbrudt uddannelse og den anden en påbegyndt uddannelse. Dette er naturligvis korrekt, men gør det vanskeligt at opgøre, hvor mange der i en given periode *netto* har påbegyndt, hhv. afbrudt en uddannelse.

Det komprimerede elevregister

I praksis har det vist sig, at det er ganske vanskeligt at arbejde med Det integrerede register af de ovennævnte grunde. For at kunne leve op til kravet om entydighed og samtidig afhjælpe ulemperne, lader man registret gennemgå en bearbejdning, hvor man samler uddannelsesdele, der hører sammen, i en record. Det betyder at man:

- samler forskellige institutioners indberetning vedrørende den samme uddannelse i én record og lader den seneste institution bære indberetningen
- samler deluddannelser (især på de videregående uddannelsers område) til én record

For universitetsuddannelsernes vedkommende er dette en vanskelig sag, da man her opererer med basisuddannelser, grunduddannelser, hovedfag, bifag, sidefag, suppleringsfag m.m. Disse skal stykkes sammen til hele uddannelser, hvorved man får noget der svarer til den almindelige opfattelse af en universitetsuddannelse, nemlig at den er en helhed med et start- og et sluttidspunkt. De uddannelsesenheder, der hermed fremkommer, er nemmere at forstå for brugerne end de teknisk betonedede deluddannelser.

Man er med Det komprimerede elevregister i den gunstige situation, at alle opgørelser nu er enkle, således at

- man kan lave bestandsopgørelser uden at skulle korrigere for dobbelttælling
- man kan lave tilgangs-/afgangsopgørelser, som uden videre giver mening
- overgange mellem uddannelser er reelle overgange og ikke interne overgange mellem uddannelsesdele

Anvendelser af registret

Registret anvendes til opgørelser over elevbestand på tællingstidspunktet og bevægelserne i tællingsperioden: tilgang og afgang fordelt på afbrudt/fuldført. Strømmene mellem hovedgrupper af uddannelser beskrives.

På grundlag af de fremadrettede aldersbetingede overgange inden for en given periode og en aldersfordelt befolkningstilgang i bunden af modellen foretages fremskrivninger af uddannelsespopulationerne.

Det er endvidere muligt at følge fx en række afgangsårgange fra gymnasiet/hf og sammenligne deres uddannelsesstatus på forskellige måletidspunkter.

Registret bruges endelig til at danne Uddannelsesklassifikationsmodul, UKM, som er en opgørelse over befolkningens gennemførte eller igangværende uddannelse på seneste tællingstidspunkt. UKM benyttes som baggrundsoplysninger i mangfoldige statistikker.

Kapitel 6. Personstatistikens samspil med andre statistikområder

Inddeling af statistik i områder er vanskeligere end det lyder, og der gives næppe noget entydigt princip for adskillelse af de forskellige statistikområder fra hverandre. Et blik på statistiske centralbureauers organisationsplaner vil illustrere dette. I registerstatistisk sammenhæng er det imidlertid naturligt at definere statistikområderne med udgangspunkt i den type af registrerede enheder eller "tællingsenheder", som de pågældende opgørelser bygger på.

Statistikområder kan afgrænses på mange måder

Følgende grupper af enheder betragtes almindeligvis som dem, der afgrænser statistikområderne:

- personer, familier, husstande (personstatistikken)
- virksomhedsenheder af forskellige typer (erhvervsstatistikken)
- ejendomme, bygninger, boliger (areal-, bygnings- og boligstatistik)

Man kunne tilføje andre enhedstyper - fx motorkøretøjer eller skibe, der i øvrigt også kan opfattes som en slags erhvervsenheder.

Når statistikområderne adskilles på denne måde, er *relationerne* mellem disse enhedstyper bestemmende for forbindelsen mellem statistikområderne.

I det følgende behandles forholdet mellem personstatistikken og andre statistikker ud fra denne tankegang. Afsnit 6.1 indeholder en kort gennemgang af "the state of the art" med hensyn til opbygning og anvendelse af centrale enhedsrelationer. Afsnit 6.2 opridser betydningen af anvendelsen af fælles data på forskellige statistikområder, og afsnit 6.3 omtaler de anvendte enheders validitet og betydningen heraf. Endelig indeholder afsnit 6.4 et enkelt perspektiv angående de fremtidige muligheder i registerstatistikken.

6.1 Status for enhedsrelationer

Under den registerstatistiske udvikling i Danmark er to fundamentale enhedsrelationer blevet opdyrket, yderligere én har været i konstant venteposition og en fjerde kunne anvendes mere systematisk end tilfældet er for øjeblikket.

6.1.1 Relationen person/bolig

Person/bolig er hovedelement i folke- og boligtællinger

Relationen mellem person og bolig er det mest afgørende element i den teknik, der tillader, at der foretages folke- og boligtællingsopgørelser på registerbasis her i landet. Muligheden blev skabt i forbindelse med BBR's oprettelse, idet der dengang blev etableret fælles adresseidentifikationer for de enkelte boliger i BBR og CPR. En oprindelig tanke var, at man til formålet anvendte fysiske boligidentifikationer i form af lejlighedsnumre på entredørene, men dette nød ikke fremme⁶⁹.

Præcisionen er ikke ideel

Erfaringerne har stort set vist, at den valgte, mere begrænsede løsning har været holdbar, uden dog at være ideel. Skarpheden og præcisionen i de anvendte specifikationer er næppe så stor som ønskeligt.

Antallet af personer (fra CPR), som ikke kan forbindes med en bolig (fra BBR) ved hjælp af en nøjagtig adresse, giver en indikation af sikkerheden i den valgte løsning. Denne andel, som ideelt set skulle være 0⁷⁰, var 4,0 pct. ved den første boligtælling pr. 1. juli 1979. Allerede ved den næste opgørelse pr. 1. januar 1980 var den nede på 2,1 pct., og siden midten af 1980'erne har den ligget nogenlunde konstant på 1,0 pct.

6.1.2 Relationen person/arbejdssted

Person/arbejdssted en forudsætning for pendlings- og erhvervsbeskæftigelsesstatistik

Etableringen af relationen person/bolig var en af forudsætningerne for, at der kunne udarbejdes "syntetiske" folke- og boligtællinger. En anden var, at en relation kunne etableres mellem beskæftigede personer og de erhvervsenheder, hvor de arbejder. Det var nærmere betegnet den lokal-faglige erhvervsenhed, arbejdsstedet, man skulle bruge. Relationen skulle nemlig indgå i opgørelser dels af erhvervspendlingen⁷¹, dels af den erhvervsfordelte beskæftigelse. Dette kræver, at enhederne kan henføres til en geografisk beliggenhed og grupperes i brancher på en nogenlunde præcis måde.

Oplysningssedlerne anvendes som instrument til at konstatere relationen

Hovedinstrumentet til at konstatere relationen var en lille udvidelse af informationen på de oplysningssedler, som arbejdsgiverne skulle indsende til skattevæsnet hvert år, og som i forvejen viste de beskæftigedes tilknytning til overordnede erhvervsenheder af typen firma eller juridiske enheder. Danmarks Statistik opnåede denne udvidelse ved stor imødekommenhed fra det daværende Statsskattedirektorat, og

⁶⁹ Jf. *Jensen & Thygesen (1985)*.

⁷⁰ Antallet af personer uden registreret bolig vil dog være større end nul, fordi der findes et begrænset antal hjemløse, som registreres med bopæl "på folkeregistret", og fordi danske diplomater i udlandet registreres med uspecificeret bolig i Danmark.

⁷¹ Statistik over den geografiske placering af bolig i forhold til placering af arbejdssted.

ideen var, at arbejdsgivere med flere lokal-faglige enheder (herefter kaldet arbejdssteder) på oplysningssedlerne for hver af deres beskæftigede skulle anføre et forud aftalt løbenummer, der svarede til personens arbejdssted. Nummeret fik navnet *Arbejdsstedskode*, jf. bilag 4. Den statistiske effekt af denne strategiske detalje var stor, og ordningen har i væsentlig grad levet op til forventningerne, selv om de praktiske vanskeligheder også har haft en anselig størrelse, jf. afsnit 4.3. Dette har bl.a. bidraget til en lang bearbejdnings tid, som det ville være særdeles ønskeligt at kunne afkorte.

6.1.3 Relationen arbejdssted/erhvervsarealenhed

En erhvervsarealenhed kan defineres som det ejendoms- eller bygningsareal, som et arbejdssted beslaglægger til sine aktiviteter. En sådan enhed går noget på tværs af de enheder, der anvendes i BBR-systemet (ejendomme, bygninger, boliger), og erhvervsarealenheder er kun i visse tilfælde afgrænset i BBR.

Erhvervsarealenhed anvendes til arbejdsstedets aktiviteter

Hvis man imidlertid systematisk kunne afgrænse erhvervsarealenheder i ejendoms- og BBR-systemerne, og hvis de kunne identificeres i disse systemer på samme måde som i erhvervsregistersystemet, kunne der etableres en kobling svarende til dem, der forbinder personer med boliger, hhv. personer med arbejdssteder, jf. afsnit 6.1.1 og 6.1.2.

Koblingen er ikke foretaget endnu...

Man kan sige, at denne problemstilling ikke har så meget med personstatistik at gøre, men den må alligevel omtales på grund af sin strategiske betydning for registerstatistikken.

Problemet har været under overvejelse i en del år, men umiddelbart synes der ikke at være nogen nem løsning af en karakter, som svarer til de to forannævnte tilfælde.

...og bliver det heller ikke foreløbigt

Mindst to delproblemer skal i givet fald klares:

- Det ene er, at der må autoriseres et fælles adressereferencesystem for alle landets ejendomme og bygninger. Der findes to gode udgangspunkter herfor: CPR's adresse- og vejcodesystem og det såkaldte *Krydsreferenceregister* i boligministeriets regi. Det vil imidlertid være ønskeligt, at der blev etableret ét integreret og autoriseret system, som formentlig i forhold til de nuværende skulle udbygges noget.
- Mere vanskeligt vil det være at etablere et system til afgrænsning af erhvervsarealenheder, der er kongruente med arbejdssteder eller eventuelt et andet lokalt virksomhedsbegreb.

6.1.4 Relationen mellem ejendomme og ejer

Koblingen mellem ejer og ejendomme delvist foretaget

En mere fremkommelig koblingsmulighed, hvis grundlag til en vis grad allerede eksisterer i dag, er at forbinde oplysninger om ejendomme fra *Ejendomsregistret* med oplysninger om ejer (dvs. person og/eller erhvervsenhed) via de *ejeroplysninger*, der er opført i Ejendomsregistret, identificeret ved person- eller virksomhedsnumre. Denne koblingsmulighed vil blive yderligere styrket, når et standarderhvervsnummer (CVR-nummeret, jf. nedenfor) bliver etableret. På et sådant grundlag ville man kunne udvikle en økonomisk ejendomsstatistik med størrelser som ejendomsvurderingsværdier, salgspriser og omsætningsfrekvenser suppleret med person- og/eller virksomhedsoplysninger som baggrundsvariable. Overvejelser om en sådan statistik er endnu på et meget foreløbigt stade.

6.2 Anvendelse af fælles data

Sammenkobling udvider rammerne...

Forbindelserne mellem de forskellige enheder betyder, at oplysninger fra et statistikområde kan trækkes over i et andet og omvendt. Det betyder igen, at forskellige typer statistik kan udarbejdes på samme grundlag, men i forskellige rammer.

Hovedeksemplet herpå er de tæt forbundne statistikker:

- registerbaseret arbejdsstyrkestatistik (RAS)
- erhvervsbeskæftigelsesstatistikken
- pendlingsstatistikken

...og forskelle i resultater kan forklares

Alle disse statistikker bygger på nogle få, stort set fælles grundoplysninger, herunder den strategiske kobling mellem beskæftigede og arbejdssted. Det er denne kobling, der bl.a. skaber mulighed for dobbeltlokalisering af de beskæftigede efter deres bopæl og deres arbejdssteds beliggenhed, dvs. pendlingsstatistikken. De grundlæggende informationer er målingsteknisk de samme, medens begrebernes afgrænsning og anvendelse er forskellig. Det fælles produktionssystem tillader, at forskelle i resultaterne kan forklares. Fx er forskellen mellem antallet af beskæftigede personer i arbejdsstyrken og antallet af "jobs" i erhvervsbeskæftigelsesstatistikken forklaret ved antallet af personer med bierhverv.

6.3 Begrebernes validitet

Forskel på register- og statistikbegreber

På godt og ondt er registerstatistikken bundet til de begreber, der faktisk findes i registrene. Disse begrebers kvalitet bliver derfor afgørende for statistikens kvalitet. Det gælder selvfølgelig også, når oplysninger fra ét område ved hjælp af forbindelser mellem enhederne

trækkes over i et andet. Begreberne i registrene bliver ofte debatteret, fordi de ikke svarer nøjagtigt til begreberne i traditionel statistik. Når overgangs- og indkøringsperioden er overstået, bør det være begrebernes egne, kvalitative egenskaber, der er afgørende - også for samspillet mellem statistikområderne. I det følgende omtales kort de mest centrale oplysningers karakteristika.

Personoplysningerne er kvalitetsmæssigt i en klasse for sig. Begrebet "en person" er så veldefineret, at det slet ikke behøver definition, og personnummeret udgør en unik identifikation, der følger personen på tværs i informationssystemerne og på langs over tiden.

Personoplysninger er de mest valide...

Noget vanskeligere er sagen med hensyn til de begreber, der bygger på grupperinger af personer: *familie* og *husstand*. Disse begreber kan forholdsvis let afgrænses på et givet tidspunkt ved hjælp af personrelationer og adressefællesskab. Med hensyn til identifikation over tid er sagen vanskeligere. Er der fx tale om den samme familie, hvis ét eller flere medlemmer fraflytter, eller hvis et antal af medlemmerne sammen flytter til en anden adresse? Det bidrager yderligere til problemerne, at adressekoden kan forandre sig i forbindelse med vejnummeromlægning o.l.

...men grupperinger kan volde problemer

Adressens manglende stabilitet spiller også en rolle, hvis man vil følge det i øvrigt ret veldefinerede begreb *bolig* over tid. Dette løses ikke af, at også ejendomme og bygninger som fysiske fænomener er forholdsvis veldefinerede.

Det er imidlertid inden for *erhvervsområdet*, at de i særklasse vanskeligste problemer med hensyn til begrebsafgrænsningen findes.

Erhvervsområdet et knudeproblem

I den danske erhvervsorienterede statistik arbejder man med tre grundlæggende enhedstyper. De kan imidlertid sammensættes til andre enhedstyper, såfremt behov herfor skulle opstå, fx i forhold til EU-statistik.

De tre enhedstyper er:

- Den *juridiske enhed*, som er den grundlæggende enhed; de andre enhedstyper bygger i princippet på dette hovedbegreb.
- Den *økonomiske enhed (firmaet)* er det samme som den juridiske enhed, undtagen i visse, forholdsvis få tilfælde, hvor en opdeling i flere juridiske enheder ikke afspejler en reel arbejdsdeling.
- Den *lokalfaglige enhed (arbejdsstedet)*, der er en lokalt og fagligt afgrænset del af et firma.

Danmarks Statistiks erhvervsregistersystem, der betjener både Det centrale erhvervsregister og erhvervsstatistikken, har i længere tid lidt

Mangel på systematik i afgrænsningen

under manglen på systematik i de juridiske enheders afgrænsning og nummerering. Et andet problem har været ufuldstændig/langsom ajourføring af arbejdsstederne og manglen på selvstændig nummerering af disse på en sådan måde, at de kan følges over tid.

Det er på denne baggrund et hovedpunkt i en igangværende reform af den generelle erhvervsstatistik at skabe et registersystem, hvori disse mangler i væsentlig grad er udbedret.

Oprettelse af et nyt centralt basisregister CVR...

Mens denne reform har været undervejs, er der imidlertid overvejelser om oprettelse af et nyt centralt basisregister, Det centrale virksomhedsregister, CVR, i Told- og Skattestyrelsens regi. Dette kan dog først træde i funktion i løbet af et par år efter en endelig beslutning. Det er en selvfølge, at der i forbindelse med oprettelse af et sådant register skal ske en kraftig sanering af juridiske enheder, herunder en tiltrængt oprydning i numrene. Disse skal i øvrigt bygge på de såkaldte SE-numre, der også skal afløse de nuværende aktieselskabsnumre.

...indeholder nye perspektiver for statistik på erhvervsområdet...

Danmarks Statistik har ud fra egne og andre potentielle CVR-brugeres interesser set det som sin opgave at argumentere for, at det nye register systematisk kommer til at indeholde arbejdsstedsniveau, at det bliver overladt Danmarks Statistik at vedligeholde oplysningerne om disse enheder i samarbejde med andre hovedbrugere, og at disse enheder forsynes med et eget uafhængigt nummersystem således, at de kan følges over tid, uden påvirkning af ændringer i de numre, som identificerer den juridiske enhed. Sammen med Arbejdsministeriet, der er administrativ storbruger af oplysninger om sådanne enheder, har Danmarks Statistik fremsat et forslag herom, der tilgodeser begge institutioners interesser, og som i alt væsentligt bygger på Danmarks Statistiks hidtil anvendte arbejdsstedsbegreb.

...men man skal være varsom

At registerbegrebernes egenskaber er centrale for statistikken anvendelighed, indebærer ikke nødvendigvis, at man skal videreføre den traditionelle statistiks kendte begreber i uændret form. Det er også vigtigt, at begreberne er stabile, velafgrænsede og egnede til administrative formål, således at oplysninger fra forskellige administrationer kan indpasses i samme helhed.

Registerstatistikken bliver ofte debatteret i internationalt regi

I den internationale diskussion anføres det ofte som et hovedargument mod registerstatistik, at disse hensyn er vanskelige at forene med de krav, som anerkendte internationale definitioner stiller til begreberne. Dette synspunkt kan selvfølgelig ikke uden videre afvises, men da grundregistrene og den generelle statistik i nogen grad har fælles interesser med hensyn til de grundlæggende begreber, kan man alligevel nå et stykke vej, såfremt problemet prioriteres ved planlægningen af statistikken.

På dette område har Danmarks Statistik med sit omfattende erfaringsgrundlag og en tilgang til begrebsproblematikken, som ikke er farvet af bestemte administrative hensyn, en betydningsfuld rolle at spille - en rolle som den fremsynede lovgivning har forudset i §1, stk. 1 nr. 3 i Lov om Danmarks Statistik⁷².

6.4 Perspektiv

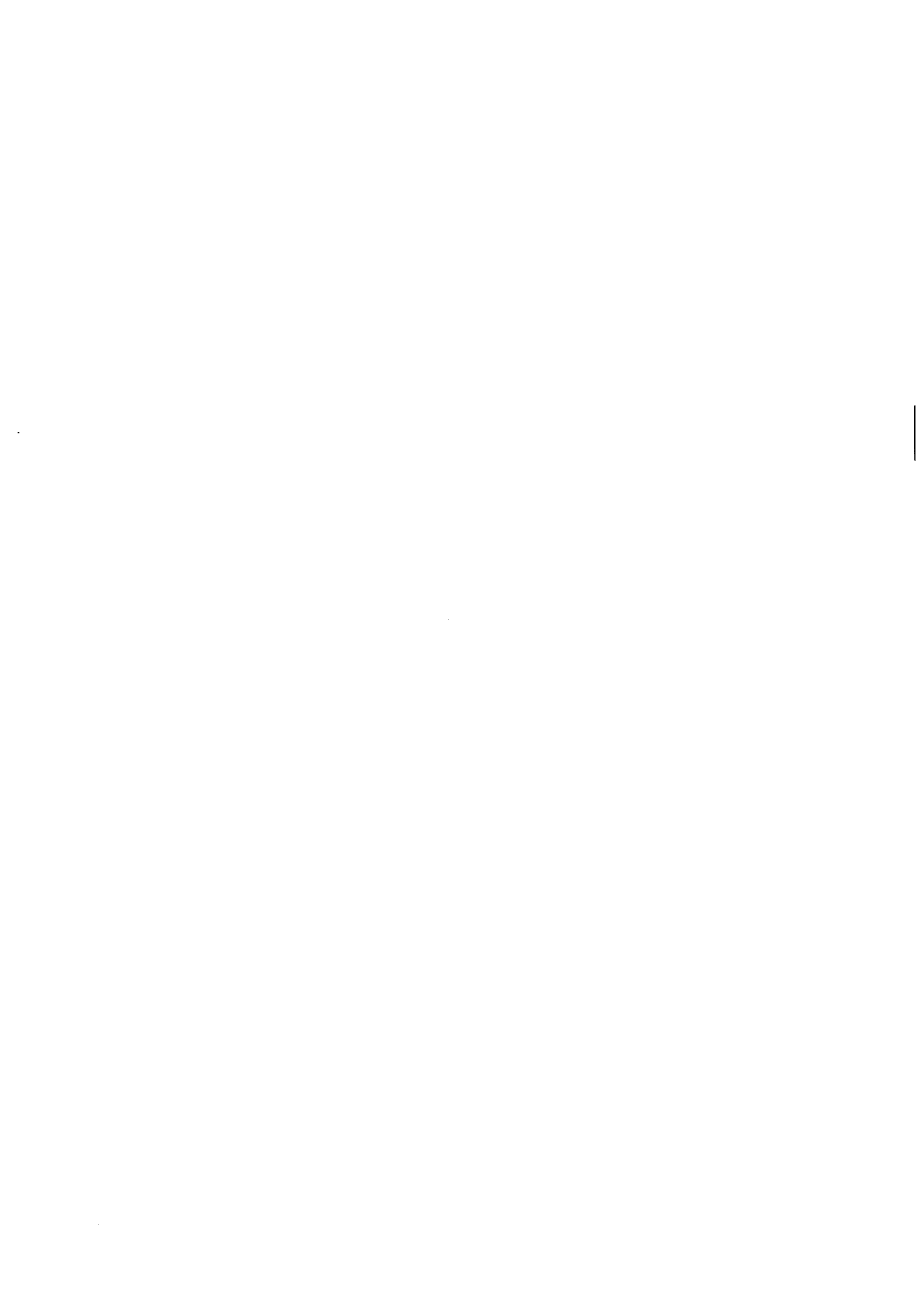
Allerede nu indebærer den registerstatistiske udvikling på personstatistikens område et stort spring fremad, blandt andet fordi den muliggør, at man opgør bruttoforskydninger i stedet for nettobevægelser. Det er en gammel, men stadig gyldig sandhed, at "ingen hidtil har set en nettoindvandrer", og det er derfor en væsentlig landvinding, at statistikken kan belyse bruttovandringerne. Noget andet er, at den fulde udnyttelse af registerstatistikens potentiale næppe er opnået endnu.

Med den rette fastlæggelse af begreberne inden for erhvervsområdet opstår der tilsvarende muligheder for at imødekomme behovene for erhvervsdemografiske opgørelser, dvs. opgørelser over virksomhedernes opståen og død, flytninger, brancheskift og ejerskift. Mere relevant i sammenhæng med dette kapitel er opgørelser af samspillet mellem på den ene side virksomhedernes demografi og på den anden side personernes vandring mellem beskæftigelse i forskellige virksomheder, arbejdsløshed samt ind og ud af arbejdsstyrken. Det vil være meget værdifuldt at kunne opgøre: Beskæftigede i hhv. ophørte og nye enheder, bruttotilgang og bruttoafgang af beskæftigede i bestående erhvervsenheder, etc. Som det fremgår af forrige kapitel, har Danmarks Statistik allerede gjort en del for at belyse dette samspil gennem etableringen af IDA, men der er behov for en mere løbende belysning. Det er de enkelte komponenter - ikke nettobevægelserne - der kan påvirkes af erhvervspolitiske og arbejdsmarkedspolitiske tiltag, og de må derfor belyses i statistikken, hvis den skal være et godt grundlag for beslutningerne.

Perspektiver i registerstatistikken

Perspektiver for det erhvervsstatistiske område

⁷² Jf. Økonomiministeriet (1992a).



Kapitel 7. Registerlovgivning og datapolitik

De mange oplysninger, som indgår i statistikregistrene, skal behandles og opbevares efter lovbestemte regler. Endvidere skal det på overbevisende måde sikres, at oplysninger om enkeltpersoner ikke bliver misbrugt, når oplysningerne stilles til rådighed for statistisk anvendelse. Denne vanskelige balance opnås gennem et sæt regler, som udtrykker den fastlagte diskretionspolitik. I dette kapitel gennemgås lovgivningens særbestemmelser for statistikoplysninger samt Danmarks Statistiks sikkerhedsorganisation og diskretionspolitik.

7.1 Registerloven

Danmarks Statistiks personstatistikregistre er omfattet af *Lov om offentlige myndigheders registre*⁷³. Denne lov gælder for edb-registre, der føres for den offentlige forvaltning, og som indeholder personoplysninger. Ved edb-registre forstås registre eller andre systematiske fortegnelser, hvor der gøres brug af elektronisk databehandling. Ved personoplysninger forstås oplysninger, som kan henføres til bestemte personer, eventuelt via kendskab til personnummer, registreringsnummer eller lignende særlige identifikationer.

Generelle bestemmelser

Loven indeholder en række bestemmelser om oprettelse af registre, om registrering og opbevaring af oplysninger, registrerede personers adgang til oplysninger om sig selv, videregivelse af registeroplysninger, Registertilsynet og straf.

For statistik og forskning gælder enkelte undtagelser i forhold til de almindelige regler. Det drejer sig om adgangen til at videregive oplysninger (§18 og §21, stk. 5), adgangen til egenindsigt (§13, stk. 5), samt adgangen til at samkøre registre (§4, stk. 3 og §8d).

Undtagelser for statistik

Som hovedregel må personoplysninger afgivet til statistikformål ikke videregives. Loven åbner dog mulighed for, at sådanne oplysninger kan videregives til andre statistiske eller videnskabelige formål under forudsætning af, at der forud indhentes tilladelse hertil fra Registertilsynet. Endvidere kan videregivelse ske, hvor der foreligger en særlig hjemmel. EF-forordningen (nr. 1588/90) om fremsendelse af fortrolige oplysninger til De Europæiske Fællesskabers Statistiske Kontor er eneste eksempel på en sådan hjemmel, for så vidt angår personstatistikregistrene. Forordningen bemyndiger statistikbureauerne til at fremsende fortrolige statistiske oplysninger til Eurostat⁷⁴. Fra det erhvervsstatistiske registersystem finder en begrænset videregivelse sted

Videregivelse af personoplysninger

⁷³ Jf. *Justitsministeriet (1992a)*.

⁷⁴ Den europæiske kommissions generaldirektorat for statistik.

til Det centrale erhvervsregister med hjemmel i Lov om Det centrale erhvervsregister⁷⁵.

Adskillelse af statistik- og personoplysninger

De restriktive videregivelsesbestemmelser er en væsentlig forudsætning for de to øvrige undtagelser. Med disse bestemmelser sikres det, at statistikoplysninger ikke får indflydelse på den administrative behandling af personoplysninger. Når det drejer sig om overordnede systemmæssige fejl i forbindelse med udtræk og levering af oplysninger til Danmarks Statistik, kan Danmarks Statistik gøre opmærksom på fejlen, men det sker uden henvisning til enkeltpersoner. Hvis Danmarks Statistik ved behandlingen af statistikoplysningerne konstaterer fejl eller andre uregelmæssigheder om enkeltpersoner, kan disse kun i meget begrænset omfang rettes ved henvendelse til den pågældende administrative myndighed.

Adgang til egenindsigt

Reglerne om videregivelse fastlægger på en vis måde, at personstatistikregistre i princippet ikke indeholder egentlige personoplysninger. Indholdet må kun være grundlag for anonyme statistiske opgørelser. Den enkelte person vil således ikke i nogen sammenhæng blive konfronteret med oplysninger fra et statistikregister og har derfor ingen egeninteresse i at kende indholdet af et sådant register. Det er baggrunden for, at loven i §13, stk. 5 indeholder en undtagelsesbestemmelse med hensyn til den enkelte persons adgang til at få oplyst, hvad der er registreret om den pågældende selv i et statistikregister.

