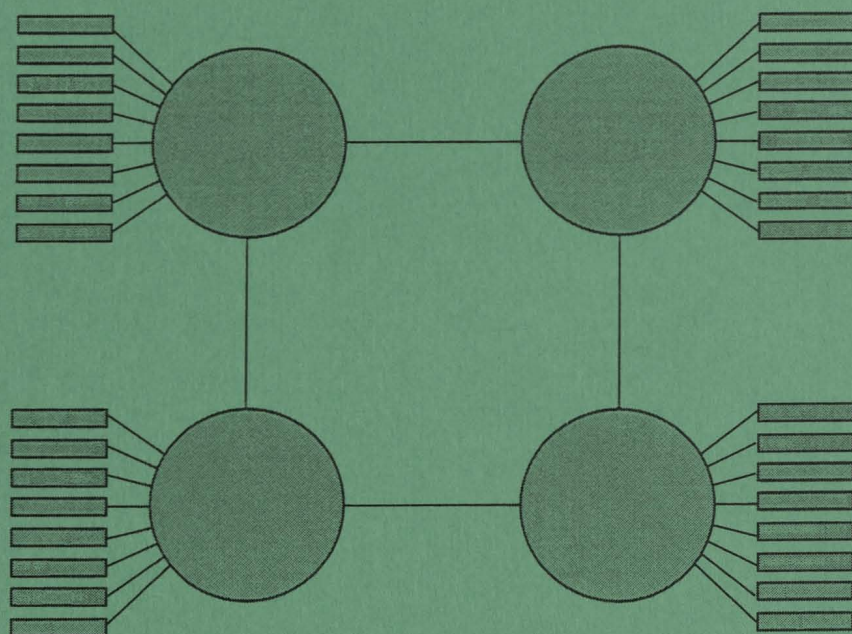


Det statistiska registersystemet

Utvecklingsmöjligheter och förslag



Slutrapport från Registerprojektet

GD-beslut med anledning av Registerprojektets slutrapport

INLEDNING

TILL

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.

Häri ingår Abstracts : sammanfattningar av metodrapporter från SCB med egen numrering.

Föregångare:

Metodinformation : preliminär rapport från Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån. – 1984-1986. – Nr 1984:1-1986:8.

U/ADB / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1986-1987. – Nr E24-E26

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Efterföljare:

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report 2001:1. Det statistiska registersystemet. Utvecklingsmöjligheter och förslag. Digitalt skapad fil, anpassad efter de digitaliserade delarna i serien och kompletterad med ett rättelseblad. Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-2001-X101OP0101



Det statistiska registersystemet

Utvecklingsmöjligheter och förslag

Slutrapport från Registerprojektet

Följande två rättelser gäller för R&D Report 2001:1 –

1) Författare skall vara: *SCB, Registerprojektet*

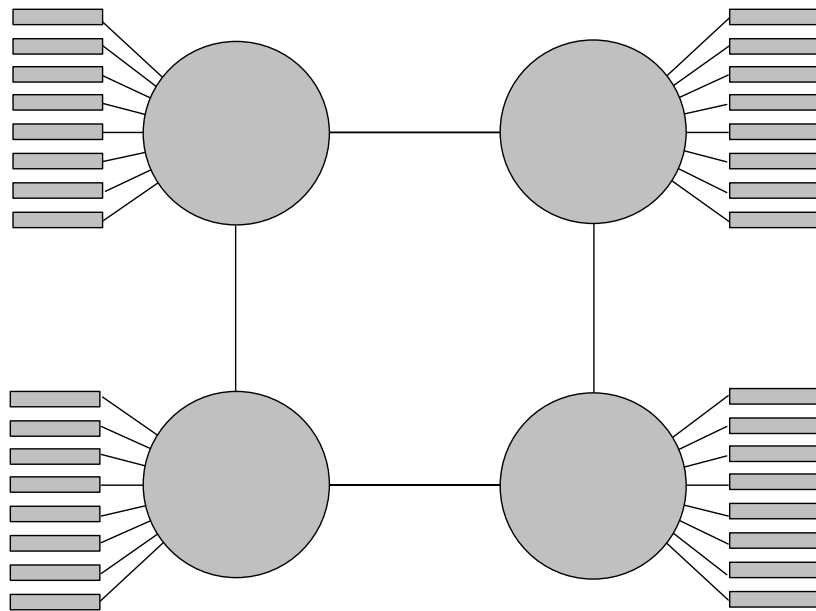
2) Sidan 19, andra strecksatsen:

– *Övertäckning* finns för vissa kategorier av invandrare. Detta upptäcktes inom SCB och RSV informerades. RSV förbättrar nu kontrollen av folkbokföringen för dessa invandrarkategorier så att *övertäckningen* kommer att minska.

Redaktionen ber om överseende med det inträffade.

Det statistiska registersystemet

Utvecklingsmöjligheter och förslag



Slutrapport från Registerprojektet

GD-beslut med anledning av Registerprojektets slutrapport

R&D Report 2001:1

Research - Methods - Development

The Swedish register system - possibilities for development with proposals

Från trycket
Producent

Maj 2001
Statistiska centralbyrån, *Statistics Sweden*, metodenheten
Box 24300, SE-104 51 STOCKHOLM

Förfrågningar

Anders Wallgren och Britt Wallgren
anders.wallgren@scb.se
britt.wallgren@scb.se
telefon 019-17 61 98

Förord

Svensk officiell statistik bygger sedan länge på en omfattande användning av administrativa data för att bygga upp olika statistiska register. Dessa register används flitigt både som bas för våra urvalsundersökningar men också för att ta fram totalräknad registerstatistik av olika slag.

Trots att denna registertradition har funnits länge har man på SCB inte alltid tänkt i termer av ett *registerstatistiskt system* utan varje register har fungerat var för sig. Registerprojektets uppgift har varit att synliggöra det registerstatistiska metodarbetet, att utveckla ett tydligare systemtänkande och att öka samverkan mellan systemets olika delar. Tillsättningen av projektet var en markering av att registersystemet skulle ses som en strategisk resurs.

Projektgruppen, som nu överlämnar sin slutrapport, har bestått av Elisabet Andersson, Evert Blom (projektledare), Linda Ryen samt Anders och Britt Wallgren. Under en kortare inledningsperiod, fram till april 1997, leddes projektet av Staffan Wahlström.

Till projektet har vi knutit en kontaktgrupp med ett 20-tal deltagare från olika register inom SCB. Gruppen har varit mycket betydelsefull för projektarbetet. Vi har via kontaktgruppen kunnat diskutera olika registerproblem med personer som har mångåriga erfarenheter av registerstatistik och spridit kunskaper om systemet och en rad olika register av allmänt intresse genom att låta deltagare i kontaktgruppen och andra inbjudna berätta om sina register och erfarenheter.

Projektet redovisar senare en *handbok i registerstatistiskt arbete*, som skall kunna användas vid kurser i registerstatistik.

Rapporten har remissbehandlats inom SCB och diskuterats i Verksdirektionen. Vissa korrigeringar har gjorts med anledning av detta. GD-beslutet med anledning av rapporten och dess behandling återfinns på nästa sida.

GD-beslut, 2001-05-07 med anledning av Registerprojektets slutrapport

Bakgrund:

Registerprojektets slutrapport behandlades i VD 2001-03-13 efter remissbehandling inom verket, se *Beslutsunderlag med anledning av Registerprojektets rapport, VD-underlag 2001-02-28*. GD konstaterade att det varit en bra beredning och att förslagen till beslut (avsnitt 3.1 – 3.8 i VD-underlaget) med vissa modifieringar stöds av VD. Modifieringarna gäller främst förslaget till allmänna riktlinjer, vilka också bör reduceras till antalet. GD konstaterade också att handboken om registerfrågor bör vara klar före 2001 års slut och att U-avdelningen har ansvaret för den fortsatta hanteringen av de övergripande registerfrågorna.

