



Statistiska centralbyrån Statistics Sweden

Befolknings-
registret

Kontrolluppgift från arbetsgivare m.fl. KU10 för inkomståret 2004

Uppgiftslämnarens namn och postadress		Inkomsttagarens person-/organisationsnr	
Uppgiftslämnarens person-/organisationsnr		08 Fr.o.m.	09 T.o.m.
Rättsledatum om rätt till uppgift sänds in (ex. 05/0301)	40	Del av år (ex. 05 - 11)	
Delägare m.fl. i fåmansföretag	61	Inkomsttagarens fullständiga namn och postadress	
Arbetsställennummer från SCB			
Skatt	Belopp anges i hela kronor.	Förmåner m.m.	
Avdragen A-skatt	01	41 Bostad småhus	42 Kost
Tjänst	Siffrorna i vänster kanten motsvarar de siffror som finns i inkomstdeklarationen under avsnittet Tjänst.	43 Bostad lej småhus	44 Ränta
Kontant lön m.m.		45 Parkering	47 Annan förmån
03 Kontant bruttolön m.m.	11	48 Förmån har justerats	14
		Kod för förmånsbil	15

Företags-
registret

Registerstatistik

– en introduktion

Handbok 2004:1

Registerstatistik

– en introduktion

Av Anders Wallgren & Britt Wallgren

Handbok 2004:1

Registerstatistik

– en introduktion

Register statistics

– an introduction

Producent: Statistiska centralbyrån, Utvecklingsavdelningen
Producer 701 89 Örebro

Förfrågningar: Anders Wallgren, 019-17 61 98
Inquiries anders.wallgren@scb.se

Britt Wallgren, 019-17 61 98
britt.wallgren@scb.se

Omslag: Ateljén, SCB

Om du citerar ur denna publikation, var god uppge: Källa: SCB, + publikationens namn
©2004 Statistiska centralbyrån

ISSN: 1652-5981
ISBN: 91-618-1235-8

Printed in Sweden

SCB-Tryck, Örebro 2004.07  MILJÖMÄRKT Trycksak 341590

Förord

Registerstatistik – en introduktion har tagits fram inom SCB:s Utvecklingsavdelning för att bidra till förståelsen av SCB:s system av statistiska register.

Administrativa data från olika myndigheter och organisationer används inom SCB för att bygga upp ett system av statistiska register. Utan att kräva en ökad uppgiftslämnarböroda kan dessa administrativa data användas för statistiska ändamål.

Denna handbok beskriver det *system av statistiska register* som är grunden för huvuddelen av den samhällsstatistik som produceras i Sverige. Handboken kan användas av alla som vill förstå principerna för SCB:s registersystem.

Arbetet med denna handbok har utförts av Anders och Britt Wallgren på SCB:s Utvecklingsavdelning.

Ulf Jorner

Chef för Utvecklingsavdelningen

Författarnas förord

Under 1996 startades ett verksprojekt, ”*Registerprojektet*”, med följande huvuduppgifter:

”SCB:s olika register som bygger på administrativa källor skall utvecklas till ett väl fungerande registersystem. Systemets olika delar skall vara klart definierade och genom ökad samverkan skall systemet effektiviseras så att nya integrationsregister kan skapas för att möta nya krav från användarna. Det registerstatistiska metodarbetet skall synliggöras och utvecklas så att registerstatistikens kvalitet kan beskrivas och förbättras.”

Till Registerprojektet knöts en grupp med ca 20 deltagare från olika register inom SCB. Denna kontaktgrupp var mycket betydelsefull för projektarbetet. De som arbetade inom projektet kunde via kontaktgruppen diskutera olika registerproblem med personer som hade mångåriga erfarenheter av registerstatistik.

Som resultat av Registerprojektets arbete startades arbetet med att beskriva registersystemet och det registerstatistiska metodarbetet. I denna handbok ges en översiktlig beskrivning av registersystemet, medan en mer omfattande R&D Report, Wallgren och Wallgren (2004), behandlar registerstatistiskt metodarbete och kvalitet, metadata, IT-verktyg, integritetsskydd och samordning genom standardiserade variabler och populationer.

Anders och Britt Wallgren

Innehåll

Kapitel 1. Vad är registerstatistik?	6
1.1 Vad är en statistisk undersökning?	6
1.2 Vad är ett register och vad är registerstatistik?	7
1.2.1 Registerstatistik – ett exempel	8
1.2.2 Hur gjordes denna registerbaserade undersökning?	9
1.3 En modell av SCB:s statistikproduktion	11
1.4 Hur kan man använda SCB:s registersystem?	12
1.4.1 Systemet används för registerstatistik	13
1.4.2 Registersystemet används för urvalsundersökningar	13
1.4.3 Register och urval kompletterar varandra	14
1.5 Varför skall statistik baseras på administrativa data?	14
1.5.1 För- och nackdelar med surveystatistik och registerstatistik	15
1.5.2 Kostnadsaspekten – är registerstatistik billigt?	16
Kapitel 2. Hur strukturera ett registersystem?	17
2.1 Objekttyper och relationer	17
2.2 Systemet av basregister	21
2.2.1 Befolkningsregistret	22
2.2.2 Aktivitetsregistret	22
2.2.3 Företagsregistret	24
2.2.4 Fastighetsregistret	26
2.3 Grundläggande variabler i registersystemet	27
2.3.1 Länkar	27
2.3.2 Klassifikationer	27
2.4 Hela registersystemet	28
Kapitel 3. Hur skapas ett registersystem?	31
3.1 Administrativa register omvandlas till statistiska register	31
3.2 Systemet får data från många administrativa källor	34
3.3 Helheten skapas genom samordning och samarbete	36
3.4 Skapa ytterligare register och användningar	38
3.4.1 Personstatistik baserad på SCB:s registersystem	38
3.4.2 Varför är denna del av systemet så bra?	39
3.5 SCB:s och övriga SAM:s register samverkar	40

Kapitel 4. Hur skapas ett statistiskt register?	42
4.1 Hur strukturera registerbaserade undersökningar?	42
4.1.1 Metodarbete på olika nivåer	43
4.1.2 Arbetsgången då ett statistiskt register skapas	44
4.2 Att avgränsa registrets objektmängd	45
4.2.1 Att definiera en population	46
4.2.2 Att definiera en population – primära statistiska register	47
4.2.3 Att definiera en population – integrationsregister	48
4.2.4 Basregistren skall användas då populationer avgränsas	49
4.2.5 Kraven på ett basregister	50
4.2.6 Alla skall stötta basregistren!	52
Bilaga 1. Riktlinjer för statistiska register på SCB	55
Bilaga 2. Referenser	56
Bilaga 3. Registerstatistiska termer	57

Kapitel 1. Vad är registerstatistik?

I detta och kommande kapitel introduceras en rad begrepp och principer som kommer att användas i denna bok när vi diskuterar *registerstatistik*, dvs. statistik som bygger på användningen av administrativa data.

1.1 Vad är en statistisk undersökning?

Utgångspunkten för en undersökning är ett antal frågor med anknytning till ett visst problemområde och för att besvara dessa frågor görs en undersökning. Undersökningsprocessen kan beskrivas mer eller mindre detaljerat. Arbetet med en statistisk undersökning består av följande tre faser:

1. problemformulering och planläggning av undersökningen
2. införskaffande av data samt bearbetning av data
3. skattningar bildas, analys och presentation av undersökningens resultat.

Görs undersökningen med statistiska metoder talar vi om en statistisk undersökning. Statistisk metodik är en sammanhängande strategi för att samla in, bearbeta och analysera data. Även i icke-statistiska undersökningar samlar man in, bearbetar och analyserar data, men i statistiska undersökningar samlas data in för att kunna bearbetas och analyseras med *kvantitativa* metoder.

Inom SCB arbetar man som regel med löpande undersökningar – samma undersökning upprepas varje år, kvartal eller månad. Vid sådana undersökningar arbetar man i huvudsak med fas 2 och 3. Men även dessa löpande undersökningar har haft en problemformuleringsfas, som kan ligga långt tillbaka i tiden.

En enstaka undersökning kan vara ett uppdrag där man på SCB får göra hela undersökningen, och då arbetar man med alla tre faserna. Men i många uppdrag gör kunden fas 1 och 3, och på SCB arbetar man endast med fas 2.

Undersökningens fas 2, införskaffande av data, kan göras på olika sätt:

- a. Redan befintliga makrodata används vid en *skrivbordsundersökning*. Makrodata avser sammanfattande uppgifter för grupper av objekt och brukar presenteras som innehållet i statistiska tabeller.
Exempel: Detta är den vanligaste typen av undersökning som SCB:s användare gör när de utnyttjar tabellerna i SCB:s publikationer eller makrodatabaser. Även Nationalräkenskaperna och Miljöräkningskaperna är exempel på undersökningar som bygger på befintliga makrodata.
- b. Redan befintliga mikrodata används vid en *registerbaserad undersökning*. Mikrodata avser uppgifter för enskilda objekt. Befintliga administrativa eller statistiska register med uppgifter som t.ex. avser enskilda personer eller företag används för den registerbaserade undersökningens syften.
Exempel: Ungdomars övergång från studier till arbetsmarknaden. Detta exempel beskrivs i avsnitt 1.2.1 nedan.
- c. Med egen datainsamling vid en *totalundersökning*.
Exempel: Folk- och bostadsräkningen 1990, där alla hushåll och fastighetsägare fick fylla i blanketter som sedan bearbetades av SCB. Eftersom total-

En mer utförlig beskrivning finns i SCB (2001a)

Makrodata

Registerbaserad undersökning

Mikrodata

undersökningar resulterar i att register skapas, kommer även mikrodata från totalundersökningar att ingå i SCB:s registersystem och därigenom vara underlag för registerbaserade undersökningar.

- d. Med egen datainsamling vid en *urvalsundersökning*.
Exempel: Arbetskraftsundersökningen med nytt urval, ny datainsamling och rapportering månadsvis.
- e. Dessutom kan man göra egen datainsamling vid ett *experiment*. Denna insamlingsmetod används inom SCB för att testa olika mätmetoder vid intervjuer eller enkäter.

Eftersom de fyra typerna av icke-experimentella undersökningar (a – d ovan) skiljer sig åt metodmässigt är det lämpligt att begreppsmässigt skilja dem åt. Registerbaserade undersökningar, total- och urvalsundersökningar är de undersökningstyper som viktigast inom en statistisk byrå.

1.2 Vad är ett register och vad är registerstatistik?

Ett *register* avses vara en *fullständig förteckning* över objekten i en viss objekt-mängd eller population, men på grund av kvalitetsbrister kan uppgifter om vissa objekt saknas. Uppgifter om objektens identitet skall finnas så att registret kan uppdateras och utökas med nya variabelvärden för respektive objekt.

Termen register används i denna handbok för register inom SCB:s registersystem. Sådana register baseras antingen på en totalundersökning genomförd av SCB eller på administrativa register från myndigheter och organisationer utanför SCB. Datainsamlingen i en urvalsundersökning ger inte upphov till ett register i denna mening, då mikrodata avseende urvalet endast innehåller en liten del av den undersökta populationens objekt-mängd.

Termerna *observationsregister* och *slutligt observationsregister* används både för urvals-, total- och registerbaserade undersökningar. De används bl.a. inom SCB:s dokumentationssystem SCBDOK.

I bilden nedan jämförs de tre slag av undersökningar som dominerar inom statistikbyråer.

Bild 1.1 Jämförelse mellan de tre typerna av undersökningar

Urvalsundersökning	Totalundersökning	Registerbaserad undersökning
Ingår inte i registersystemet	Ingår i registersystemet	
Egen datainsamling		Använder befintliga register

Med termen *registerstatistik* avses statistik som bygger på registerbaserade undersökningar. När vi diskuterar registersystem, som i kapitel 2 och 3, skiljer vi inte på totalundersökningar och registerbaserade undersökningar. I kapitel 4 där metodfrågor diskuteras, innebär termen registerstatistik enbart registerbaserade undersökningar. Termen *surveystatistik* avser statistik baserad både på total- och urvalsundersökningar.

1.2.1 Registerstatistik – ett exempel

Unga personers möte med arbetsmarknaden efter avslutad utbildning är numera ett viktigt område för olika undersökningar. Sådana studier bör göras som *longitudinella undersökningar*, där grupper av personer följs under en följd av år. Om man gör dessa studier som urvalsundersökningar görs ett urval varje år bland personer som avslutar viss utbildning och varje urval intervjuas eller besvarar enkäter en gång per år under t.ex. sju år. Denna undersökningsmetod har nackdelar. Dels blir uppgiftslämnarbördan tung – de utvalda måste besvara en stor uppsättning frågor varje år, dels kommer bortfallet att gradvis öka under perioden. Har man dessutom inte tillgång till ett bra register över avslutade utbildningar måste man först göra ett stort urval av personer inom vissa åldersgrupper för att hitta de som avslutat gymnasie- respektive högskolestudier.

En alternativ undersökningsmetod är att basera studien enbart på befintliga register. SCB publicerar sådan registerstatistik som beskriver övergången till arbetsmarknaden. Denna statistik är baserad på administrativa källor men för att producera denna statistik är det inte tillräckligt med tillgång till de administrativa källorna. *Det är också nödvändigt ha tillgång till ett fungerande system av statistiska register.* Exemplet används här för att illustrera viktiga egenskaper hos registerstatistik och registersystem.

I tabellen och diagrammen nedan kan vi jämföra personer med lägre och högre utbildning med avseende på deras övergång till arbetsmarknaden. Sex årgångar (*kohorter*) som består av alla personer som slutat gymnasiet 1987-1992 följs under åren 1988-1993 och deras övergång till förvärvsarbete kan jämföras med motsvarande sex kohorter av dem som tagit grundexamen på högskola. Denna tidsperiod är särskilt intressant då arbetsmarknaden förändrades dramatiskt. Alla de personer som tillhör dessa tolv kohorter undersöks genom att man bildar longitudinella register som sedan analyseras. Varje kolumn i tabellen nedan beskriver utvecklingen för en sådan kohort.

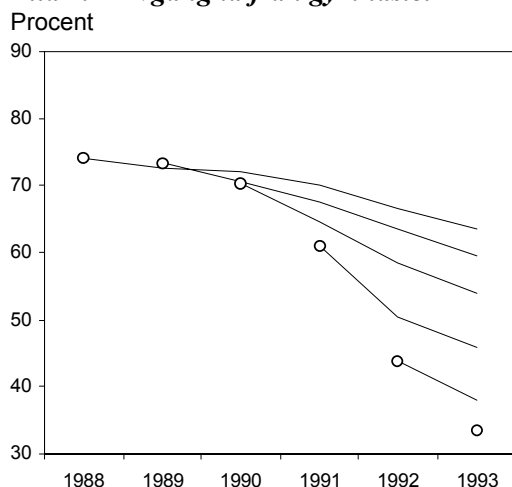
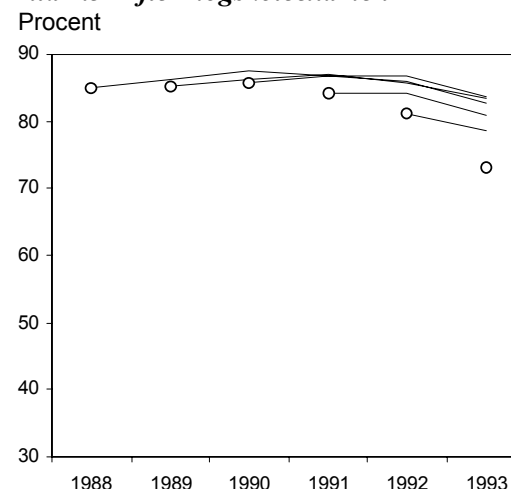
Kohort: en grupp av objekt som har sitt ursprung från en viss period (år)

Tabell 1A. Andel förvärvsarbetande 1988-1993 efter avslutad utbildning 1987-1992

	<i>Avgångna från gymnasiet</i>						<i>Efter högskoleexamen</i>					
	År då utbildningen avslutades:						År då utbildningen avslutades:					
	1987	1988	1989	1990	1991	1992	1987	1988	1989	1990	1991	1992
1988	74,0						85,1					
1989	72,6	73,5					86,2	85,1				
1990	72,1	70,6	70,5				87,4	86,3	85,8			
1991	70,0	67,6	64,5	60,9			86,8	87,0	86,6	84,2		
1992	66,5	63,5	58,5	50,4	43,9		86,8	85,7	86,0	84,3	81,1	
1993	63,5	59,5	54,0	46,0	38,0	33,5	83,8	83,4	82,7	81,0	78,6	73,2
Antal personer i resp. kohort	88 955	88 772	89 407	86 187	88 923	91 987	22 738	22 521	23 335	21 964	23 539	24 850

Anm: Förvärvsarbetande avser endast förvärvsarbetande, de som kombinerar studier och förvärvsarbete ingår inte

Cirklarna i bild 1.2 och 1.3 nedan representerar andelen förvärvsarbetande *ett* år efter avslutad utbildning. Kurvorna visar utvecklingen av andelen förvärvsarbetande inom respektive kohort.

Andel förvärvsarbetande efter avslutad utbildning 1987-1992**Bild 1.2 Avgångna från gymnasiet****Bild 1.3 Efter högskoleexamen**

I början av 1990-talet inträffade den allvarligaste krisen på arbetsmarknaden sedan 1930-talet. Bild 1.2 och 1.3 visar konjunkturedgången i början av 1990-talet som förändrade unga personers möjligheter att komma ut på arbetsmarknaden. Andelen som förvärvsarbetade ett år efter avslutade gymnasiestudier har förändrats dramatiskt från 74 % till 34 % under dessa år.

Enligt andra uppgifter från undersökningen ökade andelen som varken studerade eller förvärvsarbetade från 14 % till 42 % under dessa år bland dem som avslutat gymnasiestudierna. För de som har examen från högskola har utvecklingen inte varit lika dystert – det verkar som att högskolestudier lönar sig.

Detta exempel är baserat på longitudinella data. Varje år är varje person klassificerad som förvärvsarbetande, högskolestuderande eller varken förvärvsarbetande eller studerande. För de förvärvsarbetande är den årliga arbetsinkomsten registrerad och också inom vilken näringsgren varje person arbetar. Personerna är dessutom klassificerade efter kurs/studieprogram, kön och region vilket medger detaljerade redovisningar. Årskostnaden för denna produkt som finns redovisad i Statistiska meddelanden U 81 SM 9501 var 0,4 miljoner kronor varav hälften utgjordes av tryckkostnader.

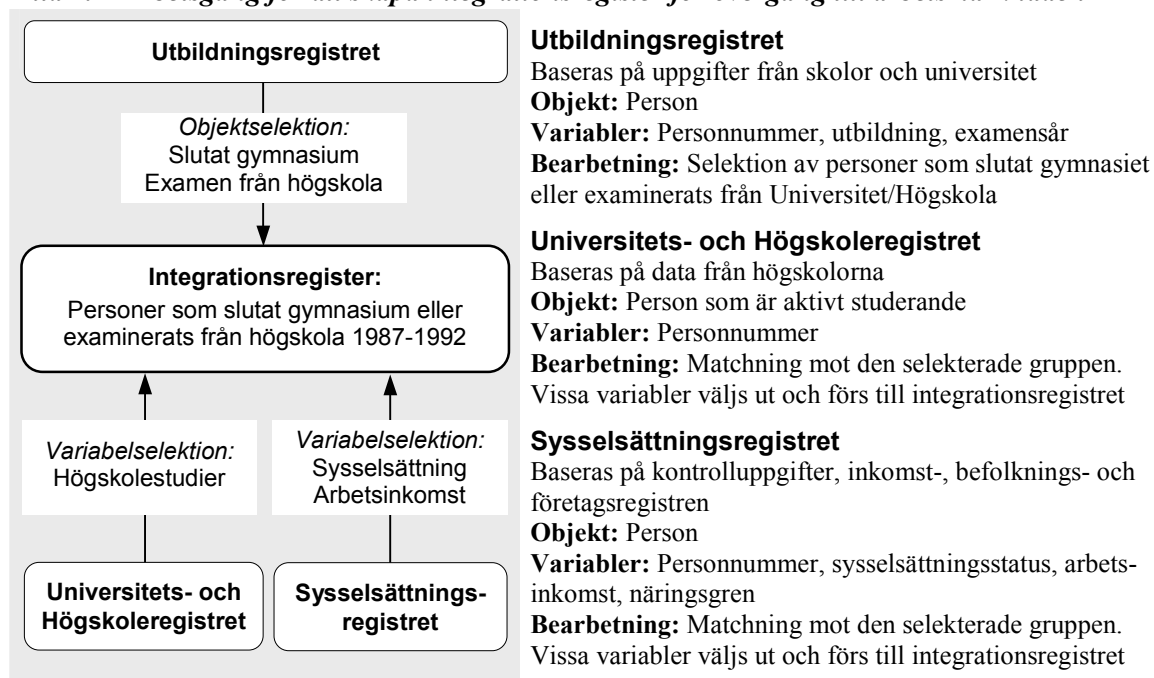
1.2.2 Hur gjordes denna registerbaserade undersökning?

Diagrammen ovan visar unga personers övergång från studier till arbetsmarknaden. För att skapa det longitudinella register som diagrammen bygger på har man gjort på följande sätt:

- Genom att kombinera information i tre SCB-register för åren 1987-1993 (dvs. totalt 21 olika register) skapas ett nytt *integrationsregister*, som markeras med kraftigare ram i bild 1.4 nedan.
- Objekten i det nya registret skapas genom *selektion av vissa objekt* i Utbildningsregistret.

- Genom att *matcha* objekten i det nya registret med motsvarande objekt i Utbildnings-, Universitets- och Högskoleregistret samt Sysselsättningsregistret hämtas variabelvärden till det nya registret. Arbetsgången illustreras i diagrammet nedan.

Bild 1.4 Arbetsgång för att skapa integrationsregister för övergång till arbetsmarknaden



Integrationsregister

Objektselektion

Variabelselektion

Matchning

Lucas

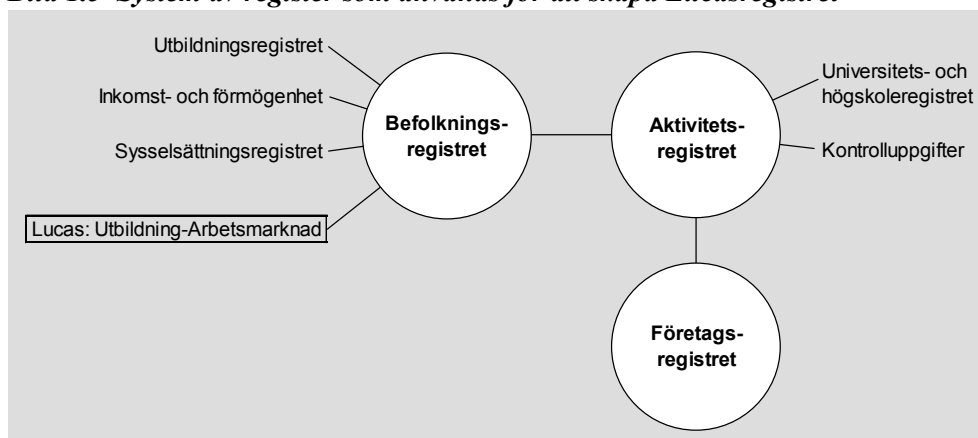
Longitudinellt register
för utbildnings- och
arbetsmarknadsstatistik

Här har en rad viktiga termer införts. Genom att sammanföra eller integrera information från flera register skapas ett *integrationsregister*. Vi skiljer mellan *objektselektion* där vissa objekt väljs ut och *variabelselektion* där vissa variabler väljs ut. *Matchning* innebär att länkarna i två eller flera register jämförs, resultatet blir träff eller icke-träff.