Adgang til samkøring

Den sidste undtagelse gælder som nævnt adgangen til at samkøre registre. Efter de almindelige regler skal der oprettes såkaldte forskrifter (jf. afsnit 7.1.1) for en samkøring af registre, med mindre formålet er statistik eller forskning. Begrundelsen er igen, at det samkørte produkt ikke må danne grundlag for afgørelser vedrørende enkeltpersoner og derved ikke kan skade disse. Det skal dog tilføjes, at undtagelsen kun gælder, hvis samkøringen resulterer i et anonymt datasæt, ellers vil der være tale om et nyt register, for hvilket der skal foreligge forskrifter.

Videregivelsen af oplysninger fra de administrative registre til Danmarks Statistik er omfattet af reglerne i (jf. §21, stk. 2 og 3). For særligt følsomme data gælder den hovedregel, at videregivelse ikke må finde sted, (jf. §21, stk. 1). Blandt undtagelserne fra denne regel er lovhjemlet adgang til videregivelse, eller hvis formålet med videregivelsen er statistik eller forskning. *Lov om Danmarks Statistik* er således et tilstrækkeligt grundlag for indsamling af administrative oplysninger til statistik. Den generelle statistikbestemmelse gør det i forbindelse med serviceopgaver, muligt at anvende supplerende oplysninger fra administrative registre til samkøring i Danmarks Statistik, når bru-

⁷⁵ Jf. *Økonomiministeriet (1992b)*.

geren ønsker det. Tilsvarende bestemmelser findes i *Lov om private registre*⁷⁶, (jf. §4 stk. 1 og 3).

7.1.1 Registerforskrifter

Registerloven indeholder en bestemmelse om, at der skal fastsættes forskrifter for hvert register. Der er undtagelser fra denne regel, men de er uden betydning for statistikregistrene. Registerforskrifterne skal indeholde de regler, som gælder for driften af det enkelte register. Indholdet omfatter følgende punkter:

- formålet med oprettelsen af registret
- registrets indhold
- opdatering
- anvendelsen af registret
- sikkerhedsforanstaltninger
- eventuelle videregivelsesbestemmelser
- opbevaring

I bilag til forskrifterne gives en specificeret beskrivelse af hvilke typer af oplysninger, der indgår i det pågældende register, og hvor de kommer fra. I et andet bilag beskrives anvendelsen af registret. Endelig er sikkerhedsforanstaltningerne beskrevet i *Danmarks Statistiks Datasikkerhedsreglement*⁷⁷, der ligeledes er bilag til forskrifterne.

Videregivelsesbestemmelserne er fastsat således, at det for de fleste registre gælder, at videregivelse ikke kan finde sted. Danmarks Statistiks datapolitik er på dette område mere restriktiv end loven. I nogle konkrete sager har man endog afvist anmodninger om videregivelse til statistikformål, selv om forskriftens regler ikke var til hinder for dette. Det er derfor kun i meget få tilfælde, at videregivelse har fundet sted, og det er kun sket, hvis videregivelsen har været åbenbart rimelig og uden sikkerhedsrisiko.

Kun videregivelse i meget begrænset omfang

Forskrifterne sætter også regler for, hvor længe oplysningerne må opbevares. Generelt må de opbevares så længe det er nødvendigt for at tjene det formål, hvortil de er indsamlet. Dog er der fastsat en bestemt frist for, hvornår et register eller en registerversion skal slettes eller arkiveres i Rigsarkivet. Danmarks Statistik har traditionelt anvendt en hovedregel om 5 års opbevaring, med enkelte undtagelser på op til 20 år. Registerversioner, der opbevares i Rigsarkivet, kan, efter tilladelse fra Registertilsynet, stilles til rådighed for konkrete statistikopgaver ud over den fastsatte frist.

Oplysningerne opbevares kun så længe det er nødvendigt

⁷⁶ Jf. *Justitsministeriet (1992b)*.

⁷⁷ Jf. *Danmarks Statistik (1991c)*.

7.2 Datasikkerhedsorganisationen

Sikkerhedsansvarlig Rigsstatistikerens er øverste ansvarlige for datasikkerheden i Danmarks Statistik. Til støtte for rigsstatistikerens er nedsat et datasikkerhedsudvalg, som løbende behandler spørgsmål i forbindelse med datasikkerheden og kommer med indstillinger om eventuelle ændringer. Udvalget skal mindst én gang om året gennemgå sikkerhedsforanstaltningerne og afgive beretning til Rigsstatistikerens. Endvidere forestår udvalget udarbejdelsen af en datasikkerhedshåndbog, som beskriver alle regler og foranstaltninger i forbindelse med datasikkerheden. Udvalget skal sikre, at en ajourført datasikkerhedshåndbog altid er let tilgængelig for alle ansatte i Danmarks Statistik.

7.3 Danmarks Statistiks diskretionspolitik

De mange muligheder for videregående analyser og forskningsprojekter, som statistikregistersystemets omfattende dataindhold og data-sammenhænge giver, har skabt behov for, at der stilles detaljerede oplysninger til rådighed for eksterne brugere. Det er Danmarks Statistiks opgave at fremskaffe de efterspurgte oplysninger til brugerne i en efter formålet hensigtsmæssig form. Samtidig skal Danmarks Statistik sikre, at man ikke af oplysningerne kan genkende enkeltpersoner, således at der opstår en risiko for misbrug. Det sidste hensyn er ufravigeligt, men det kan være vanskeligt at afgøre, hvornår der opstår en risiko. Brugerens ønsker kan være meget detaljerede oplysninger, der kan være sagligt velbegrundede i forhold til den påtænkte anvendelse. Der opstår derfor nemt en konflikt mellem de to hensyn, og Danmarks Statistik har på denne baggrund formuleret en diskretionspolitik, som udstikker retningslinier for, hvornår en opgave må afvises. I det følgende gengives hovedtrækkene.

En opgave skal være lovlig og mulig

En opgave skal være lovlig, hvilket betyder, at eventuelt supplerende oplysninger, som indberettes af en opgavestiller, ikke kan indgå, hvis det er i strid med registerlovgivningen. Endvidere skal opgaven være mulig i den forstand, at de for opgaven relevante oplysninger skal være til rådighed i et omfang, som sikrer en tilfredsstillende behandling af emnet. Endelige skal som nævnt enkeltpersoners anonymitet sikres.

Danmarks Statistiks styrelse bestemmer, hvilke emner der skal behandles

I almindelighed er det Danmarks Statistiks styrelse, som gennem den årlige arbejdsplan fastlægger, hvilke emner der skal behandles i statistikken. I forbindelse med serviceopgaver er det imidlertid normalt opgavestilleren, der bestemmer emnet. Et sådant emne kan være af en karakter, som vil være generende for visse grupper. Særlige interesser kan være truet, hvis resultaterne fører til uønskede beslutninger eller viser en koncentration af uheldige egenskaber i en særlig mindre gruppe af befolkningen. Behandlingen af sådanne emner kunne medføre

utilfredshed med Danmarks Statistik i visse kredse. Trods dette, at resultaterne kan være ubekvemme for Danmarks Statistik, må det ikke i sig selv føre til, at en opgave afvises. Til gengæld forbeholder Danmarks Statistik sig ret til at kritisere undersøgelsens konklusioner.

Den mest velkendte form for statistik er tabellerne, som offentliggøres i Danmarks Statistiks publikationer. Der vil normalt ikke være problemer forbundet med de aggregerede oplysninger, som fremgår af disse tabeller. Anderledes stiller det sig med tabeller, der fremstilles efter særlig aftale. Her kan kravene til detaljeringsgrad være sådan, at der opstår en risiko for genkendelse af enkeltpersoner. Det gælder specielt, hvis der indgår opdelinger i mindre geografiske områder, hvorfor der må stilles krav om en mindstegrænse for størrelsen af disse. Målt i antal indbyggere sættes denne grænse som hovedregel til 1000 personer, men den må naturligvis vurderes i forhold til specificationsgraden af de øvrige oplysninger i tabellen.

Normalt ingen problemer med tabeloplysninger

Hovedreglen for hvad der kan videregives af oplysninger fra Danmarks Statistik er, at de skal kunne offentliggøres uden risiko for genkendelse af enkeltpersoner. Det betyder ikke nødvendigvis, at offentliggørelse faktisk finder sted. Undtagelser fra denne regel er mulig på særlige betingelser.

Modeldata

Modeldatasæt, dvs. individualoplysninger uden identifikation, kan stilles til rådighed for statistik eller forskningsprojekter. Det kan ske gennem levering af et datasæt til brugeren, der så kan behandle oplysningerne på eget edb-anlæg, eller det kan ske ved at give brugeren adgang til at behandle datasættet på Danmarks Statistiks edb-anlæg. Sidstnævnte kan ske gennem en fast opkoblet linje fra en arbejdsstation i Danmarks Statistik på særlige forskerplaceringsvilkår, hvilket bl.a. indebærer, at forskeren selv gennemfører behandlingen af oplysningerne under opsyn og efter særlig aftale, eller ved at forskeren indsender sit edb-program, som afvikles på nærmere aftalte modeldatasæt ved Danmarks Statistiks foranstaltning, "Den elektroniske postkasse".

Modeldata, som stilles til rådighed for eksterne brugere, er omfattet af det overordnede princip, at der ikke må være en reel risiko for, at modtageren kan identificere enkeltpersoner og på den måde opnå ny viden om sådanne. For at sikre dette er der en række regler for, under hvilke omstændigheder modeldata kan stilles til rådighed:

Sikring af enkeltpersoners anonymitet

- Hvis et modeldatasæt omfatter identificerede personoplysninger, der er indleveret af opgavestilleren til *samkøring* med oplysninger i Danmarks Statistik, kan det kun stilles til rådighed under *forskerplaceringsordningen* i Danmarks Statistik.
- *Dataindholdet* i et modeldatasæt må kun indeholde de oplysninger, som er *nødvendige* for at opfylde formålet med *dannelsen* af da-

tasættet. Der kan kompenseres for forekomsten af mange data gennem en reduktion af stikprøvestørrelsen eller ved at reducere specificationsgraden for udvalgte variable.

- Der må ikke forekomme oplysninger med høj identifikationskraft, med mindre der er tale om tilbagelevering af oplysningerne til den, som oprindeligt har tilvejebragt alle personoplysningerne, eller oplysningerne skal indgå i et udviklingsprojekt om hvilket der er indgået en aftale om samarbejde med Danmarks Statistik som deltager.
- Modeldata kan som hovedregel kun stilles til rådighed under forskerplaceringsordningen eller den elektroniske postkasse. Kun hvis der er tale om en stikprøve med en meget lille udvalgsandel eller en forholdsvis lav specificationsgrad, kan et modeldatasæt leveres til behandling på brugerens eget edb-anlæg.

Aftale om levering af modeldata

I forbindelse med levering af modeldatasæt aftales de nærmere vilkår for behandlingen af oplysningerne. En sådan aftale skal omfatte en specifikation af formålet, en angivelse af, hvem der har adgang til at bruge datasættet, samt en præcisering af, hvor databehandlingen må finde sted. Endvidere skal det indgå, at videregivelse af oplysninger kun må ske i en bearbejdet form, der er i overensstemmelse med kravene til, hvad der kan publiceres. Endelig forbeholder Danmarks Statistik sig ret til efterkritik, da oplysningerne ikke er offentligt tilgængelige, og der derfor ikke umiddelbart er mulighed for, at andre kan vurdere brugerens anvendelse af de pågældende oplysninger.

I bilag A, B og C til dette kapitel vises indholdet af aftalerne vedrørende adgangen til at anvende modeldatasæt hos brugeren selv, under forskerplaceringsordningen og via den elektroniske postkasse.

Lovmodellen modtager specielle modeldata

En særlig form for modeldata stilles til rådighed for regeringens Lovmodel⁷⁸. Data er organiseret i flere såkaldte modelbefolkninger, og gennem disse har brugerne adgang til at anvende meget detaljerede oplysninger fra mange forskellige registre. Adgangen sker fra arbejdsstationer hos brugeren selv over fast opkoblede linjer, og Danmarks Statistik har lagt følgende klausuler på benyttelsen af disse modeldata:

- At de udelukkende vil blive anvendt til lovmodelberegninger, herunder afprøvning af sådanne beregninger.
- At kun autoriserede lovmodelbrugere får adgang til oplysningerne.
- At brugerne ikke er berettiget til at offentliggøre egentlige statistiske opgørelser, der måtte blive fremstillet på grundlag af modelbefolkningerne, uden forudgående aftale med Danmarks Statistik.

⁷⁸ Lovmodellen er et redskab i centraladministrationens lovforberedende arbejde, et edb-baseret beregningssystem, der gør det muligt at belyse de umiddelbare økonomiske konsekvenser af gældende og påtænkt lovgivning vedrørende skatter og en række indkomstoverførsler, jf. *Økonomiministeriet (1993)*.

Risikoen i forbindelse med Lovmodellen kan være, at man af hensyn til udefinerede behov ophober alt for mange data. Der gælder derfor følgende begrænsende regler:

- Der må ikke leveres data om stikprøver større end 3 pct. af hele befolkningen.
- Der må normalt ikke leveres forløbsdata om enkeltpersoner. Kun i særlige tilfælde, hvor der kan påvises specifikke behov i forbindelse med lovforberedende arbejde, vil det være muligt. Der må ikke indlægges oplysninger, der ikke kan bruges på grund af kvalitetsproblemer.

Levering af oplysninger om enkeltpersoner med formel identifikation er normalt ikke tilladt. Undtaget fra denne regel er levering af tilfældigt udvalgte stikprøver til Socialforskningsinstituttet. Stikprøverne kan benyttes til undersøgelser, der udføres af instituttet. Stikprøverne må kun afgrænses efter køn, alder og bopæl. Hvis en stikprøve ønskes stratificeret efter andre kendetegn end de nævnte, skal databehandling af hele undersøgelsen foregå i Danmarks Statistik.

Levering af identificerede oplysninger

En anden undtagelse gælder videregivelse af identificerede personoplysninger vedrørende den grønlandske befolkning til det grønlandske hjemmestyres statistiske kontor.

7.4 Reaktioner fra statistikbrugere

Reaktioner på Danmarks Statistiks diskretionspolitik kommer især fra brugere af modeldatasæt. De kan have svært ved at forstå nødvendigheden af den restriktive politik. Set fra en forskersynsvinkel kan det være af væsentlig betydning at have detaljerede oplysninger til rådighed, da det bl.a. kan være bestemmende for, hvilke analysemetoder der kan anvendes. I visse tilfælde kan den restriktive politik betyde, at projekter må begrænses eller helt opgives.

Især reaktioner fra brugere af modeldata

Adgangen til at anvende modeldata er blevet lempet gennem de senere år. Den tidligere omtalte forskerplaceringsordning har således åbnet mulighed for at anvende meget detaljerede oplysninger, men det opleves ikke af alle som tilstrækkeligt. Det føles af nogle som en betydelig ulempe, at databehandlingen skal foregå fra en arbejdsstation i Danmarks Statistik, og det må da også erkendes, at den geografiske afstand i nogle tilfælde kan være en væsentlig hindring for at benytte ordningen.

Lempet adgang til modeldata gennem forskerplaceringsordningen

For at afhjælpe afstandsproblemet er den såkaldte elektroniske postkasseordning indført. Gennem denne ordning har brugeren mulighed for at lave sine edb-programmer hjemme og derefter indsende dem

Den elektroniske postkasse

**Forskerne,
Registertilsynet og
befolkningens tillid
til Danmarks
Statistik**

over nettet til afvikling på aftalte datasæt i Danmarks Statistik. Brugeren har ikke direkte adgang til data, og det kan betyde en lidt tung kommunikation. Dette skal så afvejes mod muligheden for at arbejde i Danmarks Statistik under forskerplaceringsordningen, eller at arbejde "hjemme" på et stærkt begrænset datamateriale.

I almindelighed vil forskerne af forståelige grunde foretrække at få detaljerede datasæt stillet til rådighed. De støttes til en vis grad af registerlovens videregivelsesbestemmelser, på baggrund af hvilke noget sådant er muligt med tilladelse fra Registertilsynet. Danmarks Statistik på sin side ser det som et meget væsentligt led i institutionens virksomhed at stille oplysninger til rådighed, og man bestræber sig på at finde acceptable løsninger på de problemer, som den førte diskretionspolitik rejser. Der er imidlertid et overordnet og meget tungtvejende hensyn, som er baggrund for denne politik. Det gælder sikringen af, at der i befolkningen er tillid til, at oplysninger i Danmarks Statistik ikke bliver misbrugt. Dette er af afgørende betydning for Danmarks Statistiks fortsatte virksomhed og dermed også for forskernes muligheder for at få datamaterialer fremover.

7.5 EU-direktiv om personbeskyttelse

Med henblik på at skabe grundlag for informationers frie bevægelighed i EU har Kommissionen udarbejdet et forslag til *"Rådets direktiv om beskyttelse af fysiske personer i forbindelse med behandlingen af personoplysninger og om fri udveksling af sådanne oplysninger"*. Forslaget er gennem længere tid blevet behandlet, dels i en arbejdsgruppe under Rådet og dels i de enkelte medlemslande.

Forslaget vil, hvis det gennemføres i den senest kendte form, betyde væsentlige begrænsninger i mulighederne for statistisk anvendelse af personoplysninger. Problemerne er især forbundet med adgangen til at genbruge administrative data til andre formål end det oprindelige, fx statistikformål, men det handler også om adgangen til egenindsigt, form af og varighed for opbevaring af statistikoplysninger m.m. Fra dansk og fra anden side er der gjort opmærksom på konsekvenserne for statistik og forskning, hvis det foreliggende forslag gennemføres, og der synes da også at være bevægelse i retning af at tilgodese statistikernes synspunkter. Det er imidlertid for tidligt at sige noget om, hvordan det ender.

Bilag A.

AFTALE
OM UDLEVERING AF MODELDATA
TIL FORSKNINGSPROJEKT

Mellem Danmarks Statistik og

forskerens navn (fn.)

er indgået følgende aftale om udlevering og brug af modeldata til forskningsprojektet:

(projektets titel og indhold)

For aftalen gælder følgende:

1. *fn.* er ansvarlig for de leverede data og for deres anvendelse
2. Data skal opbevares på et edb-anlæg under sådanne sikkerhedsforhold, at kun *fn.* har adgang til data. *fn.* har dog ret til at give andre projektdeltagere adgang til data. Denne regel gælder også for opbevaring på PC.
3. Datasættets grundoplysninger skal betragtes som fortrolige oplysninger, jf. Forvaltningslovens §27, stk. 3 og Straffelovens §152.
4. Resultater af bearbejdningen af grunddata på papir eller andre medier skal være af en sådan karakter, at det enkelte individ eller den enkelte virksomhed ikke kan identificeres.
5. Skriftlige arbejder, der er baseret på data fra projektet, skal inden offentliggørelse forevises Danmarks Statistik
6. Modeldatasættet tilbageleveres til Danmarks Statistik senest ved projektets afslutning, dog senest den _____. Der må ikke efter tilbageleveringen forefindes kopier af datasættet udenfor Danmarks Statistik. Det samme gælder for afledte datasæt med oplysninger på individ- eller virksomhedsniveau.
7. Aftalen træder i kraft den _____

Bilag B**AFTALE OM
FORSKERPLACERING I DANMARKS STATISTIK**

Mellem Danmarks Statistik og

forskerens navn (fn.)

er indgået nedenstående aftale om forskerplacering i Danmarks Statistik.

Aftalen vedrører projektet:

projektets titel og indhold

For aftalen gælder følgende:

1. Danmarks Statistik giver *fn.* adgang til at arbejde i Danmarks Statistiks lokaler med de aftalte oplysninger, alt under hensyntagen til Danmarks Statistiks diskretionsprincipper, datasikkerhedsregler og ressourcemæssige situation.
2. *fn.* skal følge de i nærværende aftale fastsatte bestemmelser samt Danmarks Statistiks datasikkerhedsreglement, som er bilag til denne aftale. Projektets grundoplysninger skal betragtes som fortrolige oplysninger, jf. Forvaltningslovens §27, stk. 3 og Straffelovens §152.
3. *fn.* må kun bearbejde de nævnte oplysninger i Danmarks Statistiks lokaler. Resultater af bearbejdningen på papir eller andre medier må kun medtages ud fra Danmarks Statistiks lokaler efter tilladelse fra forskningskonsulenten eller dennes stedfortræder.
4. Skriftlige arbejder, der er baseret på data fra projektet skal inden offentliggørelse forvisse Danmarks Statistik. Der må ikke offentliggøres oplysninger, hvor det enkelte individ kan identificeres.
5. Danmarks Statistik stiller arbejdsplads til rådighed i et lokale med kontorinventar, telefon og skærmterminal.
6. Projektets omkostninger betales efter de af Danmarks Statistik til enhver tid fastsatte servicetakster.
7. Aftalen træder i kraft den _____

Bilag C**AFTALE OM ANVENDELSE
AF DEN "ELEKTRONISKE POSTKASSE"**

Mellem Danmarks Statistik og

forskerens navn (fn.)

er indgået nedenstående aftale om adgang til data til forskningsprojektet:

projektets titel og indhold

1. Datasættet beror i Danmarks Statistik og *fn.* har adgang til data via Danmarks Statistiks PC-ordning for forskere.
2. Projektets grundoplysninger skal betragtes som fortrolige oplysninger, jf. Forvaltningslovens §27, stk. 3 og Straffelovens §152.
3. Der må ikke udskrives individoplysninger (enkelte records) fra grunddata uden tilladelse fra Danmarks Statistik.
4. Skriftlig arbejde, der er baseret på data fra projektet, skal inden offentliggørelse forevises Danmarks Statistik.
5. Der må ikke offentliggøres oplysninger, hvor det enkelte individ kan identificeres.
6. Aftalen træder i kraft den _____

Kapitel 8. Formidling og beredskab

De foregående kapitler har gjort rede for, hvordan betydelige mængder af oplysninger indsamles fra administrative registre og organiseres i statistiske registre til brug for personstatistikken. Formålet med hele denne aktivitet er naturligvis, at brugerne af statistikken skal få ny viden om samfundet. Statistikken skal altså ud til brugerne, den skal formidles. Det er ikke noget specielt for registerbaseret statistik, det gælder alle slags statistik. Men der er alligevel nogle særlige forhold, der gør sig gældende for formidlingen, når man har adgang til så omfattende mængder grunddata, som der her er tale om.

Al statistik skal formidles...

Man må først og fremmest konstatere, at videnspotentialet er meget stort. Ved hjælp af traditionelle formidlingskanaler er det kun muligt at stille en yderst begrænset del af den relevante viden til rådighed for brugerne. Derfor er det nødvendigt at gøre intensiv brug af mere fleksible metoder, hvor brugeren selv kan gå ind og medvirke i udsøgningen og færdiggørelsen af statistikken. Her bliver løsningen elektronisk formidling af forskellig slags.

...men videnspotentialet er enormt

De enkelte brugeres specifikke behov kan umiddelbart betyde, at grundlæggende processer med udtræk og samkøring af oplysninger fra statistikregistrene ofte skal gentages, hvilket er meget uhensigtsmæssigt. Det medfører en unødigt stor ressourceindsats i forbindelse med den enkelte opgave og dermed høje omkostninger. Der er derfor brug for et beredskab, som er forberedt til at imødekomme de væsentligste ønsker med hensyn til statistikkens indhold og leveringsform.

Beredskab skal lette mulighederne

Overvejelserne om formidling og beredskab hænger nøje sammen med hensynet til datasikkerhed og diskretion, som blev omtalt i forrige kapitel. Brugerne ønsker ofte at få data stillet til rådighed, der er så detaljerede, at man ikke kan garantere anonymiteten for de personer eller virksomheder, der indgår i statistikken. Derfor er det nødvendigt at finde løsninger, der på den ene side opfylder de sikkerhedsmæssige krav og på den anden side giver brugerne relativt tilfredsstillende muligheder.

I de følgende afsnit omtales de formidlingsmåder, som for tiden anvendes af Danmarks Statistik, samt nogle perspektiver for forbedringer. Der er ikke tale om en detaljeret gennemgang af hele formidlingsaktiviteten i almindelighed, idet hovedvægten er lagt på omtale af de forhold, der er specielle for formidling af registerbaseret statistik.

8.1 Bøger

Traditionel formidling

Den klassiske og traditionelle måde at formidle statistik på er gennem bøger og andre publikationer på papir, og det gælder naturligvis også for personstatistikken. Det er et medium, som er almindeligt accepteret af og let tilgængeligt for alle brugere. Det er da også en helt fast politik, at *alle grundlæggende statistikker* skal publiceres på papir og stilles til rådighed for offentligheden for en lav pris, og sådan vil det formentlig være i mange år fremover.

Det er imidlertid umuligt at trykke mere end en brøkdel af de mangfoldige sammenstillinger af data, som kan dannes ud fra registrene, og som efterspørges af større eller mindre grupper af brugere. Selv om det havde været muligt, ville mængden blive så stor, at ingen ville kunne finde rundt i den. Derfor må de trykte medier suppleres med andre formidlingsmåder.

8.2 Ad hoc service

Specialtabeller

Det har i mange år været almindeligt, at brugere får foretaget særlige bearbejdnings af de grundoplysninger, Danmarks Statistik råder over. Det drejer sig om nye grupperinger, samkøringer, statistiske modelberegninger mv. Resultaterne skal leveres i form af statistiske tabeller på papir, diskette eller magnetbånd.

Ad hoc service er naturligvis meget fleksibel, idet alt, hvad der ikke er forbudt af hensyn til datasikkerheden, kan leveres. Problemet er imidlertid, at det kan være ganske ressourcekrævende og dermed dyrt for brugeren.

8.3 Standard-servicesystemer

Abonnement

For at undgå nogle af problemerne i forbindelse med ad hoc service kan man udnytte, at der på nogle områder findes store brugergrupper med ret ensartede behov. I sådanne situationer har man indrettet standard-servicesystemer, som kan producere de pågældende typer af statistik relativt billigt samtidig med, at brugerne har visse muligheder for at specificere, hvordan deres statistik skal se ud. Som eksempel kan nævnes de abonnementsystemer til levering af detaljeret udenrigshandelsstatistik, som tilbydes af mange landes statistikbureauer.

Fleksibel geografisk opdeling

Inden for den registerbaserede personstatistik er der navnlig grund til at nævne standardssystemer, som tilbyder forskellige statistikker for vilkårlige geografiske områder. Den offentliggjorte statistik findes som regel specificeret på kommuneniveau, men mange brugere har behov for mindre områder eller områdeinddelinger, der går på tværs af kommunegrænserne. Ved hjælp af adressekoden for bopæl eller arbejds-

sted er det muligt for brugeren at afgrænse et hvilket som helst område og modtage statistik for dette område. Det gælder både befolkningsstatistik, indkomster, boliger, beskæftigelse m.m. Brugere af disse standardsystemer er både kommuner og private virksomheder.

8.4 Databanker

Danmarks Statistik åbnede i midten af 1980'erne en on-line service, der giver brugerne direkte adgang, fra deres egen terminal eller PC, til en meget stor mængde detaljeret statistik. Denne service bygger på nogle centralt opbyggede *databanker*, som er samlinger af store, multidimensionale tabeller, som til enhver tid er fuldt opdateret med de nyeste oplysninger. Databankerne dækker næsten al den statistik, Danmarks Statistik fremstiller, men de er naturligvis begrænsede mht. sammenstillingsmuligheder. Brugere kan selv udsøge de ønskede statistikker og gruppere de anvendte variable. Endvidere er der mulighed for eksport af tabeldata til brugernes egne standardprogrammer.

On-line service

8.5 Lovmodellen

Behovet for at kunne forudsige konsekvenserne af ny eller ændret lovgivning på skatte- og socialområdet har gjort det nødvendigt at oprette Lovmodellen⁷⁹, som en helt speciel on-line service for regeringens departementer til brug for lovkonsekvensberegninger. Der er behov for detaljerede statistikker, som skal leveres på meget kort tid, når behovet opstår i Folketinget eller dets udvalg. Dette kan ikke klares ved ad hoc service.

Lovgivernes særlige behov

Lovmodellen indeholder udsnitsdata⁸⁰ fra en lang række af personstatistikens registre. Desuden findes der et analyseapparat, der er forbedret til at bearbejde oplysningerne til konsekvensberegninger. Lovmodellen omfatter en model af den gældende lovgivning på de forskellige områder.

8.6 Forskningsdatabaser

Som tidligere nævnt har forskerne i stigende grad behov for at analysere multidimensionale problemstillinger, hvor data samkøres fra mange forskellige statistikregistre og/eller fra mange tidsversioner af samme register. Det resulterer i nogle høje startomkostninger på et forskningsprojekt.