Beslut:

Mot bakgrund av den interna remissen och diskussionen i VD beslutas:

- att allmänna riktlinjer enligt bifogad text (se avsnitt 1 i denna rapport) skall gälla
- att registermodellen enligt den konceptuella modell som presenteras i slutrapportens avsnitt 3 i princip fastställs som underlag för beskrivning och fortsatt utveckling av registersystemet
- att avdelningarna och programmen ansvarar för utveckling och förbättring av registersystemet med utgångspunkt från förslagen i slutrapportens avsnitt 5. Varje avdelning skall ha minst en metodstatistiker med inriktning på registerfrågor
- att en central registerfunktion inrättas vid U-avdelningens Metodenhet. Avdelningen svarar för den närmare utformningen av denna
- att frågan om åtgärder med anledning av förslagen till ökat skydd av registeruppgifter hänskjuts till VL/Stab
- att U-avdelningen genom Metodenheten och IT-enheten, i samarbete med avdelningar och enheter, skall beakta framförda förslag till fortsatt utveckling. Avdelningen skall se till att dessa frågor behandlas i den normala processen för upprättande av kommande metod- och IT-planer
- att handboken i registerfrågor skall vara klar före 2001 års slut
- att U-avdelningen har ansvar för den fortsatta hanteringen av de övergripande registerfrågorna inklusive implementering och uppföljning av nu fattade beslut.

Innehåll

Förord.....	1
GD-beslut, 2001-05-07 med anledning av Registerprojektets slutrapport.....	3
Innehåll.....	4
1. Sammanfattning	5
2. Bakgrund, syfte och arbetsformer	8
3. Registersystemet – register och länkar.....	9
4. Samordning och standardisering av variabler	16
5. Metodarbete på systemnivå.....	18
6. Metodarbete och kvalitetssäkring på registernivå.....	22
7. Produktionsprocess och driftsmiljö.....	25
8. Dokumentation och metadata.....	28
9. Sekretess och integritetsskydd	31
10. Registerfunktion och nätverk	32

1. Sammanfattning

I detta avsnitt sammanfattas projektets förslag. Mer utförliga skrivningar och slutsatser av projektets arbete finns i de olika avsnitten av rapporten. Förslagen till beslut är ordnade efter de instanser som bör ta hand om och realisera dessa.

Riktlinjer för statistiska register på SCB

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.

Förslag till beslut

Central registerfunktion

En central registerfunktion med ansvar för utvecklings- och samordningsfrågor inrättas inom U-avdelningen. Till denna knyts ett nätverk av personal från övriga verket med metod- och produktionserfarenheter. Den centrala registerfunktionen skall ansvara för:

- Övergripande samordnings- och standardiseringsfrågor.
- Metadatasystemet och utveckling av dokumentationsstöd.
- Att ett nätverk för erfarenhetsutbyte bildas och att utveckling av nya sambearbetningar och analyskompetens sker inom nätverkets ram.

Organisationsfrågorna kring denna funktion behandlas i samband med U-avdelningens utredning angående en central registerservicefunktion och frågan om en förstärkning av den centrala standardfunktionen.

IT-enheten

Produktions- och driftsmiljön behöver förbättras för att stödja utvecklingen av registersystemet. IT-enheten får i uppdrag att:

- Ta ett helhetsgrepp och minska floran av verktyg för registerbearbetningar genom tydliga satsningar. Nya generella verktyg utvecklas.
- Utveckla tillgängligheten till SCB:s datalager och öka förmågan att hantera stora datamaterial.
- Utveckla metadatasystemet så att dokumentation och användning av metadata fungerar integrerat med produktionsledet.
- Utveckla tekniska lösningar så att metadata för standardiserade variabler blir allmänt åtkomliga och kan införas i produktionen.
- Ordna seminarier som behandlar erfarenheter från olika produktionsmoment och register.

Metodenheten

Det registerstatistiska metodarbetet ges berättigad uppmärksamhet. Metodenheten får i uppdrag att:

- Utveckla metoder för granskning och rättning av stora register.
- Utveckla metoder för bortfallhantering inom registerstatistiken.
- Utveckla metodiken för sambearbetningar.
- Utveckla evalveringsmetoder avseende objekt och variabler i registersystemet.
- Implementera ett system för kvalitetssäkring.
- De metodproblem som berör basregistren bör särskilt uppmärksammas.

Utvecklingen av metodfrågorna sker inom SCB:s gemensamma metodprogram.

Ämnesavdelningar och program

Arbetet med att realisera förslagen till ett förbättrat registersystem sker på avdelnings- och programnivå.

- Varje ämnesavdelning har ansvar för sitt basregister och övriga register inom sitt ansvarsområde och har ett särskilt ansvar för ett antal standardiserade variabler.
- Varje ämnesavdelning svarar för länkarna mellan sina statistiska register och respektive basregister.

- Den del av registersystemet som avser personstatistik används som modell och inspirationskälla för att utveckla registerstatistik inom hela systemet.
- Avdelningarnas metodgrupper skall innehålla åtminstone en metodstatistiker med ansvar för registerfrågor.

De förslag till förbättringar av register som finns i *avsnitt 5* genomförs successivt med syftet att förbättringarna skall realiseras i god tid inför FoB2005. Detta kan tas upp i styrdokument till avdelningar och program. Vissa förbättringar kräver samverkan mellan olika avdelningar och program.

Sekretess- och integritetsskydd

SCB:s sekretess- och integritetsskydd inom registersystemet bör förstärkas.

- Det bör snarast utredas hur spärrmarkerade personer kan få ökat skydd inom SCB:s registersystem.
- Det bör utredas om personnummer kan ersättas med löpnummer i något eller några känsliga SCB-register.
- De nuvarande reglerna för utlämnande av mikrodata bör preciseras: Användare och personal måste göras uppmärksamma på att samma regler gäller för personnummer inom både företags- och individregister. Reglerna bör göras mer konkreta: Vilka variabler är känsliga och vilka är mindre känsliga?

Ekonomienheten

Ekonomienheten bör förtydliga och sprida reglerna för intern och extern prissättning av registertjänster.

Lansering

Projektet har redan lagt ner ett omfattande arbete på lansering. Många seminarier och diskussionsmöten har hållits där tankarna kring registerstatistik har presenterats.

Projektet överlämnar senare en Handbok i registerstatistiska frågor, ett slags registerstatistikens CBM. Den är tänkt som en beskrivning av det registerstatistiska systemet och skall kunna användas i introduktions- och vidareutbildning av personal med registerstatistiska uppgifter och förhoppningsvis som en inspirationskälla för personer som arbetar med register.

Lanseringen kommer att fortsätta med metodseminarier och kurser.

2. Bakgrund, syfte och arbetsformer

Bakgrund

Praktiskt taget alla undersökningar som SCB genomför, bygger på att ett eller flera av SCB:s statistiska register används. Genom att utveckla registersystemet skall vi kunna göra nya registerbaserade undersökningar till låga marginalkostnader och förbättra våra nuvarande undersökningar.

Omvärlden ställer nya krav på SCB. De kraftiga samhällsförändringar som pågår ökar behovet av ny statistik. Men eftersom bortfallen och kraven på minskad uppgiftslämnarbördan ökar, samtidigt som det är svårt att få ökade anslag, kan vi inte alltid möta dessa krav med nya urvalsundersökningar. Vi måste istället lära oss att utnyttja redan befintliga registerdata effektivare. FoB2005 är ett bra exempel på dessa nya krav. Utan traditionell datainsamling måste vi kunna genomföra en helt registerbaserad folk- och bostadsräkning.

Vi har försökt skatta den ekonomiska andel som register utgör och bedömer att den utgör cirka hälften av det ekonomiska värdet av SCB:s produktion. Registerstatistisk produktion och metodarbete har trots detta fört en relativt undanskymd tillvaro. Jämfört med den urvalsbaserade surveystatistiken, som har årtionden av metodutveckling bakom sig har registerstatistikens behov av systematik och genomtänkta processer inte fått den uppmärksamhet den förtjänar. Vi bedömer att registerstatistiken har en stor utvecklingspotential, om våra förslag genomförs finns det goda möjligheter att öka SCB:s produktivitet.