För att skapa det nya integrationsregistret, *Lucas*, behövs ingen ny datainsamling, utan registret bygger på ökad användning av redan befintliga register. Även Sysselsättningsregistret bygger på redan befintliga register i SCB:s registersystem. Däremot bygger Utbildnings-, Universitets- och Högskoleregistret på att SCB får administrativa data från myndigheter utanför SCB.

För att skapa och samordna dessa register fordras dessutom omfattande bearbetningar av registerdata i flera register. Detta innebär att tillgången på administrativa källor inte räcker för att kunna framställa denna statistik. Utan SCB:s samordnade system av statistiska register skulle det vara omöjligt att producera denna statistik.

Åtta register används för att skapa integrationsregistret Lucas i vårt exempel med övergång från studier till arbetsmarknaden. Detta illustreras i bilden nedan.

Bild 1.5 System av register som används för att skapa Lucasregistret

Personnummer är länken mellan Befolnings-, Utbildnings-, Sysselsättnings-, Inkomst-, Lucas-, Aktivitetsregistret och Universitets- och högskoleregistren. *Arbetsställesnummer* är länken mellan Aktivitets- och Företagsregistret. De två kopplingsvariablerna *Personnummer* och *arbetsställesnummer* är tillsammans länken mellan Aktivitetsregistret och Kontrolluppgiftsregistret. Med hjälp av länkarna kan Lucasregistret påföras variabler från alla övriga register i register-systemet.

Här har termerna länk och kopplingsvariabel införts. En *länk* mellan två register består av en eller flera kopplingsvariabler som finns i båda registren. En *kopplingsvariabel* består av information som identifierar enskilda objekt eller grupper av objekt.

Länk

Kopplingsvariabel

De register som används i detta exempel är delar av det registersystem som kommer att presenteras mer ingående i kapitel 2 och 3. Tre av systemets basregister används, Befolnings-, Aktivitets- och Företagsregistret. Dessutom måste det finnas fungerande länkar mellan de inblandade registren för att det skall vara möjligt att skapa Lucasregistret. Den viktiga kopplingsvariabeln personnummer finns i alla registren i bild 1.5 utom Företagsregistret. Länken mellan Aktivitets- och Företagsregistret är företagets organisationsnummer (PeOrgNr) eller arbetsställets identitet (CFAR-nr).

1.3 En modell av SCB:s statistikproduktion

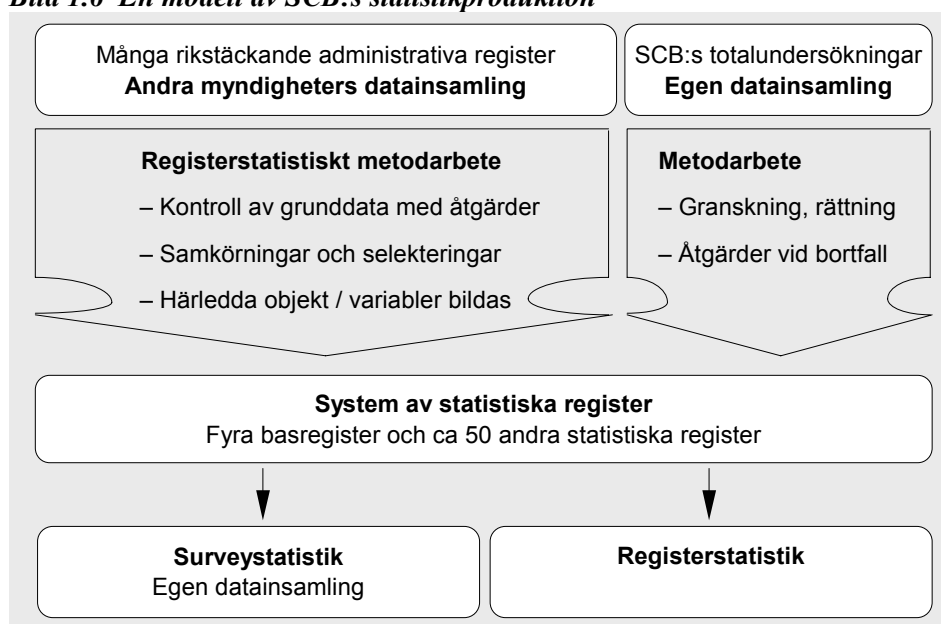
Här måste man skilja mellan de *administrativa register*, som andra myndigheter lämnar till SCB, och de *statistiska register* som SCB har skapat av dessa, efter i många fall omfattande *registerstatistiska bearbetningar*.

Administrativa register

Statistiska register

Registerstatistiska bearbetningar

Systemet av statistiska register är grunden för all statistik som SCB producerar. I alla urvals- och totalundersökningar används register som urvalsram, registervariabler används i stället för att fråga intervjupersonerna samt som hjälpvariabler vid estimationen då hjälpvariabler används för att minska slumpmässiga och systematiska fel. Förutom denna statistik baserad på egen datainsamling producerar SCB också *registerstatistik* som direkt baseras på bearbetningar av uppgifter i SCB:s register.

Bild 1.6 En modell av SCB:s statistikproduktion

Ett statistiskt registersystem består av en rad olika komponenter, förutom ett samordnat system av register krävs också etablerade metoder samt ett fungerande dokumentationssystem:

- *basregister* består av objektmängder av grundläggande betydelse, används för att bilda *standardiserade populationer*
- övriga *statistiska register* innehåller statistiskt intressanta variabler
- *länkar* mellan objekt i olika basregister samt *länkar* mellan basregister och övriga statistiska register
- *registerstatistiska metoder* och rutiner för kvalitetssäkring
- *metadata*, registerdokumentation av definitioner, kvalitet, jämförbarhet
- *IT-verktyg* för registerbearbetningar
- *rutiner för integritetsskydd*
- *standardiserade variabler*, variabler av grundläggande betydelse.

Registerstatistiskt metodarbete och kvalitet, metadata, IT-verktyg, integritetsskydd, standardiserade variabler och populationer diskuteras i Wallgren (2004)

1.4 Hur kan man använda SCB:s registersystem?

Ett statistiskt registersystem kan användas på tre sätt:

- registersystemet används för att framställa registerstatistik
- registersystemet används för urvals- och totalundersökningar
- de statistiska registren och urvalsundersökningarna ses som *ett* system där registerstatistiken och urvalsundersökningarna kompletterar varandra.

1.4.1 Systemet används för registerstatistik

Registersystemet kan användas för ren registerstatistik. Med ett befintligt statistiskt register kan tabeller med *skattningar* produceras och publiceras. Termen skattning används allmänt för urvalsundersökningar, men bör även användas inom registerstatistiken. Även här måste man skilja på det verkliga värdet och den skattning som registret ger.

Exemplet med unga personers övergång från studier till arbetsmarknaden visar hur sådan registerstatistik kan framställas. I avsnitt 3.4 ges exempel på en rad andra registerstatistiska tillämpningar inom individstatistiken.

1.4.2 Registersystemet används för urvalsundersökningar

I samband med urvalsundersökningar kan man använda registersystemet på följande sätt:

- I samband med urvalet används register som urvalsram och registervariabler används för att stratifiera populationen.
- Mätningarna underlättas av att man slipper fråga efter de uppgifter som redan finns i befintliga register.
- Under estimationsfasen kan man använda registervariabler som hjälpinformation för att öka precisionen och kompensera för bortfallet.

Urval

Basregistren används som *urvalsram* och registervariabler kan användas för att göra ett stratifierat urval.

Exempel: För Arbetskraftsundersökningarna används Befolkningsregistret för att göra urval bland befolkningen i åldrarna 16-64 år, där urvalet stratifieras efter region, kön och medborgarskap enligt Befolkningsregistret samt förvärvsarbetande/ej förvärvsarbetande enligt Sysselsättningsregistret.

Datinsamling – mätningar

Att inte fråga efter uppgifter som redan finns i SCB:s register innebär dels att uppgiftslämnarbördan minskar, dels att intervju tid sparas och dels att intervjuarna slipper ställa känsliga och svåra frågor om t.ex. inkomst och utbildning.

Exempel: Inkomstfördelningsundersökningen använder i huvudsak inkomstvariabler från register. De utvalda personerna behöver bara besvara frågor om hushållets sammansättning, yrke, om de arbetar hel- eller deltid och om vissa slag av inkomster som inte finns särredovisade i registren.

Ett annat sätt att kombinera en urvalsundersökning med registerdata är att undersöka nuläget med enkät- eller intervjufrågor och att använda registerdata för att undersöka historiken. *Exempel:* I en hälsoundersökning ställs frågor om det aktuella hälsoläget och aktuella levnadsförhållanden. Registerdata kan användas för att belysa utbildningsbakgrund, förvärvsarbetshistorik etc.

Estimationsfasen

Registren har också en viktig roll under estimationsfasen. Registersystemet innehåller många variabler vilka kan användas som hjälpinformation för att göra bättre skattningar. Dessa hjälpvariabler minskar felmarginalerna för skatt-

ningarna vid given urvalsstorlek eller minskar urvalsstorleken och kostnaderna vid ett givet precisionskrav.

Exempel: I estimationsfasen i Arbetskraftsundersökningarna, utnyttjas registervariablerna kön och ålder enligt Befolkningsregistret, förvärvsarbetande efter näringsgren enligt Sysselsättningsregistret samt arbetssökandekategori enligt AMS:s sökanderegister. En rimlig bedömning är att dessa hjälpvariabler minskar felmarginalerna med cirka 20 % vid given urvalsstorlek vilket motsvarar en minskad urvalsstorlek med cirka 36 % för samma precisionskrav.

Källa:
Hörngren (1992)

Bortfallet i urvalsundersökningar är ett stort problem och *bortfallets storlek* ökar. Registerdata kan användas för att minska bortfallets storlek då man slipper fråga efter vissa känsliga variabler som inkomst, utbildning och ålder. Dessutom blir frågeformuläret kortare om man använder registervariabler i stället för att fråga. Även detta kan minska bortfallets storlek.

Källa:
Lundström och Särndal
(2001)

Registervariabler kan också användas för att minska effekten av *bortfallsfel*. Det är då en fördel att ha tillgång till många variabler så att man med ledning av bortfallets struktur och samband med undersökningsvariablerna kan välja den korrigeringsmetod som passar bäst för den aktuella undersökningen.

1.4.3 Register och urval kompletterar varandra

En registerbaserad undersökning och en urvalsundersökning kan komplettera varandra innehållsmässigt. Registerstatistiken ger grundläggande uppgifter om skillnader mellan olika kategorier och om förändringar över tiden. Det är sedan möjligt att göra en urvalsundersökning som ger en utförligare beskrivning av orsakerna till dessa mönster. Med hjälp av registren kan urvalet avgränsas till särskilt intressanta delgrupper.

Exempel: Med Mervärdesskatteregistret skulle man kunna göra ett urval av företag som minskat sin omsättning under det senaste året. De utvalda företagen intervjuas om orsakerna till nedgången och hur man bedömer framtiden.

Exempel: Med Aktivitetsregistret (detta register beskrivs i avsnitt 2.2.2) skulle man kunna göra ett urval av personer som bytt arbetsgivare under de två senaste åren och intervjua dem om orsakerna till bytet.

1.5 Varför skall statistik baseras på administrativa data?

Det finns både fördomar och berättigad kritik mot registerstatistik. Enligt fördomarna är statistik baserad på administrativa data billig men dålig jämförd med ”riktig” surveystatistik. Berättigad kritik kan avse relevansfel, jämförbarhetsproblem och att kvaliteten inte är under kontroll. Tyvärr kan liknande kritik riktas mot urvalsundersökningar, bortsett från att urvalsfelen kontrolleras, kan mätfel och bortfallsfel vara stora och okända.

Svaret på synpunkter av detta slag är att samhällets statistik skall bestå av *både* registerstatistik och surveystatistik. Det är inte fråga om att den ena metodiken är bättre än den andra, utan i vissa situationer är registerstatistik effektivast, medan i andra fall är urvalsundersökningar lämpligast.

1.5.1 För- och nackdelar med surveystatistik och registerstatistik

Det finns en vanlig men ofta diffus uppfattning om att statistik baserad på administrativa data har låg kvalitet: ”Vi har ju ingen kontroll över datainsamlingen.” Är denna uppfattning befogad för de administrativa källor som SCB använder? En mycket stor del av SCB:s registersystem baseras på data från folkbokföring och skatteadministration. Skulle denna statistik ha högre kvalitet om SCB gjorde en egen datainsamling parallellt med Skatteverkets insamling av administrativa data? Så är knappast fallet – SCB:s egna försök att samla in dessa uppgifter skulle vara dyra, öka uppgiftslämnarbördan och ge data med större mätfel.

Från och med den 1/1 2004 bytte Riksskatteverket (RSV) namn till Skatteverket (SKV)

Bild 1.7 För- och nackdelar med surveystatistik och registerstatistik

	Fördelar	Nackdelar
Undersökningar baserade på egen datainsamling	Du kan ställa de frågor du vill Kan vara aktuell	Vissa respondenter förstår inte frågan ... har glömt hur det var ... svarar inte (bortfall) ... svarar slarvigt Uppgiftslämnarbördan kan bli stor Dyrt Låg kvalitet på skattningar för små redovisningsgrupper (gäller urval)
Registerbaserade undersökningar	Ingen ytterligare uppgiftslämnarbörda för statistiken Låga kostnader God täckning av populationens objekt Data för varje år Respondenterna svarar omsorgsfullt på administrativt viktiga frågor Goda möjligheter till redovisning för små grupper, regional statistik och longitudinella studier	Du kan inte ställa frågor själv Du är beroende av de administrativa systemens populations- och objektavgränsningar samt variabeldefinitioner I vissa fall lång tid från händelse till att data blir tillgängliga Förändringar i de administrativa systemen försvårar jämförelser Variabler som är mindre viktiga för det administrativa arbetet kan ha lägre kvalitet

I denna tabell har vi sammanfattat fördelar och nackdelar som finns med de två undersökningsmetoderna. Hur allvarliga nackdelarna är kan vara olika vid olika undersökningssituationer.

Exemplet med ungdomars övergång från studier till förvärvsarbete kan illustrera fördelar och nackdelar med registerstatistik. Vi har redan nämnt att registeransatsen är lämplig vid longitudinella studier. En annan fördel är att man kan redovisa resultat på många undergrupper, olika kurser, studieprogram och regioner. Men en nackdel med vissa typer av administrativa källor är att de administrativa systemen behöver lång tid från händelse till att data blir tillgängliga, detta försenar registerstatistiken. I exemplet publicerades rapporten som beskrev perioden 1987-1993 i juni 1995, dvs. efter ett och ett halvt år.

Det gäller inte enbart *om* man skall utnyttja administrativa data, utan i högsta grad *hur* man skall utnyttja dessa. Svaret på *hur* är att administrativa data som regel inte skall användas som de är, utan skall bearbetas för statistiska syften.

Man skall sträva efter att bygga ett system av samverkande register – detta ger både kvalitetsmässiga och ekonomiska fördelar, och kvalitetssäkring skall finnas som en viktig komponent i systemet.

Det finns många utvecklingstendenser som hotar surveystatistikens kvalitet. Ökande användning av telemarketing och nummerpresentatörer gör det svårare att genomföra telefonintervjuer. Om uppgiftslämnarnas motivation minskar så ökar bortfallet och svars kvaliteten minskar. Det blir allt svårare att motivera dubbelt uppgiftslämnande – varför svara på en enkät om företagets omsättning då man också lämnar momsdeklaration till skattemyndigheten där uppgift om omsättning ingår? Allt detta talar för att registerstatistikens betydelse kommer att öka.

1.5.2 Kostnadsaspekten – är registerstatistik billigt?

”It is quite clear that the Member States find themselves in the paradoxical situation of having to face a number of budget cutbacks at the same time as providing users with an increasing volume of high-quality relevant information.” Med dessa ord inledde Y. Franchet ett seminarium (Eurostat, 1997) som behandlade användningen av administrativa källor för statistiska ändamål. Citatet illustrerar kraven på mer effektiva statistiska system. Dessa krav på ökad effektivitet kan mötas genom att kombinera två strategier:

- Genom att fler administrativa källor används kan kostnaden för datainsamling reduceras både för uppgiftslämnare och statistikproducent.
- Genom mer effektivt och flexibelt utnyttjande av redan befintliga data kan ny efterfrågan mötas utan dyrbar insamling av nya data. Detta uppnås genom att de administrativa källorna används för att skapa ett system av samordnade statistiska register.

Att bygga upp ett sådant registersystem kostar, men då systemet används blir marginalkostnaden låg. I exemplet med ungdomars övergång från studier till arbetsliv var produktens marginalkostnad 0,4 miljoner kronor. Uppgiftslämnarbördan är också en samhällsekonomisk kostnad som kan minskas genom att gå över från surveystatistik till registerstatistik.

Urvalsundersökningar är en allt dyrare metod, då kostnaderna för bortfallsuppföljning ökar. Kostnaderna för urvalsundersökning per invånare är höga för små länder – för samma precision krävs lika stort urval i ett litet land som ett stort land. Det blir därför särskilt viktigt för små länder att bygga upp ett statistiskt registersystem. I Danmark har man konsekvent följt denna strategi.

Kapitel 2. Hur strukturera ett registersystem?

En statistikproducent som SCB behöver en modell som beskriver de register som finns inom verksamheten och som också beskriver sambanden mellan dessa register.

Mängden av alla register betraktas som ett statistiskt informationssystem och förståelsen av detta system underlättas av en *konceptuell modell* som utan att vara datateknisk beskriver systemets objekttyper och relationer på det sätt som anges i Sundgren (1992). Dels behövs modeller som beskriver hur systemet faktiskt ser ut för närvarande och dels behövs modeller som anger en tänkt förbättrad struktur som kan ligga till grund för utvecklingen av systemet.

2.1 Objekttyper och relationer

Den färdiga modellen har tidigare beskrivits av Registerprojektet (2001) och beskrivs närmare i kapitel 3. Arbetet med att ta fram denna konceptuella modell av registersystemet utgick från en inventering av de register och registerprodukter som fanns inom SCB. En statistisk undersökning börjar med att man definierar populationens objekt. På motsvarande sätt började man inom Registerprojektet det konceptuella modellarbetet med att sortera de inventerade registren efter typ av objekt. Det är uteslutande statistiskt betydelsefulla objekttyper som studeras, dvs. objekttyper som ingår i undersökningspopulationer.

Vissa register innehåller hierarkier av objekttyper och det är därför naturligt att gruppera på följande sätt:

- personer och familjer/hushåll (idag familj, i framtiden hushåll)
- arbetsställen och företag (även organisationer/myndigheter kallas företag)
- fastigheter, värderingsenheter och taxeringsenheter
- fordon

Mellan olika typer av objekt kan det finnas *relationer* av olika slag:

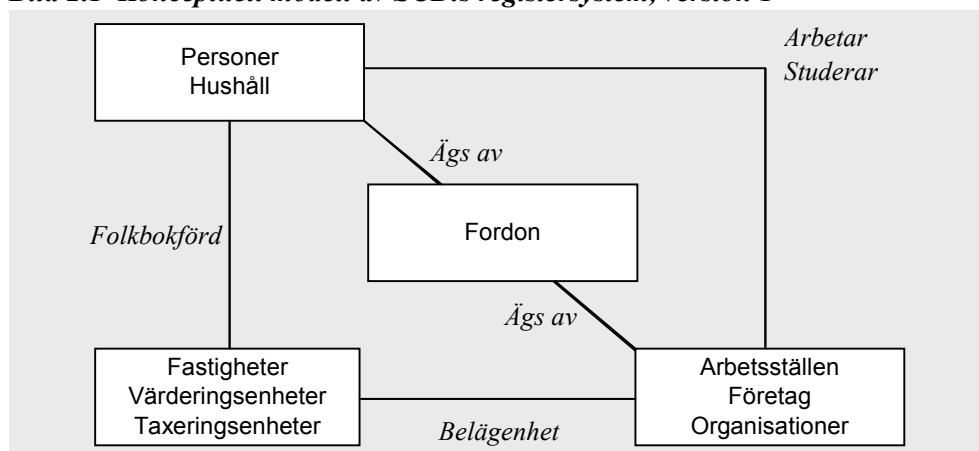
- Mellan person och arbetsställe: En person *är anställd* av ett visst företag eller organisation och *arbetar* på ett visst arbetsställe. En liknade relation gäller mellan studerande och utbildningsställe: En person *deltar i en utbildning* som anordnas av en viss organisation eller ett visst företag och undervisningen sker på ett visst utbildningsställe/arbetsställe.
- Mellan person och fastighet: En person *är folkbokförd* på en viss fastighet.
- Mellan arbetsställe och fastighet: Ett arbetsställe *är beläget* på en viss adress och adressen anger en viss fastighet.
- Mellan person och fordon respektive företag och fordon: Ett visst fordon *ägs av* en person eller ett företag.

Dessa relationer är mycket viktiga ur administrativ och juridisk synvinkel. Det är därför som de administrativa källorna innehåller information av god kvalitet om dessa relationer. Detta är också mycket viktigt för ett fungerande registersystem – när de administrativa källorna innehåller uppgifter om relationerna får

man *länkar* i systemet. En länk mellan två objekt består av en eller flera gemensamma *kopplingsvariabler* som innehåller den information som behövs för att identifiera relationen mellan objekten. Eftersom de relationer som anges ovan avser relationer mellan olika typer av objekt har dessa länkar en mycket viktig roll i registersystemet – i den färdiga modellen utgör de länkarna mellan systemets basregister.

Med dessa objekttyper och relationer har modellen nedan skapats. Rektanglarna representerar *objektmängder* och linjerna representerar *relationer* mellan objekt. Objektmängderna ger upphov till register i systemet och relationerna ger upphov till länkar.

Bild 2.1 Konceptuell modell av SCB:s registersystem, version 1



I bilden ovan finns fyra register som innehåller olika typer av objekt av stort statistiskt intresse. De fyra rektanglarna motsvaras av fyra SCB-register: *Befolknings-*, *Fordons-*, *Fastighets-* och *Företagsregistret*.

När man talar om t.ex. Befolkningsregistret (singularis) avses egentligen en mängd register (pluralis) då Befolkningsregistret finns i många versioner: en personversion, en familjeversion, olika versioner för befolkningen vid olika tidpunkter etc.