⁷⁹ Jf. Økonomiministeriet (1993).

⁸⁰ Der henvises til omtale i kapitel 7.3 af diskretionsovervejelser i forbindelse med Lovmodellens datagrundlag.

Beredskab på et højere niveau

Det har imidlertid i nogle tilfælde vist sig, at der i forskellige projekter er en fælles mængde af baggrundsplysninger, der går igen. Det drejer sig fx om arbejdsmarkedsdata og uddannelsesdata. For at udnytte dette forhold, til at nedbringe omkostningerne for det enkelte forskningsprojekt, har Danmarks Statistik oprettet forskningsdatabaserne Integreret database for arbejdsmarkedsforskning (IDA) og Fertilitetsdatabasen. Her er en stor del af arbejdet med at etablere grunddata løst én gang for alle, og forskerne kan altså starte fra dette udgangspunkt. Databaserne opdateres løbende. Omkostningerne til at etablere og dokumentere disse databaser har været høje, og Statens Samfundsvidenskabelige Forskningsråd har bidraget kraftigt til finansieringen.

Der er overvejelser om at etablere sådanne databaser inden for andre forskningsområder.

8.7 Elektroniske publikationer

Ved elektroniske publikationer forstås faste, overskuelige samlinger af statistisk information af principielt samme indhold som en bog, men på elektronisk medium. De elektroniske publikationer kan minde om databanker, men vil typisk indeholde et større element af analyse end databanker, som udelukkende består af tabeller og dokumentation af deres indhold. Publikationerne kan stilles til rådighed on-line eller på medier som diskette eller CD-ROM.

Begrænsede erfaringer...

Der er ikke de store erfaringer i Danmarks Statistik med udsendelse af sådanne publikationer. Siden 1992 har man udsendt en årlig CD-ROM⁸¹ i samarbejde med de andre nordiske landes statistikbureauer. Pladen indeholder en stor mængde aktuel statistik, der er udtrykket fra databanker. Der er ikke indlagt tekstafsnit eller andre analytiske elementer.

Også nogle af de løbende papir-publikationer udgives sideløbende på diskette, så brugerne lettere kan anvende data i deres egen PC. Heller ikke disse udgivelser indeholder tekster, og de kan i realiteten kun bruges i sammenhæng med papirudgaven.

...men fortsat udvikling

Der er behov for en videreudvikling af konceptet elektronisk publikation, så man kan kombinere det analytiske element, som kendes fra bøgerne, med tilgangen til betydeligt større mængder af statistik.

⁸¹ *Statistik uden grænser - nordisk statistik på CD-ROM.*

Kapitel 9. Dokumentation

Et væsentligt element i registersystemet er dokumentationen. I dette kapitel gives en generel status over dokumentationen af i dag samt de væsentligste principper og synspunkter omkring dokumentation af statistiksystemer. Afslutningsvis præsenteres fremtidens dokumentationssystem.

9.1 Status

På nuværende tidspunkt er den overvejende del af det personstatistiske registersystem dokumenteret med anvendelse af et system ved navn TIMES, der omfatter både en teknisk og en indholdsmæssig dokumentation. Derudover findes der dokumentation, der beskriver, hvordan de enkelte statistikregistre er blevet dannet ud fra de administrative registre.

Selvom den foreliggende dokumentation kan siges at leve op til de formelle krav, er der en række svagheder. Dokumentationen er ikke samlet for hele registersystemet, men forefindes separat for det enkelte register. Dokumentationen er ofte opdelt på enkeltversioner af registret, således at der ikke findes en oversigt over ændringer fra en version til den næste. Dokumentationen er normalt et produkt af planlægningen af registret, hvilket betyder, at der ikke findes nogen måde, hvorpå erfaringsmateriale og andre former for kvalitetsdata automatisk opsamles.

Det er planen i løbet af de kommende år at opbygge et nyt dokumentationssystem til afløsning af TIMES. Projektet, der stadig er i støbeskeen, har fået navnet TIMES-2000. Som for TIMES, vil der være tale om et system, der skal dække Danmarks Statistiks samlede dokumentationsbehov, altså ikke et system der udelukkende retter sig imod det personstatistiske registersystem.

9.2 Data i centrum

I forbindelse med udarbejdelse af edb-systemer indtager dokumentation en central rolle. Ofte har analyse- og programmeringsmetoder til udvikling af systemer været de elementer i udviklingen, der har tiltrukket sig hele opmærksomheden i metodearbejdet. I denne sammenhæng har det været antaget, at dokumentationen i stort omfang var lig output fra analysefasen, som så eventuelt kunne suppleres med driftshåndbøger og brugervejledninger. Resultatet heraf har blandt andet været, at dokumentationen af dataflow og dataprocesser har haft en tendens til at blive det centrale, mens dokumentationen af selve data er reduceret

Det nuværende dokumentationssystem...

...har en række svagheder...

...som fremtidens system forsøger at tage højde for

Dokumentation af data har været reduceret til en teknisk beskrivelse

til en teknisk beskrivelse. Med relationsdatabasens fremkomst er der dog sket en udvikling i retning af, at dokumentation af data er blevet en selvstændig disciplin.

Administrative data dokumenteres fra en administrativ synsvinkel

Når administrative registre anvendes, står man imidlertid i en anden situation. Data er ikke indsamlet med statistik som primært formål og data er dokumenteret fra en administrativ synsvinkel. I praksis betyder det ofte, at dokumentationen består af en teknisk dokumentation og en kortfattet beskrivelse af de enkelte typer af data og deres værdisæt. Denne dokumentation er som hovedregel taget som udpluk fra den samlede systemdokumentation for det administrative register. Problemet med denne dokumentationsform er, at den er systemorienteret. Formålet er at danne grundlaget for konstruktionen af et edb-system og i denne sammenhæng sikre, at de nødvendige informationer er til stede. Derimod er informationsværdien af de enkelte data kun beskrevet indirekte.

Når et administrativt register skal udnyttes som basis for statistik, vil det derfor være nødvendigt, at der laves en ny dokumentation af selve registret og det afledte statistikregister. Denne nye dokumentation må selvfølgelig tage sit udgangspunkt i den foreliggende dokumentation, men fokus ændrer sig, idet det centrale bliver dokumentation af registret og dets informationsindhold, kvalitet m.m. set fra en statistisk synsvinkel.

Hvilke krav der må stilles til denne dokumentation samt hvem der skal bruge dokumentationen, behandles i det følgende.

9.3 Hvorfor dokumentation

Anvendelse af statistikregistre kræver dokumentation

Set ud fra en datateknisk synsvinkel kan man sige, at hvis der ikke foreligger dokumentation for et edb-baseret statistikregister, kan man ikke anvende registret. Det er klart, at hvis man ikke kender navnet på det datasæt, der indeholder oplysninger fra et bestemt register, kan man ikke få fat i data. Hvis man ikke kender strukturen af de records datasættet består af, er det ikke muligt at hente information ud af data, ligesom man i den sidste ende må kende betydningen af bl.a. de koder, der anvendes, for meningsfyldt at kunne anvende data i registret.

Man kan imidlertid også sige, at dokumentationen skal foreligge, fordi det er dokumentationen, der gør det muligt at omsætte data til information. Med dette udgangspunkt lægges vægten i højere grad på at dokumentere den indholdsmæssige side af data. Som eksempel kan nævnes oplysninger om, hvorledes populationen i registret er afgrænset mht. tid og rum, en væsentlig oplysning ved fortolkningen af data.

Et vigtigt aspekt ved dokumentationen er, at registre er en fælles ressource. Mange forskellige projekter (og personer) kan trække på det samme register, og anvendelse kan foregå over et ganske langt tidsrum⁸². Kun hvis der foreligger en god og uddybende dokumentation af data, både hvad angår teknik og informationsindhold, er dette muligt. Specielt er det vigtigt at huske på, at ikke alle anvendelser af registret ud i fremtiden er kendt, dvs. dokumentationen kan ikke begrænses til at dække de øjeblikkelige behov.

Registre er en fælles ressource

9.4 Hvornår skal dokumentationen skabes

Helt generelt kan man sige, at dokumentationen skal skabes så tidligt som muligt i forløbet. Den ideelle situation er, at dokumentationen for informationsindholdet af et register er udgangspunktet for al udvikling af det edb-system, der skaber registret. I praksis er det dog som oftest en analyse af de administrative processer, som registret indgår i, der er udgangspunktet, og som derfor bestemmer registrets indhold og opbygning, samt dokumentationens form og indhold.

Dokumentationen skal skabes så hurtigt som muligt...

Ved omformningen af et administrativt register til et statistikregister, er det muligt at starte med at dokumentere informationsindholdet i statistikregistret. Hvilke spørgsmål ønsker man at kunne besvare? Hvilke tabeller og tidsserier ønsker man at kunne skabe? Hvilken information er nødvendig for at kunne gøre dette? Man tager på denne måde udgangspunkt i registrets informationsmodel, og herudfra bestemmer man den nødvendige datamodel under hensyntagen til de tilgængelige datakilder. Således kan man sikre sig det nødvendige informationsindhold samtidig med, at man har grundlaget for de transformationer af rådata, der udgør en væsentlig del af det system, som skal danne statistikregistret ud fra de administrative registre.

...med indholdet af information i statistikregistret

9.5 Hvem anvender dokumentationen

I forbindelse med udarbejdelse af dokumentation er det, som ved alle andre dokumenter, vigtigt at gøre sig klart, hvem målgruppen er, dvs. hvem der skal anvende informationen.

I forbindelse med udviklingen af det system, der producerer registret, opstår den første anvendelsesmulighed af dokumentationen. Foreløbige versioner kan anvendes til at teste, at det påtænkte register indeholder den ønskede informationsmængde. Der kan på basis af sådanne foreløbige versioner opstå en dialog mellem de personer, der planlægger registret og forskellige interessenter, fx fremtidige brugere.

Dialog mellem planlæggere og brugere

⁸² Det skal dog bemærkes, at selve adgangen til registre er begrænset af datasikkerhedsmæssige årsager.

Applikationsudviklere er en vigtig bruger

Når registret er endelig fastlagt, er dokumentationen af dette et væsentligt dokument for dem der skal udvikle de applikationer, der skal danne registret. Selv om der skal suppleres med andre dokumenter, vil det være af stor betydning, at alle hele tiden kan sammenligne med de fastlagte regler i registret.

Den primære målgruppe er statistikere, statistikbrugere og forskere

Den primære målgruppe er dog brugerne af registret, dvs. statistikere, statistikbrugere og forskere, der ønsker at skabe information ud fra data. Det primære mål med dokumentationen er at tilgodese deres behov for information om registret og dets indhold. Nogle af disse brugere vil selvfølgelig allerede være med under udviklingen af registret, mens andre først vil komme til senere. I denne sammenhæng skal man gøre sig klart, at registret faktisk er et arkivmateriale, og at dokumentationen må tilrettelægges derefter.

En ofte overset brugergruppe er dem, der anvender resultaterne af statistiske undersøgelser baseret på registre. Her bør dokumentationen af de bagvedliggende data være en tilgængelig kilde til nærmere studier af, hvad det egentlig er, der ligger til grund for en tabel eller tidsserie. Noget vil selvfølgelig blive formidlet i forbindelse med offentliggørelsen af de statistiske resultater, men i forbindelse med mere detaljerede analyser er dokumentationen af statistikregistret en vigtig kilde.

9.6 Hvad skal dokumenteres

To synsvinkler på dokumentation...

Hvad der skal dokumenteres, og indholdet af dokumentationen, må selvfølgelig ses i lyset af, hvad dokumentationen skal anvendes til. Der kan her sondres mellem to forskellige synsvinkler, en *datalogisk synsvinkel* og en *infologisk synsvinkel*.

...den datalogiske synsvinkel og...

Den datalogiske synsvinkel handler om den fysiske opbygning af registret. Er der tale om såkaldte flade filer eller en database? Hvad hedder filerne/tabellerne? Hvordan er opbygningen af records og formatet af de enkelte felter? Denne type information er nødvendig for at man kan skrive programmer, der kan håndtere registret.

...den infologiske synsvinkel

Den infologiske synsvinkel handler om informationsindholdet i registret. Hvordan er populationen og tiden afgrænset? Hvilke objekter er beskrevet? Hvilken information findes om den enkelte enhed? Hvordan er denne information repræsenteret, dvs. hvilke kodesystemer er anvendt til at beskrive de enkelte informationsdele? Hvad er kvaliteten af data, både for registret som helhed og for enkeltinformationer? Denne type af information er nødvendig ved tilrettelæggelsen af en statistisk undersøgelse med udgangspunkt i et register. Generelt kan man sige, at enhver udnyttelse af data til at skabe information skal ske ud fra den infologiske model.

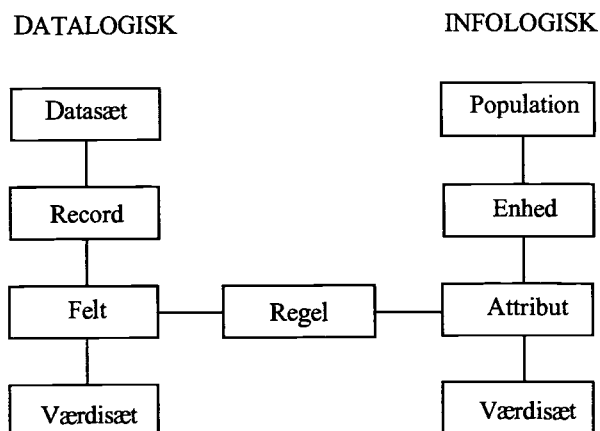
Spørgsmålet er derfor, om det er nødvendigt at lave to forskellige dokumentationssystemer, et datalogisk og et infologisk. Fordelen ved at bygge to systemer er indlysende, idet systemerne umiddelbart vil kunne tilrettelægges efter de tilstedeværende behov. Ulempen er, at begge systemer skal vedligeholdes, og at der er stor risiko for, at systemerne ikke hænger sammen; at ændringer i det ene system ikke slår igennem i det andet.

Der er imidlertid en tæt sammenhæng mellem de to systemer. Et system, der baseres på den infologiske model, må indeholde en beskrivelse af, hvordan et infologisk begreb udledes af de fysiske data. Hvis man fx har et fysisk dataelement som PNR (personnummer), vil man have de infologiske begreber KØN og ALDER pr. dato, der dannes fra det fysiske datum ved hjælp af simple regler. Tilsvarende kan man have et fysisk datum, INDKOMST (i hele kroner), og ud fra dette danne et infologisk begreb INDKOMSTGRUPPE ud fra en simpel regel. En illustration af denne tankegang gives i nedenstående figur.

Skal der laves to dokumentationssystemer ?

Sammenhængen mellem datalogisk og infologisk synsvinkel.

Figur 9.1



Figuren er et forsøg på lave et hierarki for henholdsvis det datalogiske og det infologiske syn på et register. Forbindelsen mellem de to sider er som antydnet de regler, der omformer data til information.

Det er imidlertid værd at bemærke den ensartethed, der er i opbygningen af de to synsvinkler. Først og fremmest er der værdisæt, der optræder i begge hierarkier som det laveste niveau. Ikke bare navnet vil være det samme, også selve informationen vil have samme opbygning, enten som et tilladt værdiområde (numerisk, positiv eller nul, mellem 0 og 100, etc.) eller som en værdiliste med tilhørende tekst (fx 1 = Mand, 2 = Kvinde). I de tilfælde, hvor det infologiske begreb er en

Ensartethed i opbygningen af de to synsvinkler

direkte kopi af data, fx INDKOMST (infologisk) = INDKOMST (datalogisk) vil indholdet også være det samme. Noget kunne derfor tale for at ændre modellen, således at værdisæt kun optræder én gang.

Sammenhænge i hierarkiet

Ser man på de lag, der ligger øverst i modellen, vil situationen ofte være, at der er helt banale sammenhænge mellem population og datasæt samt mellem enhed og record. Det mest almindelige er, at der er et datasæt eller en population, der indeholder en recordtype, analogt en type af enheder. Fordelen ved modellen er imidlertid, at dette ikke er påkrævet. Det vil være muligt at sammenstille informationen om enheder fra flere forskellige records i flere forskellige datasæt. Det vil endog være muligt at udlede en enkelt attribut på basis af flere felter, beliggende i forskellige records, i forskellige datasæt.

Kombination af flere statistikregistre

Det vil her være nødvendigt at komme ind på den situation, at man til en undersøgelse kombinerer data fra flere statistikregistre. Der er to mulige fremgangsmåder. Den ene er at danne et *nyt statistikregister* ved at samkøre de oprindelige registre. Den anden er at foretage en *infologisk samkørsel*, dvs. at beskrive det infologiske indhold på baggrund af de oprindelige registre. Den sidste fremgangsmåde vil, efterhånden som man flytter registre fra flade filer til databaser, blive langt mere naturlig, selvom der i teorien ikke er noget i vejen for at gøre noget sådant med den nuværende teknologi.

Udgangspunktet vil således være at danne en ny infologisk population ud fra de oprindelige, og ud fra dette skabe nye definitioner af enheder og nye attributter til disse. I praksis vil attributterne i stort omfang kunne kopieres direkte fra de oprindelige definitioner, men ofte vil også enhederne være de samme, da sammenkoblingerne ellers kan være vanskelige at foretage, ligesom populationen må være enten en delmængde af en af de oprindelige populationer eller foreningsmængden af disse.

Præcis hvordan denne problemstilling skal behandles, er ikke afklaret på nuværende tidspunkt. Noget kunne tale for at beskrive sammenhængen mellem enheder og records, hvor en sådan findes. Dette vil så betyde, at attributter til en enhed naturligt udledes af felter i den eller de records, enheden er knyttet til, og at selve sammenhængen mellem flere registre beskrives på enheds- eller recordniveau. Man kan argumentere for at udvide den datalogiske side med begrebet VIEWS⁸³, sådan som det kendes fra relationsdatabaser. Et VIEW vil arve de felter

⁸³ Med begrebet VIEW forstås en beskrivelse af et logisk datasæt. Dette datasæt kan være identisk med et fysisk datasæt eller en delmængde heraf, afgrænset ved kun at omfatte individer med bestemte egenskaber og/eller bestemte attributter. Et VIEW kan også sammenknytte data fra flere datasæt. Reglerne for sammenknytning af individer er så en del af definitionen af VIEW'et.

fra de underliggende records eller tabeller, der omfattes af VIEW'et, men samtidig kunne knyttes til enhedsbegrebet på en entydig måde.

9.7 Tidsdimensionen

Et statistikregister vil som regel blive dannet i en række versioner, der hver for sig beskriver en population på forskellige tidspunkter eller dækkende bestemte begivenheder i forskellige tidsrum. Der kan være tale om, at de enkelte versioner dannes helt uafhængigt af hinanden, eller at senere versioner dannes som opdateringer af de foregående.

I forbindelse med dokumentationen er der her særlige problemer, der må behandles, både på den datalogiske og infologiske side. Problemerne opstår, fordi det ikke er muligt at fastholde alle definitioner over et længere tidsforløb. I fx uddannelsesstatistikken vil der opstå nye uddannelser, og andre vil ophøre, hvilket må afspejle sig i værdisættet, ligesom registre vedrørende skatteforhold vil ændre sig i takt med ændrede skatteregler.

Særlige problemer ved dokumentation over tid...

Af hensyn til mulighederne for at lave tidsserier og anden statistik med sammenhæng over tid, er det af stor betydning at kunne fastholde de infologiske begreber over så lang en periode som muligt. Her hjælper den ovenfor opstillede opsplitning i en datalogisk og en infologisk del. Ofte vil et infologisk begreb kunne fastholdes på en sådan måde, at definition og værdisæt er uforandret. Det der ændres, er de regler, der anvendes til at danne informationen. Det vil ved opbygningen af et informationssystem imidlertid være vigtigt, at systemet kan håndtere denne dimension på en sådan måde, at det er muligt at fastholde begreber over tid og samtidig få overblik over, hvilke ændringer der er sket i de underliggende regler og data.

...og dokumentation af tidsseriedata

I forbindelse med dokumentation af statistikregistre kan tidsdimensionen også ansues fra en anden vinkel. Der har i litteraturen om edb-systemer ofte været fokuseret på systemernes livsforløb. Hvad der måske er overset i denne sammenhæng er, at udover systemernes livsforløb kan der opstilles et livsforløb for data, der i hvert fald fra en statistisk synsvinkel er mere interessant. Man vil ofte se, at det system, der danner et register, ændrer sig, fx via omlægning til ny teknologi, uden at dette grundlæggende ændrer på de data registret indeholder. Man vil på den anden side kunne opleve, at data ændrer sig, uden at dette fra en systemsynsvinkel er andet end en mindre justering. Der kan derfor være grund til at adskille system- og datadokumentation fra hinanden, dvs. at opfatte disse to livsforløb som uafhængige. Desuden kan der opstilles et tredje livsforløb, nemlig det infologiske. Selvom data ændrer sig, behøver dette ikke grundlæggende at ændre det infologiske syn på data, som beskrevet ovenfor.

Adskillelse af systemernes og datas livsforløb

En anden side af denne sag er, at dokumentationen af data, og specielt den infologiske dokumentation, har interesse langt udover levetiden for de systemer, der danner data, ja muligvis udover levetiden for dataene. Denne dokumentation er også en del af dokumentationen for de resultater der formidles, fx i form af tidsserier og i databanker.

9.8 Datakvalitet og informationskvalitet

Dokumentation af datakvalitet er vigtig

Et vigtigt aspekt ved dokumentationen af statistikregistre er at give information om kvaliteten af de data registret indeholder. Kvaliteten af data kan omhandle registret som helhed. Fx kan der være problemer med samtidighed, dvs. om data i registret er fuldt opdateret i forhold til den periode eller dato registret omhandler, eller der kan være problemer med fuldstændighed. Desuden bør informationer om registrets administrative anvendelse medtages her, da denne information er væsentlig ved vurderinger af, om bestemte problemstillinger med rimelighed kan belyses gennem anvendelse af registret. Disse ting vedrører normalt registret på tværs af registerversioner. Der bør derudover findes en log (en slags dagbog) over de enkelte registerversioners tilblivelse, hvor fx problemer med indflydelse på datakvaliteten registreres. Loggen bør indeholde information om alle registerversioner, dvs. også eksplicit indeholde information om, at der ikke er noget at bemærke.

Kvalitetsdata vil derudover typisk findes på felt- eller attributniveau. Her vil der normalt være tale om tidsuafhængige bemærkninger, fx at kvaliteten af data er lav, fordi data ikke har betydning for den administrative proces, eller at data er af svingende kvalitet, fordi data indberettes decentralt og instrukserne til indsamlerne ikke er tilstrækkelig klare eller ikke efterleves. Hvis der har været anvendt særlige kilder, fx stikprøveundersøgelser, til at vurdere datakvaliteten, er det endvidere vigtigt at henvise direkte til disse og de opnåede resultater.

9.9 Registrets anvendelse

Den infologiske synsvinkel er afgørende ifm. registerdata

Det er allerede nævnt, at registerdokumentation er en vigtig informationskilde, hvis man vil gå tættere på de resultater, der publiceres gennem anvendelse af registerdata. I denne forbindelse spiller det infologiske syn på registret den afgørende rolle. Alle tabeller og analyser af registret må foretages ud fra den infologiske information. Denne information bør dække alle de begreber, grupperinger m.m., der faktisk anvendes i registret. Det er vigtigt at fastholde disse krav, fordi de sikrer, at resultater fra forskellige analyser, baseret på registret, anvender de samme definitioner af forskellige begreber. En konsekvens er, at den infologiske information om registret udvides, efterhånden som registret anvendes, idet nye infologiske begreber kan tilføjes.

I forbindelse med registrets anvendelse bør der også føres log over denne, således at der i registerdokumentationen er en henvisning til databanker, publicerede tabeller mfl., der dannes på basis af registret, med information om registerversionen og de attributter, der er anvendt. Denne dokumentation er knyttet til den infologiske model. I det omfang registret, sammen med andre registre, er anvendt i andre infologiske modeller, dvs. undersøgelser der bygger på mere end ét register, bør der også være henvisninger til disse.

**Alle anvendelser
bør dokumenteres**

9.10 Specielt om det personstatistiske registersystem.

I det foregående har der været tale om en generel diskussion af dokumentationsproblematikken for statistiksystemer. I forbindelse med det personstatistiske registersystem er der imidlertid en række særlige forhold, som er værd at knytte nogle bemærkninger til.

Først og fremmest er det værd at hæfte sig ved, at systemet med sin opbygning af basisregistre og klassifikationsmoduler kan og bør betragtes både fra den infologiske og den datalogiske synsvinkel, idet den infologiske synsvinkel typisk vil inddrage data fra et eller flere klassifikationsmoduler sammen med et enkelt basisregister. Da både basisregistre og klassifikationsmoduler bygger på personnummeret som nøgle, ligger muligheden for en infologisk sammenkobling indenfor rækkevidde. Dette gælder iøvrigt også en eventuel sammenkobling mellem flere basisregistre.

**Systemet skal
betragtes både ud
fra den infologiske
og den datalogiske
synsvinkel**

Desuden er det karakteristisk for systemet, at der i stort omfang vil være 1 til 1 relationer mellem datasæt og record på den ene side og population og enhed på den anden side, idet det er op til de systemer, der danner statistikregistrene udfra de administrative data, at skabe disse 1 til 1 relationer.

Det er vigtigt, at dokumentationssystemet kan holde styr på forandringer i definitioner over tid, specielt den type, hvor den infologiske begreb ikke ændrer sig, men hvor de bagved liggende data har ændret sig. En særlig variant vil være, at der ikke er en ændring på det datalogiske niveau, men at den faktiske ændring ligger i det system, der danner statistikregistret. Det vil være vigtigt at dokumentere sådanne ændringer, da de kan have betydning for tolkningen af samme data over tid.

**Systemet skal kunne
håndtere ændringer
over tid**

9.11 TIMES-2000

Som nævnt indledningsvis arbejdes der for øjeblikket på opbygningen af et nyt dokumentationssystem, TIMES-2000, der er et eksempel på et metadatasystem, dvs. et system med data om data. Et problem med alle metadatasystemer synes at være, hvorledes man får data ind i sy-

**TIMES-2000 - et
metadatasystem**

stemet og sikrer at data er vedligeholdt. Det synes specielt at være et problem, hvis de der leverer metadata ikke er de samme, som dem der anvender metadata.

I forbindelse med TIMES-2000 vil dette problem blive forsøgt behandlet på forskellige måder. Dataindholdet kan opdeles i to klasser; *hårde data* og *bløde data*. Hårde data er recordlayouts, værdisæt på listeform og regler der er skrevet i en særlig syntaks. Karakteristisk for disse data er, at de kan behandles og anvendes maskinelt. Bløde data er beskrivelser, kvalitetsdata, bemærkninger om databrud m.m. Disse data kan opbevares maskinelt, men kan ikke gøres til genstand for en maskinel analyse.

Metadata bør kunne anvendes direkte i diverse edb-processer

De hidtidige erfaringer med TIMES og de tilhørende systemer for tabelberedskab tyder på, at et vigtigt incitament for leverandørerne vil være, at data umiddelbart kan genbruges i forbindelse med edb-opgaver, dvs. at metadata direkte kan anvendes i forbindelse med tabellering, listning og andre processer. Det vil da også være et væsentlig mål for TIMES-2000 at skabe sådanne muligheder, dels gennem forbehandling til softwarepakker som fx SAS, dels gennem at Danmarks Statistiks egne standardssystemer direkte skal kunne trække på metadata fra TIMES-2000. Målet må være, at det er muligt at bestille tabeller mv. ud fra de infologiske begreber, som metadatabasen indeholder.

Mulighederne er forskellige for hhv. hårde og bløde data

Desværre dækker ovenstående kun de hårde data. Hvad angår de bløde data vil det i praksis blive vanskeligt at få samlet alle data op. De hårde data udgør imidlertid et godt skelet for systemet. Den overordnede struktur med datasæt, record, felt, population, enhed, attribut og værdisæt skal være på plads, for at de hårde data kan beskrives.

Alle bløde data i systemet er karakteriseret ved, at de knytter an til denne struktur, idet de handler om eller beskriver et bestemt element i strukturen, hvad enten det er et datasæt eller en attribut. Det er tanken, at de bløde data i systemet skal ligge som notatfelter, knyttet til det enkelte element, idet der til hvert element kan være et vilkårligt antal notater, evt. opdelt i forskellige klasser af notater.