Registerstatistik och utveckling av registersystemet kräver samverkan över program- och avdelningsgränser. Den gamla verkskulturen med egna små registersystem avpassade för den egna verksamheten är ett hinder för ett helhetstänkande – ett sammanhängande registersystem för hela SCB!

Syfte

Registerprojektets mål och uppgift har beskrivits på följande sätt:

”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.”

De slutsatser och förslag som presenteras i rapportens rutor är Registerprojektets bidrag till att uppfylla dessa syften.

Arbetsformer:

Det inledande registerseminariet blev mycket uppskattat och gav projektet en god start. I samband med en rad olika inventeringar har projektet varit i nära kontakt med registerprogrammen. Vi har även samarbetat med andra verksprojekt: Databas-, Louise-, Metadata- och Plattformsprojektet.

En särskild *Kontaktgrupp* med ett 20-tal deltagare har fungerat som projektets referensgrupp med bred representation från olika register. De flesta register har presenterats vid Kontaktgruppens möten, i de flesta fall genom föredragningar av seminariekaraktär. Det bör nämnas att dessa möten varit mycket uppskattade av deltagarna för den information om andra register som getts och har på så sätt fungerat som ett praktiskt organ för samordning och kompetensutveckling. Dessutom har projektet arbetat praktiskt med flera register för att belysa olika metodproblem.

3. Registersystemet – register och länkar

Ett statistiskt registersystem består av en rad olika delar, i detta avsnitt diskuteras register och länkar. Objekt i olika register länkas med kopplingsvariabler.

- Basregister består av *objektmängder av grundläggande betydelse*. I bilderna nedan betecknas dessa med cirklar.
- Övriga statistiska register innehåller *statistiskt intressanta variabler*.
- Länkar mellan objekt i olika basregister samt länkar mellan basregister och övriga statistiska register. Länkar betecknas med linjer i bilderna nedan.

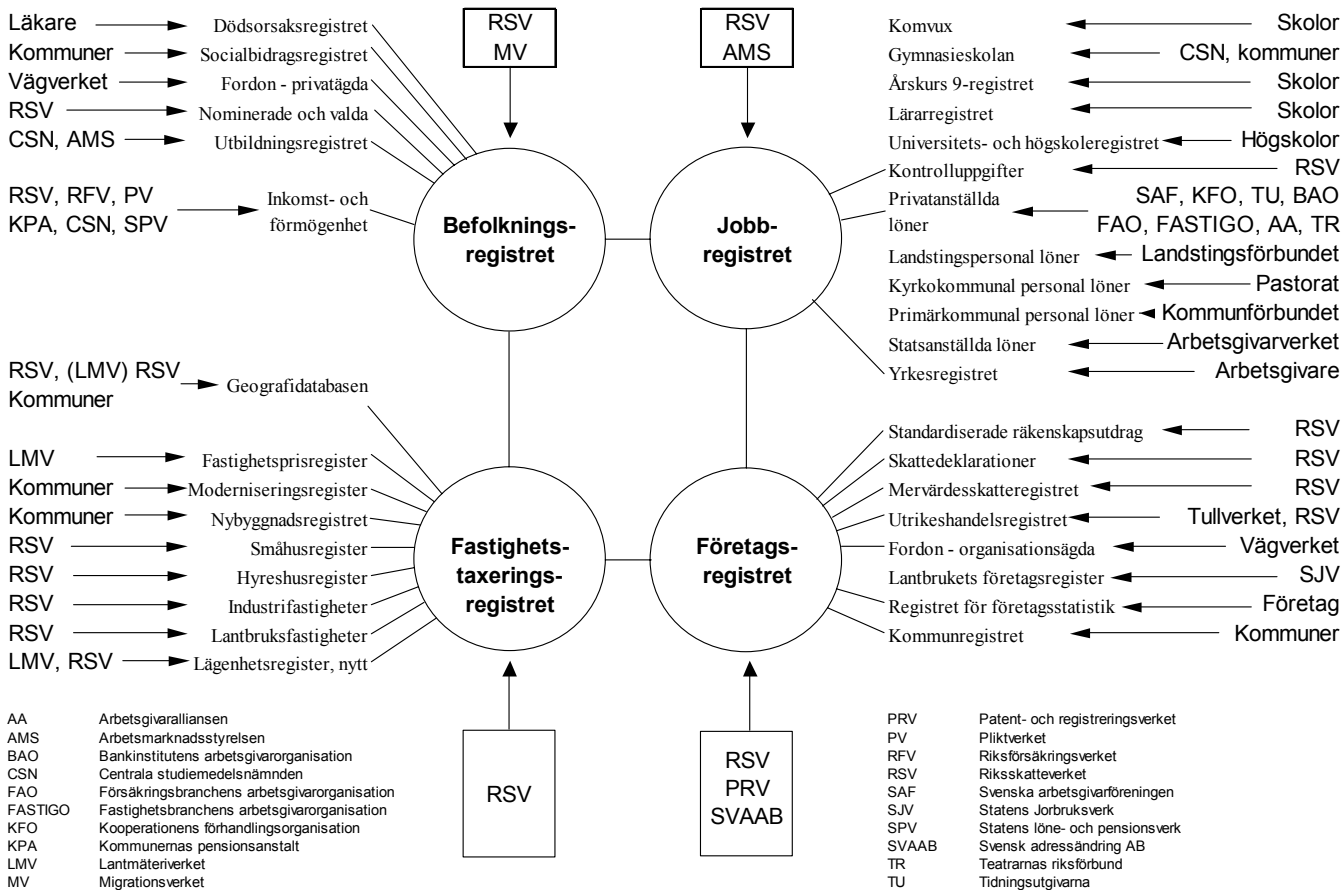
I ett komplett registersystem ingår även: *Standardiserade variabler, registerstatistiska metoder, IT-verktyg, metadata och rutiner för integritetsskydd*. Dessa delar diskuteras i avsnitt 4–9.

Inom Registerprojektet har vi under flera år arbetat med att inventera, strukturera och försöka förstå SCB:s registersystem. Begreppen ”registersystem”, ”registerstatistiskt metodarbete” etc. representerade ett nytt sätt att tänka när idén till projektet första gången presenterades för verksdirektionen i januari 1996. Vi har sedan dess utvecklat tänkandet och beskriver nu registersystemet på följande sätt där systemets uppbyggnad presenteras stegvis.

Steg 1. Ta emot data från olika administrativa källor

Från registerlämnande myndigheter får olika program inom SCB administrativa data som bearbetas så att ett antal statistiska register kan bildas. Dessa register kan sedan länkas till varandra genom att objekt i olika register kan identifieras med samma kopplingsvariabel. Vi har följande bild av systemets indatadel:

Bild 1. SCB:s registersystem – datainsamling från myndigheter och organisationer



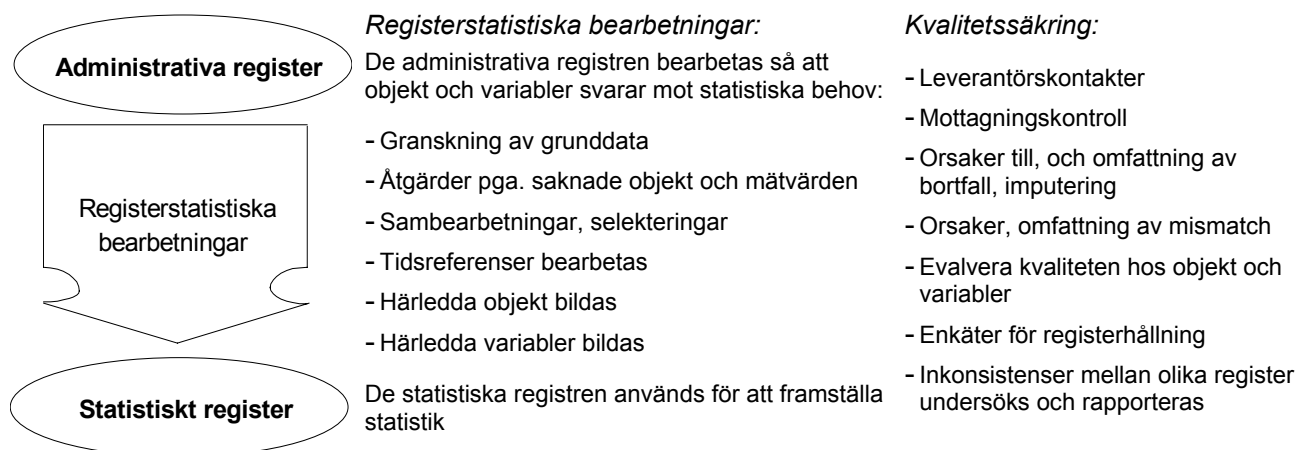
Av bild 1 framgår att många program har kontakter med registerlämnande organisationer och att många olika myndigheter och organisationer är involverade. Bild 1 motiverar följande slutsats:

1. 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 samordnas inom SCB.

Steg 2. Administrativa register omvandlas till statistiska register

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 väl genomtänkta och standardiserade *registerstatistiska metoder*. Bild 2 visar de olika moment som ingår i det metodstatistiska arbetet.

Bild 2. Från administrativa register till statistiskt register



Dessa arbetsmoment är gemensamma för många registerprogram vid SCB. Det är därför viktigt att erfarenhetsutbyte och metodutveckling kommer till stånd, så att väletablerade och väldokumenterade metoder kommer till allmän användning på det sätt som är självklart för SCB:s urvalsundersökningar. Bild 2 motiverar följande slutsats:

2. Arbetet med att ta emot stora administrativa material samt granska dessa är viktiga moment som är likartade för många program. Dessa moment effektiviseras genom erfarenhetsutbyte samt utveckling av gemensamma verktyg och granskningsmetoder. På motsvarande sätt bör metoder och verktyg för bortfallshantering, sambearbetningar och evalveringar utvecklas. Ansvar för detta bör ligga inom det gemensamma metodprogrammet och arbetet sker i nära kontakt med registerprogrammen.

Steg 3. Skapa ytterligare register och användningar

De register, som finns under a) och b) i bild 3 nedan, har bildats med administrativa data från olika registerlämnande organisationer. Dessa register är grunden för olika statistikprodukter.

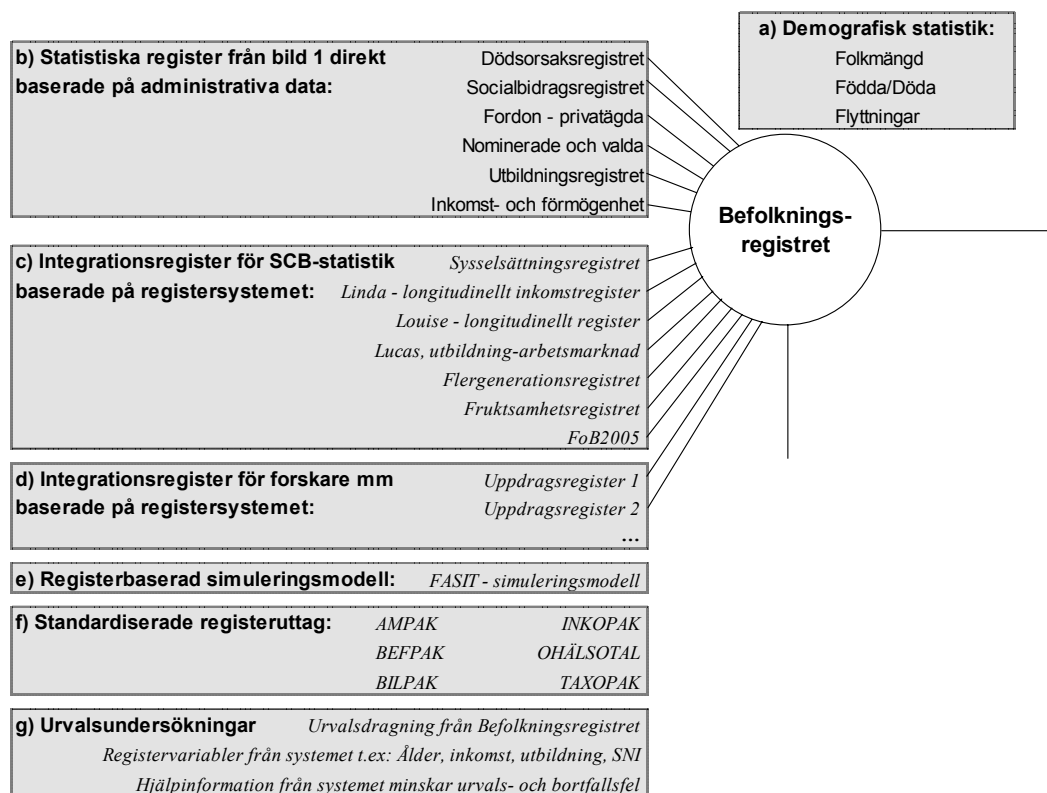
Genom att sambearbeta flera register inom systemet kan olika *integrationsregister* bildas. Dessa kan vara tillfälliga för specifika uppdrag eller permanenta för nya SCB-produkter.

De register som anges under c) och d) i bild 3 nedan är sådana integrationsregister och de har bildats utan ytterligare datainsamling.

Förmågan att göra sambearbetningar beror på hur välutvecklat SCB:s registersystem är. Datalagret skall vara bra organiserat och det skall finnas effektiva IT-verktyg för olika slag av registerbearbetningar. Registerprojektets huvudförslag är att SCB:s registersystem bör utvecklas för att underlätta arbetet med att bilda integrationsregister, både longitudinella och andra.

Den del av registersystemet som avser personstatistik är den mest utvecklade och kan användas som föredöme för systemets övriga delar.

Bild 3. Personstatistik baserad på SCB:s registersystem



Bilden ovan illustrerar de möjligheter som ett system med statistiska register kan ge. Genom att utveckla registersystemet som helhet ökas SCB förmåga att utveckla statistik baserad på integrationsregister. Ett förbättrat registersystem gör det också möjligt att förbättra SCB:s urvalsundersökningar genom att frågor kan ersättas med registervariabler och användning av hjälpinformation från register gör det möjligt att minska urvals- och bortfallsfel.

3a. Den del av registersystemet som avser personstatistik används som modell och inspirationskälla för att utveckla registerstatistik inom hela systemet. Demografisk statistik för företag, fastigheter/byggnader, förvärvsarbete och andra aktiviteter utvecklas. Simuleringsmodeller och standardiserade registeruttag utvecklas inom hela systemet. Användningen av registren skall underlättas genom att många registervariabler blir lätt åtkomliga för urvalsundersökningarnas behov.