Relationen *Folkbokförd* anges i Befolkningsregistret genom en kopplingsvariabel som anger identiteten för den bostadsfastighet, där en viss person är folkbokförd. Relationerna *Ägs av* anges i Fordonsregistret genom att fordonsägarens person- eller företagsidentitet ingår som kopplingsvariabel i registret.

Varken Befolknings- eller Företagsregistret innehåller kopplingsvariabler som anger relationerna *Arbetar/Studerar*. Men det finns andra register som innehåller information om förvärvsarbete respektive studieaktiviteter:

- Kontrolluppgifts- och lönerregistren innehåller variabler som ger information om anställningar. Objekttypen anställning definieras av relationen mellan objekttyperna anställd person, arbetsgivare och arbetsställe. För varje kombination av personidentitet, företags/organisationsidentitet och arbetsställe-identitet finns uppgifter om förvärvsinkomst/lön etc.

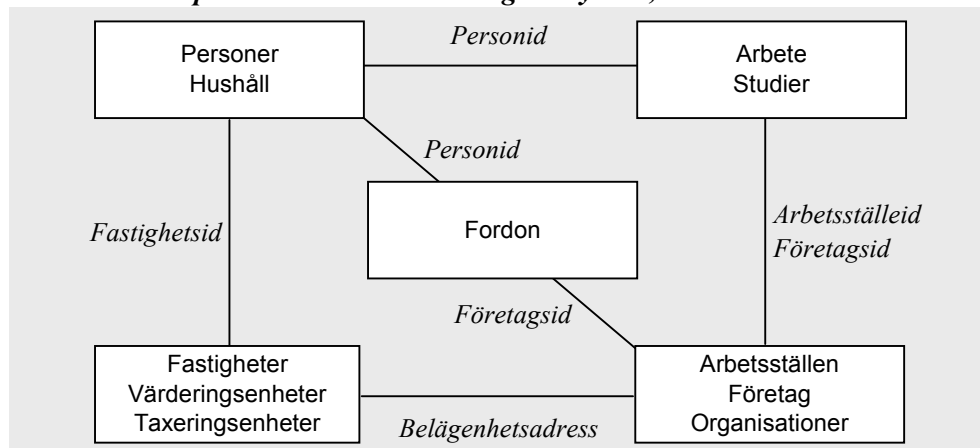
- Studeranderegistren innehåller variabler som anger relationen mellan personer som studerar, utbildningsanordnare och utbildningsställe. Utbildningsställen är arbetsplatser för skolpersonal och alltså en undertyp av objekt-typen arbetsställe.

Information om relationerna *Arbetar/Studerar* skulle kunna lagras i Befolkningsregistret men då många förvärvsarbetande har relationer till mer än en arbetsgivare skulle denna information bli ofullständig då man skulle tvingas att begränsa informationen till enbart en arbetsgivare och då tvingas att välja den huvudsakliga arbetsgivaren.

En bättre lösning är att skapa ett speciellt register som består av *relationsobjekt* dvs. objekt som beskriver relationen mellan person, företag och arbetsställe, där samma person kan ha flera relationer till olika företag och arbetsställen. I bild 2.2 införs relationsobjekten *Arbete/Studier* i ett särskilt register, som kommer att kallas *Aktivitetsregistret*. Relationerna i bild 2.1 har i bilden nedan ersatts med de variabler som fungerar som länkar mellan objekt i olika register.

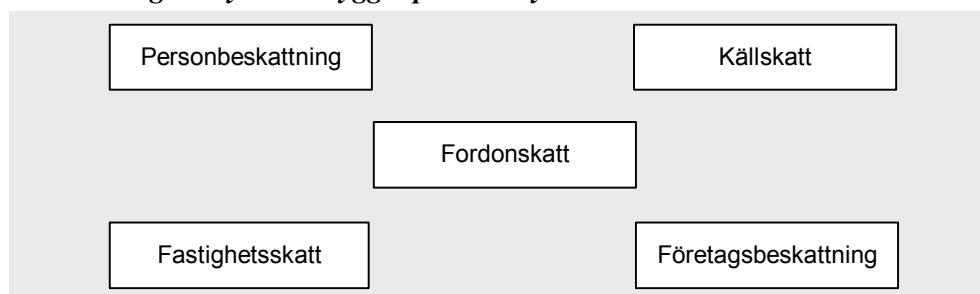
Relationsobjekt
beskrivs i
Sundgren (1992)

Bild 2.2 Konceptuell modell av SCB:s registersystem, version 2



SCB:s system av statistiska register bygger på olika administrativa källor. I bilden nedan anges de fem delar av skattesystemet som är källorna till de fem statistiska register som anges i bild 2.2 ovan.

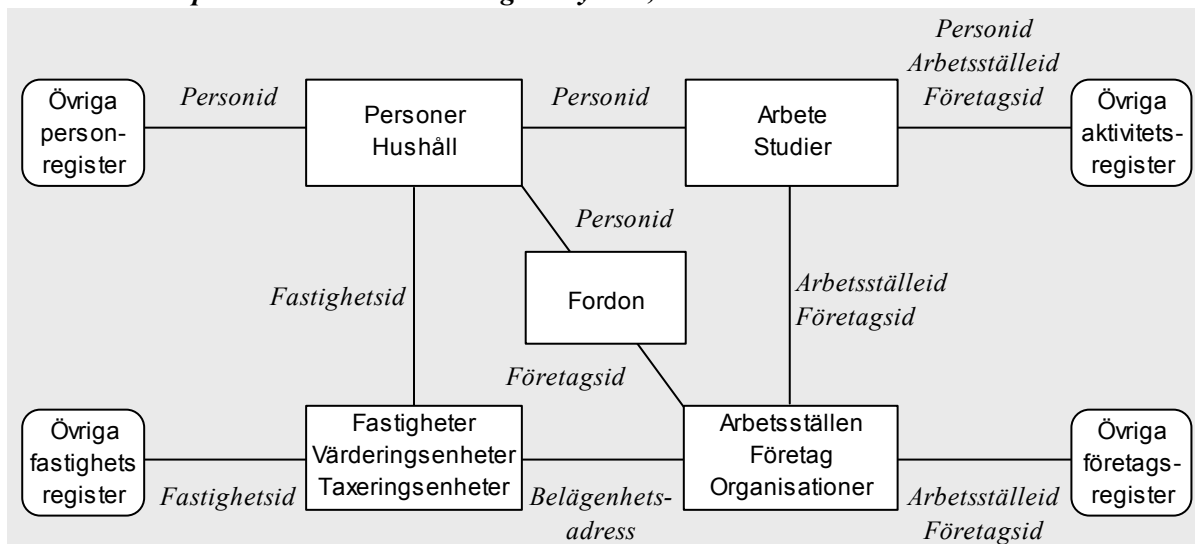
Bild 2.3 Registersystemet bygger på skattesystemets olika delar



Förutom de fem register som anges i bild 2.2 finns ytterligare cirka 50 register på SCB. Med hjälp av befintliga identitetsvariabler kan dessa kopplas till något av Befolknings-, Aktivitets-, Fastighets- eller Företagsregistret. Dessa fyra

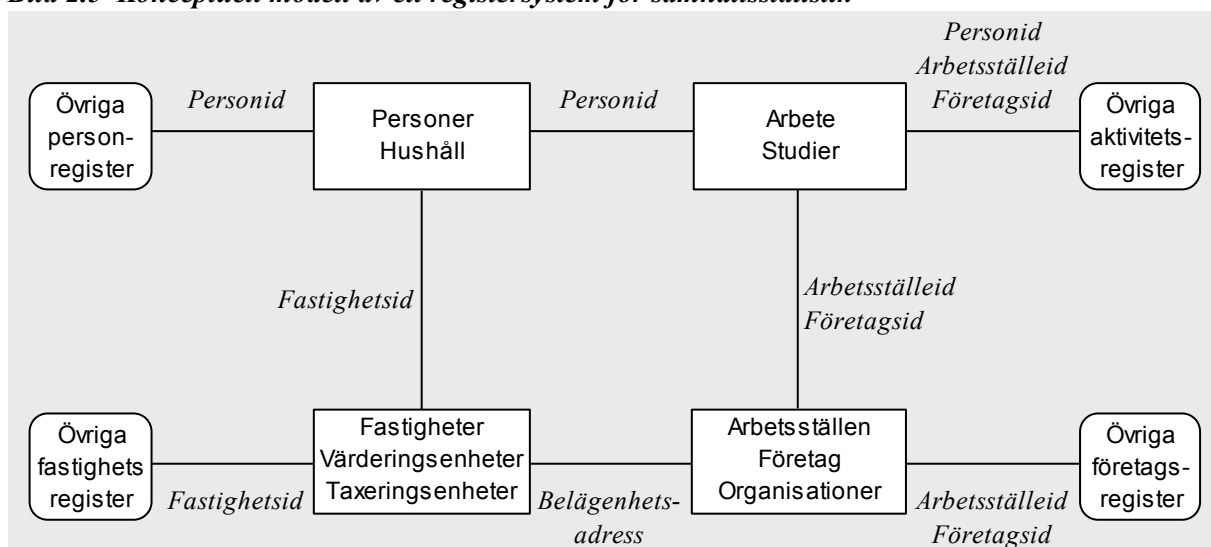
register har en särskild betydelsefull roll inom registersystemet och kallas därför för systemets *basregister*. Hela registersystemet beskrivs i den fullständiga konceptuella modellen i bild 2.4. Registersystemet innehåller ett system av objekt som förändras över tiden. Vid varje tidpunkt har varje objekt ett antal variabelvärden och står i vissa relationer till andra objekt i systemet.

Bild 2.4 Konceptuell modell av SCB:s registersystem, version 3



Fordonsregistret innehåller objekttypen fordon som har en mer begränsad roll för systemet som helhet, varför det är lämpligt att låta personägda fordon ingå bland övriga personregister och företagsägda fordon bland övriga företagsregister. Fordonsregistret skulle annars dominera bilden av registersystemet. Den slutliga modellen beskrivs nedan. Kan denna modell användas generellt? Då samma objekttyper och variabelinnehåll finns i många länder, kan modellen vara generellt användbar för att beskriva system för samhällsstatistik.

Bild 2.5 Konceptuell modell av ett registersystem för samhällsstatistik



2.2 Systemet av basregister

De fyra basregistren är grundvalen för hela registersystemet då de innehåller viktiga objekttyper och länkar. Om något av dessa basregister saknas eller har låg kvalitet skulle hela registersystemet vara betydligt mindre användbart för statistiska ändamål. Basregistrens roll är att definiera objekten i systemet där bra objektavgränsningar och bra täckning är viktiga egenskaper som är avgörande för kvaliteten i hela systemet.

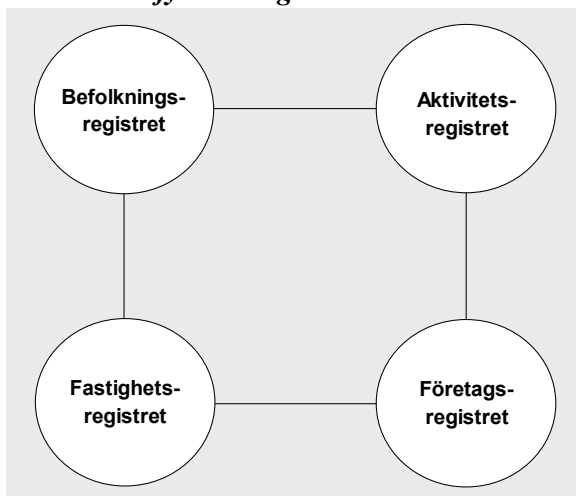
De variabler som är viktiga i ett basregister är olika variabler som identifierar objekt och kan användas för att koppla objekten till objekt i andra register. Tidsuppgifter för olika händelser som berör objekten behövs för att kunna skapa bestånd avseende viss tidpunkt/tidsperiod.

Förutom att de skall bygga på stabila och tillförlitliga administrativa källor skall basregistren ha följande egenskaper:

1. definiera viktiga objekttyper
2. definiera viktiga objekt mängder och *standardiserade populationer*, som kan följas över tiden
3. ha länkar till objekt i andra basregister
4. ha länkar till övriga register som avser samma objekttyp
5. vara viktiga för systemet som helhet – därför är det viktigt att de har hög kvalitet och är väldokumenterade
6. vara viktiga som urvalsramar
7. kunna användas för demografisk statistik avseende personer, aktiviteter, fastigheter respektive företag. På samma sätt som man beskriver befolkningens åldersfördelning och befolkningens förändringar (födda och döda), kan man beskriva åldersfördelning och förändringar bland stocken av alla jobb (förvärvsarbetsaktiviteter), byggnader eller arbetsställen.

De fyra basregistren beskrivs i bild 2.6 nedan. I fortsättningen avbildas dessa som cirklar i modellerna på det sätt som infördes av Registerprojektet 1996.

Bild 2.6 De fyra basregistren



T.ex. "Befolkningsregistret" i bilden avser många registerversioner då varje basregister kan finnas i flera versioner:

- Aktuellt bestånd
- Beståndet en viss tidpunkt, t.ex. den 31/12
- Förändringar under en viss period, t.ex. kalenderår
- Alla objekt som funnits under en viss period, t.ex. årgångsregister

Här kommer dessa fyra basregister att diskuteras och hur de befintliga register som fanns år 2003 överensstämmer med de sju krav som anges ovan.

2.2.1 Befolkningsregistret

Detta basregister har kommit längst när det gäller att utveckla registerstatistisk metodik för arbetet med ett basregister. Alla de sju krav på basregister som anges på sidan 21 ovan är uppfyllda. Befolkningsregistret kan därför användas som inspirationskälla för de övriga basregistren. De som arbetar med Befolkningsregistret kan å jourhållning, har en lång tradition av att producera avancerad registerstatistik, gör regelmässigt demografisk analys och är vana att skapa integrationsregister för olika ändamål. Man har utvecklat ett avancerat produktionssystem där objektens historik lagras.

Registret bygger på administrativa uppgifter från Skatteverkets folkbokförings-system. SCB hämtar dagligen dessa uppgifter från SKV:s elektroniska brevlåda samt vissa årliga uppgifter från Migrationsverket. Uppgifterna avser variabler för identifiering och lokalisering av personer samt demografiska variabler som ålder, kön civilstånd etc.

I samband med bytet från stordator till PC-nätverk i slutet av 1990-talet utökades variabelinnehållet i Befolkningsregistret. Bättre *adressuppgifter*, bättre uppgifter om *relationer* till andra personer samt *datum* för alla folkbokföringshändelser ger nya möjligheter för statistikproduktionen. Ändrade och tillfälliga personnummer registreras så att matchningar underlättas. Dessutom registreras olika slag av rättelser så att det är möjligt att skilja mellan verkliga händelser och rättelser. Dessa förbättringar är viktiga för registrets roll som basregister. Wilén och Johannesson (2002) beskriver det nya Befolkningsregistret.

Vissa svagheter finns emellertid:

- Hushållsavgrensningarna har ännu år 2003 brister som kan avhjälpas med ett bra lägenhetsregister och att personer är folkbokförda på lägenhet.
- Övertäckning finns för vissa kategorier av invandrare. Detta upptäcktes inom SCB och Skatteverket (SKV) informerades, se Nilsson (1995). SKV har därefter förbättrat kontrollen av folkbokföringen för att få ned övertäckningen.

2.2.2 Aktivitetsregistret

Detta basregister var år 2002 under utveckling. Registret skall bestå av tre olika delar: *förvärvsaktiviteter*, *studieaktiviteter* och *andra arbetsmarknadsrelaterade aktiviteter*. Som vi tidigare nämnt består registret av relationsobjekt. Varje sådant objekt identifieras med tre kopplingsvariabler: personnummer, företagets organisations- och arbetsställesnummer enligt Företagsregistret.

Uppgifter om *förvärvsaktiviteter* bygger på arbetsgivarnas årliga kontrolluppgifter, som ger förvärvsinkomst för varje kombination av *anställd* och arbetsställe. Självdeklarationer ger uppgifter om de personer som är aktiva som *egna företagare*. För offentliganställda får SCB månadsvisa uppgifter från arbetsgivarnas löneregister. År 2002 var registret uppbyggt för att beskriva aktiviteten förvärvsarbete och används i huvudsak för att bilda det personregister som

beskriver befolkningens sysselsättning. I princip är förvärvsinkomst, månadslön, tjänstgöringens omfattning, yrke och pendling egenskaper hos relationsobjekten i Aktivitetsregistret. I det årliga Sysselsättningsregistret är objektet däremot person och registret beskriver sysselsättning för varje person.

Uppgifter om *studieaktiviteter* bygger på uppgifter från skolor och universitet. Det finns en rad register över olika slag av studerande. Dessa register innehåller de studerandes personnummer och uppgift om skola eller studieställe, som i sin tur är arbetsställen i Företagsregistret. Genom att ge dessa studieställen arbetsställesnummer kommer t.ex. Skolregistret att kunna kopplas till Företagsregistret och aktiviteten studier kommer i Aktivitetsregistret att vara en relation mellan ett personnummer och ett arbetsställesnummer. På samma sätt som man bildar pendling för förvärvsarbetande med personnummer och arbetsställesnummer, som ger bostadens respektive arbetsställets belägenhet, kan man då bilda pendling för studerande.

Även uppgifter om *andra arbetsmarknadsrelaterade aktiviteter*, relaterade till försörjning, skall ingå i det nya registret. Olika myndigheter har information om värnplikt, sjuk- och förtidspensionering, arbetsmarknadsåtgärder, registrerad arbetslöshet och slutna sjukvård. Denna information ger en kompletterande bild av arbetsmarknadsstatus utöver förvärvsarbete och studier. Det finns uppgifter i de administrativa källorna som tidslokaliserar aktiviteterna även om kvaliteten på tidslokaliseringen ibland är låg.

Det finns en stor mängd administrativa data som innehåller kopplingar till både person och företag. Rent datatekniskt skulle dessa uppgifter kunna betraktas som antingen person- eller företagsdata. Istället betraktas dessa uppgifter som data som beskriver *aktiviteter*. Dessa aktiviteter kan uppfattas som en egen objekttyp av två skäl, de är dels en statistiskt intressant objekttyp och dels måste denna objekttyp särskiljas för att registersystemet som helhet skall få en pedagogiskt klar struktur.

Uppfyller Aktivitetsregistret de sju kraven på ett basregister (enligt sidan 21)?

1, 2, 7. Definierar viktiga objekttyper och objektmängder, populationer. Kan användas för demografisk statistik avseende aktiviteter.

När utbud och efterfrågan på arbetsmarknaden möts skapas relationer mellan personer och företag/organisationer. Dessa relationer är viktiga undersökningsobjekt för arbetsmarknadsstatistiken. Demografisk statistik, som beskriver hur t.ex. stocken av förvärvsaktiviteter förändras genom så kallad *job-creation* och *job-destruction*, måste vara mycket relevant för studier av arbetsmarknaden.

3. Har länkar till objekt i andra basregister

Kopplingsvariabeln personnummer är länken till Befolkningsregistret och kombinationen av kopplingsvariablerna organisationsnummer och arbetsställesnummer är länken till Företagsregistret.

4. Har länkar till övriga register som avser samma objekttyp

Kopplingsvariablerna person-, organisations- och arbetsställesnummer var år 2003 länken till Kontrolluppgiftsregistret, Privatanställdas löner och Yrkesregistret. När övriga löneregister och studeranderegistren standardiserats med

avseende på de variabler som identifierar arbetsställen och utbildningsställen kommer aktivitetsregistret att ha länkar till samtliga register som består av dessa relationsobjekt.

5. Viktigt för systemet som helhet

Bland användarna finns ett stort intresse för statistik där persondata och företagsdata kombineras. Aktivitetsregistret har en mycket viktig roll som brygga mellan individ- och företagsstatistiken.

6. Viktigt som urvalsram

Kanske inte direkt, men vissa kategorier av personer eller företag skulle kunna selekteras med hjälp av Aktivitetsregistret.

Aktivitetsregistret kan användas för att skapa både person- och företagsregister

Aktivitetsregistret bygger direkt på den administrativa källan – för varje kombination av person och arbetsställe finns uppgift om årets lönesumma. Genom att summera löneuppgifterna för varje person kan vi skapa ett personregister. Genom att summera löneuppgifterna i aktivitetsregistret för varje arbetsställe kan vi skapa ett arbetsställeregister med lönesummor.

Bild 2.7 Sambandet mellan person-, aktivitets- och företagsregister

Personregister		Aktivitetsregister – jobb			
Person	Årslön	Jobb	Person	Arbetsställe	Årslön
Pnr 1	450 000	J1	Pnr 1	AE 1	220 000
Pnr 2	210 000	J2	Pnr 3	AE 1	180 000
Pnr 3	270 000	J3	Pnr 1	AE 2	230 000
		J4	Pnr 2	AE 2	210 000
		J5	Pnr 3	AE 2	90 000

Företagsregister – arbetsställen	
Arbetsställe	Lönesumma
AE 1	400 000
AE 2	530 000

2.2.3 Företagsregistret

SCB får administrativa data avseende *juridiska enheter* från Patent- och registreringsverket (PRV) och Skatteverket (SKV). Data från PRV, SKV och Svensk Adressändring AB aviseras regelbundet.

Ett omfattande arbete sker inom SCB för att samla in uppgifter från företag med verksamhet vid mer än ett arbetsställe för att kunna skapa *arbetsställe-enheter*. Detta arbete är uppdelat mellan två avdelningar. Inom avdelningen för arbetsmarknadsstatistik arbetar man med kontrolluppgifter och tar många kontakter med företag i samband med att kontrolluppgifterna granskas. Kontrolluppgiften är ett exempel på *integrerad datainsamling*, där dels skatteadministrativa uppgifter samlas in men även vissa uppgifter för statistiska ändamål som bl.a. avser avgränsningen av arbetsställen. Detta arbete avser arbetsställepopulationen föregående kalenderår, till skillnad från registervårdsarbetet inom avdelningen för ekonomisk statistik, som är inriktat på det aktuella beståndet av arbetsställen. Även här gör man flerarbetsställeenkäter och tar många kontakter med företag. Inom Företagsregistret skapas även *företags-, verksamhets- och lokala verksamhetsenheter*.

Två principer genomsyrade det utvecklingsarbete som ägde rum i samband med övergången från stordatordrift till PC-nätverk:

- Inga parallella objektmängder skall förekomma inom den ekonomiska statistiken, dvs. populationen av t.ex. industriföretag skall vara densamma både inom industristatistiken och i Företagsregistret.
- Alla inom avdelningen för ekonomisk statistik skall delta i vården av Företagsregistret, dvs. de som t.ex. samlar in data från industriföretag skall vidarebefordra den information de får om industripopulationen till Företagsregistret.