Mange bløde data vil på et eller andet tidspunkt opstå i form af tekst i et notat eller tekstbehandlingsdokument. Der må skabes en mulighed for, fra tekstbehandlingsystemet, at kunne overføre et dokument eller dele heraf direkte til TIMES-2000. Man kan sige, at TIMES-2000 i denne sammenhæng skal være det naturlige arkivsystem eller journal-system for sådanne noter.

Datadokumentation skal prioriteres højt

Det er imidlertid lige så vigtigt, at der aktivt sættes en kampagne ind for, at sådanne data skal skabes, og at en sådan kampagne følges op. Det vil kræve, at man på alle niveauer i Danmarks Statistik betragter datadokumentation, som en vigtig aktivitet.

Bilag 1. Litteraturoversigt

- Arbejdsministeriet (1977): *Bekendtgørelse om virksomhedsbegreb og om antal lønmodtagere i forbindelse med foretagelse af afskedigelser af større omfang*. Arbejdsministeriets bekendtgørelse nr. 74 af 4. marts 1977.
- Arbejdsministeriet (1993): *Arbejdsmarkedspolitisk Årbog 1993*.
- Australian Bureau of Statistics (1993): *Dataediting*. Internt papir.
- Baldwin, J., Dupuy, R. & Penner, W. (1992): *Development of longitudinal panel data from business registers: Canadian experience*. Statistical Journal of the United Nations vol. 9.
- Betænkning afgivet af *folkeregisterkommissionen af 1920*, København 1922.
- Betænkning nr. 767. København 1976. *Delbetænkning om offentlige registre*.
- Betænkning om *folkeregistrenes medvirkende ved indførelse af elektronisk databehandling i den offentlige forvaltning mv.*, København 1963.
- Boligministeriet (1990): *Lov om bygnings- og boligregistrering*. Boligministeriets lovbekendtgørelse nr. 132 af 28. februar 1990.
- Danmarks Statistik (1979a): *Dødelighed og erhverv 1970-75*. Statistiske Undersøgelser nr. 37, København.
- Danmarks Statistik (1979b): *Beretning 1978 og arbejdsplan frem til 1980*.
- Danmarks Statistik (1982): *Personstatistik på registergrundlag*.
- Danmarks Statistik (1989): *Data quality and telephone interviews*.
- Danmarks Statistik (1990): *OPUS, Om Planlægning og Udvikling af Statistiksystermer*. Planlægningshåndbog - opdateres løbende.
- Danmarks Statistik (1991a): *Om Danmarks Statistiks målsætning*.
- Danmarks Statistik (1991b): *Anvendelse af oplysningssedlerne i IDA*.
- Danmarks Statistik (1991c): *Datasikkerhedsreglement for Danmarks Statistik*.
- Danmarks Statistik (1991d): *IDA - en integreret database for arbejdsmarkedsforskning. Hovedrapport*.
- Danmarks Statistik (1991e): *Om Danmarks Statistiks målsætning*. Ikke publiceret.
- Danmarks Statistik (1991f): *Service fra Danmarks Statistik - Danskerens liv og færden i tal de kan bruge*.
- Danmarks Statistik (1992a): *Arbejdsstyrkeundersøgelsen 1991*. Statistiske efterretninger, Arbejdsmarked, 1992:20.

- Danmarks Statistik (1992c): *Dansk Branchekode 1993*, 1. udgave.
- Danmarks Statistik (1992d): *Forbrugsundersøgelsen 1987. De private husstandes indkomster, opsparing og formue*. Indkomst, forbrug og priser, Statistiske efterretninger, 1992:17.
- Danmarks Statistik (1993a): *Registerbaseret arbejdsstyrkestatistik, ultimo november 1991*. Statistiske efterretninger, Arbejdsmarked, 1993:17.
- Danmarks Statistik (1993b): *Fertilitetsudviklingen i Danmark i 1980'erne (1993)*. Statistiske Efterretninger, Befolkning og valg, 1993:12.
- Danmarks Statistik (1994a): *Det personstatistiske registersystem, seminar, 10.-13. januar 1994*.
- Danmarks Statistik (1994b): *Sygehusbenyttelsesregistret - oplysninger om befolkningens brug af sygehuse*.
- Danmarks Statistik (1994c): *Sammenhængende socialstatistik 1992*. Statistiske efterretninger, Social sikring og retsvæsen, 1994:15.
- Det økonomiske Råd (1992): *Dansk Økonomi*, november 1992.
- DIKE (1982): *Registre indenfor sundhedsområdet*. Rapport udgivet af DIKE, København november 1982.
- Ferguson, Dania P. (1993): *Reviews of Methods and Software used in Dataediting*. US Dept. of Agriculture.
- Finansministeriet (1992): *Vejen til fuld beskæftigelse*. Analyser.
- Finansministeriet, Administrations- og Personaledepartementet (1993): *Oversigt over forskrifter for fælleskommunale registre*.
- Hamermesh, D.S. (1990): *A data user's look back from 2015*. Monthly Labor Review. April 1990.
- Harala, R. (1993): *Evaluation of the Results of the Register Based Population and Housing Census 1990 in Finland*. Bulletin of the ISI, Contributed Papers, 49th Session, Firenze 1993.
- ILO (1982): *Guidelines on measuring employment and unemployment. (Extract from Resolution I of the 13th International Conference of Labour Statisticians, Geneva, October 1982)*.
- ILO (1990): *International Standard Classification of Occupations, ISCO-88*.
- Indenrigsministeriet (1978): *Lov om folkeregistrering*. Indenrigsministeriets lovbekendtgørelse nr. 508 af 14. september 1978.
- Indenrigsministeriet (1991): *CPR, Danmarks folkeregister*. Notat af Henrik Nielsen.
- Indenrigsministeriet, CPR-kontoret (1993): *CPR-systemet i 1993*. Internt papir.
- Jensen, P. & Thygesen, L. (1985): *Linkage of Records on Objects of Different Kinds. Methodological Problems and Practical Experiences*.

- ce. Bulletin of the International Statistical Institute, Vol. LI, Book 2. Amsterdam.
- Justitsministeriet (1992a): *Lov om offentlige myndigheders registre*. Justitsministeriets lovbekendtgørelse nr. 654 af 20. september 1991.
- Justitsministeriet (1992b): *Lov om private registre mv.* Justitsministeriets lovbekendtgørelse nr. 622 af 2. oktober 1987.
- Knudsen, Lisbeth B. (1993a): *Fertility Trends in Denmark in the 1980s. A Register Based Socio-demographic Analysis of Fertility Trends*. Danmarks Statistik.
- Knudsen, Lisbeth B. (1993b): *Paritetsoplysningen i Sundhedsstyrelsens medicinske fødselsregister*. Validering ved hjælp af ny fertilitetsdatabase i Danmarks Statistik. Ugeskrift for Læger 1993;155:2525-9.
- Korpi, H. (1989): *Main Type of Activity and Occupational Status in the 1985 Census: Register-based Parallel Data*. Statistikcentralen, Studies no. 152. Helsinki.
- Lange, Anita (1993): *Hvordan får vi registerstatistikken anerkendt i Europa?* 19:e Nordiska Statistikermødet, Reykjavik 1992, Nordisk statistisk Skriftsserie nr. 61, Reykjavik 1993.
- Leonard, J.S. (1987): *In The Wrong Place at the Wrong Time: The Extent of Frictional and Structural Unemployment*. I Lang, K. & Leonard, J.S. (Eds). *Unemployment and the Structure of Labor Markets*. New York: Basil Blackwell Inc.
- Leth-Sørensen, S. (1993): *Jobmobilitet belyst ved hjælp af IDA-databasen*. Dansk Sociologi 1993; 4, no.1.
- Lindstrøm H.L. (1992): *Datakvalitet ved olika insamlingsmetoder*. Papir fremlagt ved det 19. nordiske statistikermøde, Reykjavik, 22.-25. juli 1992.
- Lynge, E. & Thygesen, L. (1990): *Occupational Cancer in Denmark*. Scandinavian Journal of Work, Environment and Health. Vol. 16, suppl. 2. Helsinki.
- Ministeriet for Erhvervspolitisk Samordning (1993): *Erhvervsredegørelse*.
- Mogensen, G.V. (ed.) (1990): *Time and consumption*. Danmarks Statistik, København.
- National Center for Health Statistics (1980): *The Person-Number Systems of Sweden, Norway, Denmark, and Israel*. Data Evaluation and Methods Research, Series 2, No. 84. Hyattsville, Maryland.
- Nielsen, P. (1987): *"Markedsdynamik og arbejdsformidling"*. Bilag 2 og 3, ATA-projektet, rapport nr. 7.
- Norbotten, S. (1967): *Om arkivstatistiske systemer*. Bilag 4 i Statistical Reports of the Nordic Countries, vol. 14. København 1967.

- Nordisk Statistisk Sekretariat (1981a): *Teknisk rapport nr. 26, Rapport från Nordiska Seminariet om Bortfallsproblem i statistikproduktionen*. Gausdal, Norge, 17 - 20 Februari. København 1981.
- Nordisk Statistisk Sekretariat (1981b): *The Meeting of the Chief Statisticians of the Nordic Countries in Copenhagen 1979*. København 1981.
- OECD (1987): *Employment outlook September 1987*.
- OECD (1987b): *The Process of Job Creation and Job Destruction*.
- OECD (1993): *Employment/Unemployment Study*. Interim Report by the Secretary-General.
- Redfern, P. (1987): *A Study of the Future of the Census of Population: Alternative Approaches*. Eurostat Theme 3 Ser. C. Luxembourg.
- Registertilsynet (1987): *Fortegnelse over offentlige myndigheders registre 1987*.
- Registertilsynet (1993): *Intern registerfortegnelse*.
- Rodgers, Willard L. (1984): *An Evaluation of Statistical Matching*. Journal of Business & Economic Statistics, Vol 2, no 1, pp 91-102.
- Rohwer, G. & Leth-Sørensen, S. (1993): *Upward Mobility in the Danish Labour Market*. I Jesper Lund (red.): Symposium i anvendt statistik.
- Sandgren, P. (1993): *Kvaliteten i statistik på administrativt indsamlede data*. Statistiska Centralbyrå. Stockholm (ikke publiceret).
- Singh, A. C., Mantal, H. J., Kinack, M. D. & Rowe, G. (1993): *Statistical Matching: Use of Auxiliary Information as an Alternative to the Conditional Independence Assumption*. Survey Methodology, vol 19, no 1, pp. 59-79. Statistics Canada.
- Smidt, E.I., Körmendi, E., Mogensen, G.V. & Vibe-Pedersen, J.: *24 timer i døgnet*. Systime, København.
- Thygesen, L. (1989): *Prospects for Socio-Demographic Risk Research in Denmark*. Statistical Journal of the UN ECE, vol. 6, No 1. Geneva.
- Udvalget vedrørende Helsetjenesteforskning og Medicinsk Teknologivurdering (1986): *Helsetjenesteforskning og registre*. Rapport fra Udvalget vedrørende Helsetjenesteforskning og Medicinsk Teknologivurdering. København marts 1986.
- US Department of Commerce (1980): *Report on Statistical uses of Administrative Records*. Statistical Working Paper No. 6. Washington DC.
- Van Bastelaer, A. (1993): *Differences in the designs of the labour force survey in the European Community and some consequences*. Netherlands Central Bureau of Statistics, Heerlen. Unpublished.

Vejrup-Hansen, P. (1993): *Virksomhedsdemografi: Overlevelse og vækst i nye virksomheder*. Samfundsøkonomen marts 1993.

Økonomiministeriet (1992a): *Lov om Danmarks Statistik*. Økonomiministeriets lovbekendtgørelse nr. 1189 af 21. december 1992.

Økonomiministeriet (1992b): *Lov om Det Centrale Erhvervsregister*. Økonomiministeriets lovbekendtgørelse nr. 530 af 19. juni 1992.

Økonomiministeriet (1993): *Lovmodel*.

Bilag 2. Registre i Danmarks Statistik

Dette bilag er en oversigt over aktive statistikregistre i Danmarks Statistik pr. 1. september 1994. For hvert register er anført navn samt i oversigtsmæssig form *formål, indhold, kilder* og *anvendelse*. En nærmere omtale af det enkelte register fremgår af registerforskrifterne for de pågældende registre.

Statistikregistrene er rubriceret efter hovedanvendelse i flg. grupper:

1. Personstatistik

- 1.1. Befolkning
- 1.2. Arbejdsmarked
- 1.3. Løn, indkomst, formue og forbrug
- 1.4. Uddannelse
- 1.5. Social sikring og retsvæsen
- 1.6. Sundhed
- 1.7. Andet

2. Erhvervsstatistik

3. Bygnings- og boligstatistik

1. Personstatistik

1.1. Befolkning

Befolkningsstatistikregistret

Formål	Registret danner grundlag for befolkningsstatistikken for Danmark, som omfatter opgørelser af den samlede befolkning på bestemte tidspunkter samt af befolkningsbevægelser. Registret skal i øvrigt kunne anvendes til at danne baggrundsoplysninger, der kan anvendes i opgørelser inden for andre personstatistikområder.
Indhold	Registret indeholder oplysninger om personer, der har eller har haft bopæl i Danmark. Der indgår følgende oplysninger: <p style="text-align: center;"><i>personnummer, bopælsadresse, familiehenvi­sing</i></p> <p style="text-align: center;"><i>ægteskabelig stilling, arbejdsstilling, folkekirketil­hørsforhold, fødselsregistreringssted, statsborgerskab, indvandrings- eller udvandringsland</i></p> <p style="text-align: center;"><i>fødsel, indvandring, dødsfald, udvandring, vielse, skilsmisse, adoption, flytning, tidspunkt for hændelser</i></p>
Kilde	Registret opdateres på grundlag af oplysninger fra Det centrale personregister (CPR).
Anvendelse	Den løbende befolkningsstatistik Dannelse af familier og husstande Geografisk områdekodning Befolkningsstatistiske special- og serviceopgørelser Stikprøveudvælgelse <p><i>I andre statistikregistre:</i></p> <p>Oplysninger fra dette register indgår i alle andre personstatistikregistre samt i Bygnings- og boligstatistikregistret.</p>
	Flytteundersøgelsesregistret
Formål	Flytteundersøgelsesregistret har til formål at danne grundlag for deta­iltabeller til analyse af flyttestrømme over en flerårig periode.

Indhold Registret indeholder oplysninger om personer, der i givne år er flyttet mellem to kommuner, er indvandret eller er udvandret. Der indgår følgende oplysninger:

personnummer, bopælsadresse, familiehenvi­snin­g

familieoplysninger, flytning, statsborgerskab, indkomst, boligforhold, beskæftigelsesforhold, ledighed, igangværende uddannelse, afsluttede uddannelser, ansættelsesstedets beliggenhed, branche

Kilde Danmarks Statistiks Befolkningsstatistikregister
 Danmarks Statistiks Indkomststatistikregister
 Danmarks Statistiks Bygnings- og boligstatistikregister
 Danmarks Statistiks Arbejdsklassifikationsmodul
 Danmarks Statistiks Uddannelsesstatistikregister
 Danmarks Statistiks Arbejdspladsstatistikregister

Anvendelse Registret anvendes til fremstilling af detaljerede tabeller i maskinlæsbar form. Disse tabeller stilles til rådighed for København Kommune for videre analyser.

Folke- og bolig­­tællingsregistret

Formål Registret danner grundlag for generel og specialiseret statistik over befolkningens størrelse og sammensætning med hensyn til demografiske kendetegn, beskæftigelses- og erhvervsforhold, uddannelsesforhold, familieforhold, boligforhold og pendling på opgørelsestidspunkterne 9. november 1970 og 1. januar 1981.

Indhold Registret indeholder oplysninger om personer, der har bopæl i Danmark på tællingstidspunktet. Der indgår følgende oplysninger:

personnummer, bopælsadresse (herunder tilflytningsdato til kommune)

ægteskabelig stilling (herunder evt. vielsesdato og vielsesmyndighed), stillingsangivelse/-betegnelse, familiehenvi­snin­ger, folkekirkeforhold, fødested, statsborgerskab

arbejdsstedets adresse, beskæftigelsesforhold, erhvervsforhold, indkomstforhold, afsluttet uddannelse, evt. igangværende uddannelse

adresseidentifikation, oplysninger om fx boligens art, alder, ejerforhold, størrelse, installationsforhold, opvarmning, udlejningsforhold, husleje

Kilder	Danmarks Statistiks Befolkningsstatistikregister, 1981 Danmarks Statistiks Arbejdsklassifikationsmodul (AKM), 1980 Danmarks Statistiks Uddannelsesklassifikationsmodul (UKM), 1980 Danmarks Statistiks Bygnings- og boligstatistikregister, 1981 Danmarks Statistiks Arbejdspladsstatistikregister, 1980 Spørgeskemaer udsendt til borgerne, 1970
Anvendelse	Opgørelse af befolkningens størrelse og sammensætning og boligbestandens størrelse og sammensætning samt befolkningens boligforhold. Baggrundsuplysninger til anvendelse i opgørelser inden for andre personstatistikområder.

I andre statistikregistre:

Erhvervsdødelighedsstatistikregistret
Uddannelsesklassifikationsmodulet

Medicinsk fødsels- og dødsfaldsstatistikregister

Formål	Registret har til formål at danne grundlag for statistik angående fødsler, herunder dødfødsler, og dødsfald. I denne statistik indgår fordeling efter demografiske og medicinske kendetegn gældende på tidspunktet for hændelsen.
Indhold	Registret indeholder oplysninger om fødte og mødre til fødte, der på fødselstidspunktet har bopæl i Danmark, om fædre til fødte uanset bopælen og om døde personer, der på dødsdatoen har bopæl i Danmark. Der indgår følgende oplysninger om fødte og forældre til fødte samt om afdøde:

personnummer, bopælsadresse

ægteskabelig stilling, fødselsregistreringssted, statsborgerskab, hændelsesdato, konstaterende myndighed, arbejdsstilling, erhverv, indkomst

fødselsdato og -klokkeslet, oplysning om enkeltfødsel/flerfødsel, moderens tidligere graviditeter, barnets vægt og længde

Dødsdato (samt for døde under 28 dage døds-klokkeslet), hændelsessted (fx. hjemmet), dødsårsag (primær/sekundær/tertiær, ICD-klassifikation), ulykke (arbejdsulykke eller anden ulykke), døds-måde (fx. naturlig død, selvmord), operation (i forbindelse med dødsfald)

Kilder Det centrale personregister (CPR)
Fødselsanmeldelser
Dødsattester
Danmarks Statistiks Arbejdsklassifikationsmodul (AKM)

Anvendelse Løbende befolkningsstatistik
Special- og serviceopgørelser

I andre statistikregistre:

Erhvervs/dødelighedsstatistikregistret
Færdselsholdsstatistikregistret
Sygehusbenyttelsesstatistikregistret

Statistikregistret for fertilitetsforskning

Formål Formålet med registret er at danne grundlag for personorienterede statistiske undersøgelser af fertilitetsforhold.

Indhold Registret indeholder oplysninger om kvinder født efter 1929 og mænd født efter 1915, som har, eller har haft bopæl i Danmark. Der indgår følgende oplysninger:

barns personnummer, fars personnummer, mors personnummer, personnummer ægtefælle, personnummer tidl. ægtefælle, nuværende og tidligere bopælsadresse, familiehenvisning

fødselsregistreringssted, nuværende og tidligere ægteskabelig stilling, statsborgerskab, familieoplysninger, ægteskabsvarighed, køn, enkeltfødt, flerfødt, moderens antal tidligere levendefødte/dødfødte, fødselsvægt, indkomst efter størrelse og art, beskæftigelsesforhold, ledighedsgrad,, evt. igangværende uddannelse, seneste afsluttede almenuddannelse - art og afslutningstidspunkt, højeste fuldførte erhvervsuddannelse - art og afslutningstidspunkt, boligens art, størrelse og udstyr, antal tidligere spontane og provokerede aborter, svangerskabslængde

Kilder Sundhedsstyrelsens medicinske fødselsregister
Danmarks Statistiks Befolkningsstatistikregister
Danmarks Statistiks Medicinske fødsels- og dødsfaldsstatistikregister
Danmarks Statistiks Indkomststatistikregister
Danmarks Statistiks Arbejdsklassifikationsmodul
Danmarks Statistiks Uddannelsesklassifikationsmodul
Danmarks Statistiks Uddannelsesstatistikregister
Danmarks Statistiks Statistikregister vedr. indkomsterstattendende ydelser
Danmarks Statistiks Bygnings- og boligstatistikregister

Anvendelse Registret anvendes som grundlag for analyser af fertilitetsforholdene i Danmark.

1.2. Arbejdsmarked

Arbejdsklassifikationsmodulet (AKM)

Formål Registret har til formål at give statistiske baggrundsoplysninger om personers beskæftigelsesforhold mv. til brug for de grene af den danske personstatistik, hvor sådanne baggrundsoplysninger er relevante.

Indhold Registret indeholder oplysninger om personer, som har eller har haft bopæl i Danmark, samt om personer i øvrigt, hvem det påhviler at svare indkomst- eller formueskat til staten. Der indgår følgende oplysninger:

personnummer, arbejdsstedets SE-nummer,

stillingsforkortelse, ATP-bidrag, arbejdsløshedskasse, arbejdsløshedsforsikringskategori, årsledighedsgrad, efterlønsmarkering, igangværende uddannelse, beskæftigelsesstatus (fx lønmodtager), indehavermarkering, stillingskoder, heltid/deltid

firmabranche, ejerforhold, arbejdsstedsbranche, arbejdsstedets størrelse

Endvidere indgår standardklassifikationer dannet ved kombination af ovenstående data.

Kilder Det centrale personregister (CPR)
Told- og Skattestyrelsens Centrale slutligningsregister (CSR-L)
Told- og Skattestyrelsens Centrale oplysningssedelregister (COR)
Arbejdsmarkedsstyrelsens Centrale register for arbejdsmarkedsstatistik (CRAM)
Danmarks Statistiks Indkomststatistikregister
Danmarks Statistiks Arbejdsløshedsstatistikregister
Danmarks Statistiks Uddannelsesstatistikregister
Danmarks Statistiks Løn- og personalestatistikregister for den offentlige sektor
Danmarks Statistiks Lønstatistikregister for den private sektor
Danmarks Statistiks Beskæftigelsesstatistikregister
Danmarks Statistiks Arbejdspladsstatistikregister
Danmarks Statistiks Erhvervsstatistiske registersystem

Anvendelse Special- og serviceopgaver efter samkøring med andre oplysninger

I andre statistikregistre:

Folke- og boligtællingsregistret
 Indkomststatistikregistret
 Beskæftigelsesstatistikregistret
 Medicinsk fødsels- og dødsfaldsstatistikregistret
 Kriminalstatistikregistret
 Flytteundersøgelserregistret
 Erhvervs/indlæggelsesregistret
 Velstandsstatistikregistret
 Statistikregistret for arbejdsmarkedsforskning
 Statistikregistret vedrørende tidsanvendelsen 1950-1988
 Forebyggelsesregistret

Arbejdsløhedsstatistikregistret

Formål	Registret har til formål at give statistiske oplysninger til belysning af ledighedens struktur og udvikling. De statistiske opgørelser skal give oplysning om ledigheden set i relation til såvel arbejdsmarkedsmaessige som sociale forhold.
Indhold	Registret indeholder oplysninger om personer, der er eller har været medlem af en statsanerkendt arbejdsløhedskasse samt andre personer, der er eller har været tilmeldt et offentligt arbejdsformidlingskontor som arbejdssøgende. Der indgår følgende oplysninger: <p style="margin-left: 40px;"><i>personnummer</i></p> <p style="margin-left: 40px;"><i>bopælsadresse, ægteskabelig stilling, statsborgerskab</i></p> <p style="margin-left: 40px;"><i>arbejdsløhedskasse, forsikringskategori, antal forsikrede timer, førstegangsarbejdssøgende, fag, erhverv, uddannelse</i></p> <p style="margin-left: 40px;"><i>ugevise oplysninger for seneste 80 uger om antal arbejdstimer og ikke-dagpengeberettigende syge- og ferietimer mv., ugevise oplysninger for seneste 80 uger om antal ledige timer, ugevise oplysninger for seneste 80 uger om ledighedens årsag, herunder om vejrligsledighed, afledte medlems- og ledighedsoplysninger, herunder om ledighedsgrad, ledighedsperiodens varighed m.m.</i></p>
Kilder	Arbejdsmarkedsstyrelsens Centrale register for arbejdsmarkedsstatistik (CRAM) Danmarks Statistiks Befolkningsstatistikregister
Anvendelse	Løbende arbejdsløhedsstatistik Arbejdsløhedsstatistiske special- og serviceopgørelser

I andre statistikregistre:

Beskæftigelsesstatistikregistret
 Arbejdsklassifikationsmodulet
 Udsnitsarkivet
 Arbejdsstyrkeundersøgelserregistret
 Statistikregistret vedrørende indkomsterstøttende ydelser
 Statistikregistret for arbejdsmarkedsforskning
 Indkomstforløbsregistret
 Statistikregistret vedrørende tilpasningsstrategien mellem arbejdsliv og familieliv
 Statistikregistret vedrørende erhvervsprogligt uddannede
 Statistikregistret vedrørende enkeltfagskursister
 Sygedagpengestatistikregistret
 Ressourceundersøgelserregistret
 Regionalt forløbsregister

Arbejdspladsstatistikregistret**Formål**

Registret har til formål at give statistiske oplysninger om befolkningens beskæftigelsesforhold og geografiske fordeling efter arbejdssted og om de lokale erhvervsenheders fordeling efter placering, branche og personelforhold.

Registret skal i øvrigt kunne anvendes til at danne baggrundsoplysninger om personernes beskæftigelsesforhold og arbejdsstedsplacering, der kan anvendes i opgørelser inden for andre personstatistikområder.

Indhold

Registret indeholder oplysninger om personer, der er eller har været beskæftiget som lønmodtagere i Danmark. Der indgår følgende oplysninger om ansættelsesforhold:

personnummer, SE-nummer, arbejdsstedskode, beliggenhedsadresse

beskæftigelsesperiode i året, ATP-bidrag, løn, arbejdsstedets branche

Kilder

Told- og Skattestyrelsens Centrale oplysningsseddelregister (COR)
 Danmarks Statistiks Løn- og personalestatistikregister for den offentlige sektor
 Danmarks Statistiks Erhvervsstatistiske registersystem

Anvendelse

Service- og specialopgørelser

I andre statistikregistre:

Folke- og boligtællingsregistret
 Beskæftigelsesstatistikregistret
 Arbejdsklassifikationsmodulet
 Flytteundersøgelserregistret
 Statistikregistret for arbejdsmarkedsforskning
 Det erhvervsstatistiske registersystem

Arbejdsstyrkeundersøgelserregistret**Formål**

Registret har til formål at give statistiske oplysninger om arbejdsstyrkens størrelse og sammensætning samt at belyse beskæftigelses- og ledighedsforhold. I tilknytning hertil skal det belyse særlige forhold, der vedrører deltagelse på arbejdsmarkedet, ud fra retningslinier fastlagt i det harmoniserede program for arbejdskraftopgørelser inden for De europæiske Fællesskaber. Endvidere skal det optage oplysninger til brug for andre statistiske undersøgelser indsamlet i forbindelse med oplysninger til de nævnte formål.