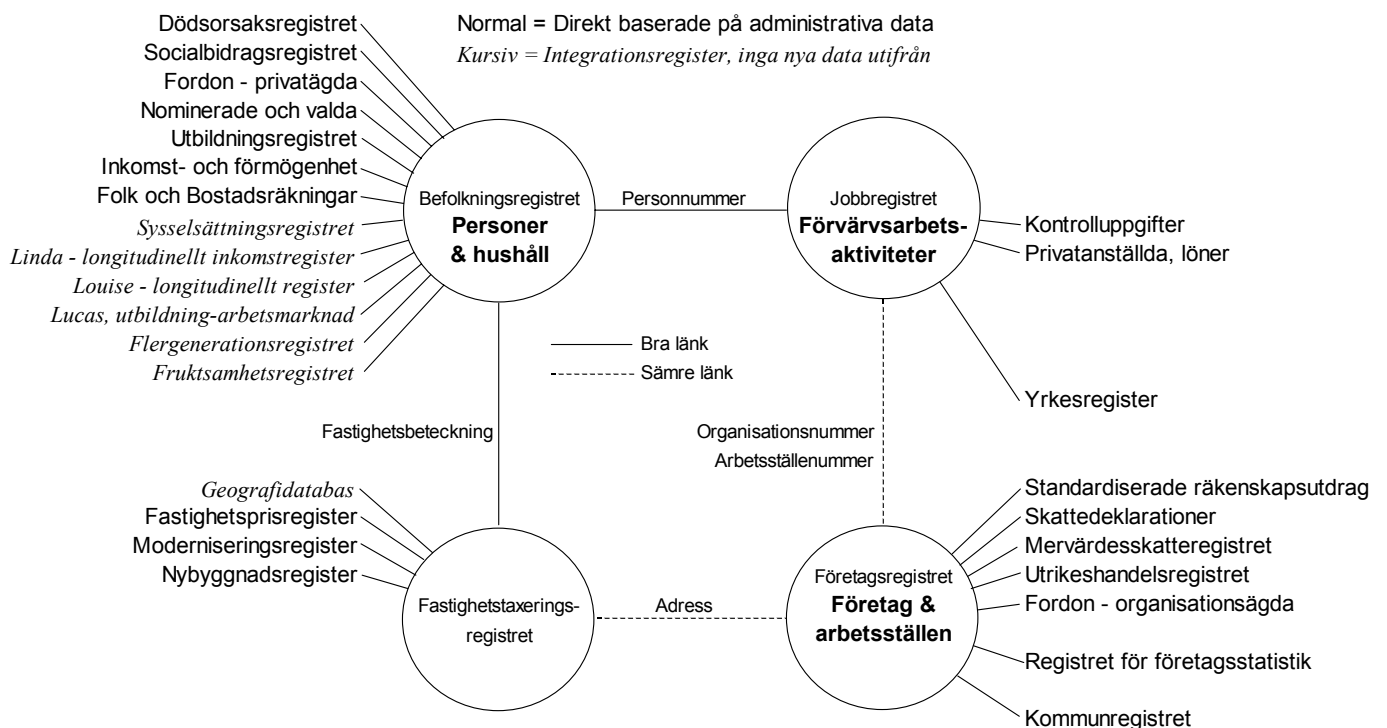
3b. Förutom att utveckla befintliga register skall vi hela tiden uppmärksamma nya administrativa källor som kan användas för att skapa nya statistiska register och nya produkter.

Steg 4. SCB:s registersystem – helheten skapas

Inom Registerprojektet utvecklade vi modellen över systemet efter att ha genomfört en inventering av de befintliga registren. Vi studerade särskilt vilka typer av objekt som ingick i olika register och vilka variabler som användes för att länka objekten i olika register med varandra. Vi fann att registren kan delas in i fyra grupper beroende på typ av objekt. Inom varje grupp finns ett basregister som definierar objekten inom hela gruppen.

Basregistrens roll är att definiera objekten i systemet där bra enhetsavgränsningar och bra täckning är viktiga egenskaper som är avgörande för kvaliteten i hela systemet. Länkarna mellan objekten i olika basregister är speciellt viktiga för möjligheterna att skapa innehållsmässigt intressanta integrationsregister. Basregistren innehåller de statistiskt intressanta objektmängderna – de statistiskt intressanta variablerna finns i de övriga registren. De variabler som är viktiga i ett basregister är variabler som identifierar objekten och kan användas för att koppla objekten till objekt i andra register. Dessutom behövs tidsuppgifter för olika händelser som påverkar objekten för att kunna skapa bestånd avseende en viss tidpunkt eller tidsperiod. Detta är en förutsättning för demografisk statistik. I bild 4 beskrivs de register som för närvarande kan länkas till sina respektive basregister.

Bild 4. SCB:s nuvarande registersystem, registren strukturerade efter objektstyp



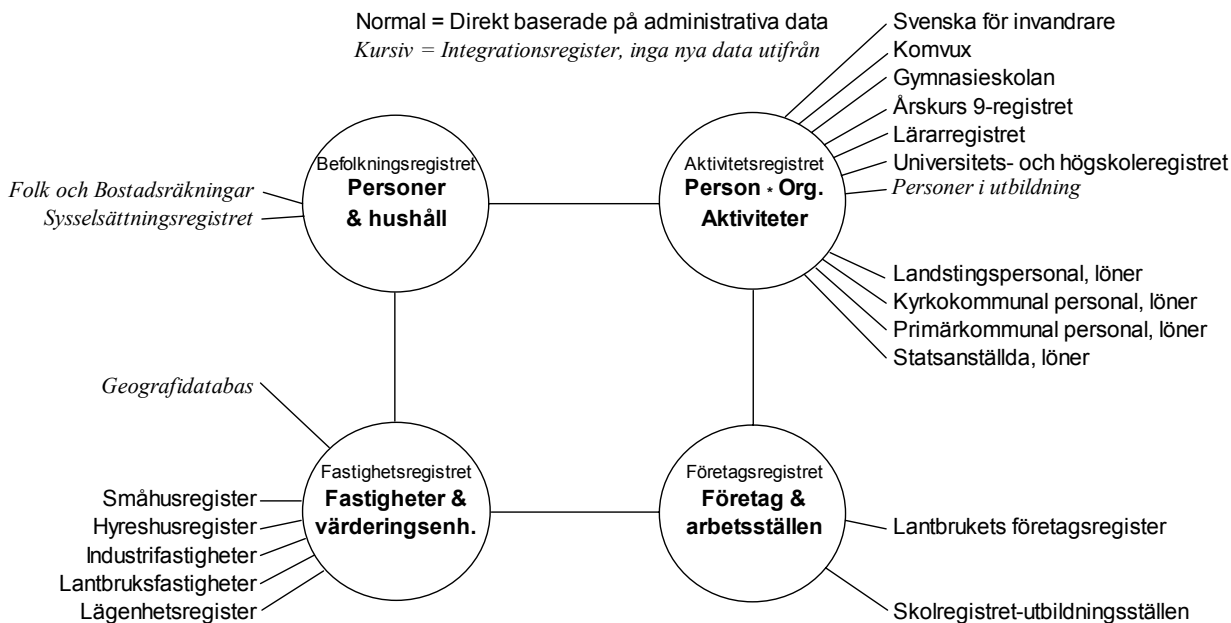
Av våra nuvarande basregister har *Fastighetstaxeringsregistret* inte den struktur som är lämplig för ett renodlat basregister för objekten fastigheter. Registret innehåller många statistiska variabler vilket gör det onödigt stort för sin roll som basregister och dessutom uppdateras det endast en gång per år vilket ger täckningsproblem.

Jobbregistret bör utvecklas till ett mer allmänt *Aktivitetsregister* som kan fungera som ett basregister för arbetsmarknadsstatistiken. Även studier och arbetsmarknadsåtgärder bör ingå i registret. På arbetsmarknaden möts utbud och efterfrågan av arbetskraft. Utbudet utgörs av personer som identifieras med personnummer, efterfrågan utgörs av arbetsställen som identifieras med arbetsställennummer.

När utbud och efterfrågan möts skapas relationsobjekt (anställning, eget företagande mm) som identifieras med kombinationen av person- och arbetsställesnummer. Aktivitetsregistret bör bland annat användas för ny demografisk statistik avseende anställningar etc. Bland användarna finns ett stort intresse för statistik där persondata och företagsdata kombineras. Genom *Jobbregistret* är denna kombination möjlig och med *Aktivitetsregistret* kan man göra ännu fler kombinationer.

De förbättringar av våra nuvarande register som Registerprojektet föreslår beskrivs i bild 5 nedan. Detaljerade förslag för respektive basregister och övriga register finns i avsnitt 5.