Dessa principer skall gälla alla basregister, inte enbart Företagsregistret. Dessutom skall alla delta i vården av ett basregister oberoende av avdelningsgränser. Konkret innebär detta att de som arbetar med Lantbrukets företagsregister skall lämna underlag för lantbruksdelen av Företagsregistret och att de som arbetar med Skolregistret lämnar underlag så att Företagsregistrets skoldel blir så bra som möjligt.

Uppfyller Företagsregistret de sju kraven på ett basregister (enligt sidan 21)?

1, 2, 7. Definierar viktiga objekttyper och objektmängder, populationer. Kan användas för demografisk statistik avseende arbetsställen och företag.

Ett flertal viktiga objekttyper och populationer för den ekonomiska statistiken finns inom Företagsregistret. Statistik som beskriver företagens och arbetsstäl- lenas demografi efterfrågas av många användare. Johansson (1997 och 2001) innehåller exempel på hur Företagsregistret kan användas för demografisk företagsstatistik inom ekonomisk forskning.

3. Har länkar till objekt i andra basregister

Kombinationen av kopplingsvariablerna organisationsnummer och arbetsstäl- lenummer är länken till Aktivitetsregistret. Arbetsställets belägenhetsadress är länken till Fastighetsregistret, fortfarande finns det dock år 2003 brister med denna kopplingsvariabel.

4. Har länkar till övriga register som avser samma objekttyp

Kopplingsvariablerna organisations- och arbetsstäl- lenummer är länken till andra register som avser företag/organisationer och arbetsställen. För person- liga företag fungerar företagarens personnummer som organisationsnummer. Även objekttypen *Företagsenheter* identifieras med en sifferkod som kan användas som kopplingsvariabel.

Lantbrukets företagsregister och Skolregistret med utbildningsanordnare och utbildningsställen hade år 2003 bristfälliga länkar till Företagsregistret.

5. Viktigt för systemet som helhet

Företagsregistret är betydelsefullt inte enbart för företagsstatistiken utan även för arbetsmarknads- och individstatistiken. Uppgifter om arbetsställets närings- gren och belägenhet används inom många statistikprodukter.

6. Viktigt som urvalsram

Används som ram för många urvalsundersökningar som avser företag, företags- enheter eller arbetsställen.

2.2.4 Fastighetsregistret

År 2003 fanns inte ett renodlat basregister för fastigheter. Det Fastighetstaxeringsregister som har funnits länge har inte samma struktur som de övriga tre basregistren. Ett basregister skall avgränsa populationer av objekt. Viktiga variabler är då sådana som identifierar och lokaliserar objekten och anger tidpunkter för viktiga händelser. Fastighetstaxeringsregistret i nuvarande form är inte något idealt basregister för fastigheter. Ett sådant bör på sikt byggas upp med hjälp av Lantmäteriverkets fastighets- och byggnadsregister och kompletteras med det planerade lägenhetsregistret samt det nuvarande Fastighetstaxeringsregistret. I ett nytt basregister för fastigheter skall det finnas länkar till taxeringsenheter och byggnader. Basregistret kan uppdateras med lagfartsuppgifter, bygglagsuppgifter och fastighetstaxeringsuppgifter.

Ett renodlat basregister för fastigheter och andra objekttyper som hör samman bör bildas. Ett sådant register skall uppdateras ofta så att uppdelningar och sammanslagningar av fastigheter kan följas. Registret skall uppfylla alla de krav vi ställer på basregister (enligt sidan 21). Följande objekttyper bör ingå:

- fastighet*
- byggnad*
- lägenhet (både bostäder och lokaler)*
- taxeringsenhet*
- värderingsenhet*

Viktiga variabler i ett sådant basregister är identiteter på de olika objekt som tillhör ovanstående objekttyper. Belägenhetsadresser och geografiska koordinater och koder skall också ingå. Adresserna skall förbättras så att fastigheter med bostäder och arbetsställen kan knytas till standardiserade adresser.

2.3 Grundläggande variabler i registersystemet

Vissa variabler används inom många register och har en viktig roll för systemet som helhet. Det register som tar in en variabel från en annan myndighet eller själva skapar en variabel ska ha ansvaret för denna på SCB. Att ha ansvar för en *standardiserad variabel* innebär följande:

- ha kontakt med registerlämnande myndighet och ansvara för att variabeln och information om variabeln hämtas från myndigheten till SCB
- ansvara för granskning, bearbetning, namngivning och dokumentation.

Andra register inom SCB som använder en standardiserad variabel skall använda den bearbetade variabeln utan ändringar, dess korrekta namn och den ursprungliga dokumentationen. På detta sätt minskas dubbelarbete och sambearbetningar underlättas samtidigt som konsistensen inom systemet förbättras. Två kategorier av standardiserade variabler har speciella roller inom systemet, länkar och klassifikationer.

2.3.1 Länkar

De länkar som systemet bygger på anges i bild 2.5 och baseras på följande identiteter: *Person-, Arbetsställe-, Företags- och Fastighetsidentitet*. Vidare är *Arbetsställets belägenhetsadress* länk mellan Företags- och Fastighetsregistret.

När ett register skapas i registersystemet görs många matchningar mellan olika register. Ett administrativt register matchas mot ett basregister för att avstämma objekt mängden, man matchar mot olika andra statistiska register för att kunna hämta variabler till det nya registret. Alla dessa matchningar förutsätter att de olika registren innehåller bra länkar. En länk består av en eller flera variabler som identifierar enskilda objekt. För att underlätta matchning av olika register fodras att samma identifierande variabler finns i många register. När man matchar register som avser olika årgångar behöver man länkar som är stabila över tiden dvs. ett enskilt objekt som är oförändrat skall ha samma värden på de identifierande variablerna.

Exempel: Personnummer är en mycket stabil variabel vilket underlättar registerbearbetningarna. Fastighetsidentitet består av län, kommun och fastighetsbeteckning. Detta är ett exempel på en mycket instabil variabel, om läns- eller kommunkod ändras kommer många fastighetsidentiteter att ändras.

2.3.2 Klassifikationer

Näringsgren, varugrupp, utbildning, yrke m.m. är exempel på viktiga statistiska *standarder* och *klassifikationer*. Dessa baseras på internationella rekommendationer och är innehållsmässigt betydelsefulla och används inom många undersökningar, både registerbaserade och andra.

De administrativa källorna innehåller uppgifter om dessa klassifikationer och denna information används för att skapa variablerna inom registersystemet. Dessutom måste de som ansvarar för en standard inom SCB kunna *koda* nya yrkesbenämningar, utbildningar etc. så att de nya benämningarna förs till lämpliga kategorier inom klassifikationen.

En *standardiserad variabel* är så viktig att ansvaret för variabeln klargörs genom särskilt beslut.

I läroböcker används samma term för ett annat begrepp: En variabel transformeras till en *standardiserad variabel* så att medelvärdet blir 0 och standardavvikelsen blir 1

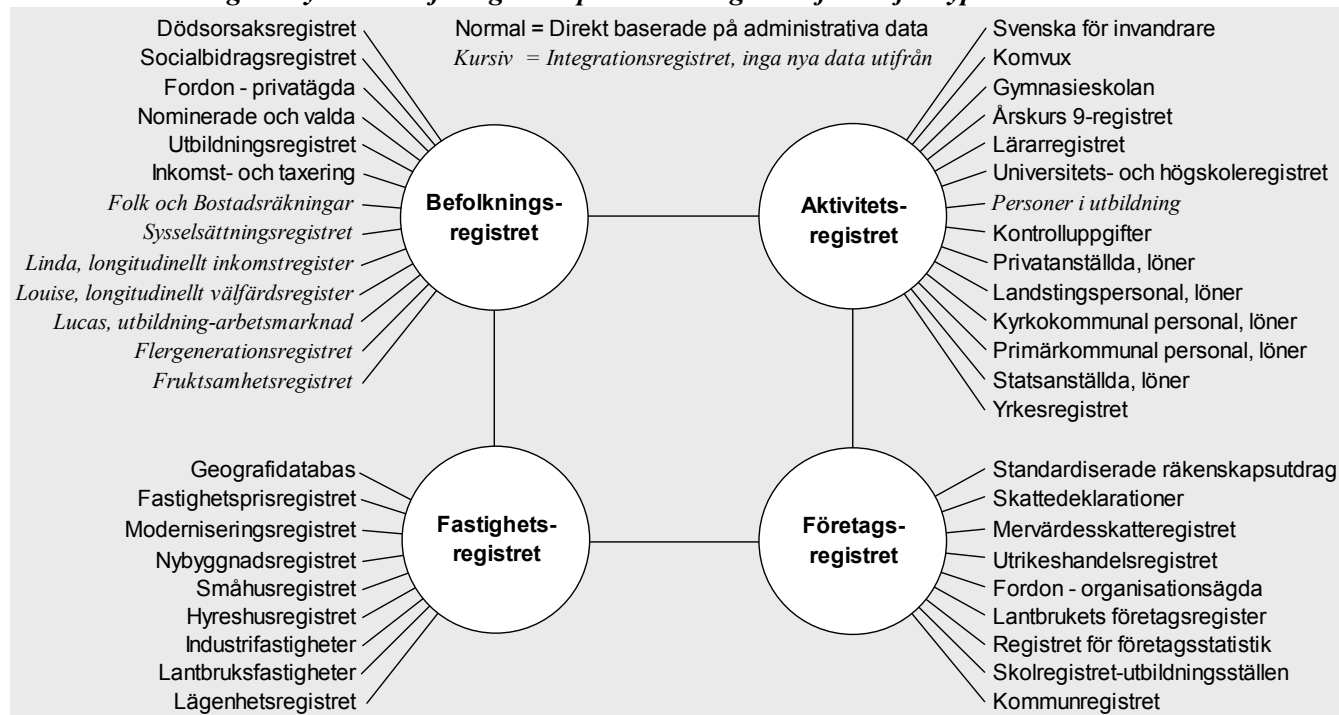
Standarder
Klassifikationer

Kodning

2.4 Hela registersystemet

Nästa steg är att koppla övriga statistiska register till respektive basregister. Cirklarna i bild 2.8 representerar basregistren och linjerna befintliga eller planerade länkar mellan objekten i olika register. Ett väl fungerande system fordrar att objektdefinitioner har samordnats, att tidsreferenser och gemensamma variabler är harmoniserade och att det finns bra identifieringsvariabler som kan användas för att länka objekt i olika register.

Bild 2.8 SCB:s registersystem – befintliga och planerade register efter objekttyp



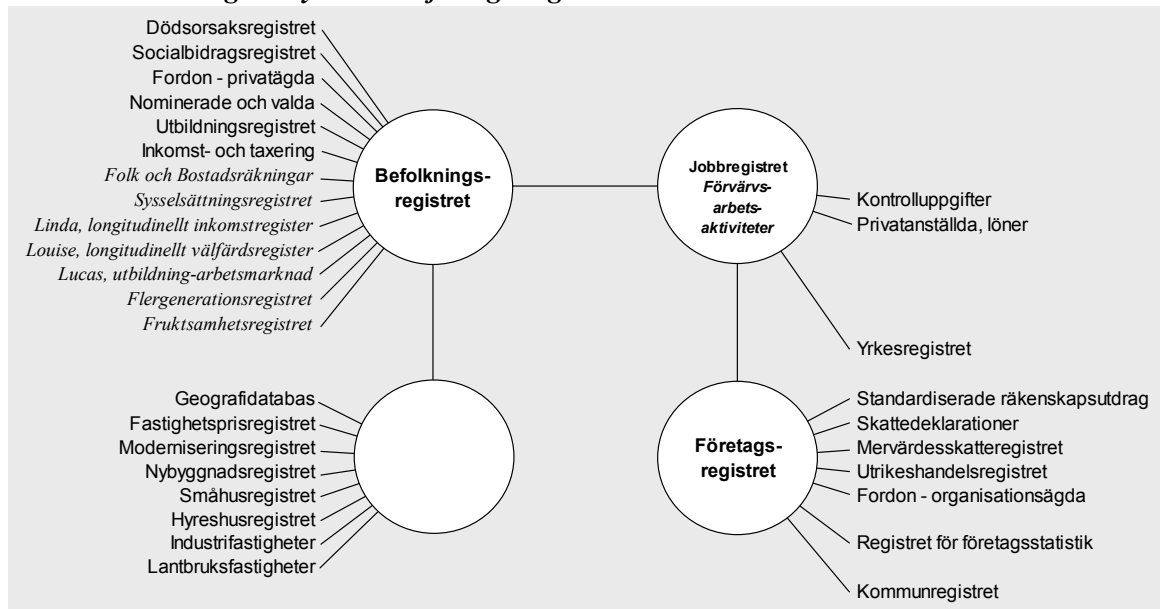
De basregister, övriga statistiska register och länkar som ingår i SCB:s registersystem visas i bild 2.8 ovan. De flesta komponenterna i systemet fanns år 2003, men ett antal register och länkar bör samordnas och utvecklas för att registersystemet skall fungera väl.

Den verklighet registersystemet avbildar förändras kontinuerligt – objekt tillkommer eller försvinner. Relationer mellan objekt förändras och egenskaper och variabelvärden hos objekten ändras. Varje register i modellen ovan finns för flera årgångar och kan finnas i flera versioner.

I bild 2.8 skiljer vi på tre typer av register: basregister (cirklar), register som baseras direkt på administrativa data (normal stil) och register som helt är baserade på befintliga register (*kursiv* stil). Även utvecklingsarbetet inom SCB medför att registersystemet förändras över tiden. Nya register baserade på nya administrativa källor kan tillkomma och nya integrationsregister kan bildas. Vidare kan befintliga register förändras och inom systemet som helhet kan samordningen förbättras.

I bilden nedan beskrivs de delar av systemet som hade fungerande länkar år 2003. Aktivitets- och Fastighetsregister finns, men har inte den struktur som är lämplig för ett basregister.

Bild 2.9 SCB:s registersystem – befintliga register som var samordnade år 2003

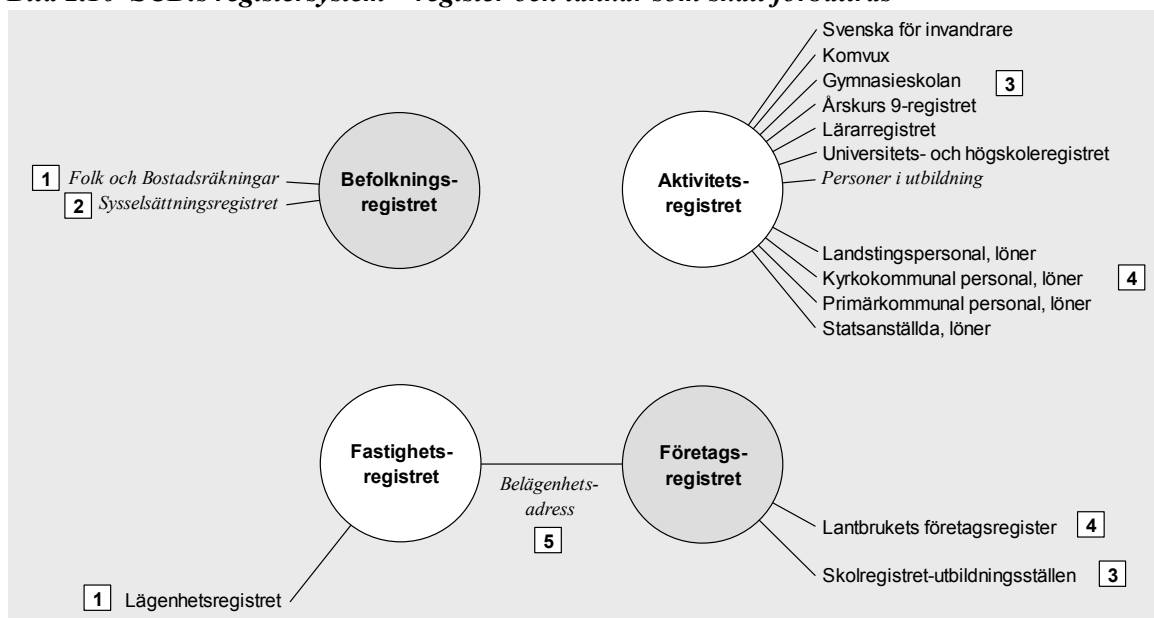


Följande delar av systemet skall förbättras enligt de förslag som presenteras i Registerprojektets slutrapport (2001).

1. För att helt ersätta de traditionella Folk- och Bostadsräkningarna med registerbaserad statistik förutsätts att ett Lägenhetsregister finns som kan användas för att avgränsa hushåll.
2. Sysselsättningsregistret innehåller en variabel som anger om en person förvärvsarbetar under november. Idag är denna variabel endast baserad på Kontrolluppgifter, men genom att använda uppgifter i löneregister skulle kvaliteten i denna variabel kunna förbättras.
3. Skolor och utbildningsställen finns både i Företagsregistret och i olika register som avser skol- och universitetsutbildningar. Men idag används olika identiteter inom dessa register. Då skolregistren innehåller samma organisations- och arbetsställesnummer som Företagsregistret blir matchningar möjliga. Även skolornas belägenhetsadresser måste förbättras.
4. På motsvarande sätt bör alla löneregister samt Lantbrukets företagsregister innehålla samma organisations- och arbetsställesnummer som i Företagsregistret.
5. Genom att förbättra adresser så att de blir entydiga kan arbetsställesnas belägenhetsadresser kopplas till rätt fastighet. Länken mellan Företags- och Fastighetsregistret kommer då att få högre kvalitet.

Siffrorna i bilden nedan refererar till dessa planerade förbättringar. Aktivitets- och Fastighetsregistret skall utvecklas som basregister, Lägenhetsregistret planeras och de övriga förbättringarna avser ändrade eller förbättrade länkar.

Bild 2.10 SCB:s registersystem – register och länkar som skall förbättras



Varför fyra basregister?

I vissa andra länder har man utgått från en struktur med tre basregister, individer, företag och fastigheter. Denna struktur med tre basregister avser de *administrativa* register som används för att administrera dessa tre slag av objekt. Då individer, företag och fastigheter administreras fristående från varandra är länkar mellan de administrativa registren inte så betydelsefullt.

Varje administrativt register är en viktig källa till motsvarande *statistiska* basregister, men dessa bygger som regel på fler källor. Dessutom är länkarna mellan basregistren mycket betydelsefulla för det statistiska registersystemet och Aktivitetsregistret behövs därför som länk mellan individer och företag. En registermodell med fyra basregister har en klar struktur och framhäver *aktiviteter* som en statistiskt intressant objektmängd som kan belysas med många administrativa källor.

Strukturen med fyra basregister utgår *inte* från SCB:s indelning i avdelningar. När strukturen förslögs av Registerprojektet 1996 fanns tre av dessa basregister inom samma avdelning.

Kapitel 3. Hur skapas ett registersystem?

I kapitel 2 går vi igenom registersystemets struktur och dess olika delar. I detta kapitel beskriver vi arbetsgången från mottagandet av administrativa data fram till skapandet av de register som är underlag för registerstatistiken.

De administrativa register, som finns hos olika myndigheter och från vilka SCB får leveranser, kallar vi för administrativa *källregister*. Med hjälp av dessa administrativa källor skapas inom SCB olika *statistiska register*. Varje sådant statistiskt register kan vara underlag för en eller flera *statistiska produkter*, dvs. undersökningar. Om dessa statistiska register ingår i ett samordnat system kan de sambearbetas och användas så att nya produkter kan framställas.

Register 1	Produkt A
Register 2	Produkt B Produkt C
Register 3	Produkt D
Register 2 o 3 sambearbetas	Produkt E

3.1 Administrativa register omvandlas till statistiska register

Den ursprungliga databildningen sker hos myndigheter och organisationer. Definitioner av objekt och variabler är avpassade för olika administrativa ändamål. Varje myndighet gör kontroller, rättelser och andra bearbetningar som passar för deras administrativa syften. När en myndighet levererar data till SCB görs eventuellt ytterligare selektioner och bearbetningar för att tillfredsställa SCB:s önskemål. Leveransen skall vara fullständig, internt konsistent och avse en viss angiven tidsperiod eller tidpunkt. Inom respektive myndighet finns även metadata i form av kunskaper om definitioner, databildning och kvalitet. Denna typ av information är också viktig för dem som tar emot data inom SCB.

Exempel: SKV:s leveranser till SCB:s Inkomst- och taxeringsregister

1. *Databildningen* hos Skatteverket (SKV)
Den årliga inkomsttaxeringen bygger på inkomsttagarnas självdeklarationer och lokala skattemyndigheters taxeringsbeslut. Både inkomsttagare och skattemyndigheter använder kontrolluppgifter avseende lön, sjukpenning etc. och räntor som arbetsgivare, försäkringskassor och kreditinstitut ansvarar för. Denna information sammanställs slutligen hos SKV. Deklarationer, kontrolluppgifter och taxeringsbeslut kan ändras och kompletteras.
2. *Mikrodataleveranser* till SCB:s Inkomst- och taxeringsregister
SKV skapar årligen ”taxeringsband” som innehåller uppgifter om Sveriges befolkning. *Taxeringsbandet* för ett år – innehåller ca 9 miljoner poster och varje post innehåller ca 300 variabler – levereras direkt till SCB:s Inkomst- och taxeringsregister. *Kontrolluppgifterna* sammanställs av SKV och levereras till SCB:s Kontrolluppgiftsregister som sedan skickar en kopia till Inkomst- och taxeringsregistret. Leveransen omfattar ca 70 miljoner poster och med över 300 variabler fördelade på 20 olika posttyper.
3. *Metadata* till SCB:s Inkomst- och taxeringsregister
Postbeskrivningar med variabelnamn och variabeldefinitioner medföljer leveranserna från SKV. Deklarationsblanketter, kontrolluppgiftsblanketter, taxeringsbestämmelser, deklarationsanvisningar och anvisningar till arbetsgivare behövs också för att kunna tolka data.

Inkomst- och taxeringsregistret (IoT) kallades tidigare för Inkomst- och förmögenhetsregistret (IoF)

Vad händer när data levererats till SCB?

Det är som regel inte lämpligt att framställa statistik direkt från de mottagna administrativa registren eftersom dessa inte är anpassade för statistikens krav. Objektmängder, objektavgränsningar och variabler behöver kontrolleras och därefter kan det vara nödvändigt med vissa bearbetningar så att registret uppfyller statistikens krav på objekt och variabler. De registerstatistiska bearbetningar, som syftar till att omvandla ett eller flera administrativa register till ett statistiskt register, skall baseras på genomtänkta och standardiserade *registerstatistiska metoder*. Följande bild visar de olika moment som ingår i det metodstatistiska arbetet.

Bild 3.1 Från administrativa register till statistiskt register



Arbetet med att ta emot stora administrativa material samt granska dessa är viktiga moment som är likartade för många registerenheter vid SCB. Det är därför viktigt med erfarenhetsutbyte, gemensam metodutveckling och att väl-etablerade och väldokumenterade metoder används allmänt på det sätt som länge har varit självklart i samband med SCB:s urvalsundersökningar. På motsvarande sätt bör gemensamma metoder och verktyg för bortfallshantering, sambearbetningar och evalveringar användas.