Indhold

Registret indeholder oplysninger om et repræsentativt udsnit blandt personer, der har bopæl i Danmark. Der indgår følgende oplysninger:

personnummer

beskæftigelsesforhold i tællingsperioden, erhvervstilhørsforhold, virksomhedsstørrelse, beskæftigelses- og erhvervstilhørsforhold et år før tællingsperioden, ledighedsforhold, uddannelsesforhold, overgang fra uddannelse til erhvervsarbejde, gennemførte efteruddannelseskurser i foregående år, transportforhold mellem hjem og arbejdssted, skifteholdsarbejde, natarbejde

familieforhold, indvandringsoplysninger, ægteskabelig stilling, statsborgerskab, ledighedsgrad, ledighedsårsag, arbejdsløsheds-kasse, forsikringskategori, alder, ægteskabelig stilling, uddannelsens art og tidspunkt for afslutning af uddannelsen, arbejdsstilling

*Oplysninger til specialanalyser på servicebasis***Kilder**

Spørgeskema- og interviewoplysninger
 Danmarks Statistiks Befolkningsstatistikregister
 Danmarks Statistiks Arbejdsløshedsstatistikregister
 Danmarks Statistiks Uddannelsesklassifikationsmodul
 Danmarks Statistiks Arbejdsklassifikationsmodulet

Anvendelse

Kvartårlige og andre arbejdsstyrkestatistiske opgørelser
 Arbejdsstyrkestatistik til EU

Arbejdsmarkedsstatistiske special- og serviceopgørelser

Beskæftigelsesstatistikregistret

Formål	Registret har til formål at give statistiske oplysninger om befolkningens beskæftigelsesforhold mv., dels til opgørelse af antal beskæftigede og beskæftigelsesomfang og dels til karakteristik af beskæftigelsesart.
Indhold	<p>Registret indeholder oplysninger om personer, som er eller har været erhvervsmæssigt beskæftiget i Danmark eller som har eller har haft bopæl her. Der indgår følgende oplysninger:</p> <p><i>personnummer, bopælsadresse, arbejdsstedsadresse, uddannelsesstedsadresse</i></p> <p><i>ansættelsesperioder, løn, ATP-bidrag, pension, arbejdsstilling, bruttoindkomst, skattepligtig indkomst, efterlønsmarkering, arbejdsløshedsforsikringskategori, arbejdsløhedskasse, oplysning om ledighed, igangværende uddannelse, afsluttet uddannelse</i></p> <p><i>antal beskæftigede ved arbejdssted, arbejdsstedets branche, virksomhedens ejerforhold</i></p> <p><i>ægteskabelig stilling, statsborgerskab</i></p>
Kilder	<p>Told- og Skattestyrelsens Centrale oplysningsseddelregister (COR)</p> <p>Danmarks Statistiks Arbejdspladsstatistikregister</p> <p>Danmarks Statistiks Arbejdsklassifikationsmodul</p> <p>Danmarks Statistiks Befolkningsstatistikregister</p> <p>Danmarks Statistiks Erhvervsstatistiske registersystem</p> <p>Danmarks Statistiks Uddannelsesstatistikregister</p> <p>Danmarks Statistiks Sygedagpengestatistikregister</p>
Anvendelse	<p>Beskæftigelsesstatistik</p> <p>Arbejdsmarkedsstatistik vedrørende nyuddannede</p> <p>Pendlingsstatistik</p> <p>Beskæftigelsesstatistiske special- og serviceopgørelser</p> <p><i>I andre statistikregistre:</i></p> <p>Statistikregistret for arbejdsmarkedsforskning</p> <p>Regionalt forløbsregister</p> <p>Forbrugsundersøgelsesregistret</p> <p>Uddannelsesklassifikationsmodulet</p>

Statistikregistret for arbejdsmarkedsforskning⁸⁴

Formål	Registret indeholder oplysninger om arbejdssteder (virksomheder) og personer. Formålet med registret er at danne grundlag for person- og virksomhedsorienterede statistiske undersøgelser af arbejdsmarkedsforhold.
Indhold	<p>Registret indeholder oplysninger om personer, der er eller har været erhvervsmæssigt beskæftiget i Danmark, eller som har eller har haft bopæl her, samt om virksomheder, der har eller har haft aktive arbejdssteder med adresse i Danmark. Der indgår følgende oplysninger:</p> <p><i>personnummer, arbejdsgivernummer, arbejdsstedskode, bopælsadresse, arbejdsstedsadresse</i></p> <p><i>baggrundsoplysninger, oplysninger om tilknytning til arbejdsmarkedet, uddannelsesoplysninger, indkomst/pensionsoplysninger, ATP-bidrag (beskæftigelsesomfang hos arbejdsgiver), løn, beskæftigelsesoplysninger pr. arbejdssted, ejerforhold, branche, stillingsoplysninger, højeste uddannelse, samlede indkomst, primær indkomster, overførselsindkomster, formueoplysninger, ledighedsgrader, ledighedsperioder, antal uger med forskellige typer af ledighed, kurser for specialarbejdere/faglærte (antal, retning)</i></p> <p><i>ægteskabelig stilling, statsborgerskab, oplysninger vedr. familieforhold, oplysninger om demografiske ændringer</i></p>
Kilder	<p>Danmarks Statistiks Beskæftigelsesstatistikregister Danmarks Statistiks Arbejdspladsstatistikregister Danmarks Statistiks Arbejdsklassifikationsmodul Danmarks Statistiks Befolkningsstatistikregister Danmarks Statistiks Uddannelsesstatistikregister Danmarks Statistiks Indkomststatistikregister Danmarks Statistiks Arbejdsløshedsstatistikregister</p>
Anvendelse	<p>Person- og virksomhedsorienterede forskningsprojekter inden for arbejdsmarkedsområdet</p> <p>Statistikregistret for arbejdsmarkedspolitiske foranstaltninger</p>
Formål	Registret har til formål at danne grundlag for statistik vedrørende de arbejdsmarkedspolitiske foranstaltninger.

⁸⁴ Også kaldet Integreret database for arbejdsmarkedsforskning (IDA, Jf. kapitel 5, eksempel 22).

Indhold Registret indeholder oplysninger om personer, der deltager eller har deltaget i en arbejdsmarkedspolitisk foranstaltning. Der indgår følgende oplysninger

*personnummer, familiehenvi-
sning, SE-nummer*

bopælskommune, statsborgerskab, ægteskabelig stilling, a-kasse-medlemskab, ledighedsoplysninger, uddannelsesbaggrund, foranstaltningens art og omfang, periode, ydelse, art og størrelse, seneste beskæftigelse, kommune hvor foranstaltningen foregår, stedet hvor foranstaltningen foregår, årsag til ophør af foranstaltning.

Kilder Arbejdsmarkedsstyrelsens Arbejdsmarkedspolitikregister, AMPO
Arbejdsmarkedsstyrelsens Centrale register for arbejdsmarkedsstatistik, CRAM
Undervisningsministeriets register over personer, der deltager i uddannelser med voksenuddannelsesstøtte
Indberetninger fra kommunerne
Danmarks Statistiks Befolkningsstatistikregister
Danmarks Statistiks Arbejdsløshedsstatistikregister
Danmarks Statistiks Uddannelsesklassifikationsmodul

Anvendelse Løbende statistisk dækning af det arbejdsmarkedspolitiske aktiveringsområde
Special- og serviceopgørelser

Statistikregistret vedrørende arbejdsmarkedspolitiske foranstaltninger

Formål Registret har til formål at danne grundlag for forskningsprojekter til belysning og vurdering af arbejdsmarkedspolitiske foranstaltninger.

Indhold Registret indeholder oplysninger om et udsnit af personer, der har deltaget i arbejdsmarkedspolitiske foranstaltninger samt et udsnit af tilfældigt udvalgte personer. Der indgår følgende oplysninger:

personnummer

*køn, alder, bopæl, familiestatus, uddannelse, arbejdsmarkedstil-
knytning, ledighed, indkomst, ledighedsoplysninger, familiefor-
hold, uddannelse og erhvervs erfaring, arbejdsforhold, viden om
uddannelses- og beskæftigelsesmuligheder, deltagelse i arbejds-
markedspolitiske foranstaltninger og erfaringer hermed*

Kilder Socialforskningsinstituttets interviewundersøgelse "Indsatsen mod langtidsledighed"

Socialforskningsinstituttets interviewundersøgelse "Evaluering af EF's socialfond"

Socialforskningsinstituttets interviewundersøgelse "Der er brug for alle"

Socialforskningsinstituttets interviewundersøgelse "Iværksætterordningen"

Danmarks Statistiks Statistikregister for arbejdsmarkedsforskning

Anvendelse

Registret anvendes som grundlag for fremstilling af statistisk materiale, herunder anonyme registeruddrag, der analyseres i Danmarks Statistik af Socialforskningsinstituttet.

Statistikregistret vedrørende den kommunale beskæftigelsesindsats

Formål

Registret har til formål at belyse situationen for personer, der har været omfattet af kommunale beskæftigelsesforanstaltninger samt effekten af disse foranstaltninger.

Indhold

Registret indeholder oplysninger om et udsnit af personer, der har deltaget i kommunale beskæftigelsesforanstaltninger samt et udsnit af personer med samme aldersstruktur som førstnævnte. Der indgår følgende oplysninger:

personnummer

familieforhold, indkomstforhold, boligforhold, arbejdsløshedsforsikring, ledighed, uddannelse, deltagelse i beskæftigelsesprojekter, ungdomsydelser, revalidering, beskæftigelse, foranstaltningsart, virksomhedens beliggenhedskommune, virksomhedens branche, virksomhedens størrelse, ægteskabelig stilling, familiesituation, antal børn, boligtype, aktuel beskæftigelse for forældre, ønsker om uddannelse, evt. hindringer for uddannelse, beskæftigelse efter deltagelse i foranstaltning, nytten af deltagelse i foranstaltningen

Kilder

Spørgeskema

Danmarks Statistiks Regionalt forløbsregister

Danmarks Statistiks Løn- og personalestatistikregister for den offentlige sektor

Arbejdsmarkedsstyrelsens register AF-MATCH

Anvendelse

Registret anvendes som grundlag for fremstilling af statistisk materiale, der analyseres i Danmarks Statistik af Amternes og Kommunernes Forskningsinstitut.

Statistikregistret vedrørende kvinder i mandefag

Formål Registrets formål er at danne grundlag for belysning af arbejdsforholdene for kvinder i såvel faglærte som ufaglærte mandefag.

Indhold Registret indeholder oplysninger om et repræsentativt udvalg af personer, der har fuldført en mandefaglig uddannelse eller et specialarbejderkursus i 1988. Der indgår følgende oplysninger:

personnummer

køn, alder, ægteskabelig stilling, bopælskommune, familieforhold

igangværende og afsluttede almenuddannelser, igangværende og afsluttede erhvervsuddannelser, gennemførte arbejdsmarkedsuddannelser, ledighedsoplysninger, arbejdsløshedsforsikringsoplysninger, erhvervsgymnasiale uddannelser/specialarbejderkurser, aktuelle og tidligere beskæftigelsessituation, arbejdsfunktioner (jobindhold), arbejdsorganisering, arbejdsmiljøet

graviditet, børn, barselsorlov, overvejelser m.h.t. at forlade det faglærte/ufaglærte mandefag og evt. begrundelser for at have forladt det

Kilder Spørgeskemaer
Danmarks Statistiks Befolkningsstatistikregister
Danmarks Statistiks Uddannelsesstatistikregister
Danmarks Statistiks Arbejdsløshedsstatistikregister
Danmarks Statistiks Arbejdsklassifikationsmodul

Anvendelse Registret anvendes som grundlag for fremstilling af statistisk materiale, der analyseres af Socialforskningsinstituttet.

Statistikregistret vedrørende ledige med kompetencegivende uddannelse

Formål Statistikregistret vedrørende ledige med kompetencegivende uddannelse har til formål at belyse tilknytning til arbejdsmarkedet efter uddannelsens afslutning.

Indhold Registret indeholder oplysninger om et repræsentativt udsnit blandt personer med en kompetencegivende uddannelse, der i juni 1989 var i alderen 25-49 år og havde været ledige i 12 af de seneste 15 måneder. Der indgår følgende oplysninger:

personnummer

køn, alder, afsluttede almenuddannelser, afsluttede erhvervsuddannelser, gennemførte arbejdsmarkedsuddannelser, ledighed, arbejdsløshedsforsikring, beskæftigelsesforhold, ægteskabelig stilling, aktuelle og tidligere beskæftigelsessituation, arbejdsøgning

Kilder

Spørgeskemaer
 Danmarks Statistiks Befolkningsstatistikregister
 Danmarks Statistiks Uddannelsesstatistikregister
 Danmarks Statistiks Arbejdsløshedsstatistikregister
 Danmarks Statistiks Arbejdsklassifikationsmodul

Anvendelse

Registret anvendes som grundlag for fremstilling af statistisk materiale for Arbejdsdirektoratet.

1.3. Løn, indkomst, formue og forbrug

Ejendomsstatistikregistret

Formål

Registrets formål er at danne grundlag for statistiske opgørelser vedrørende fast ejendom i Danmark.

Indhold

Registret indeholder oplysninger om faste ejendomme i Danmark samt om personer, som ejer fast ejendom i Danmark. Der indgår følgende oplysninger:

For hver ejer: *Personnummer eller SE-nummer, købt andel, ejers andel, ejers status, overtagelsesdato, ejerforholdskode, lejeværdikode, pensionistkode*

For hver ejendom: *Ejendomsnummer, adresse, kode for flere adresser, moderejendomsnummer, vurderingsdata (specificeret), koder for skattefritagelse mv., afgiftspligtige værdier (specificeret), ejendomsskatter (specificeret), benyttelseskode, samlet vurderet areal, kode for ej solgt ejerlejlighed, kode for landbrugsstatistik*

For hver matrikel: *Ejerlavskode, matrikelnummer/bogstav, arts-kode, zonekode, arealtypekode, matriklens areal og vejareal*

Kilde

Statsskattedirektoratets salgs- og vurderingsregister vedrørende fast ejendom

Anvendelse

Ejendomsskattestatistik

Specialopgørelser og serviceopgørelser
Lovmodellen

I andre statistikregistre:

Indkomststatistikregistret
Velstandsstatistikregistret

Forbrugsundersøgelserregistret

Formål	Registret har til formål at oparbejde et statistisk materiale om befolkningens forbrugs-, indkomst- og opsparingsforhold.
Indhold	<p>Registret indeholder oplysninger om en stikprøve af personer, som på undersøgelsestidspunktet har bopæl i Danmark. Der indgår følgende oplysninger:</p> <p><i>adresse, personnummer, navn (udgår, når registret er endeligt dannet)</i></p> <p><i>beholdning af varige forbrugsgoder, forbrug af offentlige ydelser, specificerede oplysninger om forbrug af varer og tjenester med angivelse af art, mængde og udgift</i></p> <p><i>stilling, ægteskabelig stilling, fødselsår, husstandssammensætning, erhverv, uddannelse, bolig, indkomster, formue, skatter, arbejdsstedsbranche, pendling, boligforhold, evt. igangværende uddannelse, senest afsluttede erhvervsuddannelse, evt. ydelser efter bistandsloven (art og størrelse)</i></p>
Kilder	<p>Interviews med husstande Regnskabshæfter, som føres i de medvirkende husstande Danmarks Statistiks Beskæftigelsesstatistikregister Danmarks Statistiks Bygnings- og boligstatistikregister Danmarks Statistiks Indkomststatistikregister Danmarks Statistiks Uddannelsesklassifikationsmodul Danmarks Statistiks Bistandslovsstatistikregister</p>
Anvendelse	<p>Statistik vedrørende forskellige befolkningsgruppers levevilkår Opgørelser af forbrug af varer og tjenester i forskellige befolkningsgrupper Grundmateriale til nationalregnskabsstatistikens opgørelser af befolkningens forbrugs- og indkomstforhold Statistik vedrørende befolkningens anvendelse af de offentlige ydelser samt overførsler til det offentlige Lovmodellen Internationale sammenligninger af befolkningens levevilkår</p>

Indkomstforløbsregistret

Formål	<p>Indkomstforløbsregistret har til formål at danne grundlag for dannelse af modeldata for en flerårig periode til økonomiske analyser, herunder navnlig:</p> <ul style="list-style-type: none"> – konsekvensberegninger i forbindelse med konkrete lovforslag – analyser og fremskrivninger til belysning af udvalgte emner – analyser i forbindelse med makroøkonomiske fremskrivninger
Indhold	<p>Registret indeholder oplysninger om et repræsentativt udvalg af personer, der har eller har haft bopæl i Danmark. Der indgår følgende oplysninger:</p> <p style="text-align: center;"><i>personnummer, bopælsadresse, familiehenvísninger</i></p> <p style="text-align: center;"><i>familieoplysninger, indkomst, formue, skatteforhold, boligforhold, arbejdsløhedskasse, forsikring, ledighed</i></p>
Kilder	<p>Danmarks Statistiks Befolkningsstatistikregister Danmarks Statistiks Indkomststatistikregister Danmarks Statistiks Bygnings- og boligstatistikregister Danmarks Statistiks Arbejdsløhedsstatistikregister</p>
Anvendelse	<p>Registret anvendes til fremstilling af en modelbefolkning i form af anonymiserede individualoplysninger.</p>

Indkomststatistikregistret

Formål	<p>Registret danner grundlag for statistiske opgørelser vedrørende befolkningens indkomst-, fradrags-, formue- og skatteforhold. Registret skal i øvrigt kunne anvendes til at danne baggrundsuplysninger, der kan indgå i opgørelser inden for andre personstatistikområder.</p>
Indhold	<p>Registret indeholder oplysninger om personer, der har eller har haft bopæl i Danmark, samt om personer i øvrigt hvem det påhviler at svare indkomst- eller formueskat til staten. Der indgår følgende oplysninger:</p> <p style="text-align: center;"><i>personnummer</i></p> <p style="text-align: center;"><i>kommunekoder, ligningstekniske oplysninger (fx skattepligtens omfang og sambeskatning), indkomstarter, fradragsarter, formuearter, skattearter, ATP-bidrag</i></p> <p style="text-align: center;"><i>beskæftigelsesstatus (fx lønmodtager), stillingsgrupperinger, arbejdsstedets branche og ejerforhold</i></p>

ægteskabelig stilling, statsborgerskab

Kilder	<p>Told- og Skattestyrelsens Centrale slutligningsregister (CSR-L) Told- og Skattestyrelsens Centrale oplysningsedelregister (COR-S) Danmarks Statistiks Arbejdsklassifikationsmodul (AKM) Danmarks Statistiks Befolkningsstatistikregister Danmarks Statistiks Register vedrørende indkomsterstattendende ydelser Danmarks Statistiks Boligstøttestatistikregister Danmarks Statistiks Børnetilskudsstatistikregister Danmarks Statistiks Pensionsstatistikregister</p>
Anvendelse	<p>Årlige indkomst- og formuestatistik Indkomststatistiske special- og serviceopgørelser Modelbefolkninger til Lovmodellen Baggrundsoplysninger på en række områder Stikprøveudvælgelse</p> <p><i>I andre statistikregistre:</i></p> <p>Velstandsstatistikregistret Arbejdsklassifikationsmodulet Udsnitsarkivet</p> <p>Løn- og personalestatistikregistret for den offentlige sektor</p>
Formål	<p>Registret har til formål at give statistiske oplysninger om antallet af ansatte og sammensætningen af arbejdsstyrken inden for stat og kommuner, herunder tilskudsområdet samt om løn- og beskæftigelsesforhold for ansatte i stat og kommuner. Registret skal i øvrigt kunne anvendes til at danne baggrundsoplysninger, der kan indgå i opgørelser inden for andre personstatistikområder.</p>
Indhold	<p>Registret indeholder oplysninger om personer, der er eller har været beskæftiget som lønmodtagere i den offentlige sektor, herunder tilskudsområdet. Der indgår følgende oplysninger:</p> <p><i>personnummer</i></p> <p><i>ansættende myndighed, arbejdssted, stilling, overenskomstforhold, formål (der anvendes de i de offentlige regnskabsystemer anvendte formålsskategorier), beskæftigelses omfang, løn (opdelt på løn i hovedstilling og visse tillæg)</i></p>
Kilder	<p>Offentlige lønsystemer</p>
Anvendelse	<p>Den løbende løn- og personalestatistik for den offentlige sektor</p>

I andre statistikregistre:

Arbejdspladsstatistikregistret
 Arbejdsklassifikationsmodulet
 Regionalt forløbsregister
 Social- og arbejdsforløbsregistret

Lønstatistikregistret for den private sektor**Formål**

Registrets formål er give oplysninger til statistiske opgørelser vedrørende lønforhold inden for den private sektor i Danmark.

Registret skal i øvrigt kunne anvendes til at danne baggrundsoplysninger, der kan anvendes i opgørelser inden for andre personstatistikområder.

Indhold

Registret indeholder oplysninger om personer, der er eller været beskæftiget som lønmodtager i den private sektor. Der indgår følgende oplysninger:

personnummer, firmaets SE-nummer

firmaets brancheforhold, størrelse og arbejdsstedets beliggenhed

firmaanciennitet, stillingsbetegnelse, uddannelse, månedsløn, virksomhedens bidrag til pension, overarbejdsbetaling, tantieme, gratiale, deltid/heltid, antal arbejdstimer, oplysninger om uddannelse og tidspunktet for uddannelsens afslutning

Kilder

Spørgeskemaer til virksomhederne
 Danmarks Statistiks Erhvervsstatistiske registersystem
 Danmarks Statistiks Uddannelsesklassifikationsmodul

Anvendelse

Lønstatistik
 Service- og specialopgørelser

I andre statistikregistre:

Arbejdsklassifikationsmodulet

Ressourceundersøgelsesregistret**Formål**

Ressourceundersøgelsesregistret har til formål at give en tilstandsbeskrivelse af ressourcefattige gruppers situation i samspil med forhold, som ligger forud for den fattigdomstruede situation.

Indhold Registret indeholder oplysninger om en stikprøve af personer, der har eller har haft bopæl i Danmark. Der indgår følgende oplysninger:

personnummer

fra interviewundersøgelse: ægteskabelig stilling, beskæftigelse, uddannelse, bolig, sundhed, sociale relationer, familiens økonomiske muligheder, familiens sociale muligheder, familiens kulturelle muligheder, sociale begivenheder i en periode forud for interviewingen

fra registre for en 6-års periode forud for interviewundersøgelse: familieforhold, indkomst og formue (art og størrelse), beskæftigelsesforhold, ledighed, oplysninger om ydet bistand, arten af denne og evt. beløb

Kilder Interviewundersøgelse (Socialforskningsinstituttet)
Danmarks Statistiks Befolkningsstatistikregister
Danmarks Statistiks Indkomststatistikregister
Danmarks Statistiks Arbejdsløshedsstatistikregister
Danmarks Statistiks Bistandslovsstatistikregister

Anvendelse Registret er anvendt som grundlag for fremstilling af statistisk materiale, der er analyseret i Socialforskningsinstituttet.

Velstandsstatistikregistret

Formål Registrets formål er at danne grundlag for statistiske opgørelser vedrørende befolkningens velstand.

Indhold Registret indeholder oplysninger om personer, der har eller har haft bopæl i Danmark. Der indgår følgende oplysninger:

*personnummer, familiehenvi-
sning, bopælsadresse, ejendomsnum-
mer*

*indkomst (art og størrelse), formue (art og størrelse), bidrag til
pensionsordninger*

*branche, stilling, ledighedsgrad, familiestatus, ægteskabelig stil-
ling, højeste afsluttede uddannelse, igangværende uddannelse*

*ejendomsværdi, beliggenhedskommune, ejendomsnummer, ejer-
andel, overtagelsesdato, ejendomsværdi, benyttelseskode, ejerfor-
holdskode*

antal biler/ejer, antal biler/bruger

boligart, opførelsesperiode, installationsforhold, opvarmningsforhold, ejer/lejer

Kilder Danmarks Statistiks Indkomststatistikregister
 Danmarks Statistiks Arbejdsklassifikationsmodul
 Danmarks Statistiks Befolkningsstatistikregister
 Danmarks Statistiks Ejendomsstatistikregister
 Danmarks Statistiks Motorkøretøjsstatistikregister
 Danmarks Statistiks Pensionsstatistikregister
 Danmarks Statistiks Børnetilskudsstatistikregister
 Danmarks Statistiks Boligstøttestatistikregister
 Danmarks Statistiks Bygnings- og boligstatistikregister
 Danmarks Statistiks Uddannelsesklassifikationsmodul

Anvendelse Årlig person- og husstandsorienteret indkomst- og formuestatistik

1.4. Uddannelse

Statistikregistret vedrørende enkeltfagskursister

Formål Registret har til formål at danne grundlag for statistiske undersøgelser vedr. enkeltfagskursisters sociale rekruttering og mobilitet.

Indhold Registret indeholder oplysninger om personer, der har modtaget undervisning ved et amtskommunalt enkeltfagskursus samt et udsnit på højst 10 pct. af andre 18-79 årige personer med bopæl i Danmark. Der indgår følgende oplysninger:

personnummer, bopælsadresse, arbejdsstedsadresse, familiehenvisning

uddannelse, institutionskode, koder for fag, forløb og prøver, bopælsadresse, familieoplysninger, indkomster, beskæftigelsesoplysninger, arbejdsstedets ejerforhold, arbejdsstilling, branche, beskæftigelsesomfang, forsikringsomfang, ledighedsgrad på årsbasis, husstandstype, boligforhold

Kilder Danmarks Statistiks Uddannelsesstatistikregister
 Danmarks Statistiks Befolkningsstatistikregister
 Danmarks Statistiks Indkomststatistikregister
 Danmarks Statistiks Beskæftigelsesstatistikregister
 Danmarks Statistiks Arbejdsløshedsstatistikregister
 Danmarks Statistiks Bygnings- og boligstatistikregister

Anvendelse

Registret anvendes til fremstilling af opgørelser vedrørende kursisternes baggrund, kursets indhold og resultat samt arbejdsmarkedsmobilitet.

Uddannelsesklassifikationsmodulet (UKM)**Formål**

Registret har til formål at give statistiske baggrundsoplysninger om den danske befolknings uddannelsesmæssige baggrund til brug inden for de personstatistiske områder, hvor sådanne oplysninger er relevante. Registret har endvidere til formål at danne grundlag for specialstatistikker om befolkningens beskæftigelses og uddannelsesforhold, om de studerendes tilknytning til arbejdsmarkedet samt om overgangen fra uddannelse til arbejdsmarked.

Indhold

Registret indeholder oplysninger om personer, der har eller har haft bopæl i Danmark eller Grønland. Der indgår følgende oplysninger:

personnummer, bopælskommune

køn, alder, ægteskabelig stilling, statsborgerskab, igangværende uddannelse med angivelse af institution, senest afsluttede almenuddannelse med angivelse af dato og institution, højeste afsluttede erhvervsrettede uddannelse med angivelse af dato og institution

forsikringskategori, ledighedsgrad, beskæftigelsesomfang, heltid/deltid, pensionsmarkering, arbejdsstilling, branche, arbejdsstedsadresse, arbejdsgivernummer, ejerforhold, skattepligtig indkomst, bruttoindkomst, stillingskode, socioøkonomisk gruppe, formålskode for kommunalt ansatte, arbejdsløshedskasse tilhørsforhold

Kilder

Danmarks Statistiks Uddannelsesstatistikregister
Danmarks Statistiks Folke- og bolig-tællingsregister
Danmarks Statistiks Beskæftigelsesstatistikregister

Anvendelse

Opgørelse af befolkningens fordeling efter uddannelse
Special- og serviceopgaver efter samkøring med andre oplysninger

I andre statistikregistre:

Folke- og bolig-tællingsregistret
Kriminalstatistikregistret
Beskæftigelsesstatistikregistret
Sygehusbenyttelsesstatistikregistret

Statistikregistret for fertilitetsforskning
 Arbejdsstyrkeundersøgelserregistret
 Forbrugsundersøgelserregistret
 Regionalt forløbsregister
 Lønstatistikregistret for den private sektor
 Velstandsstatistikregistret
 Social- og arbejdsforløbsregistret

Uddannelsesstatistikregistret⁸⁵

Formål	Registret danner grundlag for statistik om uddannelsessøgendes bevægelser gennem uddannelsessystemet.
Indhold	<p>Registret indeholder oplysninger om personer, som er eller har været indskrevet ved en uddannelsesinstitution i Danmark. Der indgår følgende oplysninger:</p> <p><i>personnummer, bopælsadresse</i></p> <p><i>ægteskabelig stilling, statsborgerskab, statuskode (aktiv, død, udvandret)</i></p> <p><i>uddannelse, uddannelsesinstitution, påbegyndelsestidspunkt for uddannelsen ved institutionen, afslutningstidspunkt for uddannelsen ved institutionen, trin i uddannelsen, retning, uddannelse afbrudt/fuldført, forventet afslutningstidspunkt for elever ved erhvervsfaglige uddannelser</i></p>
Kilder	Indberetninger fra uddannelsesinstitutionerne og elevregistre Danmarks Statistiks Befolkningsstatistikregistre
Anvendelse	<p>Den løbende uddannelsesstatistik Uddannelsesstatistiske special- og serviceopgørelser</p> <p><i>I andre statistikregistre:</i></p> <p>Beskæftigelsesstatistikregistret Arbejdsklassifikationsmodulet Flytteundersøgelserregistret Statistikregistret for fertilitetsforskning Statistikregistret for arbejdsmarkedsforskning Statistikregistret vedrørende erhvervsprogligt uddannede Statistikregistret vedrørende kvinder i mandefag Statistikregistret vedr. ledige med kompetencegivende uddannelse Statistikregistret vedrørende enkeltfagskursister</p>

⁸⁵ Også kaldet Det integrerede elevregister.