Bild 5. Register och länkar som bör förbättras



Följande åtgärder syftar till att förbättra samordningen inom systemet:

4a. Basregistren utvecklas till renodlade basregister, som skall vara systemets bas avseende objektsavgränsningar och objektmängder. Basregistren skall innehålla tidsmässigt olika versioner – aktuell version, bestånd vid årsskifte, förändring under året och objekt som funnits någon gång under året.

4b. Länkarna mellan basregistren förbättras. Kopplingsvariablerna i basregistren skall vara standardiserade och träffprocenten skall vara hög när basregister matchas.

4c. Alla övriga register skall kunna länkas till sitt basregister med de standardiserade kopplingsvariablerna. Objektmängderna skall vara avstämda med respektive basregister.

4d. Variablerna i systemets olika register skall samordnas. Variabelnamn och definitioner skall överensstämja för samma variabel i olika register. Metadata samutnyttjas för att undvika dubbelarbete.

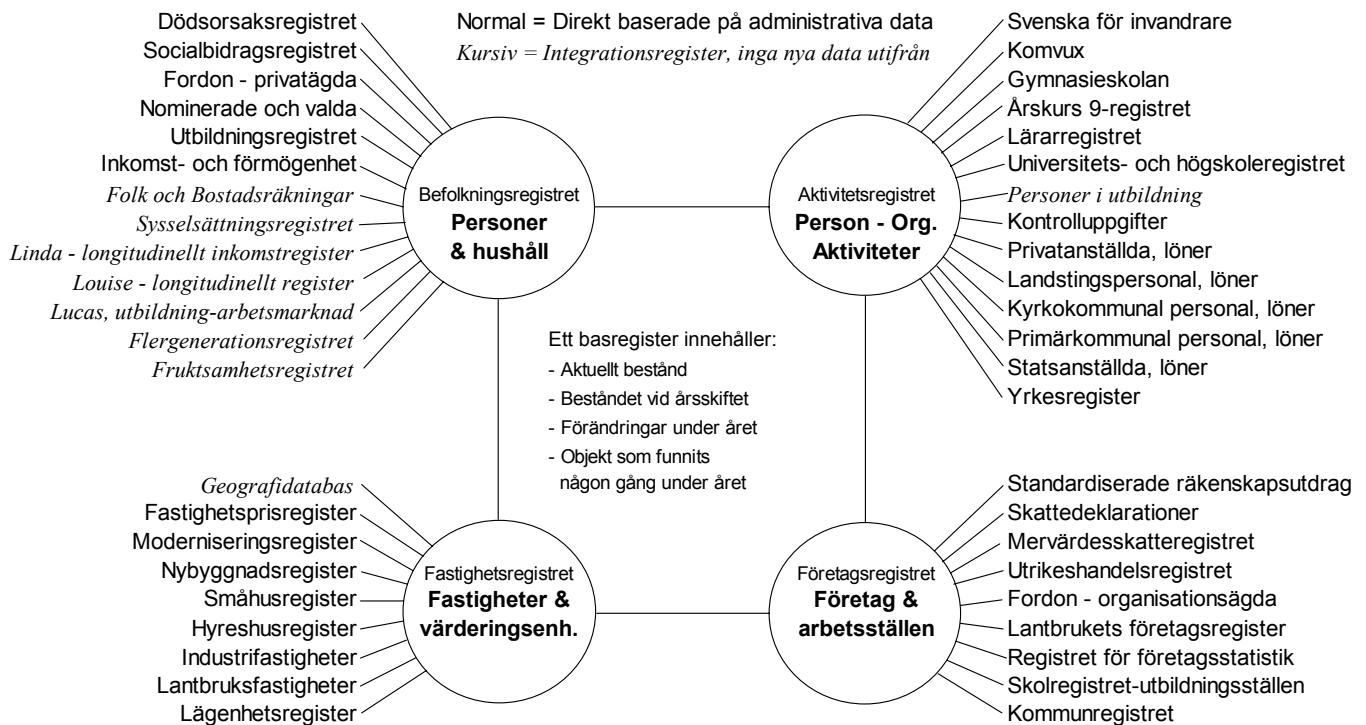
4e. Gemensamma statistiska metoder och IT-verktyg används.

4f. Sekretess- och integritetsskyddet inom systemet förbättras.

Dessa allmänt formulerade förslag behandlas mer detaljerat i avsnitten 4–9.

Efter att ha förbättrat befintliga register och länkar enligt Registerprojektets förslag kommer SCB:s registersystem att få följande utseende:

Bild 6. SCB:s framtida registersystem då befintliga register och länkar har förbättrats



Samordnings- och förbättringsarbete kommer att ge oss ett bättre fungerande registersystem med större möjligheter. Förbättringen av registersystemet skall göras samtidigt som integritets- och sekretesskyddet förbättras enligt förslagen i avsnitt 9. Befintlig statistik kommer efter förbättringarna att bli mer konsistent och får högre kvalitet, samtidigt som möjligheterna till ny registerbaserad statistik kommer att öka.

Gamla versioner av register arkiveras hos Riksarkivet. För vissa uppdrag behöver SCB få tillgång till dessa arkiverade register. För att underlätta arbetet med arkiverade register bör smidiga rutiner utarbetas i samråd med Riksarkivet.

Nya integrationsregister innebär att vissa organisatoriska frågor blir aktuella. För att undvika konflikter skall regelmässigt samarbete ske mellan berörda program. Ingen skall ha monopol på att använda vissa data. Kostnadsdebitering internt och externt skall följa gemensamma och kända regler.

5a. För att vi inom SCB skall kunna utnyttja det förbättrade systemets möjligheter bör en grupp med kompetens från olika delar av registersystemet arbeta med att utveckla innehållsmässigt nya sambearbetningar och integrationsregister. Denna grupp kommer att arbeta med kvalificerat utvecklingsarbete med syfte att skapa ny statistik.

5b. Bekvämlig tillgång till långtidsarkiverade register skapas.

5c. Samarbete över programgränser sker då nya integrationsregister bildas.

5d. Reglerna prissättning av registertjänster bör förtydligas och spridas.