Exempel: SCB:s Inkomst- och taxeringsregister (IoT)

1. *Leverantörskontakter, mottagningskontroll och granskning av grunddata*
IoT-registret får data från elva olika leverantörer både utanför och inom SCB. Data som kommer utifrån granskas och åtgärdas. Data som levereras från andra SCB-register är redan granskade och åtgärdade. Leverantörskontakterna är viktiga för att få kunskap om förändringar i de administrativa systemen, vilket är viktigt för att ha kontroll av kvaliteten hos registerstatistiken – administrativa förändringar får inte misstolkas som verkliga inkomstförändringar.
2. *Sambearbetningar, selektioner*
Det är ett stort antal register som skall sambearbetas för att skapa de olika delregister som ingår i IoT-registret. Här sker då matchningar med hjälp av personnummer och samtidigt aggregeras t.ex. alla kontrolluppgifter för en viss person så att personens arbetsinkomst bildas. Taxeringsmaterialet inne-

håller data som avser ett helt år. En typ av bearbetning är att selektera personer 16 år och däröver som ingick i befolkningen den 31/12.

3. Härledda objekt bildas

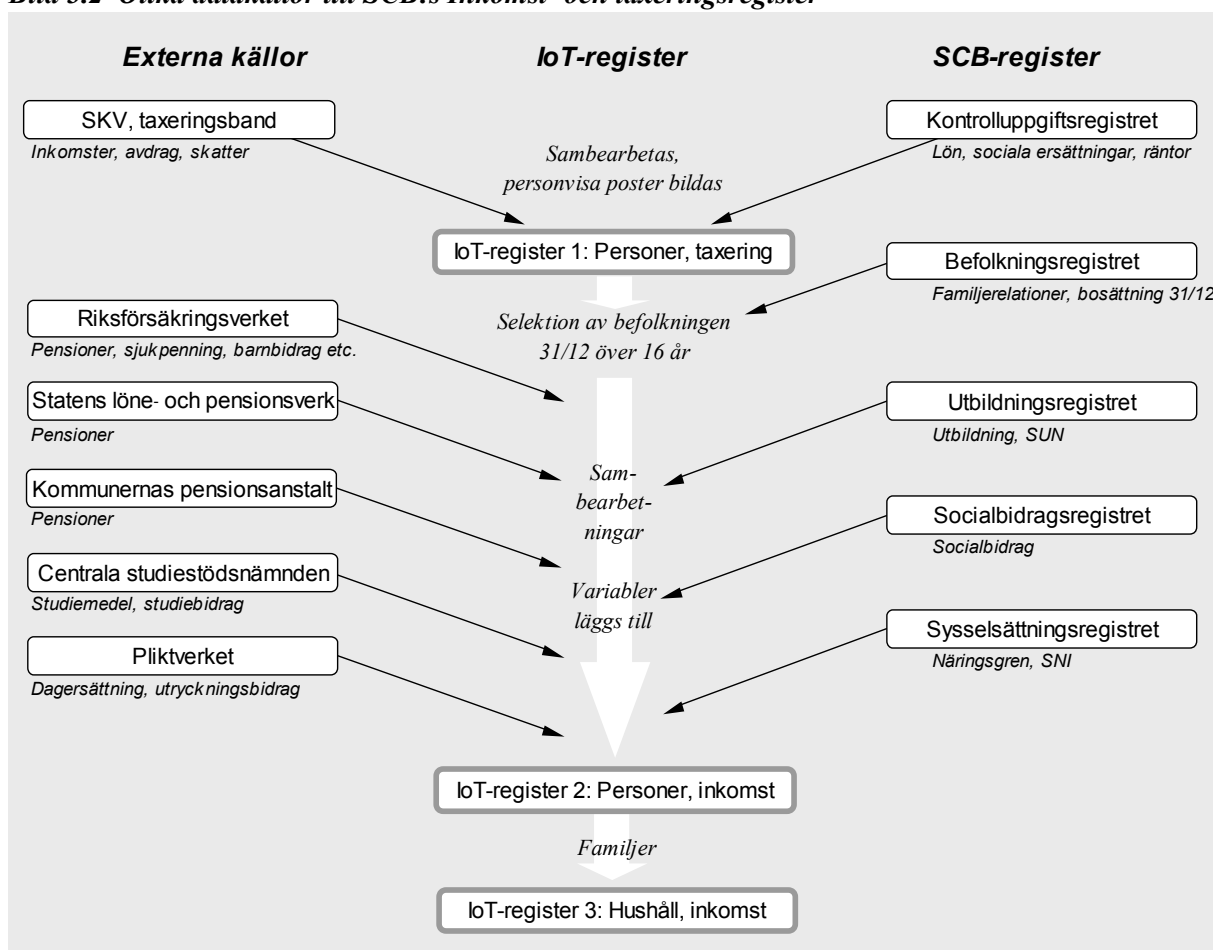
Med information om vissa relationer bildas familjer. Mellan vuxna medför relationerna *gifta* eller *sammanboende med gemensamma barn* att man förs till samma familj. Dessa relationer anges med de anhörigas personnummer som finns i taxeringsmaterialet och i SCB:s befolkningsregister.

4. Härledda variabler bildas

Ett stort antal härledda inkomstvariabler bildas. T.ex. aggregeras lönebelopp på olika kontrolluppgifter till personens *arbetsinkomst*. Varje persons samtliga inkomster och sociala ersättningar minus skatter blir personens *disponibla inkomst*. För familjer bildas bl.a. variablerna *familjetyp*, *antal konsumtionsenheter* och *disponibel inkomst*.

Bilden nedan visar hur man på Inkomst- och taxeringsregistret tar emot administrativa data från en rad olika externa källor och vissa SCB-register. Med *källregister* avser vi både de administrativa källorna och de SCB-register som används för att skapa det nya registret. De olika arbetsmomenten när källregistren omvandlas till nya statistiska register anges i bildens mittfält.

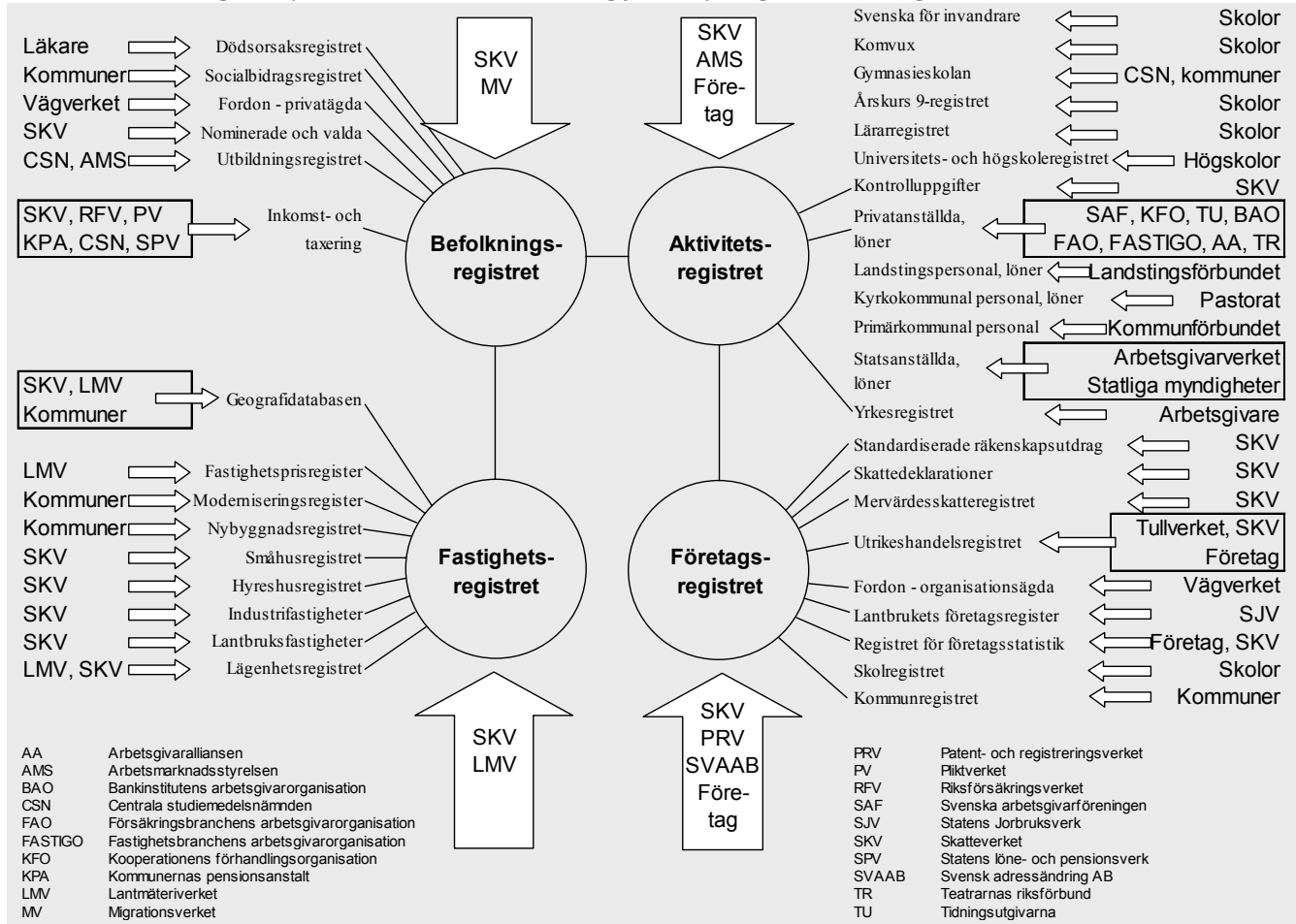
Bild 3.2 Olika datakällor till SCB:s Inkomst- och taxeringsregister



3.2 Systemet får data från många administrativa källor

Från ett stort antal registerlämnande myndigheter och organisationer får olika registerenheter inom SCB administrativa data som bearbetas så att ett antal statistiska register bildas. Bilden nedan beskriver registersystemets indatadel.

Bild 3.3 SCB:s registersystem 2003 – datainsamling från myndigheter och organisationer



Av bilden framgår att många enheter har kontakter med registerlämnare och att många olika myndigheter och organisationer är involverade. SCB:s försörjning av administrativa data organiseras och tryggas genom goda och välorganiserade kontakter med registerlämnande organisationer. Dessa kontakter sker på olika nivåer men bör vara samordnade inom SCB. Huvuddelen av SCB:s datalager kommer ursprungligen från de datainflöden som markeras med vita pilar i bilden.

Exempel på olika sätt att ta emot administrativa data

I avsnitt 3.1 beskrivs hur man inom SCB:s Inkomst- och taxeringsregister tar emot administrativa data från olika myndigheter. Dessa leveranser är *årsvisa* och kommer från ett *begränsat antal centrala myndigheter* och leveranserna sker på *magnetband*. För andra SCB-register kan förhållandena vara helt annorlunda och sättet att ta emot data kan struktureras på följande sätt:

1. Leveranserna kan vara löpande och avser då förändringar i beståndet av objekt, eller avse månadsvisa eller årsvisa uppgifter. Det är intressant att så stor del av registersystemet får data månadsvis eller oftare – registerstatistik kan alltså ha hög aktualitet.

Bild 3.4 Registersystemets indata efter den administrativa källans periodicitet

Får data löpande	Får data månadsvis	Får data årsvis
Befolkningsregistret	Jobbregistrets AMS-del	Utbildningsregistret
Dödsorsaksregistret	Universitets- och	Inkomst- och taxeringsregistret
Fordonsregistret	högskoleregistret	Svenska för invandrare (3 ggr/år)
Företagsregistret	Statsanställda, löner	Komvux (2 ggr/år)
Skolregistret	Skattedeklarationer	Gymnasieskolan
Fastighetsprisregistret	Mervärdesskatterregistret	Årskurs 9-registret
<i>Det planerade</i>	Utrikeshandelsregistret	Lärrregistret
<i>Fastighetsregistret kommer</i>	Geografidatabasen	Kontrolluppgifter
<i>att få löpande aviseringar</i>	Moderniseringsregistret	Privatanställda, löner
	Nybyggnadsregistret	Landstingspersonal, löner
		Kyrkokommunal personal, löner
		Primärkommunal personal, löner
		Standardiserade räkenskapsutdrag
		Lantbrukets företagsregister
		Registret för företagsstatistik
		Kommunregistret
		Fastighetstaxeringsregistret
		(Småhus, Hyreshus, Industri- och
		Lantbruksfastigheter ingår här)

2. SCB kan få data på tre olika sätt:
 - a) Som *rikstäckande register* från ett begränsat antal centrala myndigheter, t.ex. SKV, RFV, PV, KPA, CSN och SPV lämnar data till Inkomst- och taxeringsregistret.
 - b) Som data från ett *stort antal lokala register*. När det inte finns ett rikstäckande register kan data hämtas från de lokala register som finns i olika regioner. T.ex. så lämnar varje högskola data från sitt lokala studie-dokumentationssystem till SCB:s Universitets- och högskoleregister.
 - c) Som uppgifter på *blanketter som är ifyllda av tjänsteman* med hjälp av data från organisationens administrativa system. T.ex. så fyller läkare i dödsorsaksintyg som skickas till SCB:s Dödsorsaksregister. Företag, kommuner och skolor fyller i enkäter från SCB för olika totalundersökningar. Arbetet med att fylla i blanketterna kan ibland vara omfattande, det går t.ex. åt ca 40 timmar för en uppgiftslämnare att räkna om uppgifterna i kyrkokommunernas bokföring till nationalräkenskapernas ändamålsgruppering.
3. Leveransernas form kan vara någon form av datamedium eller pappersblankett. I vissa fall kan telefonkontakt förekomma. De SCB-register som tar emot data från många lokala myndigheter kan få data på flera slag av datamedier och samtidigt även på pappersblankett. T.ex. får SCB-registret Svenska för invandrare data på blankett, diskett med data på textfil eller Excel-fil. Filerna kan även skickas via e-post.

3.3 Helheten skapas genom samordning och samarbete

Det exempel med *Inkomst- och taxeringsregistret* (IoT) som diskuteras i avsnitt 3.1 visar att ett statistiskt register på SCB kan innehålla tre typer av variabler:

- variabler hämtade direkt från de administrativa källorna
- variabler hämtade från andra statistiska register inom SCB
- variabler som har härletts inom det egna statistiska registret.

Man gör även medvetna avgränsningar av objektmängden – inom IoT-registret avgränsas populationen till personer 16 år och däröver som ingick i befolkningen den 31/12 enligt Befolkningsregistret.

Vissa register är så kallade integrationsregister som *uteslutande* innehåller variabler från andra statistiska register. Man tar inte själv emot några data från administrativa myndigheter och organisationer. Lukas-registret som beskrivs i avsnitt 1.3.2 är ett exempel på ett sådant integrationsregister.

Bild 3.5 De olika variabeltyperna inom ett statistiskt register

Variabler från:						
Administrativ källa		Andra SCB-register			Egna härledda variabler	
Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7

Bild 3.6 De olika variabeltyperna inom ett integrationsregister

Variabler från:						
Andra SCB-register					Egna härledda variabler	
Var 11	Var 12	Var 13	Var 14	Var 15	Var 16	Var 17

Detta visar på betydelsen av samordning och samarbete mellan olika register inom SCB. Inom ett visst register behövs variabler från andra register och inom andra register kan man behöva variabler från det aktuella registret. T.ex. så finns det flera register inom SCB, där man behöver olika inkomstvariabler från IoT-registret.

För att registersystemet skall fungera effektivt fordras att:

- Kopplingsvariablerna i basregistren är standardiserade och att träffprocenten är hög när basregister matchas.
- Alla övriga register skall kunna länkas till sitt basregister med de standardiserade kopplingsvariablerna.
- Objektavgränsningar och objektmängderna är avstämda mot respektive basregister.
- Variablerna i systemets olika register är samordnade. Variabelnamn och definitioner överensstämmer för samma variabel i olika register. Metadata samutnyttjas för att undvika dubbelarbete.
- Dubbelarbete undviks genom att statistiska metoder och IT-verktyg utvecklas gemensamt.
- Gemensamma sekretessregler tillämpas av alla.

Samordning och samarbete ger oss ett bättre fungerande registersystem med större möjligheter. Befintlig statistik blir mer konsistent och får högre kvalitet, samtidigt som möjligheterna till ny registerbaserad statistik ökar.

Det som är utmärkande för statistiska register är att de skall kunna användas för många syften. Ett visst register skall dels användas för de olika produkterna på den egna enheten men även kunna användas av andra enheter för undersökningar med helt andra syften. All SCB-statistik bygger på registersystemet – även urvalsundersökningarna och nationalräkenskaperna!

Exempel 1: Populationsavgränsningar

Många produkter inom SCB avser årlig personstatistik där populationen består av befolkningen den 31/12. Denna avgränsning görs inom Befolkningsregistret som sedan används av olika produkter. Detta medför att det blir konsistens mellan olika statistikprodukter med avseende på antal personer totalt och uppdelat efter ålder, kön och region. På motsvarande sätt kan fastighets-, företags- och arbetsställepopulationer definieras för någon viss tidpunkt under året.

Nationalräkenskaperna behöver en annan typ av populationsdefinition. Deras årliga statistik avseende befolkningens inkomster avser de personer som ingick i befolkningen någon gång under året. I ett flexibelt registersystem måste det vara möjligt att bilda både årsskiftes- och årgångspopulationer.

Exempel 2: Andelen träff vid matchning

När man framställer det årliga Sysselsättningsregistret matchas kontrolluppgiften för varje persons huvudsakliga arbetsgivare mot Företagsregistret. Avsikten är att hämta uppgifter om arbetsställets näringsgren och belägenhet från företagsregistret.

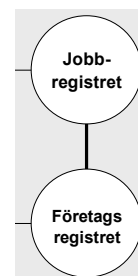
För den första årgången av Sysselsättningsregistret som avsåg 1985 var andelen träff vid denna matchning 93,6 %. För 1999 års Sysselsättningsregister var andelen träff 98,3 %. Länken mellan Jobb- och Företagsregistret har förbättrats mellan 1985-1999 tack vare av det kompletteringsarbete som har lagts ner. Arbetet består av att ta fram listor över förvärvsarbete som inte fått träff vid matchning mellan kontrolluppgifter och Företagsregistrets arbetsställen och att göra enkäter till de berörda företagen.

Exempel 3: Objektavgränsningar

Inom företagsstatistiken är det ett centralt problem att kunna avgränsa statistiskt intressanta objekt. De administrativa materialen avser i första hand juridiska enheter (JE) medan den statistiskt intressanta verksamheten bedrivs inom en organisation där flera juridiska enheter kan ingå. Inom Företagsregistret läggs mycket arbete ner på att bilda företagsenheter (FE) och arbetsställen (AE) – två statistiskt betydelsefulla objekttyper.

Exempel 4: Samutnyttja metadata

Metadata behövs både i statistikproduktionen och när resultaten används. För att undvika dubbelarbete bör variabler namnges och dokumenteras endast en gång hos det ansvariga registret som har tagit emot eller skapat variablerna. När andra register hämtar variabler från det ansvariga registret bör metadata överföras samtidigt.



Exempel 5: Konsistens vid granskning och rättning

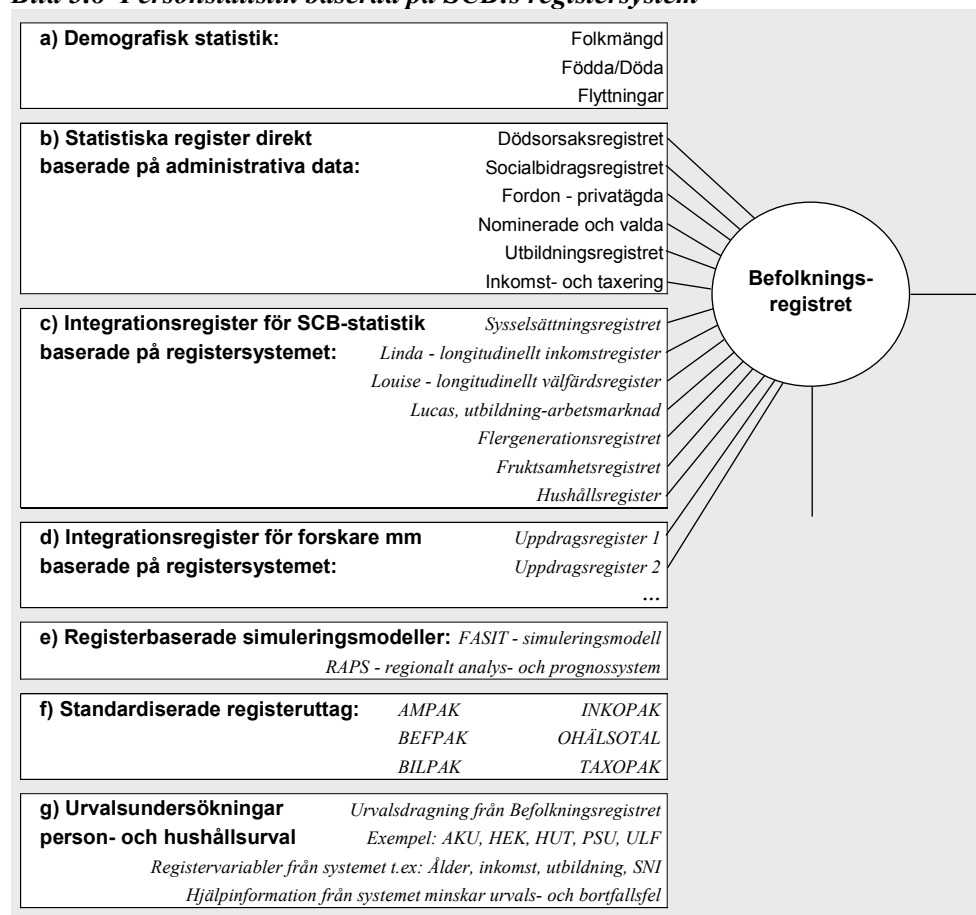
Hur skall granskning och rättning av administrativa data ske? Flera register kan bygga på samma administrativa källor och man gör var sin granskning. Om man då hittar olika fel och åtgärdar felen på olika sätt, får registersystemet som helhet sämre kvalitet och onödiga konsistensbrister. Genom samarbete bör upptäckta fel och lämpliga åtgärder förmedlas till berörda register. Av bild 3.2 framgår att kontrolluppgiftsdata levereras till IoT-registret från Kontrolluppgiftsregistret som tagit emot data från SKV. Dessa två SCB-register bör alltså samarbeta om granskning och rättning av detta administrativa material.

3.4 Skapa ytterligare register och användningar

Den del av registersystemet som avser personstatistik är den mest utvecklade och kan användas för att visa på de möjligheter som ett samverkande system kan ge.

3.4.1 Personstatistik baserad på SCB:s registersystem

De produkter som anges i ruta a) i bild 3.8 bygger helt på data från Befolkningsregistret.