Statistikregistret vedr. tilpasningsstrategien mellem arbejdsliv og familiev
 Uddannelsesklassifikationsmodulet

1.5. Social sikring og retsvæsen

Bistandslovsstatistikregistret

Formål Registret danner grundlag for statistik vedrørende modtagere af visse ydelser efter bistandsloven.

Indhold Registret indeholder oplysninger om personer (familier), der modtager eller har modtaget kontanthjælp eller bistand vedrørende børn og unge i henhold til bistandsloven. Der indgår følgende oplysninger:

personnummer, bopælsadresse

ydelserart, -størrelse og varighed

ægteskabelig stilling, statsborgerskab, familiesammensætning

Kilder Kommuneindberetninger
 Indberetninger fra revaliderings- og pensionsnævn
 Danmarks Statistiks Befolkningsstatistikregister
 Det fælleskommunale økonomisystem

Anvendelse Løbende bistandslovsstatistik
 Special- og serviceopgørelser
 Lovmodellen

I andre statistikregistre:

Statistikregistret vedrørende indkomsterstøttende ydelser
 Ressourceundersøgelsesregistret
 Indkomststatistikregistret
 Statistikregistret vedrørende børns opvækstforhold

Boligstøttestatistikregistret

Formål Registret danner grundlag for statistiske opgørelser vedrørende udbetaling af boligsikring og boligydelse i Danmark.

Indhold Registret indeholder oplysninger om personer, der modtager eller har modtaget boligsikring eller boligydelse i Danmark. Der indgår følgende oplysninger:

*personnummer, bopælsadresse, familiehenvi-
sning*

*stilling i sagen (ansøger, modtager, ægtefælle, barn), sagstype,
boligforhold (Størrelse, antal personer, leje), statsborgerskab*

Kilder Det fælleskommunale Boligsikringssystem på Kommunedata I/S og Århus kommunes edb-central samt fra Københavns kommunes System for Boligsikring
Det fælleskommunale Boligydelssystem på Kommunedata I/S og Århus kommunes edb-central, samt fra Københavns kommunes System for Boligydelse
Danmarks Statistiks Befolkningsstatistikregister

Anvendelse Løbende boligstøttestatistik
Specialopgørelser og serviceopgørelser
Lovmodellen

I andre statistikregistre:

Indkomststatistikregistret
Velstandsstatistikregistret
Statistikregistret vedrørende sociale processer og boligforhold

Børnetilskudsstatistikregistret

Formål Registrets formål er at danne grundlag for statistiske opgørelser vedrørende udbetaling af børnetilskud og børnefamilieydelse i Danmark.

Indhold Registret indeholder oplysninger om familier med hjemmeboende børn under 18 år, herunder familier bestående af en eller flere personer under 18 år. Registret omfatter dels personer med bopæl i Danmark, dels personer, som bor i udlandet, og som i henhold til EU-regler har ret til at modtage dansk børnetilskud og børnefamilieydelse. Der indgår følgende oplysninger:

Tilskudsmodtagere: *personnummer, adresse, ægteskabelig stilling, udbetalte ydelser efter arten, sagskoder (kode for ekstra børnetilskud, skattepligtsgode mv.)*

For mor og far (evt. plejeforældre): *personnummer, adresse, ægteskabelig stilling, sagskoder (pensionskode, statsborgerretskode, skattepligtsgode mv.)*

For børnene: *personnummer, adresse, ægteskabelig stilling, koder vedr. børnetilskud og børnefamilieydelse mv., betalingskommune*

Kilder	<p>Det fælleskommunale Børnetilskudssystem Statsskattedirektoratets børnefamilieydelsessystem på Kommunedata I/S og Århus kommunes edb-central Københavns kommunes Børnetilskudssystem Danmarks Statistiks Befolkningsstatistikregister</p>
Anvendelse	<p>Løbende statistik om børnetilskud og børnefamilieydelse Specialopgørelser og serviceopgørelser Lovmodellen</p> <p><i>I andre statistikregistre:</i></p> <p>Befolkningsstatistikregistret Indkomststatistikregistret Velstandsstatistikregistret</p> <p>Daginstitutionstatistikregistret</p>
Formål	<p>Registrets formål er at danne grundlag for statistiske opgørelser vedrørende børn og unges benyttelse af daginstitutioner, dagpleje og skolefritidsordninger, der drives i henhold til Bistandsloven, Folkeskoleloven og Lov om friskoler og private grundskoler.</p>
Indhold	<p>Registret indeholder oplysninger om familier med børn i daginstitution eller dagpleje i Danmark. Der indgår følgende oplysninger:</p> <p style="padding-left: 40px;"><i>For børnene: personnummer, institutionsnummer, sagskoder vedrørende beregning af betaling og friplads, herunder indtægtsgrundlag, forældres personnummer</i></p> <p style="padding-left: 40px;"><i>For kommunen: oplysning om generelle daginstitutionstakster</i></p> <p style="padding-left: 40px;"><i>For institutionen: institutionsnummer, navn og adresse, institutions art og ejerforhold, institutionstakster, pladser, åbningstid</i></p>
Kilder	<p>Det fælleskommunale daginstitutionssystem på Kommunedata I/S og det fælleskommunale daginstitutionssystem på Århus kommunes edb-central Venteliste- og betalingssystem for Københavns børneinstitutioner på Københavns kommunes edb-central Danmarks Statistiks Befolkningsstatistikregister</p>
Anvendelse	<p>Den årlige ressourceopgørelse Løbende statistik om familiers daginstitutionsbenyttelse Specialopgørelser og servicekørsler Lovmodellen</p>

Kriminalstatistikregistret

Formål Registret har til formål at give statistiske oplysninger om personer, der sigtes for lovovertrædelser og personer, om hvem der i øvrigt træffes strafferetlige afgørelser.

Indhold Registret indeholder oplysninger om personer, der sigtes for lovovertrædelser, og personer, om hvem der i øvrigt træffes strafferetlige afgørelser. Der indgår følgende oplysninger:

personnummer, bopælsadresse, evt. virksomhedsnummer

ægteskabelig stilling, statsborgerskab, senest afsluttede skoleuddannelse, senest afsluttede erhvervsuddannelse, beskæftigelsesstatus (selvstændige mv.), stillingsklassifikation, branche, evt. dødsdato

Oplysninger vedr. sigtelser
sigtelsesdato, gerningstidspunkt eller -tidsrum, gerningsindhold

Oplysninger vedr. afgørelser
afgørelsesdato, afgørende myndighed, afgørelsens art og omfang, gerningsindhold, dato for evt. anke, ophør af evt. betingelse

Oplysninger vedr. fængslinger
fængslings- eller løsladelsesdato og dato for ophør af evt. prøvetid, fængsel eller arresthus, afgørelsens art og længde, gerningsindhold, oplysninger om evt. anholdelse og/eller evt. varetægtsfængsling

Kilder Det centrale kriminalregister
Indberetninger fra Direktoratet for Kriminalforsorgen
Danmarks Statistiks Befolkningsstatistikregister
Danmarks Statistiks Uddannelsesklassifikationsmodul
Danmarks Statistiks Arbejdsklassifikationsmodul

Anvendelse Løbende kriminalstatistik
Recidivstatistik
Kriminalstatistiske special- og serviceopgaver

Pensionsstatistikregistret

Formål Registrets formål er at danne grundlag for statistiske opgørelser vedrørende udbetaling af pensioner og efterløn i Danmark.

Indhold Registret indeholder oplysninger om personer, der modtager eller har modtaget pension eller efterløn i Danmark. Der indgår følgende oplysninger:

personnummer, bopælsadresse, familiehenviisning

ægteskabelig stilling, statsborgerskab

beregningsgrundlag, pensionsbeløb m.v. (udbetalte ydelser efter art, fx grundbeløb, invaliditetsbeløb, ventetillæg), skatteoplysninger, beregningsoplysninger (virkningsdato, hændelsestype, pensionsstand, udbetalingsoplysninger), pensionsbeløb (akkumuleret A-indkomst, grundbeløb), beskæftigelsesforhold, udbetalte beløb

Kilder Det fælleskommunale Pensionssystem på Kommunedata I/S og Århus kommunes edb-central samt fra Københavns kommunes System for sociale pensioner
 Det fælleskommunale delpensionssystem på Kommunedata I/S og Århus kommunes edb-central, samt fra Københavns kommunes System for delpension
 Danmarks Statistiks Indkomststatistikregister
 Danmarks Statistiks Befolkningsstatistikregister

Anvendelse Løbende pensionsstatistik
 Specialopgørelser og serviceopgørelser
 Lovmodellen

I andre statistikregistre:

Indkomststatistikregistret
 Statistikregistret vedrørende indkomsterstattende ydelser
 Velstandsstatistikregistret

Statistikregistret vedrørende børns opvækstvilkår

Formål Registret har til formål at belyse den nuværende situation for unge, der har været anbragt uden for hjemmet, sammenlignet med en gruppe af jævnaldrende, hvor familien har fået langvarig bistandshjælp, og en gruppe af tilfældigt udvalgte jævnaldrende.

Indhold Registret indeholder oplysninger om et udsnit af personer, der har eller har haft bopæl i Danmark. Alle identifikationer er slettet. Der indgår følgende oplysninger:

Anbringelse, bistandshjælp, familieforhold, børn, køn, boligforhold, skolegang og skoleforhold, omfang og karakter af erhvervsarbejde, erhvervsuddannelse, erhvervsevne (fysisk helbred), op-

lysninger om forældre, aktuelle psykiske situation, barndom, oplevelser af evt. anbringelse

Kilder	Spørgeskema Danmarks Statistiks Bistandslovsstatistikregister
Anvendelse	Registret anvendes som grundlag for fremstilling af statistisk materiale, der analyseres af Socialforskningsinstituttet.

Statistikregistret vedrørende indkomsterstattende ydelser⁸⁶

Formål	Registrets formål er at danne grundlag for statistiske opgørelser vedrørende udbetaling af indkomsterstattende ydelser.
Indhold	Registret indeholder oplysninger om personer, der modtager eller har modtaget indkomsterstattende ydelser i Danmark. Der indgår følgende oplysninger:

personnummer, bopælsadresse, familiehenvi­sing

pensionsstand, udbetalte beløb efter art, start- og slutdatoer

ledighedsgrad pr. uge, ledighedsårsag pr. uge, årets, ledighedsgrad, årets antal ledige timer

kontanthjælp, sagsart, udbetalt beløb efter art, start- og slutdatoer

arbejdsløshedsdagpenge, samlet indkomst, beskæftigelsesforhold, statsborgerskab, familiestatus, familietype

Kilder	Danmarks Statistiks Pensionsstatistikregister Danmarks Statistiks Arbejdsløshedsstatistikregister Danmarks Statistiks Sygedagpengestatistikregister Danmarks Statistiks Bistandslovsstatistikregister Danmarks Statistiks Indkomststatistikregister
Anvendelse	Løbende socialstatistik Specialopgørelser og serviceopgørelser.

I andre statistikregistre:

Statistikregistret for fertilitetsforskning

Sygehusbenyttelsesstatistikregisteret

Regionalt forløbsregister

Statistikregistret vedrørende sociale processer og boligforhold

Forebyggelsesregisteret

⁸⁶ Også kaldet sammenhængende socialstatistik, jf. Kapitel 5, eksempel 19.

Sygedagpengestatistikregistret

Formål Registrets formål er at danne grundlag for statistiske opgørelser vedrørende udbetaling af sygedagpenge i Danmark.

Indhold Registret indeholder oplysninger om personer, der modtager eller har modtaget sygedagpenge i Danmark. Der indgår følgende oplysninger:

*modtagers personnummer, bopælsadresse, familiehenvi-
sning, arbejdsgivers SE-nummer*

*første fraværsgang (sygedag), første beregningsdato for sygedag-
penge, sidste beregningsdato for sygedagpenge, koder for sagsart,
fraværsgang, ophørgang, indkomstgrundlag for beregning af
dagpenge, udbetalte dagpengebeløb*

*ægteskabelig stilling, statsborgerskab, indkomster, beskæftigel-
sesoplysninger*

Kilder Det fælleskommunale Dagpengesystem på Kommunedata I/S og Århus kommunes edb-central
Københavns kommunes System for dagpenge
Danmarks Statistiks Befolkningsstatistikregister
Danmarks Statistiks Indkomststatistikregister

Anvendelse Løbende sygedagpengestatistik
Specialopgørelser og serviceopgørelser
Lovmodellen

I andre statistikregistre:

Arbejdsløshedsstatistikregistret
Beskæftigelsesstatistikregistret
Statistikregistret vedrørende indkomsterstøttende ydelser

Sygesikringsstatistikregistret

Formål Registrets formål er at danne grundlag for statistiske opgørelser vedrørende sygesikrede og vedrørende ydelser i henhold til sygesikringsordningen i Danmark.

Indhold Registret indeholder oplysninger om personer, der modtager eller har modtaget sygesikringsydelser i Danmark. Der indgår følgende oplysninger:

*sikredes personnummer, bopælsadresse, familiehenvi-
sning*

sikredes amtskommune og kommune, sikredes sikringsgruppe, barnemarkering, ydelsens speciale, ydelsens gruppe, ydelsens løbenummer, patient- og behandlingstype, antal ydelser

ægteskabelig stilling, statsborgerskab, indkomster, beskæftigelsesoplysninger

Kilder Det fælleskommunale sygesikringssystem på Kommunedata I/S
Danmarks Statistiks Befolkningsstatistikregister
Danmarks Statistiks Indkomststatistikregister

Anvendelse Løbende sygesikringsstatistik
Specialopgørelser og serviceopgørelser
Lovmodellen

I andre statistikregistre:

Sygehusbenyttelsesstatistikregistret
Forebyggelsesregistret

1.6. Sundhed

Forebyggelsesregistret

Formål Registrets formål er at danne grundlag for statistiske analyser og forskning vedrørende den danske befolknings sundhedstilstand.

Indhold Registret indeholder oplysninger om personer, der har eller har haft bopæl i Danmark siden 1. januar 1977. Der indgår følgende oplysninger:

personnummer, familiehenviisning, bopælsadresse

sygehus, indlæggelse, diagnoser, operationsoplysninger, udskrivning, oplysninger om indgreb, oplysninger om fødsel, oplysninger om misdannelse, dødsårsager, oplysninger om svangerskab, fødestedsoplysninger, oplysninger om barn, oplysninger om dødfødte og døde under 1 år

statsborgerskab, ægteskabelig stilling, familietype/status, husstandstype, fælleshusholdning, oplysninger om demografiske ændringer, oplysninger om flytninger, oplysninger om familieforhold, antal sygesikringsydelse, ydelseernes art, honorar, sikringsgruppe, praksisoplysninger, barnemarkering, boligforhold, so-

ciala ydelser, uddannelse, indkomst, beskæftigelsesforhold, ledighed

Kilder	Sundhedsstyrelsens Landspatientregister Sundhedsstyrelsens register over legalt provokerede aborter Sundhedsstyrelsens misdannelsesregister Sundhedsstyrelsens dødsårsagsregister Sundhedsstyrelsens medicinske fødselsregister Danmarks Statistiks Befolkningsstatistikregister Danmarks Statistiks Sygesikringsstatistikregister Danmarks Statistiks Bygnings- og boligstatistikregister Danmarks Statistiks Register vedrørende indkomsterstøttede ydelser Danmarks Statistiks Statistikregister for arbejdsmarkedsforskning Danmarks Statistiks Arbejdsklassifikationsmodul
Anvendelse	Løbende sundhedsstatistik Dannelse af en modelbefolkning, som danner grundlag for statistiske analyser og forskning. Modelbefolkningen stilles til rådighed for Sundhedsministeriet og Sundhedsstyrelsen, som gennemfører analysekørsler i Danmarks Statistik.

Erhverv/cancerstatistikregistret

Formål	Registeret har til formål at danne grundlag for statistiske undersøgelser af den talmæssige sammenhæng mellem cancerforekomst og erhvervs- og andre samfundsmæssige forhold i Danmark.
Indhold	Registret indeholder oplysninger om personer, der var 15-74 år gamle og havde bopæl i Danmark ved folke- og boligtællingen 1970. Der indgår følgende oplysninger:

personnummer

Oplysninger vedr. personen pr. 9.11.70:

Bopælskommune og -område (sogn, by og ø), familie- og husstandsforhold, ægteskabelig stilling, beskæftigelses- og erhvervsforhold, beskæftigelses- og erhvervsforhold for evt. ægtefælle, arbejdssted, uddannelsesforhold, boligforhold

Oplysninger om perioden 9.11.70 - 8.11.80

Oplysning om udgang af befolkningen:

årsag, dødsårsag, døds måde

dødssted og anmeldelsesform. cancerdiagnosetidspunkt (måned og år), erhverv på diagnosetidspunktet, diagnose, behandling

Kilder Cancerregisteret
Danmarks Statistiks Erhvervs/dødelighedsstatistikregister

Anvendelse Analyse af cancer og erhverv
Service- og specialopgørelser vedr. cancer og erhverv

Erhverv/dødelighedsstatistikregistret

Formål Registret danner grundlag for undersøgelser af den talmæssige sammenhæng mellem beskæftigelsesforhold og dødsrisiko i den danske befolkning.

Indhold Registret indeholder oplysninger om personer, der var 15-74 år gamle og havde bopæl i Danmark ved en folke- og boligtælling. Der indgår følgende oplysninger:

personnummer, bopælsadresse, familiehenvisning

familie- og husstandsforhold, ægteskabelig stilling, beskæftigelses- og erhvervsforhold, arbejdssted, uddannelsesforhold, boligforhold

dødsfald, udvandring, forsvinding, dato for afgang fra befolkningen

døds Kommune, institution på død tidspunktet (sygehus o.l.), dødsårsag (primær/sekundær/tertiær, ICD-klassifikation), hændelsessted (fx hjemmet), ulykke (arbejdsulykke el. anden ulykke), døds måde (fx naturlig død, selvmord), operation (i forbindelse med dødsfald), autopsi, attestens art (medico-legal o.l.), løbenummer for dødsattest

Kilder Det centrale personregister (CPR)
Danmarks Statistiks Folke- og bolig tællingsregister
Danmarks Statistiks Medicinske fødsels- og dødsfaldsstatistikregister

Anvendelse Analyse af sammenhængen mellem dødelighed og erhverv
Special- og serviceopgørelser vedr. dødelighed og erhverv

Erhvervsindlæggelsesregistret

Formål Registret har til formål at danne grundlag for statistiske undersøgelser af den talmæssige sammenhæng mellem sygdomsforekomst og erhvervs- og andre samfundsmæssige forhold i Danmark.

Indhold Registret indeholder oplysninger om personer, der var 15-64 år gamle og havde bopæl i Danmark den 1. januar 1981 og personer, der var 15-

64 år gamle og havde bopæl i Danmark den 1. januar hvert 5. år herefter. Der indgår følgende oplysninger:

personnummer

bopælskommune pr. 1. januar i femårsperioden, urbanisering, ægteskabelig stilling, oplysning om evt. død i femårsperioden, dødsdato, oplysninger om evt. (første) udvandring i femårsperioden, udvandringsdato, beskæftigelsesstatus, stillingsgrupper, årsledighedsgrad, patienttype, indlæggelsesdato, ulykke (art), udskrivningsdato, hvortil udskrevet, kode for diagnoser, modifikation af diagnoser

Kilder

Sundhedsstyrelsens Landspatientregister
Danmarks Statistiks Befolkningsstatistikregister
Danmarks Statistiks Arbejdsklassifikationsmodul

Anvendelse

Arbejds miljøinstituttets undersøgelse af beskæftigelsesforholdenes indvirkning på antal hospitalsindlæggelser.
Registret danner grundlag for statistiske opgørelser til forskere og andre interesserede.

Sygehusbenyttelsesstatistikregistret

Formål

Registrets formål er at danne grundlag for statistiske opgørelser vedrørende personer, der har været indlagt på somatiske sygehusafdelinger i Danmark.

Indhold

Registret indeholder oplysninger om personer, der har været indlagt på somatiske sygehusafdelinger. Der indgår følgende oplysninger:

personnummer

sygehusnummer, afdelingsnummer, patienttype, indlæggelsesdato, indlæggelsestidspunkt, indlæggelsesmåde, udskrivningsdato, kontaktårsagskode, ulykkeskode, hoveddiagnose, hoveddiagnose

bopæl, statsborgerskab, ægteskabelig stilling, familietype/status, husstandstype, antal børn, yngste og ældste barns alder, antal personer i familien, personens indkomst (brutto/skattepligtig), familiens indkomst (brutto/skattepligtig), branche, stillingsgruppe, højeste igangværende/afsluttede uddannelse, boligens art, installationsforhold, opførelsesår, udlejningsforhold

antal sygesikringsydelse, udbetalt bruttohonorar, ydelseernes art og varighed, udbetalte beløb, dødsårsag, dødstidspunkt

Kilder	Sundhedsstyrelsens Landspatientregister Danmarks Statistiks Befolkningsstatistikregister Danmarks Statistiks Indkomststatistikregister Danmarks Statistiks Uddannelsesklassifikationsmodul Danmarks Statistiks Bolig- og bygningsregister Danmarks Statistiks Sygesikringsstatistikregister Danmarks Statistiks Statistikregister vedrørende indkomstersttende ydelser Danmarks Statistiks Medicinske fødsels- og dødsfaldsstatistikregister
Anvendelse	Løbende sygehusbenyttelsesstatistik Specialopgørelser og serviceopgørelser

1.7. Andet

Færdselsuhedsstatistikregistret

Formål	Registret har til formål at danne grundlag for færdselsuhedsstatistikken for Danmark omfattende opgørelser af det samlede antal færdselsuheld og personskader som følge heraf og deres fordeling efter trafikale og demografiske kendetegn samt efter skadernes følgevirkninger.
Indhold	Registret indeholder oplysninger om personer, som har været impliceret i færdselsuheld. Der indgår følgende oplysninger: <i>personnummer, motorkøretøjets registreringsnummer</i> <i>trafikal uheldssituation, stedsbeskrivelse, fysiske forhold, vejr- og lysforhold, køretøjstype/fodgænger, vejtype, færdselsregulering, hastighedsskøn, kørekortsforhold, personskade, skadens årsag, hospitalsbehandling, påvirkning af spiritus, medicin eller narkotika, sele- eller hjelmbrug, indlæggelse, anden behandling, diagnoser, dødsårsag</i>
Kilder	Indberetninger fra politiet Landspatientregistret Danmarks Statistiks medicinske fødsels- og dødsfaldsstatistikregister
Anvendelse	Løbende færdselsuhedsstatistik Forskningsmæssige formål mv.

Motorkøretøjsstatistikregistret

Formål Registrets formål er at danne grundlag for statistiske opgørelser vedrørende ejerforhold og anvendelse af motorkøretøjer.

Indhold Registret indeholder oplysninger om personer, der er eller har været registreret som ejere (brugere) af motorkøretøjer i Danmark. Der indgår følgende oplysninger:

Oplysninger vedr. registreret ejer og bruger: *Personnummer, SE-nummer, navn, bopæl.*

Oplysninger vedr. køretøjet: *Tekniske data, afgiftsoplysninger, forsikringsoplysninger, registrerings-/afmeldingstidspunkt, oplysning om ændringer i registreringsforhold.*

Kilde Centralregistret for motorkøretøjer (CRM)

Anvendelse Nationalregnskabsstatistik
Transportstatistik.
Service- og specialopgørelser vedr. motorkøretøjer
Lovmodellen

I andre statistikregistre:

Velstandsstatistikregistret

Omnibusundersøgelsesregistret

Formål Registret har til formål at opsamle oplysninger, der i Danmarks Statistiks omnibusundersøgelser tilvejebringes til brug for statistiske og videnskabelige undersøgelser, baseret på interviews med et mindre udvalgt af den danske befolkning⁸⁷.

Indhold Registret indeholder oplysninger om en repræsentativ stikprøve af personer, der på undersøgelsestidspunktet har bopæl i Danmark. Der indgår følgende oplysninger:

personnummer, bopæl

familieforhold, statsborgerskab, arbejdsstilling, erhverv, uddannelse, boligforhold, erhvervsarbejdets omfang, indkomst, urbaniseringsgrad,

⁸⁷ Visse omnibusundersøgelser er gennemført i samarbejde med Socialforskningsinstituttet, men registret indeholder kun Danmarks Statistiks del af disse ældre undersøgelser.

interviewpersonens forventninger med hensyn til den økonomiske udvikling, herunder til arbejdsløshed og priser, og til familiens egen økonomi, herunder planer eller ønsker med hensyn til opsparring og anskaffelse af formuegoder og varige forbrugsgoder

interviewpersonens daglige transport og rejser over længere afstande

andre specialoplysninger, indsamlet som led i en serviceopgave i forbindelse med en enkelt undersøgelse

Kilder Interviews med udvalgte personer
Det centrale personregister (CPR)
Danmarks Statistiks Befolkningsstatistikregister

Anvendelse Statistik vedrørende forbrugerforventninger.
Statistik vedrørende forskellige befolkningsgruppers levevilkår.
Statistiske opgørelser og analyser vedrørende specielle emner
Transportstatistik

Regionalt forløbsregister

Formål Regionalt forløbsregister har til formål at danne grundlag for dannelse af modeldata for en flerårig periode til analyser af regionale sammenhænge mellem sociale forhold, arbejdsmarkeds-, indkomst- og boligforhold i befolkningen.

Indhold Registret indeholder oplysninger om personer, der har eller har haft bopæl i Danmark. Der indgår følgende oplysninger:

personnummer, bopæl, familiehenvísninger

familieoplysninger, indkomst, formue, skatteforhold, boligforhold, arbejdsløshedskasse, forsikring, ledighed, afsluttet erhvervsuddannelse, igangværende uddannelse, arbejdsmarkedsuddannelser, deltagelse i beskæftigelsesprojekter, ungdomsydelser, revalidering, arbejdsløshedsdagpenge, sygedagpenge, bistandshjælp, efterløn, sociale pensioner, beskæftigelsesforhold

Kilder Danmarks Statistiks Befolkningsstatistikregister
Danmarks Statistiks Indkomststatistikregister
Danmarks Statistiks Bygnings- og boligstatistikregister
Danmarks Statistiks Arbejdsløshedsstatistikregister
Danmarks Statistiks Uddannelsesklassifikationsmodul
Danmarks Statistiks Løn- og personalestatistikregister for den offentlige sektor

Danmarks Statistiks Statistikregister vedrørende indkomsterstøttende ydelser
Danmarks Statistiks Beskæftigelsesstatistikregister

Anvendelse Registret anvendes til fremstilling af en modelbefolkning, evt. suppleret med data, der er indhentet til et konkret projekt, i form af anonymiserede individualoplysninger. Modelbefolkningen stilles til rådighed for Amternes og Kommunernes Forskningsinstitut, som gennemfører analysekørslerne i Danmarks Statistik.

Social- og arbejdsforløbsregistret

Formål Registret har til formål at danne grundlag for dannelse af modeldata for en flerårig periode til analyser af sammenhænge mellem beskæftigelsesforhold og modtagelse af sociale ydelser i befolkningen.