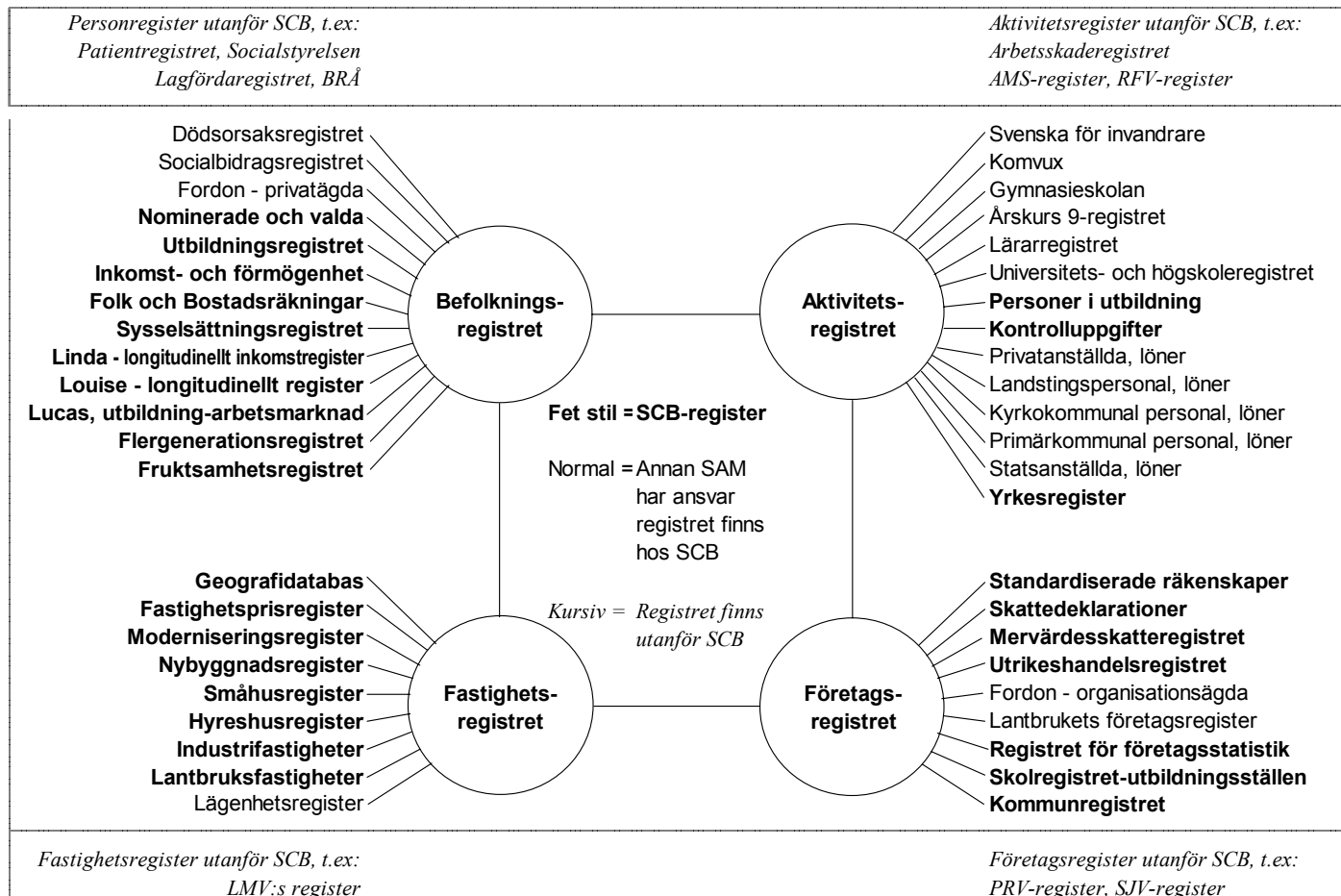
Steg 5. SCB:s och övriga SAM:s registersystem samverkar

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.

Steg 4 ovan, där SCB:s registersystem som helhet skapas, förutsätter ökat samarbete mellan avdelningar och program inom SCB. På motsvarande sätt bör SCB och övriga SAM samarbeta i registerfrågor för att skapa ett registersystem för hela det statistiska systemet.

Det är viktigt för SCB att ha tillgång till bra SAM-register. På motsvarande sätt är det viktigt för andra SAM att ha tillgång till SCB:s registersystem. Ett samarbete kring registerfrågor är därför av gemensamt intresse. Motsatsen till samarbete är att revirtänkande hindrar utvecklingen av registerbaserad statistik.

Bild 7. SCB:s och övriga statistikansvariga 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 kunna göra ny egen statistik.

De olika utbildnings- och löneregistren kommer att få ökad betydelse för SCB:s statistik enligt Registerprojektets förslag. De borde enligt våra förslag användas för Aktivitets- och Sysselsättningsregistrets behov och därigenom vara viktiga för kvaliteten hos kommande registerbaserade FoB:ar.

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.

6. SCB utarbetar en strategi för samarbetet kring registerfrågor med alla berörda statistikansvariga myndigheter.

4. Samordning och standardisering av variabler

SCB:s datalager är i dag fysiskt och organisatoriskt starkt decentraliserat med mer än 1000 databaser lagrade på ett 60-tal servrar. En förutsättning för en effektiv statistikproduktion är att de register som ingår i datalagret är innehållsmässigt och tekniskt samordnade. Onödiga särlösningar, som annars kan ställa till oreda när data tolkas, matchas eller jämförs, kan då undvikas. När det gäller den innehållsmässiga delen är det viktigt med samordning avseende både variabler och klassifikationer. Ett viktigt samordningsinstrument är Klassifikationsdatabasen (KDB) som innehåller statistiska klassifikationer och som används inom SCB och det svenska statistiska systemet. Här ingår klassifikationer och nycklar som har fastställts som standard i Sverige eller internationellt. Ett exempel på en sådan klassifikation är näringsgren som anges enligt standard för svensk näringsgrensindelning (SNI).

Fram tills nu har inga stora insatser gjorts för att skapa enhetlighet mellan variabler i olika register på SCB. Plattformsbytet medförde ett omfattande omlägningsarbete och en genomgång av produktionen och det råder knappast någon tvekan om att SCB missat ett bra tillfälle att i samband med detta genomföra en samordning av variabler. Bristen på dokumentation har tidigare gjort att det har varit svårt att göra jämförelser av definitioner och beskrivningar av variabler. Det finns exempel på att variabler med samma innehåll kan ha olika namn i olika register. Likaså finns exempel på att variabler med olika definitioner har samma namn.

Situationen idag

Under 1999 påbörjade Registerprojektet en första inventering av variabler med utgångspunkt från Befolkningsregistret, RTB. Jämförelser gjordes med variabler i en rad andra register och ett antal olikheter upptäcktes. Ett förslag till beslut om samordning av variabler togs fram av projektet och utifrån detta fattades under våren 2000 ett beslut om standardisering av vissa viktiga variabler av SCB:s generaldirektör (GD-beslut 2000-04-03).

De variabler som är aktuella för standardisering är sådana som återkommande används i statistikproduktionen av andra register. Det register som tar in en variabel från en annan myndighet eller själva skapar en variabel ska ha ansvaret för denna på SCB. I enstaka fall, såsom när ett register direkt slussar en variabel vidare till ett annat, kan ansvaret läggas hos det senare registret.

Registerprojektet har gjort en genomgång av variabler i SCB:s basregister som är lämpliga för standardisering. Detta har skett i samarbete med representanter för de berörda registren. Samordningsarbetet inleddes med en genomgång av variablerna i RTB vilket resulterade i att sex variabler föreslogs för standardisering i en första omgång. Variablerna är personnummer, kön, ålder, civilstånd, födelseland samt medborgarskap. Vi bedömer att det totalt för basregistren rör sig om ett tiotal variabler som är aktuella till att börja med.

Projektet har också sett över förteckningen med de tretton centrala kopplingsvariablerna eftersom dessa också är en form av standardiserade variabler. För att innehållet i förteckningen ska ligga i linje med ovannämnda beslut bör de variabler som finns med vara sådana som används av flera register. Vid kontakter med berörda ämnesprogram har det visat sig att det i dagsläget även ingår några variabler i förteckningen som endast används inom det egna programmet eller som har blivit inaktuella. När det gäller dessa variabler är det inte berörd avdelningschef som fattar beslutet eftersom den ursprungliga listan kommit till i form av ett GD-beslut. Även om inte

kopplingsvariablerna ändras bör information om dessa gå ut till personalen eftersom projektets erfarenhet är att alla inte känner till dem.

Vilken teknisk lösning som är lämplig för att tillgängliggöra de standardiserade variablerna bör utredas närmare. Tidigare bestod KDB också av en variabeldel som bl.a. innehöll kopplingsvariabler. Nu pågår nyutveckling av KDB och ingen sådan del är planerad. En möjlighet är att ändå på sikt lägga in variabler i KDB p.g.a. det nära sambandet mellan variabler och klassifikationer. Alternativt skapas en separat databas som har koppling till klassifikationerna i KDB. Ytterligare en tänkbar lösning är en sökfunktion i Metadok. Som ett första steg kan de aktuella variablerna med tillhörande information sammanställas i en enkel lista, vilket bör vara tillräckligt tills ett större antal variabler har utsetts.

Vid skapandet av helt nya register kan standarder enkelt följas, men för att uppnå full samordning krävs att ändringar av t.ex. variabelnamn och koder kan göras i redan befintliga register och i de procedurer som används för att skapa nya årgångar. Det är därför svårt att ta fram en realistisk tidsplan för hur lång tid det kan ta att genomföra de ändringar det gäller.