Bild 3.8 Personstatistik baserad på SCB:s registersystem

AKU=Arbetskraftsundersökningen

HEK=Hushållens ekonomi

HUT=Hushållens utgifter

PSU=Partsympatiundersökningen

ULF=Undersökningen av befolkningens levnadsförhållanden

De register, som finns i ruta b), har bildats med administrativa data från olika registerlämnande myndigheter. Inkomst- och taxeringsregistret som har tagits upp i flera exempel ingår i denna grupp. Varje sådant register ensamt är grunden för olika statistikprodukter.

Genom att sambearbeta flera register inom systemet kan olika *integrationsregister* bildas. Dessa kan vara tillfälliga för specifika uppdrag eller permanenta för nya SCB-produkter. De register som anges under c) och d) i bild 3.8 är sådana integrationsregister och de har bildats utan ytterligare datainsamling. Förmågan att göra sambearbetningar beror på hur samordnat och välutvecklat SCB:s registersystem är. Datalagret skall vara bra organiserat och det skall finnas effektiva IT-verktyg för olika slag av registerbearbetningar.

Registerbaserade simuleringsmodeller (ruta e) har utvecklats för att bl.a. visa hur förändringar i skattesystemet påverkar ekonomin för personer och hushåll. De standardiserade uttagen av regionala tabeller (ruta f) har stor efterfrågan. Slutligen baseras alla urvalsundersökningar (ruta g) på registersystemet.

3.4.2 Varför är denna del av systemet så bra?

Basregistret har under en lång tid använts för att framställa förstklassig demografisk statistik. Detta har stimulerat framväxten av en god registerstatistisk tradition som i sin tur har bidragit till att basregistret har hög kvalitet. I samband med plattformbytet från stordator till den nya databastekniken 1998 fick Befolkningsregistret en effektiv IT-struktur som underlättar framställningen av olika registerversioner. Historiska uppgifter från tidigare årgångar finns lagrade så att longitudinella register kan skapas relativt enkelt. Registrets struktur beskrivs av Wilén och Johannesson (2002).

Personnumret har använts länge inom folkbokföringen och kvaliteten på denna kopplingsvariabel är hög. Personnumret är därför en bra länk mellan Befolkningsregistret och alla andra personregister och även till Aktivitetsregistret. Även länken till Fastighetsregistret (fastighetsbeteckning) har god kvalitet så att personer kan knytas till bostadsfastigheten.

De andra registren i denna del av systemet använder Befolkningsregistret som avser befolkningen den 31/12 varje år för att definiera sina populationer. Detta ger konsistens mellan olika register produkter med avseende på population.

Befolkningen den 31/12 är ett exempel på en standardiserad population

Ansvar för viktiga variabler av gemensamt intresse är decentraliserat så att man inom olika register ansvarar för "sina" variabler. Dessa variabler samutnyttjas, vilket leder till olika register är konsistenta med avseende på variabler.

Befolkningsregistret	Demografiska variabler, civilstånd, födelseland
Utbildningsregistret	Utbildning enligt SUN2000
Sysselsättningsregistret	Sysselsättningsstatus
Inkomstregistret	Olika inkomstvariabler

Dessa variabler är exempel på variabler som bör vara standardiserade

Registerstatistikens fördelar utnyttjas genom att många kvalificerade integrationsregister har bildats helt med befintliga data. Många av dessa register är longitudinella och har stor betydelse för samhällsvetenskaplig och medicinsk forskning.

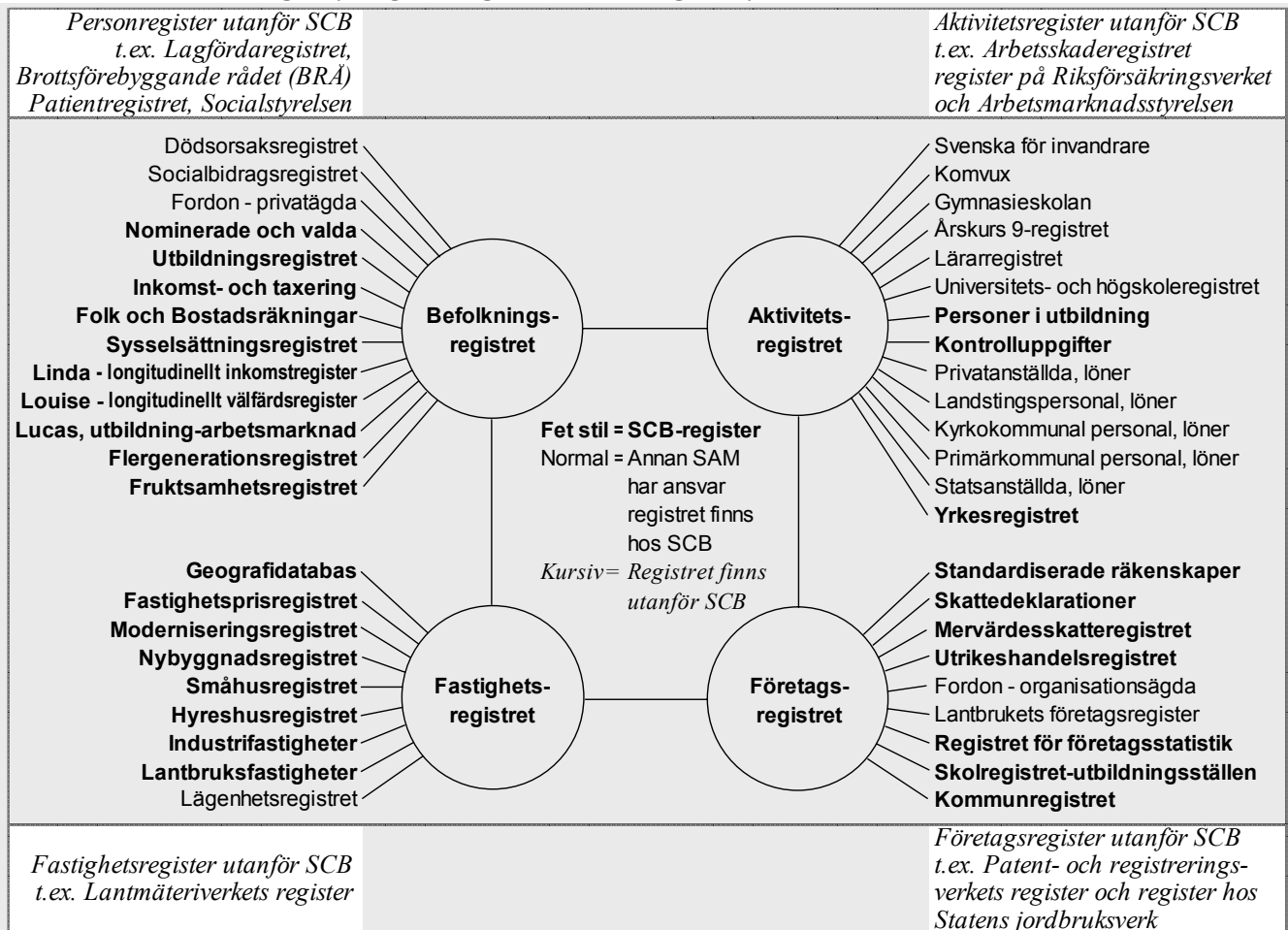
3.5 SCB:s och övriga SAM:s register samverkar

Vi har hittills enbart diskuterat SCB:s registersystem. De register som ingår i detta system kan dels vara register som SCB har ansvar för och dels register som andra *statistikansvariga myndigheter* (SAM) ansvarar för. Dessutom finns utanför SCB många register som skulle kunna kopplas till SCB:s basregister.

I kapitel 2 och 3 där vi beskriver hur SCB:s registersystem som helhet skapas, påpekar vi, att ett effektivt system förutsätter samordning och samarbete mellan avdelningar och enheter inom SCB. På motsvarande sätt bör SCB och övriga statistikansvariga myndigheter samarbeta i registerfrågor för att skapa ett bra registersystem för hela det svenska statistiska systemet.

Det är viktigt för SCB att ha tillgång till bra register från andra myndigheter. På motsvarande sätt är det viktigt för andra myndigheter att ha tillgång till SCB:s registersystem. Samarbete kring registerfrågor är därför av gemensamt intresse.

Bild 3.9 SCB:s och övriga myndigheters gemensamma registersystem



Fordonsregistret (privat- och organisationsägda fordon) används för SCB:s nationalräkenskaper och skulle kunna användas i andra sammanhang t.ex. inom välfärdsstatistiken. Genom att knyta bilägaren till utbildning, inkomst mm. från SCB:s system skulle SIKÄ, som är den myndighet som ansvarar för Fordonsregistret, kunna göra ny egen statistik.

SIKÄ
Statens institut för
kommunikationsanalys

De olika utbildnings- och lönerregistren kommer att få ökad betydelse för SCB:s statistik. De kan användas för Aktivitets- och Sysselsättningsregistrets behov så att kvaliteten hos den registerbaserade sysselsättningsstatistiken förbättras. Detta är exempel på register som andra SAM än SCB har ansvaret för – ett gott samarbete med dessa myndigheter är därför viktigt för många statistikprodukter inom SCB.

För att Lantbrukets företagsregister skall kunna kopplas till Företagsregistret måste enhetsbegreppen stämma överens. Metodiken för att avgränsa jordbruksföretag blir då inte endast betydelsefull för Statens jordbruksverk utan är även en viktig fråga för SCB och hela det statistiska systemet.

Statistiken över nystartade företag är en del av den demografiska företagsstatistiken. Företagsregistret har en nyckelroll för denna typ av statistik. Även om en annan SAM-myndighet (ITPS) ansvarar för denna officiella statistik måste statistiken vara avstämd mot förändringar i SCB:s Företagsregister.

ITPS
Institutet för tillväxt-
politiska studier

Ofta behöver olika forskningsinstitut möjlighet att analysera data som finns i SCB:s registersystem. Dessa behov skulle kunna tillgodoses antingen genom att forskarna analyserar data som finns hos SCB eller att forskningsinstitutet får aidentifierade kopior av SCB-register för att bygga upp ett eget system. Detta aktualiserar tre problem:

- onödiga kostnader och risk för kvalitetsbrister på grund av ett omfattande dubbelarbete med att hålla många registerkopior aktuella och konsistenta om flera institut har egna versioner av SCB-register
- sekretess- och integritetsproblemen ökar om kopior på register finns utanför SCB
- det måste finnas ett nära samarbete mellan SCB och forskarna, så att forskarna förstår de möjligheter som SCB:s registersystem erbjuder och att SCB förstår forskarnas behov.

I flera länder lämnas inte data ut, utan analyserna får ske hos den statistiska myndigheten. I Sverige lämnas ibland data ut, ett exempel på detta är Institutet för arbetsmarknadspolitisk utvärdering (IFAU), som bland annat har delar av fyra aidentifierade SCB-register, Kontrolluppgiftsregistret, Linda, Louise och Sysselsättningsregistret.

Kapitel 4. Hur skapas ett statistiskt register?

Metodik betyder enligt ordboken metodlära

Behovet av metodik är lika stort för registerstatistiken som för surveystatistiken. Inom många registerenheter vid SCB görs sedan länge kvalificerade bearbetningar och man har byggt upp stor kompetens när det gäller hur registerstatistik skall produceras. Man använder visserligen metoder, men har ingen generell formulerad metodik att stödja sig på. I detta kapitel formuleras vissa generella begrepp och principer som kan tjäna som en början till en allmän registerstatistisk metodik.

Det är viktigt att skilja på register och registerbaserade undersökningar

Surveymetodiken ger svar på frågan "hur gör man en urvalsundersökning respektive en totalundersökning?" På liknande sätt skall den registerstatistiska metodiken ge svar på frågan "hur gör man en registerbaserad undersökning?" Men eftersom ett visst statistiskt register inte endast är avsett för en viss undersökning, utan skall ensamt eller i kombination med andra register kunna användas till flera olika undersökningar, måste man tänka mer flexibelt. Den registerstatistiska metodiken ska därför även ge svar på frågan "hur bygger man upp ett statistiskt register, som kan användas för vissa givna undersökningar, men som även bidrar till möjligheterna att göra andra undersökningar med registersystemet?" Detta är frågor som både skall diskuteras när man överväger ett nytt register, men även när man vidareutvecklar befintliga register.

4.1 Hur strukturera registerbaserade undersökningar?

För att inte blanda samman olika begrepp är det viktigt att tydligt skilja på de tre typerna av undersökningar som förekommer inom SCB. Här jämförs likheter och skillnader i förutsättningar för urvals-, total- och registerbaserade undersökningar. De senare som diskuteras i detta kapitel har skuggats i bild 4.1.

Bild 4.1 Likheter och skillnader mellan olika typer av undersökningar

Urvalsundersökning	Totalundersökning	Registerbaserad undersökning
Ingår inte i registersystemet	Ingår i registersystemet – kan användas för andra registerbaserade undersökningar	
Använder registersystemet för att avgränsa populationer och som variabelkälla		
Viktigt med urvalsdesign, estimation, osäkerhetsmått	Viktigt med systemtänkande och samordning med andra registerbaserade undersökningar	
Egen datainsamling – gör egna frågeformulär	Använder andras administrativa register	
Granskning – kan kontakta uppgiftslämnare	Granskning - kan kontakta registerlämnande myndighet	
Bortfall – påminnelser, när avbryta datainsamlingen?	Bortfall/undertäckning kan synas som mismatch	
Kvalitetsbrister - urvalsfel, mätfel	Kvalitetsbrister - mätfel	Kvalitetsbrister – relevansfel, bristande jämförbarhet
Små tabeller – kan inte dela upp urvalet i för små grupper	Presentation – stora flervägsindelade tabeller	

Urvalsundersökningar görs idag av objekt inom alla basregister. Personer, företag/arbetsställen och fastigheter kan väljas direkt medan hushåll och aktiviteter väljs genom tvåstegsurval:

- först väljs person, sedan undersöks hela den valda personens hushåll
- först väljs företag, sedan undersöks alla anställningar inom det valda företaget med avseende på t.ex. lön eller yrke.

Totalundersökningar görs idag endast av objekt inom Företagsregistret. Kommuner, skolor respektive företag totalundersöks med postenkäter eller elektroniska frågeformulär. Undersökningsmetoden medför en tung uppgiftslämnarbörda och en förutsättning är att uppgiftslämnandet är obligatoriskt. Inom företagsstatistiken förekommer *kombinationsundersökningar* där register används för små företag, mellanstora företag undersöks genom urval och de stora företagen totalundersöks. Syftet med detta är att minska kostnader och uppgiftslämnarbörda.

Totalundersökning är en undersökning med egen datainsamling

Registerbaserade undersökningar görs av objekt inom alla basregister. Det är metodiken för denna typ av undersökningar som beskrivs i detta kapitel. De skuggade delarna i bild 4.1 är viktiga förutsättningar för de registerstatistiska metoder som diskuteras.

4.1.1 Metodarbete på olika nivåer

Registerstatistiskt metodarbete bedrivs på tre nivåer:

- *Arbetet på systemnivå.* Arbetet innebär att strukturera och samordna ett stort antal register till ett fungerande system av statistiska register. Systemet som helhet skall utvecklas för att möjliggöra ny och förbättrad statistik. Detta diskuteras i de tidigare kapitlen.
- *Arbetet med att skapa ett statistiskt register inom det befintliga systemet.* Arbetet innebär ansvar för de metoder som används när ett statistiskt register skapas. Hur använder man administrativa data för att bygga upp ett primärt statistiskt register? Hur bygger man upp ett basregister? Hur använder man registersystemet för att bygga upp ett integrationsregister? Arbetet syftar till att skapa den eller de datamatriser som registret skall bestå av. Denna typ av frågor diskuteras kortfattat i detta kapitel och utförligare i Wallgren (2004).
- *Arbetet med en undersökning baserad på ett statistiskt register.* Arbetet syftar till att med utgångspunkt från undersökningsproblemen analysera registrets datamatriser genom olika beräkningar, aggregeringar och skattningar. Hur skall beräkningarna göras för att ta hänsyn till olika metodfrågor som täckningsproblem, bortfall eller tidsseriebrott? Hur bearbetas register med flervärda variabler? Som resultat av registerbearbetningarna får man ett antal statistiska tabeller för en viss undersökning. Denna typ av frågor diskuteras i Wallgren (2004).

Administrativa data skall bearbetas för att svara mot statistiska behov. Med hjälp av alla tillgängliga källor som har anknytning till det statistiska register som skall skapas, avgränsas registrets objektmängd, objekt och variabelinnehåll på bästa sätt. Vilka registerbearbetningar skall göras? Hur skall de göras?

4.1.2 Arbetsgången då ett statistiskt register skapas

Hur bygger man upp ett statistiskt register som dels skall användas för vissa givna undersökningar och dels bidrar till möjligheterna att med registersystemet göra andra undersökningar? På liknande sätt som arbetet med en urvalsundersökning brukar delas upp i olika steg bör även arbetet med att skapa ett statistiskt register struktureras i olika faser:

1. Problemformuleringsfasen: Vilka är de statistiska behov och önskemål som skall uppfyllas med registret?
2. Inventeringsfasen: Vilka källor finns när man skall skapa det nya registret? Vilka administrativa källor finns och vilka befintliga statistiska register i systemet kan användas?
3. Planeringsfasen kan delas in i tre delar:
 - a. Hur skall man avgränsa registrets objektmängd?
 - b. Hur skall man avgränsa registrets objekt?
 - c. Hur bestämmer man registrets variabelinnehåll?
4. Leverantörskontakter och mottagande av administrativa data, mottagningskontroll och granskning.
5. Bearbetningsfasen kan delas in i tre delar:
 - a. Hur bearbetas befintliga källor så att det nya registret innehåller den önskade objektmängden? Administrativa data kontrolleras, granskas och åtgärdas så att objektmängden blir den önskade. Olika källor matchas och objekt selekteras. Tidsreferenserna bearbetas för att skapa objektmängden för den avsedda tidpunkten eller tidsperioden.
 - b. Vilka bearbetningar skall göras för att kontrollera och ändra objektavgränsningarna? Administrativa data kontrolleras, granskas och åtgärdas så att objektavgränsningarna blir de önskade. Härledda objekt bildas i det nya registret.
 - c. Vilka bearbetningar skall göras för att skapa de avsedda variablerna? Variablerna i de administrativa källorna kontrolleras, granskas och åtgärdas. Åtgärder vidtas om variabelvärden saknas. Olika källor matchas och variabler selekteras och importeras till det nya registret. Härledda variabler bildas i det nya registret.

Arbetet med en registerbaserad undersökning bör struktureras så att man är medveten om tre processer som pågår samtidigt under arbetet med undersökningen. Den första processen består av arbetet med att skapa det statistiska registret och beskrivs i punkterna 1–5 ovan. Parallellt med detta arbete så skall samtidigt registret och bearbetningarna *kvalitetssäkras* och *dokumenteras*.

Kvalitetssäkring

Arbetet med att skapa statistiska register skall kvalitetssäkras. Registrets kvalitet bör bedömas och även beskrivas med olika kvalitetsindikatorer. Även dokumentationsarbetet är en viktig del av kvalitetssäkringen. Felaktig och okritisk användning av administrativa data kan förhindras med metadata som ger information om jämförbarhetsproblem.

Dokumentation

Eftersom ett statistiskt register skall kunna användas av flera användare som utnyttjar registersystemet, måste alla register vara dokumenterade så att alla kan utnyttja dokumentationen. Metadata har en mycket betydelsefull roll vid arbetet med registerstatistik. När man sambearbetar olika register måste man känna till definitioner och jämförbarhetsproblem. Det är viktigt att även bearbetningsmetoderna dokumenteras så att metodutveckling och erfarenhetsutbyte underlättas. Dessa frågor diskuteras i Wallgren (2004).

Bild 4.2 De tre parallella arbetsprocesserna med en registerbaserad undersökning

Skapa och använda ett register	Kvalitetssäkring	Dokumentation
<p>Att skapa ett register (kap 4.2)</p> <ol style="list-style-type: none"> 1. Problemformulering 2. Inventering 3. Planering 4. Mottagande av data 5. Bearbetningar: <ul style="list-style-type: none"> - Skapa population - Skapa objekt - Skapa variabler <p>Beräkningar och tabeller</p> <p>Val av skattningsmetoder för att göra beräkningar och ta fram tabeller</p>	<p>Leverantörskontakter</p> <p>Mottagningskontroll</p> <p>Bortfall, omfattning och orsaker</p> <p>Mismatch, omfattning och orsaker</p> <p>Evalvera kvaliteten hos objekt och variabler</p> <p>Enkäter för registerhållning</p> <p>Inkonsistenser mellan olika källor undersöks</p> <p>Dokumentation</p>	<p>Samtidigt som mikrodata hämtas från olika källregister skall man samtidigt hämta metadata för källorna:</p> <p>Registerpopulation för varje källa</p> <p>Definitioner av importerade variabler</p> <p>Vilka granskningar och bearbetningar har gjorts i olika källor?</p> <p>Vad vet man om de olika källornas kvalitet?</p> <p>Därefter skall det färdiga registret dokumenteras</p>

4.2 Att avgränsa registrets objektmängd

I varje statistisk undersökning skall man definiera undersökningens population. Här diskuteras avgränsningen eller definitionen av objektmängden och i nästa avsnitt behandlas avgränsningen av objekt.

Då ett register skall kunna användas för flera undersökningar bör man med registret kunna definiera olika populationer. För att möta dessa krav skall man kunna göra olika objektselektioner från registret så att olika delmängder passar för olika undersökningar. En annan möjlighet är att göra olika versioner av registret för olika tillämpningar.

Begreppet *population* reserveras till att avse en objektmängd som hör till en viss undersökning. När ett register beskrivs utan att avse en viss undersökning används begreppet *registrets objektmängd*. Men varje statistiskt register skapas för en eller flera huvudsakliga användningar eller undersökningar. Det är därför vanligt att registrets objektmängd överensstämmer med huvudundersökningens population. När man skapar ett nytt register för en viss undersökning skall det nya registrets population avgränsas. Varje källregister har sin objektmängd som helt eller delvis kommer att ingå i det nya registret.

Population är en objektmängd för en viss undersökning

Problemformulerings-
fasen i en under-
sökning diskuteras i
Jeppesen (1973)

4.2.1 Att definiera en population

Varje undersökning utgår från en frågeställning som formuleras i teoretiska eller allmänna termer. De teoretiska begreppen i frågeställningen måste sedan *operationaliseras*, dvs. översättas till mätbara begrepp. När man definierar dessa mätbara begrepp avgör man *vad* som skall undersökas.

Populationsdefinitionen skall entydigt ange vilka objekt som ingår i populationen. Objekttypen skall vara klart angiven. Dessutom ingår alltid en tidsreferens och en geografisk avgränsning. I den geografiska avgränsningen ingår också att ange vilken relation som finns mellan objekten och det geografiska området.