Indhold Registret må kun indeholde oplysninger om et repræsentativt udsnit af personer, der har eller har haft bopæl i Danmark. Der indgår følgende oplysninger:

personnummer, familiehenviisning

bopælskommune, familieoplysninger, indkomst, formue, skatteforhold, boligforhold, ledighed, afsluttet erhvervsuddannelse, igangværende uddannelse, arbejdsmarkedsuddannelser, deltagelse i beskæftigelsesprojekter, ungdomsydelser, revalidering, arbejdsløshedsdagpenge, sygedagpenge, bistandshjælp, efterløn, sociale pensioner, beskæftigelsesforhold, deltagelse i arbejdstilbud, uddannelse med uddannelsesydelse, tildelt iværksætterydelse

Kilder Arbejdsmarkedsstyrelsens Centrale register for arbejdsmarkedsstatistik (CRAM)
Danmarks Statistiks Indkomstforløbsregister
Danmarks Statistiks Uddannelsesklassifikationsmodul
Danmarks Statistiks Løn- og personalestatistikregister for den offentlige sektor
Danmarks Statistiks Beskæftigelsesstatistikregister

Anvendelse Registret anvendes til fremstilling af en modelbefolkning i form af anonymiserede individualoplysninger. Modelbefolkningen stilles til rådighed for Amternes og Kommunernes Forskningsinstitut, som gennemfører analysekørslerne i Danmarks Statistik.

Statistikregistret vedrørende erhvervsprogligt uddannede

Formål Registret har til formål at belyse de erhvervsprogligt uddannedes faglige organisationsforhold og tilknytning til arbejdsmarkedet.

Indhold Registret indeholder oplysninger om et repræsentativt udsnit af personer, der pr. 1. oktober 1988 var erhvervsprogligt uddannet. Der indgår følgende oplysninger:

personnummer

afsluttede almenuddannelser, afsluttede erhvervsuddannelser, tidspunkt for afsluttede uddannelser, arbejdsløshedsforsikringsoplysninger, køn, alder, familie- og boligforhold, beskæftigelsessituation og arbejdsforhold, indkomst, læsevaner og fritidsinteresser, indstilling og forventninger til fagforeningsarbejde, indstilling til arbejdssituationen og erhvervsmæssige ambitioner

Kilder Spørgeskema
Danmarks Statistiks Uddannelsesstatistikregister
Danmarks Statistiks Arbejdsløshedsstatistikregister

Anvendelse Registret anvendes som grundlag for fremstilling af statistisk materiale i form af grupperede oplysninger til Erhvervsprogligt Forbund.

Statistikregistret vedrørende sociale processer og boligforhold

Formål Registret har til formål at danne grundlag for dannelse af modeldata for en flerårig periode til analyser af sammenhænge mellem udviklingen i arbejdsmarkedstilknøytning, indkomstsammensætning, uddannelsesforhold, familie- og husstandsmæssige forhold og boligforhold.

Indhold Registret indeholder oplysninger om et repræsentativt udsnit af personer i alderen 15 år eller mere, der har eller har haft bopæl i Danmark. Der indgår følgende oplysninger:

personnummer, familiehenviisning, bopæl

familie- og husstandsoplysninger, fødested, statsborgerskab, indvandringsdato, dødsdato, indkomst, art og størrelse, formue, skatteforhold, beskæftigelsesforhold, ledighedsoplysninger, virksomhedsstørrelse, uddannelsesoplysninger, ATP-indbetalinger, sygedagpenge, bistandshjælp, efterløn, sociale pensioner, revalidering, boligens standard karakteriseret ved boligens art, type, installationer mv. samt husleje, boligstøttens størrelse og art

Kilder Danmarks Statistiks Befolkningsstatistikregister
Danmarks Statistiks Statistikregister for arbejdsmarkedsforskning
Danmarks Statistiks Indkomststatistikregister
Danmarks Statistiks Statistikregister for indkomsterstøttende ydelser

Danmarks Statistiks Bygnings- og boligstatistikregister
 Danmarks Statistiks Boligstøtteregister

Anvendelse Registret anvendes til fremstilling af en modelbefolkning i form af anonymiserede individualoplysninger. Modelbefolkningen stilles til rådighed for Amternes og Kommunernes Forskningsinstitut, som gennemfører analysekørslerne i Danmarks Statistik.

Statistikregistret vedrørende tidsanvendelsen 1950-88

Formål Registret har til formål at belyse befolkningens tidsanvendelsesmønstre.

Indhold Registret indeholder oplysninger om et udsnit af personer, der i 1988 havde bopæl i Danmark. Der indgår følgende oplysninger:

personnummer

køn, alder, bopælskommune, familieforhold, erhverv, indkomstforhold

andel af husligt arbejde, evt. ægtefælles erhvervsaktivitet, egen arbejdsstilling, arbejdsstedets placering i land eller by, ejerbolig, gør-det-selv arbejde, betalt beskæftigelse med og uden beskatning

Kilder Spørgeskemaer
 Danmarks Statistiks Befolkningsstatistikregister
 Danmarks Statistiks Arbejdsklassifikationsmodul
 Danmarks Statistiks Indkomststatistikregister

Anvendelse På grundlag af edb-kørsler i Danmarks Statistik analyserer Rockwool Fondens forskningsprojekt "Tid og forbrug", der er lokaliseret i Danmarks Statistik som forskerkunde, materialet.

Statistikregistret vedrørende tilpasningsstrategien mellem arbejdsliv og familieliv

Formål Registret har til formål at belyse familielivets påvirkning af arbejdslivet og arbejdspladsernes tilpasning til familielivet.

Indhold Registret indeholder oplysninger om 8 deludsnit af lønmodtagere. Hver af disse udsnit omfatter personer med en bestemt uddannelsesmæssig baggrund. Der indgår følgende oplysninger:

personnummer

køn, alder, bopælskommune, ægteskabelig stilling, familieforhold, afsluttede erhvervsuddannelser, tidspunkt for afsluttede uddannelser, ledighedsoplysninger, arbejdsløshedsforsikringsoplysninger, beskæftigelsesoplysninger, socioøkonomisk gruppe, branche, indkomst- og beskæftigelsesforhold for hovedperson og ægtefælle, uddannelsesforhold for hovedperson og ægtefælle, arbejdspladsforhold for hovedperson og ægtefælle, holdning til arbejds- og familieliv, børnepasningsordninger

Kilder	Spørgeskema Danmarks Statistiks Befolkningsstatistikregister Danmarks Statistiks Uddannelsesstatistikregister Danmarks Statistiks Arbejdsløshedsstatistikregister Danmarks Statistiks Beskæftigelsesstatistikregister
Anvendelse	Registret anvendes som grundlag for fremstilling af statistisk materiale, der analyseres i Socialforskningsinstituttet.

Udsnitsarkivet

Formål	Udsnitsarkivet danner grundlag for at udnytte allerede indsamlet data-materiale i nye sammenhænge eller efter nye metoder og udgør beredskab for tidsrække- og tidsforløbsanalyser.
Indhold	Registret indeholder oplysninger om udsnit af personer fra andre af Danmarks Statistiks statistikregistre. Der indgår følgende delregistre: Beskæftigelsesundersøgelser 1969-78 Arbejdsstyrkeundersøgelser 1975- Indkomststatistikudsnit 1970- Selvangivelsesundersøgelser 1972-75 Arbejdsløshedsstatistik 1976- Befolkningsstatistik 1976- Uddannelsesstatistik 1976- Beskæftigelsesstatistikregisterudsnit 1991- Arbejdsklassifikationsmodul 1976-

Kilder	Registre i Danmarks Statistik, jf. ovenfor
---------------	--

Anvendelse	Statistiske analyser
-------------------	----------------------

Valgstatistikregistret

Formål	Registret har til formål at danne grundlag for statistikken vedrørende valg til amtskommunale og kommunale råd.
---------------	---

Indhold	<p>Registret indeholder oplysninger om personer, der opstilles ved valg til amtskommunale og kommunale råd i Danmark. Der indgår følgende oplysninger:</p> <p><i>personnummer, bopæl</i></p> <p><i>liste, amt (for kommunevalg), kommune/Amtskommune, kode for valgt, dato for tilflytning til adressen, ægteskabelig stilling, fødested, statsborgerskab, leje- eller ejerbolig, antal personer på adressen, antal børn på adressen, indkomst efter art, stilling, socioøkonomisk gruppe, branche, offentlig/privat ansættelse for lønmodtagere</i></p>
Kilder	<p>Kandidatlister/valgbøger Danmarks Statistiks Befolkningsstatistikregister Danmarks Statistiks Bygnings- og boligstatistikregister Danmarks Statistiks Indkomststatistikregister</p>
Anvendelse	Kommunalvalgsstatistik.

2. Erhvervsstatistik

Det erhvervsstatistiske registersystem

Formål	Registret har til formål at danne grundlag for varetagelsen af Danmarks Statistiks erhvervsstatistiske opgaver samt opgaver for Det Centrale Erhvervsregister.
Indhold	<p>I registret indgår oplysninger om såvel juridiske enheder (firmaer) som fysiske enheder (filialer, fabrikker og andre arbejdssteder). Registret må om personer kun indeholde oplysninger, såfremt de pågældende som erhvervsdrivende eller arbejdsgivere er ansvarlige indehavere af registrerede virksomheder. Der indgår følgende oplysninger:</p>

1. Basisregistret

enhedens navn, adresse(r), SE-nummer, VIRK-nummer (identifikationsnummer i momsafregningssystemet), aktieselskabs- og anpartsselskabsnummer mv., erhvervsregisternummer (JUR-nummer), arbejdsstedskode, ansvarlige indehaveres navne, ansvarlige indehaveres personnumre, bedriftsnummer

ejerform, art, samlet beskæftigelse, omsætning, Told- og Skatte-region

2. Delregistrene

a. Vedrørende virksomheder inden for flere erhvervsområder

løn mv., ATP/AUD-bidrag, momsafregninger, eksport og import, resultatopgørelse, balance, investeringer, driftsregnskabsoplysninger, mellemværender med udlandet, offentlige og koncessionerede virksomheder, lønomkostninger, funktionærlønninger, udenlandsk ejede firmaer

b. Vedrørende industri

beskæftigelse og løn, ordrer og omsætning, salg af enkeltvarer, råvarekøb, energiforbrug, investeringer, lagerværdi, konjunkturbedømmelse, investeringsforventninger, foreningsbekæmpende foranstaltninger, resultatopgørelse og balance

c. Vedrørende engroshandel

priser, lagerværdi

d. Vedrørende detailhandel

omsætning, priser

e. Vedrørende bygge- og anlægsvirksomhed

beskæftigelse, bruttolønsummer, konjunkturforventninger

f. Vedrørende transport

transportydelser på lastbiler, danske skibes fart på udlandet, regnskabsoplysninger vedrørende rederier

g. Vedrørende turisme

overnatninger i hoteller og på campingpladser, erhvervmæssig udlejning af danske feriehuse, rejsebureauer

h. Vedrørende øvrige erhvervsområder

smede- og maskinreparationsværksteder, maskinstationer, kornhandlere, foderstofhandlere, producenter af foderblandinger, skove, landbrugskonsulenter

i. Vedrørende alle delregistrene

identifikationsoplysninger, registertekniske koder og andre oplysninger til brug ved skemaudsendelse mv.

Kilder

Spørgeskemaer
 Told- og Skattestyrelsens moms-afregningssystem
 Told- og Skattestyrelsens import- og eksportregistre
 Told- og Skattestyrelsens Oplysningsseddelregister (COR)
 Told- og Skattestyrelsens Salgs- og Vurderingsregister (SVUR)
 Told- og Skattestyrelsens Selskabsskatregister (SSR)
 Told- og Skattestyrelsens ligningssystem for erhvervsdrivende (SLS-E)
 ATP-arbejdsgiverregister
 Det centrale erhvervsregister
 Centralregistret for motorkøretøjer
 Danmarks Nationalbank
 Danmarks Statistiks Landbrugsstatistikregister
 Danmarks Statistiks Arbejdspladsstatistikregister

Anvendelse

Løbende erhvervsstatistik
 Erhvervsstatistiske specialopgørelser og serviceopgørelser
 Udbygning og opdatering af Det centrale erhvervsregister

I andre statistikregistre:

Arbejdsklassifikationsmodulet
 Beskæftigelsesstatistikregistret
 Lønstatistikregistret for den private sektor
 Arbejdspladsstatistikregistret

Landbrugsstatistikregistret

Formål

Registret danner grundlag for indhentning af oplysninger vedrørende landbrug og gartneri samt for statistiske opgørelser vedrørende samme.

Indhold

I registret indgår oplysninger om teknisk-økonomiske driftsenheder, der frembringer landbrugs- eller gartneriprodukter (bedrifter). Registret må om personer kun indeholde oplysninger, såfremt de pågældende er ejere eller brugere af registrerede enheder. Der indgår følgende oplysninger:

bedriftens nummer, personnummer, kommunkode, jordtilliggende, skovareal, landbrugsarealet og dets anvendelse, diverse ejendomsoplysninger fra de kommunale ejendomsregistre, arbejdskraft, driftsform (branche), husdyr, maskiner og anlæg, nyopførte væksthuse, kornbeholdninger, produktion af spiselige væksthuseprodukter

Kilder	Spørgeskemaer De kommunale ejendomsregistre Det centrale erhvervsregister (CER) Det centrale personregister (CPR)
Anvendelse	Den løbende landbrugsstatistik Landbrugsstatistiske specialopgørelser og serviceopgørelser <i>I andre statistikregistre:</i> Det Erhvervsstatistiske registersystem

3. Bygnings- og boligstatistik

Bygnings- og boligstatistikregistret

Formål	Registrets formål er at danne grundlag for statistik vedrørende bygninger samt erhvervs- og boligenheder i disse.
Indhold	Registret omfatter bebyggede ejendomme beliggende i Danmark. Oplysningerne i registret er på ejendoms-, bygnings- og enhedsniveau. Der indgår følgende oplysninger:

Ejendomsoplysninger:

ejerforhold, vand- og afløbsforhold, størrelse i kvm. (bygningernes samlede areal til beboelse og/eller erhverv), antal boliger

Bygningsoplysninger:

hovedsagelig anvendelse, opførelsesår, varmeinstallation og opvarmningsmiddel, antal etager, antal boliger, samlet bygningsareal, boligareal, erhvervsareal, bebygget areal, tagetageareal, kælderareal, tagdæknings- og ydervægsmateriale

Enhedsoplysninger:

adresseidentifikation, ejerlejlighed, boligenhedens anvendelse, boligtype, dato for enhedens ibrugtagning, samlet areal, areal til beboelse, erhvervsareal, samlet antal værelser, antal værelser til erhverv, toiletforhold, badeforhold, køkkenforhold, installationsforhold (kombinerede), udlejningsforhold (Ejer-/lejerbolig)

Husstandsoplysninger:

adresseidentifikation, husstandstype, antal børn i forskellige aldersklasser, antal voksne i forskellige aldersklasser, antal perso-

ner i husstanden, yngste voksne persons alder og køn, ældste voksne persons alder og køn

Personoplysninger:

adresseidentifikation, køn, alder, ægteskabelig stilling, statsborgerskab, husstandstype

Kilder

Kort- og Matrikelstyrelsens Bygnings- og Boligregister (BBR)
Det kommunale ejendomsstamregister
Danmarks Statistiks Befolkningsstatistikregister

Anvendelse

Løbende boligstatistik
Løbende statistik vedr. bygningsanvendelse
Fertilitetsforskning
Special- og serviceopgaver

I andre statistikregistre:

Oplysninger fra registret indgår i en lang række personstatistikregistre

Bilag 3. Milepæle i registerstatistikken siden 1968

Danmarks Statistiks virksomhed samt de vigtigste ændringer og planer omtales i Danmarks Statistiks arbejdsplan, som udgives årligt. I det følgende omtales nogle hovedpunkter i udviklingen, specielt hvad angår anvendelsen af registermetoder:

- Med virkning fra 1968 etableres en statistik vedrørende omsætning mv. i samtlige erhverv, baseret på oplysninger fra momsadministrationen kombineret med data fra Danmarks Statistiks Erhvervsregister.
- Fra 1971 iværksættes en tilsvarende statistik vedrørende erhvervenes beskæftigelse og fra 1974 tillige om udbetalt lønsum baseret på arbejdsgivernes indberetninger til ATP-ordningen henholdsvis skattemyndighederne.
- Efter etableringen af disse opgørelser er udarbejdelse af generelle erhvervstællinger, der blev udarbejdet for 1925, 1935, 1948 og 1958, ophørt.
- Pr. 1. maj 1970 udarbejdes første gang en totalopgørelse, baseret på CPR, af befolkningen i de enkelte kommuner med fordeling på køn, alder og ægteskabelig stilling. I perioden 1970-73 gennemføres en omlægning af den øvrige befolkningsstatistik til CPR-basis, således at manuelle indberetninger fra kommunerne kan ophøre.
- Fra 1970 omlægges indkomststatistikken i forbindelse med kilde-skattens indførelse til registerbasis. Fra 1976 rekonstrueres statistikken væsentligt, bl.a. ved at indkomstoplysningerne samkøres med oplysninger fra et arbejdsklassifikationssystem, som også anvendes til andre formål.
- I 1971 etablerer Danmarks Statistik et Udsnitsarkiv til brug for udviklingen af registerstatistiske metoder.
- Efter beslutning i 1971 overtager Danmarks Statistik i 1973 fra Undervisningsministeriet statistikken vedrørende elever og studerende. De individuelle oplysninger, som undervisningsinstitutionerne indberetter, indgår i et uddannelsesstatistisk register, således at institutionernes løbende indberetninger kun skal omfatte ændringer i forhold til året i forvejen.
- I 1974 beslutter økonomiministeren, at der ikke skal gennemføres en traditionel, skemabaseret folketælling i folketællingsterminen 1975/76, jf. lov om Danmarks Statistik. I stedet gennemføres for 1976 en registerbaseret folketælling på basis af CPR-oplysninger kombineret med en række forskellige andre registeroplysninger vedrørende erhvervsforhold mv.
- Fra 1977 etableres en bistaandslovsstatistik baseret på personorienterede oplysninger fra det fælleskommunale økonomisystem suppleret

med skema-indberetninger fra de ikke-tilsluttede kommuner og kombineret med familieoplysninger fra befolkningsstatistikken. Denne statistik giver anledning til megen offentlig debat om det betimelige i at benytte personnumre i statistikproduktionen og fører til en retssag mellem Danmarks Statistik og 5 kommuner, som nægter at afgive oplysningerne. I 1980 vinder Danmarks Statistik sagen ved Højesteret.

- Pr. 1.4.1977 gennemføres den første bygnings- og boligopgørelse på basis af det nyoprettede Bygnings- og boligregister, BBR. Pr. 1. januar 1980 gennemføres den første registerbaserede bolig-tælling, som kombinerer oplysninger om boligerne med oplysninger om beboerne.
- I 1979 besluttet det at oprette et Arbejdspladsstatistikregister baseret på eksisterende registeroplysninger og en begrænset supplerende dataindsamling med bistand af skattemyndighederne og arbejdsgiverne. Dette indebærer bl.a., at totale dataindsamlinger fra befolkningen i form af traditionelle folketællinger definitivt ophører.
- Med virkning fra 1979 bliver statistikken om strafferetlige afgørelser, som før var summarisk og skemabaseret, omlagt til at basere sig på årlige indberetninger fra Rigspolitiets centrale kriminalregister.
- Pr. 1.1.1981 gennemføres den første egentlige folke- og bolig-tælling på registerbasis.
- I løbet af 1982-84 oprettes en række sociale statistikregistre, som bygger på udtræk fra de landsdækkende fælleskommunale udbetalings-systemer. En betydelig manuel arbejdsbyrde ophører samtidig med, at de statistiske anvendelser bliver drastisk forøget.
- I 1992 oprettes Motorkøretøjsstatistikregistret baseret på udtræk fra Rigspolitiets centrale motorregister. Registret danner grundlag for opgørelser om ejerforhold og anvendelse af motorkøretøjer.
- I 1992 oprettes en ny sundhedsstatistik baseret på udtræk fra Landspatientregistret kombineret med en række baggrundsoplysninger fra eksisterende statistikregistre om bl.a. befolknings-, social-, indkomst-, beskæftigelses-, uddannelses- og boligforhold.
- I 1994 etableres en udvidet statistik om personer i aktiveringsordninger og tilbagetrækningsordninger baseret på personorienterede indberetninger fra de administrative myndigheder.
- I 1994 etableres, efter flere års omfattende udviklingsarbejde, et samlet, moderniseret erhvervsregistersystem med en række indholdsmæssige og tekniske forbedringer.
- I 1994-95 gennemføres en omfattende lønstatistisk reform, som vil føre til samlede og konsistente personorienterede opgørelser for alle lønmodtagere i både den private og den offentlige sektor.

Bilag 4. Oversigt over identifikationer

I dette bilag gives en oversigt over de vigtigste identifikationer, der benyttes i den registerbaserede statistik. I afsnit 1 ses en liste i alfabetisk orden over identifikationerne, og for hver af dem gives en kort beskrivelse samt en kildeangivelse. I afsnit 2 anføres de vigtigste enhedstyper med deres identifikationer.

1. Identifikationer

Adresse

Enhed: Adressen identificerer en beliggenhed og svarer til den fulde postadresse. Den kan både identificere en *bolig* og et *arbejdssted*.

Fødested: Adressen fødes primært i CPR i Indenrigsministeriet i forbindelse med registrering af bopæl for personer. Desuden fødes den i BBR i Boligministeriets regie i forbindelse med registrering af nye bygninger og boliger og i SE-registret i Told- og Skattestyrelsens regie i forbindelse med registrering af virksomheder beliggende i områder, der ikke er omfattet af CPR.

Opbygning: Adressen er et felt på 17 cifre, der består af flg. felter:

Kommunekode	3 cifre numerisk
Vejkode	4 cifre numerisk
Husnummer	3 cifre numerisk
Husbogstav	1 ciffer alfabetisk
Etage	2 cifre numerisk/blank
Side/dørnummer	4 cifre alfabetisk

Arbejdsstedskode

Enhed: Arbejdssteds-koden identificerer *arbejdsstedet*, når det suppleres med SE-nummer, Hovedkontonummer eller Kommunekode. Arbejdssteder, der består af en hel juridisk erhvervsenhed, tildeles normalt ikke nogen arbejdssteds-kode (eller har en blank arbejdssteds-kode), idet SE-nummeret eller Hovedkontonummeret i sig selv identificerer arbejdsstedet.

Fødested: Det centrale Erhvervsregister, der føres af Danmarks Statistik.

Opbygning: Arbejdssteds-koden er et 4-cifret, numerisk, informationsløst løbenummer, der tildeles inden for en juridisk enhed, en

kommune eller en hovedkonto. Numrene tildeles som hovedregel startende fra 1.

CIR-nummer

CIR-nummeret er forløberen for det nuværende SE-nummer og anvendes indtil 1. april 1985. CIR står for det Centrale Indeholdelsespligtige-Register.

Enhed: Administrativ enhed, der sædvanligvis svarer til en juridisk enhed. En juridisk enhed kan i visse tilfælde have haft flere CIR-numre.

Fødested: Skattevæsenet.

Opbygning: 7 numeriske cifre, hvoraf sidste er et checkciffer (1-9), der er udregnet efter modulus 11, jf. Personnummer.

Ejendomsnummer

Enhed: Ejendomsnummeret identificerer ejendommen som beskatningsenhed i vurderingsmæssig henseende. Arealer, som i matriklen er anført under ét matrikelnummer, får tildelt samme ejendomsnummer. Det samme gælder arealer anført under flere matrikelnumre, hvis de i følge notering i matrikel og tingbog skal holdes forenede.

Fødested: Ejendomsnummeret fødes i de kommunale forvaltninger, Det Fælleskommunale Ejendomsstamregister (ESR), ved udmatrifulering af grundarealer mv.

Opbygning: Et 9-cifret felt bestående af kommunekoden + et løbenummer på 6 cifre.

Hovedkontonummer

Enhed: Bevillingsformålene på den statslige kontoplan fastlægges på hovedkontoer, dvs. bevillingerne gives og styres på hovedkontoniveauet. Da der er placeret økonomisk styring og kontrol på dette niveau, anvendes hovedkontoen i den statslige sektor som den økonomiske enhed, hvortil de statslige arbejdssteder knyttes.

Fødested: Finansministeriet i forbindelse med kontoplanens opbygning til brug for statsbudgettet.

Opbygning: De første seks cifre på den 10-cifrede kontoplan. De to første cifre angiver ministerområdet, det tredje ciffer hovedområdet, det fjerde ciffer aktivitetsområdet og det 5.-6. ciffer hovedkontoen.

Kommunekode

Enhed: Kommunekoden identificerer *kommunen*, hvilket både kan forstås som en beliggenhedskommune (i denne betydning indgår den i adressen, jf. ovenfor) og myndighedskommunen. Det er i den sidste betydning, koden anvendes til at identificere juridisk erhvervsenhed og arbejdssted. Der anvendes forskellige kodeværdier for primærkommuner og amtskommuner.

Fødested: Koderne tildeles af Indenrigsministeriet i forbindelse med CPR.

Opbygning: Kommunekoden er et trecifret numerisk felt, som er tildelt i geografisk rækkefølge, således at kommunerne i et amt udgør et interval.

Matrikelnummer

Enhed: Matrikelnummeret identificerer den enkelte matrikel.

Fødested: Nummeret administreres af Kort- og Matrikelstyrelsen.

Opbygning: Matrikelnummeret består af:

Kommunennummer, 3 cifre

Ejerlav (herredsnummer, sognenummer, løbenummer), 5-7 cifre

Matrikelnummer, 3-5 cifre

Litra, 1-3 bogstaver

Evt. lodnummer, 3 cifre

Personnummer (PNR)

Enhed: Personnummeret identificerer enheden *person*.

Fødested: Fødes og vedligeholdes i CPR, der drives af Indenrigsministeriet. PNR benyttes som identifikation i næsten alle offentlige personregistre.

Opbygning: Personnummeret er rent numerisk og består af 10 cifre, som inkluderer et checkciffer. Nedenstående beskrives med et eksempel, hvorledes nummeret er opbygget:

Nummer: 03 06 36 1171

Løbenummer:	1171
Fødselsdato:	03 06 36
Fødselsdag:	03
Fødselsmåned:	06
Fødselsår:	36
Fødselsårhundrede:	Det første ciffer i løbenummeret
Checkciffer, køn:	Det sidste ciffer i løbenummeret

Der er her tale om en mand, der er født den 3. juni 1936.

Checkcifferet beregnes ved hjælp af modulus 11 algoritmen ud fra de 9 andre cifre.

Beregningen foretages således:

1. Cifrene multipliceres med konstanterne

4, 3, 2, 7, 6, 5, 4, 3, 2

og resultaterne adderes:

$$\begin{array}{cccccccccc}
 0 & 3 & 0 & 6 & 3 & 6 & 1 & 1 & 7 & \\
 \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
 4 & 3 & 2 & 7 & 6 & 5 & 4 & 3 & 2 & \\
 \hline
 0 & + & 9 & + & 0 & + & 42 & + & 18 & + & 30 & + & 4 & + & 3 & + & 14 & = & 120
 \end{array}$$

2. Summen divideres med 11 og divisionens rest findes:

$$120 : 11 = 10$$

$$\begin{array}{r}
 110 \\
 \underline{10} \\
 10
 \end{array}$$

3. Checkcifferet findes som (11- resten):

$$11 - 10 = 1$$

Checkcifferet er altså 1.

Visse kombinationer af de første 9 cifre vil give et checkciffer på 10, og de kan ikke benyttes. Hvis resultatet af beregningen er et checkciffer på 11 (dvs. en rest på 0), skal checkcifferet være 0.

Når checkcifferet som her er ulige, skal nummeret tilknyttes en mand. Kvinder får lige checkcifre.

Cifferet for århundrede tildeles sådan, at personer, der er født i et ulige århundrede (fx 1900-tallet), tildeles værdierne 0-4, mens personer født i et lige århundrede (1800-tallet og i nær fremtid 2000-tallet) tildeles værdier i intervallet 5-9⁸⁸.

⁸⁸ Denne regel om århundrede er fra 1993 modificeret, fordi man har fået pladmangel, dvs. man har været tæt på at komme op på flere fødte på en dag end de ca. 450, der kan rummes. Herefter må koden 4 kun tildeles til fødte i ulige århundrede, hvis deres fødselsår er (19)37-99, mens koden 9 bruges til ulige århundrede, hvis året ligeledes er (19)37-99. Man forventer, at systemet på denne måde kan holde indtil år 2036.

SE-nummer

Enhed: SE-nummeret identificerer en enhed, som er anmeldt hos Told- og Skattestyrelsen som indeholdelsespligtig for skat mv. eller i henhold til told- og afgiftslovgivningen⁸⁹. Enheden er en administrativ enhed, der i det store og hele er identisk med en juridisk enhed. En juridisk enhed kan i visse tilfælde have flere SE-numre. Der kan oprettes SE-numre til brug for afregning for flere juridiske enheder.