Exempel på populationsdefinition:

”*Permanent boende* (= relation) *individer* (= objekttyp) i *Sverige* (= geografiskt område) den *31/12 2002* (= tidpunkt). Med permanent boende avses ...”

SCB:s riktlinjer för kvalitetsbegrepp och kvalitetsdeklaration (SCB, 2001a) innehåller tre populationsbegrepp:

- *Intressepopulationen* avser populationen i den teoretiska frågeställningen.
- *Målpopulationen* avser den operationaliserade populationen, den teoretiska intressepopulationen har översatts till en konkret och undersökningsbar population, dvs. den population man har som *mål* att undersöka.
- *Rampopulationen* avser den objektmängd som *ramen* och *ramförfarandet* faktiskt leder till. Dessa begrepp hör hemma inom teorin för urvals- och totalundersökningar med egen datainsamling.

För registerbaserade undersökningar använder vi endast de två första begreppen. Då urvalsram och ramförfarande inte förekommer vid registerbaserade undersökningar måste det tredje begreppet *rampopulation* ersättas:

- *Registerpopulationen* avser objektmängden i det register man har skapat för den aktuella undersökningen, dvs. den population man *faktiskt* undersöker.

Det finns skillnader mellan ram- och registerpopulationer. En rampopulation avgränsas *före* datainsamlingen, en registerpopulation skapas *efter* referensperioden när alla administrativa uppgifter har kommit in. En ram kan bestå av annat än de objekt man vill undersöka, t.ex. en karta för att undersöka hushåll.

Exempel: Undersökning av befolkningens struktur med Befolkningsregistret

Vi vill undersöka den svenska befolkningens sammansättning den 31/12 ett visst år. Intressepopulationen är *de permanent boende i Sverige den 31/12* men detta vaga begrepp måste definieras närmare. Denna definition kan göras på olika sätt. Allmänt gäller att en bra definition skall uppfylla följande krav:

- den skall vara *adekvat*, dvs. överensstämmande med vad man vill undersöka
- den skall vara *funktionell*, dvs. den skall vara praktiskt användbar

När man gör dessa definitioner ställs man inför svåra avvägningar mellan vad man vill undersöka (adekvata definitioner) och vad man kan undersöka (funktionella definitioner). I detta exempel kan man bl.a. tänka sig följande definitioner av ”*de permanent boende i Sverige den 31/12 år t*”:

1. Folkbokförda i Sverige den 31/12 år t
2. Folkbokförda i Sverige den 31/12 år t enligt uppgifter som är tillgängliga den sista januari år $t+1$

De som arbetar med Befolkningsregistret brukar vänta till utgången av januari innan man skapar den registerversion som skall avse den 31/12. Man hoppas då att alla folkbokföringshändelser för året har hunnit rapporterats in. Det register som då har skapats får vara det som gäller trots att ett fåtal aviseringar som avser år t kan tillkomma. Definition 2 anger alltså hur man avgränsar registerpopulationen för denna undersökning.

Däremot så anges i kvalitetsdeklarationen för den årliga befolkningsstatistiken att undersökningen avser att beskriva en målpopulation enligt definition 1. Skillnaden mellan målpopulation och registerpopulation är då registrets *täckningsfel*. Dödsfall och flyttningar ur landet som inte har rapporterats före den sista januari ger *övertäckning*, medan födslar och inflyttning som inte har hunnit rapporterats ger *undertäckning*.

Förutom dessa täckningsfel får man också relevansfel om den använda definitionen inte är adekvat. Skillnaden mellan intressepopulationen och målpopulationen är ett av undersökningens *relevansfel*. Det finns troligen mellan 25 000 - 50 000 personer som är folkbokförda i Sverige, men som inte bor permanent i Sverige. Man bedömer att mellan 4 % - 8 % av de utomnordiska invandrarna har lämnat Sverige utan att rapportera detta. Detta relevansfel påverkar statistiska mått som beskriver dödlighet, medelinkomst etc. för utomnordiska invandrare så att måtten blir missvisande.

Uppskattningen avser 1998. Se Qvist (1999)

I exemplet ovan ser vi att SCB:s befolkningsstatistik använder en administrativ definition, den folkbokförda befolkningen, när undersökningens målpopulation definieras. Att använda administrativa begrepp ger alltid definitioner som är funktionella. Det är förnuftig statistisk praxis att använda dessa administrativa begrepp för att definiera målpopulationen om relevansfelen blir små. Men grundregeln är att populationsdefinitionen skall svara mot den statistiska undersökningens krav. Är de administrativa begreppen inte tillräckligt relevanta måste man göra egna definitioner och göra de registerbearbetningar som krävs för att registrets objektmängd skall avspegla den definierade populationen så väl som möjligt.

4.2.2 Att definiera en population – primära statistiska register

Hur skall de som arbetar med SCB:s primära statistiska register definiera målpopulationer och avgränsa registerpopulationer? De primära statistiska registren bygger helt eller delvis på administrativa register, och det finns då en risk för att de administrativa systemens objektmängder kommer att styra valet av registerpopulation på ett olämpligt sätt.

Objektmängden i de administrativa registren kanske inte helt täcker den målpopulation som är statistiskt intressant. Den administrativa objektmängden består endast av de objekt som ingår i det administrativa systemet och det kan då finnas både över- och undertäckning jämfört med den statistiskt önskvärda målpopulationen.

Primärt statistiskt register

Statistiskt register som direkt baseras på åtminstone en administrativ källa

Exempel: Den administrativa objektmängden – duger den som målpopulation?
Lantbruksstatistiken kommer i framtiden att baseras på de ansökningar om EU-bidrag som lantbrukarna lämnar till länsstyrelserna. Dessa ansökningar registreras i IAKS-systemet som används för att administrera lantbruksstödet. I bild 4.3 jämförs de arealuppgifter som fanns i ansökningarna om arealbidrag 1995-97 inom IAKS-systemet med motsvarande uppgifter i Lantbrukets företagsregister, LBR, som baserats på totalundersökningar gjorda av SCB. Skillnaderna mellan IAKS och LBR beror på undertäckning i IAKS-registret – vissa lantbrukare ansöker inte om arealbidrag trots att de är aktiva odlare.

Bild 4.3 Undertäckning i ett administrativt register



Slutsatser: Trots att IAKS-registret kanske anses ha god täckning så medför till synes små variationer i täckningen att tidsserien för odlad åkerareal blir helt missvisande – en verklig nedgång 1996-97 ser ut som en fortsatt uppgång.

Man skall inte bortse från bristande täckning i de administrativa registren utan här bör målpopulationen definieras enligt statistikens krav. Därefter skapar man ett nytt register som innehåller den avsedda registerpopulationen med hjälp av det aktuella basregistret, i detta fall genom att selektera objekt från Företagsregistret. Sedan kan detta nya register matchas mot IAKS-registret och då kommer man att upptäcka både eventuell över- som undertäckning i IAKS-registret. Övertäckningen är indikation på eventuella brister i Företagsregistret. Undertäckningen i IAKS-registret blir bortfall i det nya registret. Detta bortfall kan åtgärdas antingen genom bortfallskorrektion eller genom att komplettera med en speciell undersökning för att samla in data för den del av målpopulationen som inte finns med i det administrativa registret.

Integrationsregister
består uteslutande av
variabler från andra
statistiska register

4.2.3 Att definiera en population – integrationsregister

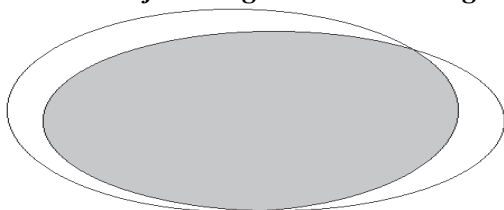
Hur skall de som arbetar med integrationsregister definiera målpopulationer och avgränsa registerpopulationer? Vi tar här upp grundprinciperna för detta.

Exempel: Registeruppdrag med sambearbetningar – val av målpopulation

I många registeruppdrag behöver flera register sambearbetas. Man gör en rad matchningar och importerar variabler från olika register. Slutresultatet blir ett integrationsregister med många variabler som är intressanta för uppdragsgivaren. Men hur har undersökningens målpopulation definierats? Det är lätt hänt

att integrationsregistrets objektmängd blir snittmängden av de matchade registrens objektmängder. Motsvarar denna mängd en adekvat målpopulation för detta uppdrag? Det är inte alls självklart att detta är fallet.

Bild 4.4 Objektmängder vid matchning av två register



Respektive ellips representerar objektmängden i två olika register.

Den skuggade ytan motsvarar de objekt som gav träff vid matchning, dvs. snittmängden.

Även vid ett sådant registeruppdrag skall man börja med att definiera målpopulationen med hänsyn till undersökningsproblemet. Sedan selekterar man en lämplig objektmängd från det aktuella basregistret. Denna objektmängd är registerpopulationen, som sedan matchas mot de register som innehåller de intressanta variablerna. För de objekt där man fått träff, importeras variabelvärdena till det nya registret. För de objekt där man inte fått träff anges partiellt bortfall, dvs. att variabelvärden saknas.

Slutsatser: Snittmängden – vilken målpopulation motsvarar den? Den kan aldrig vara målpopulation, då man inte skulle kunna ge en definition av denna. Den bör inte heller vara registerpopulation, då bortfallet på grund av mismatch skulle riskera att glömmas bort. Att vara medveten om bortfallet är viktigt och dessutom kan man göra bortfallskorrekationer då bortfallets omfattning och struktur är kända.

4.2.4 Basregistren skall användas då populationer avgränsas

Både de som arbetar med integrationsregister och de som skapar primära statistiska register med hjälp av administrativa register bör använda basregistren när objektmängderna avgränsas. Det finns två skäl till detta, dels skall basregistren vara ”bäst” på objektmängder och dels måste SCB:s registerstatistik vara inbördes konsistent. De som arbetar med primära register och integrationsregister bör utgå från respektive basregisters objektmängder och använda någon av de *standardiserade populationer* som skapats för allmän användning:

- *årsskiftesversionen* passar för årlig stockstatistik (t.ex. befolkningens förmögenhet den 31/12),
- *årgångsversionen* passar för årlig flödesstatistik (t.ex. befolkningens inkomster ett visst år),
- en *månadsversion* passar för månadsstatistik.

Generell metod:

- Definiera målpopulationen
- Selektera avsedd objektmängd från basregistret, detta ger registerpopulationen
- Matcha mot de register som innehåller intressanta variabler
- Vid träff: Importera variabelvärden till det register som skapas
- Vid mismatch: Ange partiellt bortfall

4.2.5 Kraven på ett basregister

Det ställs tre slags krav på ett basregister för att de skall kunna användas för att avgränsa registerpopulationer:

- basregistret skall innehålla tidsreferenser, dvs. alla händelser som berör registrets objekt skall vara daterade,
- basregistret skall ha god täckning,
- länkar och indelningsvariabler skall ha hög kvalitet.

Tidsreferenser

För varje objektinstans måste tidpunkten då objektet tillkom vara känd liksom tidpunkten då objektet upphör. Även tidpunkter för andra viktiga händelser, som flyttningar eller förändrad kategoritillhörighet, skall vara kända.

Med hjälp av dessa tidsreferenser kan man skapa registerpopulationer som speglar populationens tillstånd vid en viss önskad tidpunkt eller period.

Tidsreferenserna kan ha olika precision, händelser för personer är som regel kända så att en viss dag kan anges, händelser avseende företag kanske endast är kända så att månad eller endast år kan anges.

Det är viktigt att man i basregistren skiljer på statistiskt relevanta förändringar och statistiskt triviala förändringar. Byte av riktnummer eller postnummer för ett helt område skall inte uppfattas som objekt har flyttat. På samma sätt får inte ett företags byte av bolagsform uppfattas som att ett företag har upphört och ett annat företag har bildats.

Täckning, indelningsvariabler och identifierande variabler

Med hjälp av något basregister avgränsas de flesta populationer i SCB:s undersökningar. Detta ställer höga krav på att basregistrens objekt mängder har god täckning för många undersökningars behov och att länkar och variabler som används för att selektera fram delpopulationer eller dela in i redovisningsgrupper (*indelningsvariabler*) har hög aktualitet och kvalitet. Kvalitetsbrister hos dessa variabler kan ge täckningsfel i delpopulationer, t.ex. genom att uppgift om näringsgren (SNI-kod) saknas eller att SNI-koden är felaktig.

Exemplet ovan som handlar om lantbruksföretag visar vilka krav som kan ställas på ett basregister. I detta fall måste man med Företagsregistret kunna göra en selektion av lantbruksföretag så att det nya registret får en god täckning av den aktuella målpopulationen. För att lyckas med detta bör Företagsregistret dels ha en god täckning av alla aktiva företag, dels bör näringsgrenskoden ha god kvalitet så att den kan användas för att selektera den önskade registerpopulationen. Dessutom innehåller Företagsregistret identifierande variabler som person/organisationsnummer (PeOrgNr), företagsnamn, adress och telefonnummer som är viktiga när man skall matcha mot IAKS-registret.

För att få god täckning i ett basregister skall man kombinera och använda alla relevanta administrativa källor. Detta är en allmän princip för arbetet med att skapa statistiska register – *genom att kombinera många källor förbättras kvaliteten.*

Exempel: Novemberversionen av Företagsregistret anses ha god kvalitet och används som urvalsram för många undersökningar. Den goda kvaliteten behöver inte gälla småföretag och alla näringsgrenar. Matchas novemberversionen år t med IAKS-registret från sommaren år t ser man att undertäckningen är betydande. Genom att komplettera Företagsregistret med kontrolluppgifter, moms och standardiserade räkenskapsutdrag för år t blir täckningen god.

IAKS-registret innehåller lantbruksföretagens ansökningar om arealbidrag

Tabell 4A. Resultat av olika matchningar av Företags- och IAKS-registret 1995

IAKS-registret matchas mot	Kopplingsvariabler	Andel matchande i IAKS
Företagsregistret, november 1995	PeOrgNr	75,2 %
<i>Företagsregistret kompletterat med:</i>		
KU, MOMS och SRU för 1995	PeOrgNr	96,6 %

KU = Kontrolluppgifter
MOMS = Företagens momsdeklarationer
SRU = Företagens standardiserade räkenskapsutdrag

Slutsatser: En slutsats av detta exempel är att Företagsregistret hade dålig täckning av lantbruksföretag. Genom att komplettera Företagsregistret med andra källor kan täckningen förbättras avsevärt. I detta exempel finns fyra källor som skulle kunna användas för att förbättra denna del av Företagsregistret. IAKS-registret är den mest aktuella källan som blir tillgänglig cirka ett år tidigare än de övriga källorna. Dessutom vet man att IAKS-företagen är aktiva lantbruksföretag, dvs. IAKS-registret innehåller även information om SNI.

Om vissa av de variabler i ett basregister, som ofta används för att selektera eller matcha är inaktuella eller ofullständiga, bör man göra *registerhållningsenkäter*. En sådan enkät, som bör vara kortfattad för att inte medföra onödig uppgiftslämnarbörda, kan skickas till de objekt i basregistret där uppgift saknas eller där uppgiften kan misstänkas vara inaktuell.

Exempel: För att upprätthålla kvaliteten i Företagsregistret görs två enkäter per år till alla företag med mer än ett arbetsställe. Denna enkät görs visserligen för att upprätthålla kvaliteten på populationen av arbetsställen, men man frågar samtidigt efter aktuella näringsgrenar och adressuppgifter.

Det är naturligt inom Företagsregistret att man lägger ner mer registervårdsarbete på de stora företagen som ju har stor ekonomisk betydelse i de företagsundersökningar som SCB gör, men i en dynamisk ekonomi som förändras kraftigt kan detta statistiska synsätt leda fel. I Johansson (2001) finns följande exempel där tillväxten i IT-branschen studeras.

Exempel: Småttingar kan vara jätteviktiga!

Johansson beskriver tillväxten inom IT-branschen mellan 1993 och 1998. Tabellen nedan bygger på uppgifter från Företagsregistret (Johansson, s 82-83).

Tabell 4B. Sysselsättning inom IT-branschen efter företagens storleksklass

Storlek, antal anställda	0	1	2-4	5-9	10-19	20-49	50-99	100-199	200-499	500+	Totalt
Antal sysselsatta, 1993	9 157	2 219	6 724	7 548	8 082	10 678	10 123	6 544	16 597	82 588	160 230
Antal sysselsatta, 1998	17 825	3 084	8 459	9 634	11 582	16 434	13 206	12 933	21 045	85 600	199 802
Förändring 1993-98, antal	8 668	865	1 735	2 086	3 500	5 756	3 083	6 389	4 478	3 012	39 572
Förändring 1993-98, %	95 %	39 %	26 %	28 %	43 %	54 %	30 %	98 %	27 %	4 %	25 %

Anm: För företag med 0 anställda har vi räknat med 1 sysselsatt, Johansson har räknat med 0.

Slutsats: Tabell 5B visar att de allra minsta företagen hade antalsmässigt den största tillväxten. Detta visar att det kan vara riskabelt att bortse från småföretagen vid registervårdsarbetet. Viktiga tillväxtnått kan då få låg kvalitet.

Exemplet ovan innehåller flera viktiga metodproblem – hur få jämförbarhet mellan åren, hur mäta sysselsättning och tillväxt? Jämförbarheten är ett stort problem för Johanssons undersökning, då Företagsregistret har ett tidsseriebrott 1996/97 som är störande för studier av småföretag.

Exempel: Populationen i Företagsregistret

Antal företag enl Företagsregistret		
	Gammal serie	Ny serie
1995	562 765	
1996	585 571	
1997	601 385	791 385
1998		810 337
1999		797 338

Registerpopulationen i Företagsregistret består av företag som är momspliktiga, registrerade som arbetsgivare eller F-skatteskyldiga. Från 1996 är ett företag momspliktigt oavsett omsättningens storlek, tidigare var gränsen en årsomsättning på 200 000 kronor. Den nya principen innebär att antalet företag ökade med ca 190 000 till totalt 791 385 företag under 1997 då de nyblivna momspliktiga företagen infördes i registret. Man har alltså låtit administrativa momsregler styra avgränsningen av registerpopulationen. Enligt exemplet i tabell 5A på föregående sida har den gamla serien för jordbruksföretag en undertäckning på ca 25 %. Genom att kombinera flera källor skulle undertäckningen kunna minskas och samma statistiskt relevanta populationsavgränsning kanske skulle kunna ha använts både före och efter momspliktsförändringen.

Ett basregister bygger på källor som kan ha helt olika periodicitet och olika tidsfördröjningar. Registrets täckning varierar då under året.

Exempel: Varierande täckning i Företagsregistret

Företagsregistret bygger på aktuella månadsdata avseende skattedeklarationer för moms, arbetsgivaravgifter och avdragen skatt för företagens anställda. Men för små företag som inte lämnar månadsvisa skattedeklarationer används de årliga självdeklarationerna som kommer året efter inkomståret. Företagsregistret får kontinuerligt information om ny- och ombildade företag med varierande tidsfördröjning. Sådana ny- och ombildningar är särskilt vanliga vid årsskiften. Det krävs både tid och arbete innan dessa förändringar har införts.

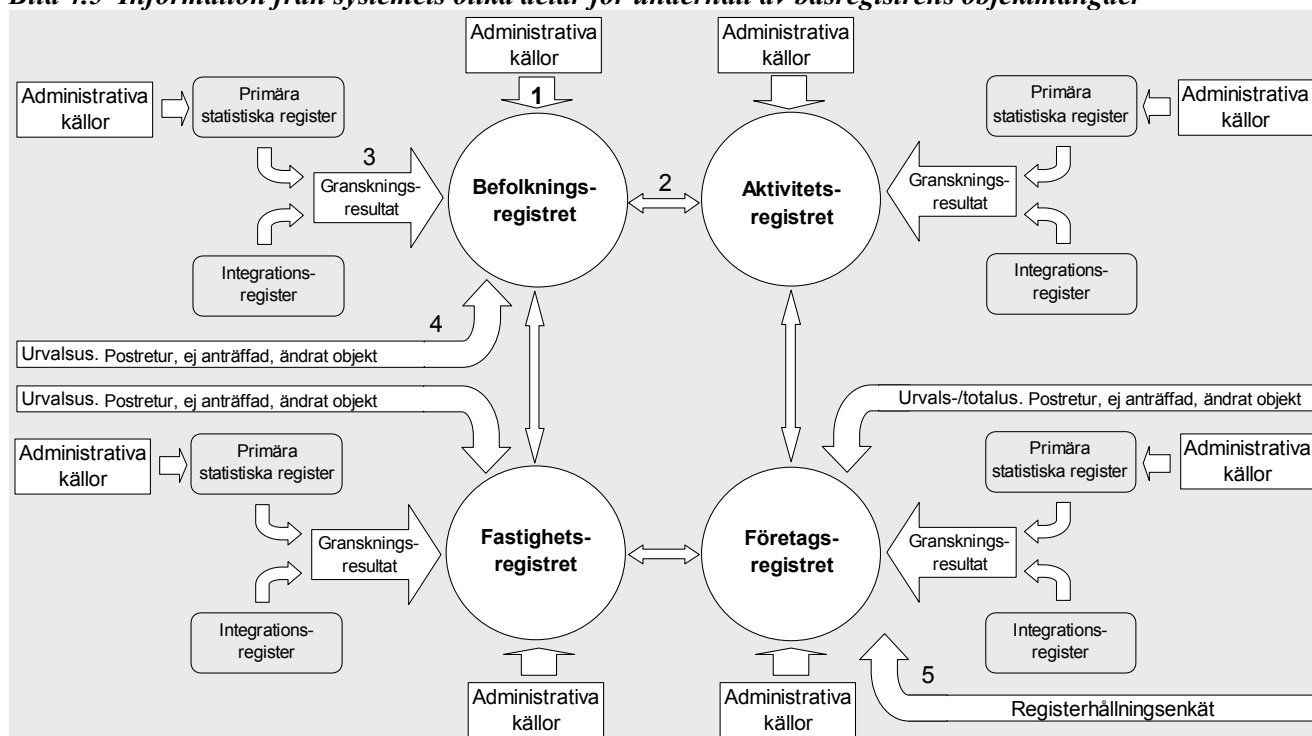
Alla basregister måste uppfylla de krav som vi behandlat i detta avsnitt även om vi mest exemplifierar med Företagsregistret.

4.2.6 Alla skall stötta basregistren!

Hela systemet skall alltså samverka när det gäller att avgränsa objektmängder och objekt. Urvals- och totalundersökningar hämtar sina rampopulationer och de registerbaserade undersökningarna hämtar sina registerpopulationer från basregistren. Alla som gör undersökningar av dessa populationer får kunskaper som måste vidarebefordras till basregistren, så att basregistren kan vara SCB:s bästa källa avseende objektmängder.

Bild 4.5 visar de olika informationsflöden som bidrar till att basregistren kan hålla hög kvalitet med avseende på objekt och objektmängder. Dels skall de som arbetar med basregistren skaffa sig information men dessutom skall de som arbetar med andra register i systemet bidra med information som kan förbättra basregistrens kvalitet med avseende på objektmängder.

Bild 4.5 Information från systemets olika delar för underhåll av basregistrens objektmängder



Siffrorna i bild 4.5 refererar till punkterna nedan:

1. Inom varje basregister skall man hålla kontakt med de *myndigheter* som levererar administrativa data. Förändringar i det administrativa systemet kan påverka registrets objekt och täckning.
2. Inom varje basregister kan man få information från *angränsande basregister* där relaterade objekt kan ge upplysningar om det egna basregistrets objekt.
Exempel: Förändringar i fastigheter med bostäder måste också innebära att Befolkningsregistret skall ändras. Om en person saknar vissa aktiviteter i Aktivitetsregistret kan detta innebära att personen är övertäckning i Befolkningsregistret.
3. Granskningsresultat från olika *statistiska register* kan ge information om täckningsfel eller felaktiga objektavgränsningar.
4. *Urvals- och totalundersökningar* kan via postreturer och orsak till "ej anträffad" ge information om övertäckning och om att objekt har förändrats.
5. Registerhållningsenkäter görs av ett basregister till delar av objektmängden där man misstänker att uppgifterna är ofullständiga eller inaktuella.

Basregistren har alltså huvudrollen när det gäller objektmängder och även när det gäller objektavgränsningar, som diskuteras i nästa avsnitt. För att basregistren skall få så hög kvalitet som möjligt skall alla som använder SCB:s basregister vidarebefordra information om upptäckta kvalitetsbrister.

Det är tre kategorier användare som skall vidarebefordra sådan information:

Information från urvals- och totalundersökningar

De som gör urvals- eller totalundersökningar har direktkontakt med uppgiftslämnare. Postreturer och spårning för telefonintervjuer kan ge information om övertäckning, inaktuella adresser, att objekt har ombildats eller andra förändringar som är intressanta för dem som arbetar med registervård av det aktuella basregistret. Hushållsavgränsningar kan kontrolleras med urvalsundersökningar riktade mot hushåll. Företagsenheter kan kontrolleras i samband med olika företagsundersökningar.

Exempel: Övertäckning i Befolkningsregistret

Greijer (1995) analyserar bortfallet i Arbetskraftsundersökningen och ger en skattning av antal utrikes födda som troligen var övertäckning i Befolkningsregistret. Greijer (1996) analyserar postreturer och bortfall i en enkät till utrikes födda och skattar övertäckningen för olika födelseländer, åldrar och kön.

Information från primära statistiska register

De som arbetar med de primära statistiska registren tar emot administrativa register. Dessa kan innehålla objekt som inte finns i basregistren. Dessutom kan objekt ha försvunnit ur det administrativa registret, vilket kan ge information om att objektet har förändrats, ombildats eller upphört.

Exempel: Övertäckning i Befolkningsregistret

Greijer (1997) analyserar de som saknade disponibel inkomst i Inkomstregistret och ger skattningar på övertäckningen för olika kategorier utrikes födda. Hon visar att man måste korrigera för övertäckning för att inte inkomststatistik efter födelseland skall bli missvisande.

Information från integrationsregister

De som arbetar med integrationsregister kan jämföra variabelvärden, som skall avse samma objekt från olika källor. Om man då upptäcker fel och inkonsistenser kan orsaken till detta vara att objekten har förändrats.

Exempel: Felaktiga objektavgränsningar i lantbruksregister

IAKS-registret innehåller lantbrukens ansökningar om arealbidrag. Wallgren (1999) matchade detta register mot SCB:s Lantbruksregister med länken PeOrgNr. Alla träffar kontrollerades genom att åkerareal från de två registren jämfördes. Ca 9 % falska träffar upptäcktes. Genom att föra samman enheter med samma telefonnummer kunde nya enheter härledas där åkerarealen stämde.

Bilaga 1. Riktlinjer för statistiska register på SCB

Enligt GD-beslut 2001-05-07 skall följande allmänna riktlinjer gälla:

1. SCB skall öka användningen av administrativa uppgifter med syfte att minska uppgiftslämnarbördan och för att kunna producera ny statistik utan att uppgiftsbördan ökar.
2. I alla undersökningar skall övervägas om och vilka registerdata som kan användas. Registerdata kan användas istället för att fråga, för att kontrollera kvalitet och som hjälpinformation för att reducera urvals- och bortfallsfelen i urvalsundersökningar.
3. SCB skall samarbeta med myndigheter som lämnar administrativa register och vid behov aktualisera behoven av statistiska variabler eller anpassning av administrativa uppgifter för att förbättra det statistiska värdet av informationen. Kontakterna skall samordnas när flera avdelningar och program är inblandade. Kvalitetsinformation skall återföras till berörda myndigheter.
4. Mottagna administrativa data skall bearbetas på ett enhetligt sätt. På motsvarande sätt som urvalsundersökningar görs enligt en vedertagen metodik, skall även SCB:s statistiska register, som bygger på administrativa material, skapas enligt vedertagna principer. Samarbete över avdelnings- och programgränser kring registerstatistiska metodfrågor skall därför ske regelmässigt.
5. All statistik skall vara kvalitetsdeklarerad och väl dokumenterad, vilket också gäller bakomliggande register. Register som ingår i registersystemet skall vara dokumenterade i Metadok och det skall finnas en fullständig SCBDOK-dokumentation.
6. Definitioner av objekt, objektmängder och variabler skall vara samordnade. Det skall finnas standardiserade variabler som är särskilt namnsatta, definierade och lätt tillgängliga. De får inte användas med avvikande definition.
7. Varje statistiskt register skall innehålla de kopplingsvariabler som krävs för att kunna länka objekten till respektive basregister.
8. Registren utgör merparten av SCB:s omfattande mikrodatalager. Detta skall förvaltas på ett sådant sätt att kraven på sekretess och integritetsskydd upprätthålls.
9. Om behov finns av att inrätta ett permanent integrationsregister skall en arbetsgrupp tillsättas med uppgift att fastställa om ett nytt register är den bästa lösningen, klarlägga ansvarsområden, skyldigheter och rättigheter för de program som berörs.
10. SCB:s registersystem tillhör det statistiska systemet och skall kunna användas som ett effektivt alternativ för SAM-myndigheter och andra till att själva bygga upp och underhålla omfattande databaser med registerinformation

Bilaga 2. Referenser

- Eurostat (1997): *Proceedings of the Seminar on the Use of Administrative Sources for Statistical Purposes*, Luxembourg, 15-16 January 1997.
- Greijer, Å. (1995): *Uppskattning av övertäckningen i RTB avseende utlandsfödda med hjälp av AKU*. Metodrapport från BoR-avdelningen 1995:3, SCB.
- Greijer, Å. (1996): *Övertäckningen i Registret över totalbefolkningen - en studie av postreturer*. Metodrapport från BoR-avdelningen 1996:7, SCB.
- Greijer, Å. (1997): *Skattning av övertäckningen i folkbokföringen med hjälp av SCB:s inkomstregister*. Metodrapport från BoR-avdelningen 1997:11, SCB.
- Hörngren, J. (1992): *The Use of Registers as Auxiliary Information in the Swedish Labour Force Survey*. R&D Report, Statistics Sweden 1992:13.
- Jeppesen, J. (1973): *Statistisk metode*. Akademiskt forlag, Köpenhamn.
- Johansson, D. (1997): *The Number and the Size Distribution of Firms in Sweden and Other European Countries*. IUI Working Paper no 483, 1997.
- Johansson, D. (2001): *The Dynamics of Firm and Industry Growth – The Swedish Computing and Communications Industry*. Kungliga tekniska högskolan, TRITA-IEO R 2001:05.
- Lundström, S. och Särndal, C-E. (2001): *Estimation in the presence of nonresponse and frame imperfections*. Statistics Sweden 2001.
- Nilsson, Å. (1995): *Brister i folkbokföringen - övertäckningen bland utomnordiska invandrare*. Metodrapport från BoR-avdelningen 1995:2, SCB.
- Qvist, J. (1999): *Täckningsproblem i Registret över totalbefolkningen RTB*. R&D Report, Statistics Sweden 1999:1.
- Registerprojektet (2001): *Det statistiska registersystemet – Utvecklingsmöjligheter och förslag. Slutrapport från Registerprojektet*. R&D Report, Statistics Sweden 2001:1.
- SCB (2001a): *Kvalitetsbegrepp och riktlinjer för kvalitetsdeklaration av officiell statistik*. MIS 2001:1, SCB.
- SCB (2001b): *SCB-data för forskare, individbaserade register med innehållsbeskrivning*. SCB:s hemsida www.scb.se
- SCB (2001c): *Statistisk röjandekontroll av tabeller, databaser och kartor*. CBM, SCB september 2001.
- SCB (2003): *Slutrapport från mikrodataprojektet*. SCB, september 2003.
- Sundgren, B. (1992): *Databasorienterad systemutveckling*. Studentlitteratur.
- Wallgren, A., Wallgren, B. (1999): *Administrative Registers in an Efficient Statistical System – How Can we Use Multiple Administrative Sources?* Statistics Sweden and Eurostat.
- Wallgren, A., Wallgren, B. (2004): *Registerstatistik – administrativa data för statistiska syften*. R&D Report, SCB.
- Wilén, A. och Johannesson, I. (2002): *Ett nytt system för Registret över totalbefolkningen – Fler möjligheter och bättre kvalitet*. Bakgrundsfakta till Befolknings- och Välfärdsstatistik 2002:2, SCB.

Bilaga 3. Registerstatistiska termer

Term	Förklaring	Sida
Administrativt register	Register som används i ett administrativt informationssystem.	7, 11, 27, 30-32, 42, 47-49, 54
Basregister	Ett statistiskt register av stor betydelse för hela registersystemet. Ett basregisters huvudsakliga uppgift i systemet är att avgränsa viktiga objektmängder och att ha länkar till andra basregister.	11, 12, 17, 19, 21-23, 25-30, 36, 40, 43, 49-53
Bortfallskorrektion	Skattningsmetod som syftar till att minska bortfallsfel.	48, 49
Granskning	Grunddata kontrolleras för att upptäcka och åtgärda fel.	27, 32, 38, 42, 44, 45, 53
Hjälpinformation	Registervariabler som används vid estimationsfasen i en urvalsundersökning.	13
Härledd variabel	Ny variabel bildas med hjälp av befintliga variabler. Härledning kan ske genom gruppering/klassindelning, beräkningar, adjungering eller aggregering. Vi skiljer även på lokala och importerade härledda variabler	32, 33, 36, 44
Integrationsregister	Genom att sammanföra eller integrera information från flera register skapas ett integrationsregister. Integrationsregister består <i>uteslutande</i> av variabler från andra statistiska register.	9, 10, 22, 28, 36, 39, 43, 48, 49, 54
Integrerad datainsamling	När en administrativ myndighet samlar in uppgifter åt en statistisk myndighet, i samband med den administrativa datainsamling, kallas detta för integrerad datainsamling. Kontrolluppgiften är ett exempel på detta, där samlar Skatteverket även in data åt SCB.	24
Integritetsskydd	Uppgifter som lämnas till SCB omfattas av statistiksekretessen i sekretesslagen. Vidare finns särskilda regler om skydd mot kränkning av enskilda personliga integritet genom behandling av personuppgifter i personuppgiftslagen som anger när behandling av personuppgifter är tillåten. För känsliga personuppgifter gäller ännu strängare regler.	12
Intressepopulation	Avser populationen i den teoretiska frågeställningen.	46, 47
Klassifikation	En kvalitativ variabel bildas genom att objekt klassificeras, dvs. indelas i ett antal grupper, sådana variabler kallas därför för klassifikationer. Exempel: Näringsgren SNI, yrke SSKYK.	27
Kodning	Ibland används uppgifter i textform för att skapa statistiskt användbara variabler. Genom en kodningsprocess överförs information i delvis ostrukturerad text till helt strukturerade variabelvärden. Sifferkoder används för att ange olika kategorier för en kvalitativ variabel t.ex. 1 för man, 2 för kvinna.	27, 32
Kohort	En grupp av objekt som har sitt ursprung från en viss period. Longitudinella register kan analyseras genom att olika kohorter följs över tiden.	8

Kombinationsundersökning	Undersökning där total-, urvals- och registerbaserad undersökningsmetodik används för olika delar av populationen.	43
Konceptuell modell	En konceptuell modell beskriver systemets objekttyper och relationer oberoende av fysiska/tekniska restriktioner.	17-20
Konsistens	Konsistens betyder överensstämmelse. Undersökningar är konsistenta om de avser samma population och objekten har samma mätvärden på gemensamma variabler och ger samma skattningar.	27, 32, 37-39, 45, 54
Kopplingsvariabel	En länk mellan två register består av en eller flera kopplingsvariabler. En kopplingsvariabel består av information som identifierar enskilda objekt eller grupper av objekt.	11, 17, 18, 22, 23, 25, 36, 39, 51
Kvalitetssäkring	Kvalitetssäkring syftar till att undersöka, åtgärda och förebygga kvalitetsbrister som kan uppstå i olika moment i arbetet med ett register.	12, 16, 32, 44, 45
Källregister	Med källregister avser vi både de administrativa källorna och de SCB-register som används för att skapa ett nytt register.	31, 33, 45
Inkonsistens	Brist på konsistens, se konsistens	
Länk	Består av en eller flera kopplingsvariabler som identifierar objekt. Om sådana variabler har hög kvalitet och finns i flera register är de lämpliga för att användas som länkar när flera register matchas. En länk består av information som identifierar enskilda objekt.	10-12, 17-19, 21, 23-30, 36, 37, 39, 50
Makrodata	Avser sammanfattande uppgifter för grupper av objekt och brukar presenteras som innehållet i statistiska tabeller.	6
Matchning	Länkarna i två eller flera register jämförs. Resultatet blir träff eller icke-träff (mismatch).	10, 22, 27, 29, 32, 37, 48, 49, 51
Metadata	Den information som behövs för att kunna använda och tolka statistik. Beskriver data genom att ange definitioner av objekt och variabler samt metodik och kvalitet.	12, 31, 36, 37, 44, 45
Mikrodata	Avser uppgifter för enskilda objekt	6, 7, 31, 45
Mismatch	Se matchning	32, 42, 45, 49
Mottagningskontroll	Granskning av administrativa data i register som kommer utifrån till SCB.	32, 44, 45
Målpopulation	Intressepopulationen avser populationen i den teoretiska frågeställningen. <i>Målpopulationen</i> avser den operationaliserade populationen, den teoretiska intressepopulationen har översatts till en konkret och undersökningsbar population.	46-50
Objekt	En population består av ett antal objekt eller enheter.	6, 7, 9-12, 17-23, 25-28, 31-33, 43-54
Objektavgränsning	Definition av objekttyp i ett register eller en population.	15, 21, 32, 36, 37, 44, 53, 54

Objektmängd	När vi beskriver ett register utan att avse en viss undersökning använder vi begreppet <i>registrets objektmängd</i> . Men varje statistiskt register skapas för en eller flera huvudsakliga användningar eller undersökningar. Det är därför vanligt att registrets objektmängd överensstämmer med huvudundersökningens <i>population</i> .	7, 12, 18, 21, 23, 25, 27, 30, 32, 36, 43-50, 52, 53
Objektselektion	Se selektion	10, 45
Objekttyp	Det slag av objekt som ett register eller en population består av. Exempel: Person, hushåll, arbetsställe, fastighet.	17-21, 23, 25, 26, 28, 37, 46
Population	Avgränsning i tid och rum av en mängd objekt av samma typ med vissa egenskaper. En population avgränsas för att passa syftet med en viss statistisk undersökning.	7, 12, 13, 15, 17, 21, 23-26, 36, 37, 39, 42, 45-52
Populations-avgränsning	Synonym till populationsdefinition.	37, 52
Post	Observation	31
Primärt statistiskt register	Statistiskt register som direkt baseras på åtminstone en administrativ källa.	43, 47, 49, 54
Rampopulation	Avser den objektmängd som ramen och ramförfarandet leder till.	46, 52
Register	En fullständig förteckning över objekten i en viss definierad objektmängd, på grund av kvalitetsbrister kan vissa objekt saknas. Objekten skall vara identifierbara så att registret kan uppdateras och utökas med nya variabelvärden för respektive objekt.	7
– <i>Administrativt register</i>	Register som används i ett administrativt informationssystem	7, 11, 27, 30-32, 42, 47-49, 54
– <i>Basregister</i>	Ett statistiskt register av stor betydelse för hela registersystemet. Ett basregisters huvudsakliga uppgift i systemet är att avgränsa viktiga objektmängder och att ha länkar till andra basregister.	11, 12, 17, 19, 21-23, 25-30, 36, 40, 43, 49-53
– <i>Integrationsregister</i>	Genom att sammanföra eller integrera information från flera register skapas ett integrationsregister. Integrationsregister består <i>uteslutande</i> av variabler från andra statistiska register.	9, 10, 22, 28, 36, 39, 43, 48, 49, 54
– <i>Källregister</i>	Med källregister avser vi både de administrativa källorna och de SCB-register som används för att skapa ett nytt register.	31, 33, 45
– <i>Primärt statistiskt register</i>	Statistiskt register som direkt baseras på åtminstone en administrativ källa.	43, 47, 49, 54
– <i>Statistiskt register</i>	Register som har bearbetats för statistiska ändamål. Genom att bearbeta administrativa register skapas ett statistiskt register.	6, 8, 10-13
Registerbaserad undersökning	Statistisk undersökning utan egen datainsamling. Istället används befintliga administrativa eller statistiska register.	6, 7, 9, 14, 15, 27, 29, 42-46, 52
Registerhållnings-enkät	En enkät som skickas till de objekt i ett basregister där uppgift saknas eller kan misstänkas vara inaktuell.	45, 51, 53

Registerpopulation	Avser objektmängden i det register som har skapats för den aktuella undersökningen, dvs. den population man faktiskt undersöker.	45-50, 52
Registerstatistik	Statistik som framställs genom registerbaserade undersökningar	6-8
Registerstatistiska bearbetningar	De bearbetningar som görs för att skapa ett statistiskt register. Bearbetningarna består bland annat av kontroll av grunddata, åtgärder pga. saknade värden, samkörningar och selekteringar, bildande av härledda objekt och härledda variabler.	11, 32
Registersystem	För att mikrodata skall kunna samutnyttjas effektivt skall de statistiska registren infogas i ett gemensamt och samordnat registersystem.	7, 8, 10-13
Registrets objektmängd	Avser objektmängden i ett register när man inte avser en viss undersökning.	43-45, 47, 49
Relation	Mellan olika objekt av samma objekttyp kan det finnas relationer av olika slag, t.ex. förälder – barn. Mellan objekt av olika objekttyp kan det också finnas relationer av olika slag, t.ex. anställd – arbetsgivare. Då många relationer är juridiskt viktiga finns uppgifter om relationer ofta i administrativa register.	17-20, 22, 23, 33, 46
Relationsobjekt	En relation mellan två objekt kan ses som ett relationsobjekt. Det är nödvändigt att objektifiera en relation när det finns variabler som måste knytas till kombinationer av objekt snarare än till enskilda objekt. Exempel: Relationen person anställd av företag, är en statistiskt viktig relation som kan beskrivas med variabler som yrke, lön etc. Då samma person kan ha flera anställningar är dessa variabler inte personvariabler.	19, 22-24
Relevansfel	Om populationens definition och/eller variabelernas definitioner inte är adekvata för undersökningens syften, så innehåller undersökningens resultat relevansfel.	14, 42, 47
Sambearbetning	En registerbearbetning där <i>flera</i> register bearbetas.	27, 32, 39, 48
Selektion	Ett nytt register skapas genom att välja vissa objekt och vissa variabler ur ett befintligt register. Vi skiljer på objektselektion, där vissa objekt selekteras och variabelselektion, där vissa variabler selekteras	9, 10, 31, 32, 45, 50
Skattning	Med ett befintligt statistiskt register kan tabeller med skattningar produceras och publiceras. Termen skattning används allmänt för urvalsundersökningar, men bör även användas inom registerstatistiken. Även här måste man skilja på det verkliga värdet och den skattning som registret ger.	6, 13, 15, 43, 54
Standarder	Vissa klassifikationer används som statistiska standarder, dvs. som enhetliga variabeldefinitioner. Se klassifikation	27

Standardiserad population	En viktig uppgift för basregistren är att skapa standardiserade populationer, som sedan används för att skapa registerpopulationer i systemets övriga register. Dessa standardiserade populationer skall ha hög kvalitet med avseende på täckning och viktiga indelningsvariabler. De skapas därför efter den tidsperiod eller tidpunkt de skall avse.	12, 21, 39, 49
Standardiserad variabel	Variabler av grundläggande betydelse som används inom flera register. Det skall finnas ett tydligt ansvar för dessa variabler med avseende på definitioner, namngivning och dokumentation.	12, 27
Statistiskt informations-system	Ett informationssystem som skall uppfylla statistiska syften. Till skillnad från enstaka undersökningar så genererar ett statistiskt (informations)system ständigt nya data.	17
Statistiskt register	Register som har bearbetats för statistiska ändamål. Genom att bearbeta administrativa register skapas ett statistiskt register.	6, 8, 10-13
Surveystatistik	Statistik baserad på urvals- eller totalundersökningar.	7, 12, 14-16, 42
Totalundersökning	Undersökning med egen datainsamling, där mätvärden för alla objekt i populationen samlas in.	6, 7, 11, 12, 35, 42, 43, 46, 48, 52-54
Urvalsram	Se rampopulation	11, 13, 21, 24, 25, 46, 51
Urvalsundersökning	Undersökning med egen datainsamling, där mätvärden för ett sannolikhetsurval av populationens objekt samlas in.	7, 8, 12-14, 16, 25, 32, 37, 39, 42, 43, 44, 54
Variabelselektion	Se selektion	10
Årgångsregister	Den registerversion som innehåller alla objekt som har funnits någon gång under ett visst år. Objekt som tillkommer eller upphör under året skall ingå med angivna händelsedatum.	21
Övertäckning	Registret innehåller objekt som inte tillhör målpopulationen.	22, 47, 48, 53, 54

Administrativa data från olika myndigheter och organisationer används inom SCB för att bygga upp ett system av statistiska register. Utan att kräva en ökad uppgiftslämnarbörd kan dessa administrativa data bearbetas för statistiska ändamål, så att administrativa register omvandlas till statistiska register.

Denna handbok beskriver det *system av statistiska register* som är grunden för huvuddelen av den samhällsstatistik som produceras i Sverige.

ISSN: 1652-5981
ISBN: 91-618-1235-8

Statistikpublikationer kan beställas från SCB, Publikationstjänsten, 701 89 ÖREBRO, e-post: publ@scb.se, tfn: 019-17 68 00, fax: 019-17 64 44. De kan också köpas genom bokhandeln eller direkt hos SCB, Karlavägen 100 i Stockholm. Aktuell publicering redovisas på vår webbplats (www.scb.se). Ytterligare hjälp ges av Bibliotek och information, e-post: information@scb.se, tfn: 08-506 948 01, fax: 08-506 948 99.