Fødested: Nummeret fødes i og vedligeholdes af Told- og Skattestyrelsen.

Opbygning: SE-nummeret er et 8-cifret numerisk felt, hvor de første 7 cifre er vilkårligt tildelt (informationsløse⁹⁰), mens det sidste ciffer er et checkciffer (0-9), der er beregnet på grundlag af de øvrige 7 ved hjælp af modulus 11, jf. Personnummer.

VIRK-nummer

VIRK-nummeret (momsnummer) benyttedes indtil 2. april 1991 i VIRK-registret. Fra 2. april 1991 benyttes SE-nummer som identifikationsnummer i VIRK-registret. VIRK-registret har fra dette tidspunkt flere data fælles med SE-registret.

Enhed: Administrativ enhed, der sædvanligvis svarer til en juridisk enhed. En juridisk enhed kan i visse tilfælde have flere SE-numre. Der kan oprettes SE-numre til brug for afregning for flere juridiske enheder under et.

Fødested: Indtil 2. april 1991; Toldvæsenet. Herefter Told- og Skattestyrelsen (Told- og Skatteregionerne).

Opbygning: 7 numeriske cifre, hvoraf sidste er et checkciffer, der er udregnet efter modulus 11, jf. Personnummer.

⁸⁹ Indtil 2. april 1991 omfattedes kun enheder, der var anmeldt til skattevæsenet, mens toldmyndighederne anvendte et særligt VIRK-nummer (momsnummer).

⁹⁰ Enheder registreret i CIR-registret eller VIRK-registret fik konverteret hhv. CIR-nummer og VIRK-nummer til SE-nummer. Hvis 7. ciffer har værdien 1 eller 2, er SE-nummer dannet på grundlag af CIR-nummer, men hvis 7. ciffer har værdien 5 eller 6, er SE-nummer dannet på grundlag af VIRK-nummer.

2. Enheder

De vigtigste enheder er:

Arbejdssted

Identificeret ved en af flg.:

SE-nummer & Arbejdsstedskode (ikke offentlige)

Hovedkonto & Arbejdsstedskode (staten)

Kommunekode & Arbejdsstedskode (kommunerne)

Bolig

Identificeret ved Adressekode

Ejendom

Identificeret ved Ejendomsnummer

Juridisk erhvervsenhed

Identificeret ved et SE-nummer (JURNR), Hovedkonto eller Kommunekode

Person

Identificeret ved Personnummer

Bilag 5. Anvendte begreber og forkortelser

Adressekode	Kode for bolig- og erhvervsenheders beliggenhed. Koden identificerer den enkelte enhed.
Afledt variabel	Variabel dannet med flere andre variable som informationsgrundlag.
AKM	Danmarks Statistiks ArbejdsKlassifikationsModul, jf. bilag 2.
Arbejdsstedskode	Identifikation af det enkelte arbejdssted i et firma eller en offentlig forvaltning.
ATP	Arbejdsmarkedets TillægsPension. En obligatorisk aldersforsikring for lønmodtagere i alderen 16-66 år med mindst 10 timers ugentlig arbejdstid.
BBR	Kort- og Matrikelstyrelsens Bygnings- og BoligRegister, jf. bilag 6.
CAPI	Computer-Aided Personal Interviewing. Samme teknik som <i>CATI</i> , men benyttet under personligt interview.
CATI	Computer-Aided Telephone Interviewing. Indebærer, at dataene registreres direkte på edb-medium under selve interviewet. Ofte kombineret med en samtidig maskinel fejlretning. Se også <i>CAPI</i> .
CER	Det Centrale ErhvervsRegister, jf. bilag 6.
Cold-deck	Imputeringsmetode (se <i>Imputering</i>), hvor man anvender data fra en tidligere gennemført tælling, jf. afsnit 4.4. Se også <i>Hot deck</i> .
COR	Told- og Skattestyrelsens Centrale OplysningsseddelRegister.
CPR	Det Centrale PersonRegister, jf. bilag 6.
CRAM	Arbejdsmarkedsstyrelsens Centrale Register for ArbejdsMarkedsstatistik, jf. bilag 6.
Databrud	Ændret definition af en statistikvariabel mellem to opgørelser, oftest forårsaget af ændringer i det administrative datagrundlag.
Dataeditering	Kodning, fejlsøgning, fejlrettelse og redigering af datasæt.
Erhvervsregistret	Almindelig anvendt betegnelse for Det centrale erhvervsregister, jf. bilag 6.

Henvisningsnummer	Personnummer for ægtefælle/registreret partner, mor og far til en given person.
Horizontal integration	Samkøring af samtidige oplysninger fra forskellige statistikområder, ofte i forbindelse med vertikal integration, jf. afsnit 5.2. Se også <i>vertikal integration</i> .
Hot-deck	Imputeringsmetode (se <i>Imputering</i>), hvor man anvender data fra den aktuelle tælling. Se også <i>Cold-deck</i> .
IDA	Danmarks Statistiks Integreerede Database for Arbejdsmarkedsforskning.
Imputering	Metode til opretning af datafejl og -mangler, hvor den fejlagtige eller manglende værdi erstattes med en anden værdi uden fornyet kontakt med dataleverandøren, jf. afsnit 4.4.
Indleverede datasæt	Data, indsamlet af statistikbrugeren selv, som samkøres med registre i Danmarks Statistik med henblik på løsningen af en konkret opgave.
Integrationsregister	Statistikregister dannet ved samkøring af andre statistikregistre.
Integreret dataindsamling	En indsamlingsform som består i, at en administrativ myndighed indsamler oplysninger for en anden myndighed <i>samtidig</i> med indsamling af oplysninger til eget brug, jf. afsnit 4.3.
Klassifikationsmoduler	Klassifikationsmoduler er statistikregistre, jf. nedenfor, som ikke i sig selv er grundlag for statistikproduktion. De er oprettet med henblik på anvendelse sammen med oplysninger fra andre kilder inden for forskellige statistikområder, jf. kapitel 3.
Metadata	Data om data, jf. kapitel 9.
Modeldata	Data om enkelte enheder (personer, virksomheder), men uden identifikation.
PNR	Forkortelse for <i>personnummer</i> , som er det i personregistre generelt anvendte identifikationsnummer for personer, jf. bilag 4.
Population	En afgrænset gruppe af enheder, fx den danske befolkning, alle kvinder mellem 25 og 49 år, personer med en lang videregående uddannelse, etc.
RAS	Danmarks Statistiks Registerbaserede ArbejdstyrkeStatistik.
Record	Fysisk samling af data (i en record), som i edb-behandlingen behandles som en enhed. Et datasæt består af et antal records.

Samkøring	Oplysninger fra flere kilder samles på enhedsniveau i ét datasæt på grundlag af fællesidentifikation af oplysningerne i de anvendte kilder, typisk PNR.
SE-nr.	Nummer, som giver en éntydig identifikation af den enkelte virksomhed. Vedligeholdes i Told- og Skattestyrelsens SE-register, jf. bilag 4.
Statistikregister	Betegnelse for et register, der er oprettet alene med henblik på anvendelse til statistiske formål. Der gælder særlige regler for sådanne registre, jf. kapitel 7.
TASU	Danmarks Statistiks Telefoninterview-baserede ArbejdsStyrkeUnder-søgelse.
TIMES	Forkortelse for Tællingernes Integrerede ME tadata System. Systemet er udviklet i Danmarks Statistik og er udover dets anvendelse som dokumentationssystem et hjælpeværktøj ved planlægning, kodegenerering, vedligeholdelse og produktionsstyring af edb-systemer, jf. kapitel 9.
UKM	Danmarks Statistiks UddannelsesKlassifikationsModul.
Vertikal integration	Samkøring af oplysninger med forskellig tidsreference, jf. afsnit 5.3. Se også <i>Horisontal integration</i> .
VIEW	En beskrivelse af et logisk datasæt, jf. kapitel 9.

Bilag 6. Administrative kilderegistre anvendt i Danmarks Statistik

Dette bilag er en oversigt over de væsentligste administrative registre, som er kilder til opdatering af personstatistikregistrene i Danmarks Statistik pr. 1. september 1994.

Det centrale personregister (CPR)

Myndighed	Indenrigsministeriet
Formål	Formålet med registret er at administrere personnummersystemet, samt at opbevare almindelige personoplysninger.
Indhold	Registret indeholder oplysninger om personer med tildelt personnummer (folkeregistertilmeldte).
Primær anvendelse i Danmarks Statistik	Befolkningsstatistikregistret Adressering i forbindelse med surveys

Det centrale oplysningssedelregister (COR)

Myndighed	Told- og Skattestyrelsen
Formål	Registret har til formål at opsamle årlige oplysninger om udbetaling af arbejds- og overførselsindkomster med henblik på opdatering af grundlaget for beregning af indkomstskat for den enkelte person.
Indhold	Registret indeholder oplysninger om udbetalte enkeltydelser mm. fra danske arbejdsgivere samt fra sociale myndigheder, pensionskasser og lignende i Danmark. Ydelserne er identificeret ved modtagers personnummer og udbetalersens SE-nummer. For oplysninger vedrørende løn er det endvidere anført, på hvilket arbejdssted modtageren har været ansat, hvis den pågældende arbejdsgiver har flere arbejdssteder.
Primær anvendelse i Danmarks Statistik	Indkomststatistikregistret Arbejdsklassifikationsmodulet Arbejdspladsstatistikregistret

Det centrale skatteyderregister - ligningsdelen (CSR-L)

Myndighed	Told- og Skattestyrelsen
Formål	Registrets formål er at danne grundlag for administration af indkomst- og formuebeskatningen i Danmark.

Indhold Registret indeholder oplysninger om personer, der er skattepligtige i Danmark, og for hvem der skal foretages forskudsregistrering eller slutligning. Der indgår bl.a. oplysninger om personnummer, skattepligtige indkomster, fradrag og formuer samt beregnede skatter.

Primær anvendelse i Danmarks Statistik Indkomststatistikregistret
Arbejdsklassifikationsmodulet

Statens salgs- og vurderingsregister vedrørende fast ejendom (SVUR)

Myndighed Told- og Skattestyrelsen

Formål Formålet med registret er at lette statens og kommunernes administration af lov om vurdering af landets faste ejendomme samt andre love vedrørende beskatningsforhold ved afståelse af fast ejendom.

Indhold Registret omfatter alle ejendomme i Danmark og indeholder identifikationer af ejendommen i form af ejendomsnummer og matrikelnummer. Endvidere indgår der oplysninger om ejendommens beliggenhed, ejeren (personnummer eller SE-nummer), størrelse, anvendelse og andre oplysninger af betydning for ejendomsvurderingen samt oplysninger om ejendomsskatter og -afgifter.

Primær anvendelse i Danmarks Statistik Ejendomsstatistikregistret

Det centrale register for arbejdsmarkedsstatistik (CRAM)

Myndighed Arbejdsmarkedsstyrelsen

Formål Formålet med registret er at tilvejebringe et datagrundlag for arbejdsmarkedsstatistikken. Endvidere tjener registret visse administrative formål.

Indhold Registret indeholder oplysninger om personer, der er eller har været medlem af en statsanerkendt arbejdsløshedskasse, samt personer, der som arbejdssøgende henvises fra kommunerne til arbejdsformidlingen. Der indgår oplysninger om personnummer, evt. medlemskab af arbejdsløshedskasse og ledighed.

Primær anvendelse i Danmarks Statistik Arbejdsløshedsstatistikregistret
Statistikregistret vedr. arbejdsmarkedspolitiske foranstaltninger

Bygnings- og boligregistret (BBR)

Myndighed Kort- og Matrikelstyrelsen

Formål Formålet med registret er ved kommunernes foranstaltning at tilvejebringe en systematisk registrering af bolig- og bygningsforholdene til brug for såvel statslige som kommunale myndigheders administration og planlægning.

Indhold Registret indeholder oplysninger om bygninger samt bolig- og erhvervsenheder i Danmark. Der indgår oplysninger om beliggenhedsadresse (identifikation), anvendelse, størrelse og indretning mv.

Primær anvendelse i Danmarks Statistik Bygnings- og boligstatistikregistret

Landspatientregistret

(Hovedregister for patienter på somatiske sygehusafdelinger)

Myndighed Sundhedsstyrelsen

Formål Formålet med registret er at fremskaffe materiale til Sundhedsstyrelsens sygehusstatistik til belysning af sygehusenes aktiviteter og befolkningsgruppernes forbrug af sygehusydelse, samt at forsyne de myndigheder, der er ansvarlige for sygehusplanlægningen, med det nødvendige beslutningsgrundlag.

Indhold Registret indeholder oplysninger om de personer, der siden 1. maj 1976 er udskrevet efter stationær behandling på en somatisk afdeling ved et dansk sygehus. Der indgår oplysninger om personnummer, sygehusnummer, afdelingsnummer, indlæggelsesperiode, diagnoser, behandling mm.

Primær anvendelse i Danmarks Statistik Sygehusbenyttelsesstatistikregistret
Forebyggelsesregistret

Det centrale erhvervsregister

Myndighed Danmarks Statistik

Formål Registret har til formål efter anmodning at give oplysninger om registrerede enheder til offentlige myndigheder og andre, at bistå offentlige myndigheder i erhvervsregistreringsspørgsmål samt at medvirke ved samordning og forenkling af den offentlige erhvervsregistrering.

Indhold Registret indeholder oplysninger om juridiske erhvervsenheder og om eventuelle afdelinger (arbejdssteder). Personer indgår kun, hvis de på-

gældende som erhvervsdrivende eller arbejdsgivere er ansvarlige indehavere af registrerede virksomheder. Der indgår en række identifikationer, bl.a. en eventuel ansvarlig indehavers personnummer samt oplysninger om beliggenhed, ejerform, branche og størrelse.

Primær anvendelse i Danmarks Statistik Det erhvervsstatistiske registersystem
Arbejdspladsstatistikregistret
Arbejdsklassifikationsmodulet

Centralregistret for motorkøretøjer (CRM)

Myndighed Rigspolitichefen

Formål Formålet med registret er at opsamle oplysninger om køretøjer til brug for administrations- og kontrolopgaver vedrørende vægtafgift, forsikringsforhold, syn mv.

Indhold Registret indeholder oplysninger om dansk indregistrerede køretøjer. Ejere er registreret med personnummer eller SE-nummer. Køretøjsbrugere kan være registreret med personnummer.

Primær anvendelse i Danmarks Statistik Motorkøretøjsstatistikregistret

Det centrale kriminalregister

Myndighed Rigspolitichefen

Formål Formålet med registret er at registrere afgørelser mv. i kriminelle sager og udskrive straffeattester og andre meddelelser til brug i strafferetsplejen samt at udskrive kriminalstatistiske oplysninger.

Indhold Registret omfatter to delregistre, et internt arbejdsregister for politiet og et afgørelsesregister. Sidstnævnte indeholder oplysninger om personer, om hvem der er truffet afgørelse i straffesager. Der indgår oplysninger om afgørelsens art og den pågældende persons personnummer.

Primær anvendelse i Danmarks Statistik Kriminalstatistikregistret

Fælleskommunale registersystemer

De enkelte registre Der er etableret fælleskommunale registre inden for en lang række områder. De væsentligste, i forhold til anvendelse som kilder til opdatering af Danmarks Statistiks statistikregistre, er

- Pensionssystemet
- Børnetilskudssystemet

- Boligsikringssystemet
- Boligydelssystemet
- Dagpengesystemet
- Sygesikringssystemet
- Daginstitutionssystemet
- Økonomisystemet

Myndighed	Kommunalbestyrelsen og i visse tilfælde det sociale udvalg er registeransvarlig myndighed for den enkelte kommunes del af registret.
Formål	Formålet med registrene er at lette kommunernes administration inden for de enkelte områder.
Indhold	De enkelte registersystemer indeholder oplysninger om de personer, som modtager ydelser af den omhandlede art, identificeret ved personnummer. Der indgår oplysninger om ydelseernes art, størrelse og tidsmæssige udstrækning.
Primær anvendelse i Danmarks Statistik	Registrene vedrørende socialstatistik

Bilag 7. Deltagerliste for seminar over emnet "Det personstatistiske registersystem", 10.-13. jan. 1994 på Gentofte Hotel.

Fra Danmarks Statistik

Otto Andersen
Lars Borchsenius
Sven Egmosé
Søren Hostrup-Pedersen
Gunvor Højberg
Leo Jensen
Anette B. Jerlach
Lisbeth B. Knudsen
Anitta Lange
Poul Henning Larsen
Lisbeth Laursen
Søren Leth-Sørensen
Kaj Kammer Madsen
Sten Mogensen
Bo Møller
Søren Netterstrøm
Leif Nielsen
Vøgg Løwe Nielsen
Jørn Daugård Pedersen
Anne Nærvig Petersen
Marius Ejby Poulsen
Anna Qvist
Palle Qvist
Gert Schmidt
Bjarne Nave Simonsen
Lone Solbjergøj
Finn Spieker
Lars Thygesen
Carsten Torpe
Thorkild Wedebye
Kirsten Wismer
Hans E. Zeuthen

Eksterne deltagere

Olaf Ingerslev, AKF
Britta Kyvsgård, Kriminalistisk Institut, Københavns Universitet
Peter Maskell, Handelshøjskolen i København
Henrik Waaben, Registertilsynet

Stikordsregister

A

Administrative registre, 15; 23; 35; 58; 219
 Adressekode, 18; 25; 38; 142; 209; 215
 Afledt variabel, 68; 215
 Afstandsmatrice, 84
 AKM, Arbejdsklassifikationsmodulet, 33; 35; 49; 69; 107; 166
 Anonymitet, 133
 Arbejdsklassifikationsmodulet, AKM, 33; 35; 49; 69; 107; 166
 Arbejdsløhedsstatistikregistret, 167
 Arbejdspladsprojektet, 76; 79; 97
 Arbejdspladsstatistikregistret, 32; 94; 168; 208
 Arbejdssted, 122; 125; 214
 Arbejdssted/erhvervsarealenhed-relation, 123
 Arbejdsstedskode, 123; 209; 215
 Arbejdsstedsmodulet, 33
 Arbejdsstyrkeundersøgelsen, 36; 59; 93; 217
 Arbejdsstyrkeundersøgelsesregistret, 169
 Arkivering, 131
 ATP-statistikken, 59

B

Baggrundsdata, 28
 Basisdata, 16; 27
 Basisregister, 16; 62; 126; 153
 BBR, Bygnings- og boligregistret, 16; 21; 122; 208; 221
 Befolkning/datomodul, 33
 Befolkningsstatistikregistret, 33; 45; 63; 105; 109; 162
 Beredskab, 141
 Beskæftigelsesstatistikregistret, 170
 Beskæftigelsesstatus, 33; 69
 Bistandslovsstatistikken, 207
 Bistandslovsstatistikregistret, 184
 Bolig, identifikation af, 214
 Boligregistret, delregister i CPR, 19
 Boligstøttestatistikregistret, 184
 Brancheoplysninger, 94
 Bygnings- og boligregistret, BBR, 16; 21; 122; 208; 221
 Bygnings- og boligstatistik, 205
 Bygnings- og boligstatistikregistret, 205
 Børnetilskudsstatistikregistret, 185

C

CAPI, 88; 215
 CATI, 88; 215
 Centralregistret for motorkøretøjer (CRM), 222
 Checkciffer, 211
 CIR-nummer, 210
 Cold-deck imputering, 86; 215
 COR, Det centrale oplysningsseddelregister, 72; 94; 219
 CPR, Det centrale personregister, 11; 18; 21; 45; 63; 70; 122; 219
 CRAM, Det centrale register for arbejdsmarkedsstatistik, 72; 82; 220

D

Daginstitutionstatistikregistret, 186
 Danmarks Statistiks styrelse, 10; 132
 Databanker, 106; 143
 Databaser, 40; 108
 Databrud, 60; 74; 108; 215
 Dataeditering, 87; 215
 Datakvalitetsprojektet, 93
 Datapolitik, 129
 Datasikkerhed, 27; 99; 131; 141
 Datasikkerhedsreglement, 138
 Det centrale erhvervsregister, 16; 22; 78; 125; 207; 221
 Det centrale kriminalregister, 222
 Det centrale oplysningsseddelregister, COR, 72; 94; 219
 Det centrale personregister, CPR, 11; 18; 21; 45; 63; 70; 122; 219
 Det centrale register for arbejdsmarkedsstatistik, CRAM, 72; 82; 220
 Det centrale skattheyderregister - ligningsdelen (CSR-L), 219
 Det erhvervsstatistiske registersystem, 202
 Det integrerede elevregister, 116
 Diskretion, 141
 Diskretionspolitik, 132
 Dokumentation, 34; 108; 145

E

Editering, 87; 215
 Egenindsigt, adgang til, 129
 Ejendom, identifikation af, 214
 Ejendomme/ejer-relation, 124
 Ejendomsnummer, 210
 Ejendomsregistrering, 21
 Ejendomsstatistik, 32; 38; 121
 Ejendomsstatistikregistret, 32; 175
 Elektroniske postkasse, den, 133; 135; 139
 Enheder, 214
 Enhedsgraf, 39; 42
 Enhedsrelation, 25; 121
 Erhverv/cancerstatistikregistret, 192

Erhverv/dødelighedsstatistikregistret, 193
Erhvervsarealenhed, 123
Erhvervsbeskæftigelsesstatistik, 124
Erhvervsdemografiske opgørelser, 127
Erhvervsindlæggelsesregistret, 193
Erhvervsregistersystem, 125; 208
Erhvervsregistrering, 22
Erhvervsregistret, 16; 22; 78; 125; 207; 221
Erhvervsstatistik, 121; 202
EU, 10; 12; 108; 129; 136

F

Familie, 125
Familie, definition af, 63
Familiestatistik, 63; 65
Fertilitetsdatabasen, 109; 144
Firma, 125
Flytteundersøgelsesregistret, 162
Folke- og boligtælling, 12; 21; 24; 51; 105; 122; 208
Folke- og boligtællingsregistret, 163
Folkeregister, 17; 23
Folkeregistreringsloven af 10. juni 1968, 18
Forbrugsundersøgelsen, 36; 89
Forbrugsundersøgelsesregistret, 176
Forebyggelsesregistret, 191
Formidling, 141
Forskerplaceringsordningen, 133; 135; 138
Forskningsdatabaser, 108; 143
Forældrehenvisningsnummer, 64; 67
Fælleskommunale registersystemer, 222
Færdselsuhedsstatistikregistret, 195

H

Henvisningsnummer, 216
Horisontal integration, 24; 31; 100; 216
Hot-deck imputering, 86; 216
Hovedkonto-nummer, 210
Husstand, 125
Husstands- og familiemodulet, 33; 63; 67
Hændelsesdato, 54

I

IDA, Integreret database for arbejdsmarkedsforskning, 112; 144; 171
 Identifikationer, 53; 209
 Identifikationsdata, 24
 Imputering, 81; 91; 216
 Indholdsanalyse, 40
 Indkomstforløbsregistret, 177
 Indkomststatistikregistret, 177
 Indleverede datasæt, 37; 216
 Integrationsregister, 32; 36; 99; 216
 Integreret database for arbejdsmarkedsforskning, IDA, 112; 144; 171
 Integreret dataindsamling, 29; 61; 75; 97; 216

J

Job-creation, 115
 Job-destruction, 115
 Juridisk enhed, 122; 125
 Juridisk erhvervsenhed, identifikation af, 214

K

Kilderegister, 31; 42; 62; 97; 101
 Kirkebøger, 11; 25; 51
 Klassifikationsmodul, 31; 33; 153; 216
 Kommunekode, 209; 211
 Korrektioner, 55; 72
 Kriminalstatistikregistret, 187
 Kvalitative egenskaber, 41
 Kvalitet, 56; 92; 152
 Kvantitative egenskaber, 41

L

Landbrugsstatistikregistret, 204
 Landspatientregistret, 44; 208; 221
 Lov om Danmarks Statistik, 9; 12; 61; 75; 127; 130
 Lov om Folkeregistre, 17
 Lov om offentlige myndigheders registre, 76; 129
 Lov om private registre, 19; 131
 Lovmodellen, 63; 134; 143
 Løn- og personalestatistikregistret for den offentlige sektor, 178
 Lønstatistik, 71; 208
 Lønstatistikregistret for den private sektor, 179

M

Matrikelnummer, 211
Medicinsk fødsels- og dødsfaldsstatistikregister, 164
Metadata, 153
Modeldata, 133; 135; 137; 216
Modulus 11 algoritmen, 212
Motorkøretøjsstatistikregistret, 196; 208

N

Notatdata, 28

O

Omnibusundersøgelsen, 36
Omnibusundersøgelsesregistret, 196
Områdekodningsmodulet, 33
Oplysningssedler, 76; 94; 122
Opsugning i et arbejdssted, 115

P

Paritet, 111
Pendlingsstatistik, 124
Pensionsstatistikregistret, 187
Person, identifikation af, 214
Person/arbejdssted-relation, 122
Person/bolig-relation, 122
Personnummer, PNR, 18; 62; 211; 216
Personregistrering, 17
Personregistret, delregister i CPR, 18
Personstatistik, 162
Personstatistiksystemet, 31
Population, 216
Postenquete, 36
Præcision, vedr. data, 58
Pålidelighed, vedr. data, 57

R

RAS, registerbaseret arbejdsstyrke statistik, 78; 84; 93; 124
Record, 216
Redundans, 37
Regionalt forløbsregister, 197
Registerbaseret arbejdsstyrke statistik, RAS, 78; 84; 93; 124
Registerbegrebet, 15
Registerforskrifter, 76; 131
Registerlovgivning, 129
Registrerede partnerskaber, 65
Registreringsdato, 54
Relationsenhed, 41
Relevans, vedr. data, 57

Ressourceundersøgelsesregistret, 179
Rigsarkivet, 131

S

Samboende par, 64
Samkøring, 29; 31; 39; 57; 83; 99; 130; 217
Samlevende par, 65
Sammenhængende socialstatistik, 101
SE-nummer, 72; 77; 126; 213; 217
SE-registret, 22
Social- og arbejdsforløbsregistret, 198
Sorte indkomster, 91
Specialdata, 27
Spørgeskemaer, 36
Statens salgs- og vurderingsregister vedrørende fast ejendom, 220
Statistikdata, 29; 51; 79
Statistikregister, 20; 31; 35; 42; 61; 99; 129; 146; 217
Statistikregistret
for arbejdsmarkedsforskning, 171
for arbejdsmarkedspolitiske foranstaltninger, 171
for fertilitetsforskning, 165
vedrørende arbejdsmarkedspolitiske foranstaltninger, 172
vedrørende børns opvækstvilkår, 188
vedrørende den kommunale beskæftigelsesindsats, 173
vedrørende enkeltfagskursister, 181
vedrørende erhvervsprogligt uddannede, 198
vedrørende kvinder i mandefag, 174
vedrørende ledige med kompetencegivende uddannelse, 174
vedrørende sociale processer og boligforhold, 199
vedrørende tidsanvendelsen 1950-88, 200
vedrørende tilpasningsstrategien mellem arbejdsliv og familieliv, 200
Stillingsoplysninger, 70; 79
Survey, 11; 82; 88
SVUR, 220
Sygedagpengestatistikregistret, 190
Sygehusbenyttelsesstatistikregistret, 40; 44; 46; 194
Sygesikringsstatistikregistret, 190
Systemdiagram, 47

T

Tidsreferencer, 26; 54
Tidsseriedata, 106; 151
TIMES, 145; 217
Transaktioner, 42

U

Uddannelseskarriere, 116
Uddannelsesklassifikationsmodulet, UKM, 33; 46; 120; 182

Uddannelsesstatistikregistret, 116; 207
Udskilning fra et arbejdssted, 115
Udsnitsarkivet, 201; 207
Udtræk, 31; 43
UKM, Uddannelsesklassifikationsmodulet, 33; 46; 120; 182

V

Valgstatistikregistret, 201
Vejkode, 209
Vejregistret, delregister i CPR, 19
Velstandsstatistikregistret, 180
Vertikal integration, 100; 106; 217
Videregivelse af data, 129
VIEW, 217
VIRK-nummer, 213
Virksomheders identitet, 113
Virksomheders livsforløb, 114

Æ

Ægtefællehenvisningsnummer, 25; 66

Ø

Økonomisk enhed, 125