7. Standardiseringen av variabler kräver utredningar av teknisk karaktär. Det behövs en bra teknisk lösning så att metadata för de standardiserade variablerna blir lättillgängliga. Dessutom måste man göra en bedömning av hur en förändring av variabelnamn, koder etc. kan införas i produktionen.

Förslag till insatser

För att variabelsamordning enligt ovannämnda GD-beslut ska kunna genomföras är en central funktion nödvändig. Som det ser ut idag finns ingen naturlig plats i organisationen för denna typ av arbete och inget enskilt program har resurser nog för att klara av detta på egen hand.

8. En central registerfunktion beskrivs i avsnitt 10 och en sådan funktions ansvarsområde bör innefatta att driva standardiseringsarbetet framåt och utforma tydliga regler och instruktioner för hur de standardiserade variablerna ska skötas och följas.

Det ansvariga registret har till uppgift att ta fram en förteckning över de variabler som ska standardiseras samt utse en ansvarig person för dessa. Variabelförteckningen ska sedan godkännas av den centrala funktionen. Definitivt beslut om de aktuella variablerna tas därefter av respektive avdelningschef. Den ansvarige personen på registret har både ett dokumentations- och informationsansvar för sina standardiserade variabler. Dokumentationsansvaret innebär att de aktuella variablerna ska vara dokumenterade i Metadok enligt anvisningar och kopieras till den centrala databasen. I informationsansvaret ingår att informera och samråda med register som regelbundet hämtar en variabel om en ändring planeras. Om fel upptäcks i data/metadata ska detta också meddelas till de berörda registren.

9. För en ökad innehållsmässig samordning ska variabler i basregistren standardiseras och de centrala kopplingsvariablerna bör aktualiseras. Som ett andra steg bör en genomgång av variabler i övriga register genomföras. Vi föreslår att den centrala registerfunktion får det övergripande ansvaret för att detta genomförs.

5. Metodarbete på systemnivå

Registerstatistiskt metodarbete kan bedrivas på två nivåer:

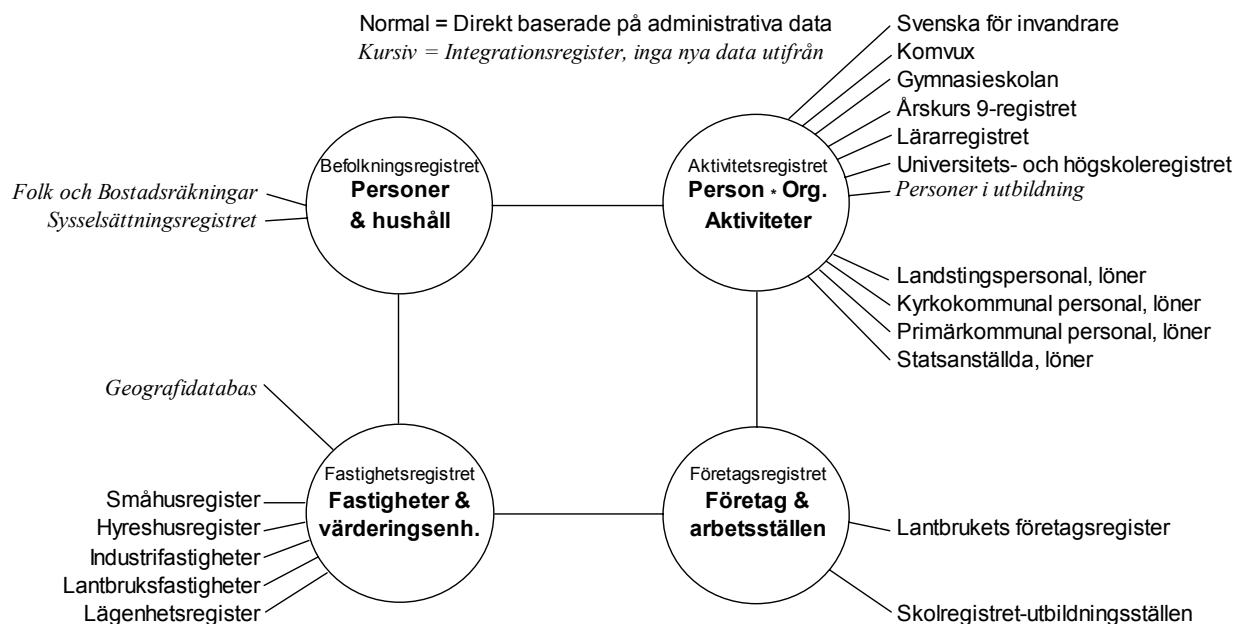
- *Metodarbete på registernivå.* Arbetet innebär ansvar för de metoder som används när ett statistiskt register skapas. Detta kommer att diskuteras i avsnitt 6.
- *Metodarbete på systemnivå.* Arbetet innebär att samordna ett stort antal register till ett fungerande system av statistiska register. Systemet som helhet skall utvecklas för att möjliggöra ny och förbättrad statistik.

Även om metodarbetet på registernivå sker på de enskilda registerprogrammen så måste samordning ske då alla register skall kunna samutnyttjas. Metodarbetet på systemnivå förutsätter kunskaper om hela systemet och kan alltså inte ske på programnivå.

10. Det övergripande ansvaret för metodarbetet på systemnivå bör ligga centralt, förslagsvis vid den centrala registerfunktionen som diskuteras i avsnitt 10. Det måste finnas fortlöpande kontakter mellan denna centrala funktion och de olika registerprogrammen. Avdelningarnas metodgrupper måste innehålla åtminstone en metodstatistiker med ansvar för registerfrågor.

Registerprojektet har tillsammans med projektets kontaktgrupp studerat systemets olika delar. Vi har även inventerat de olika register som ingår i systemet. På grundval av våra kunskaper om systemet som helhet kan vi föreslå en rad förbättringar av nuvarande register. Redan i avsnitt 3 nämndes detta och förslagen sammanfattades i följande bild:

Bild 5 (från avsnitt 3). Befintliga register och länkar som bör förbättras



Systemet av basregister

Systemet av basregister är grundvalen för hela registersystemet och består av fyra register som innehåller viktiga objektstyper. Om någon av dessa basregister saknas eller har låg kvalitet skulle hela registersystemet vara betydligt mindre användbart för statistiska ändamål.

För att SCB:s register skall fungera som ett system av register fordras att objekten i olika register kan kopplas samman med olika *kopplingsvariabler*. Ett

effektivt system kräver att kopplingsvariablerna är av god kvalitet och att samma kopplingsvariabler finns i flera register. Kopplingarna mellan basregistren är särskilt betydelsefulla.

Befolkningsregistret

Detta basregister har kommit längst när det gäller att utveckla registerstatistisk metodik för arbetet med ett basregister. Man kan å jourhållning, har en lång tradition av att producera avancerad registerstatistik, gör regelmässigt demografisk analys avseende objektpopulationen och man är van att skapa integrationsregister för olika ändamål. Man har ett avancerat produktionssystem där objektens historik lagras. Befolkningsregistret kan användas som inspirationskälla för de övriga basregistren. Vissa nackdelar finns emellertid:

- Hushållsavgrensningarna har idag brister som kan avhjälpas med ett bra lägenhetsregister.
- Övertäckning finns för vissa kategorier av invandrare. Detta upptäcktes inom SCB och RSV informerades. RSV förbättrar nu kontrollen av folkbokföringen för dessa invandrarkategorier så att övertäckningen kommer att minska.

11a. Det kan förekomma att personnummer återanvänds (samma nummer används för mer än en person) och att personnummer ändras för samma person. Detta ger vissa matchningsproblem som skulle kunna åtgärdas genom att införa SCB-löpnummer på samma sätt som Danmarks statistik har gjort. Vi föreslår i avsnitt 9 att löpnummer införs för att förbättra integritetsskyddet, men även matchningsproblemen motiverar att löpnummer införs.

11b. Ett årgångsregister med alla personer som varit folkbokförda någon gång under ett år bör skapas. När olika register matchas måste alla avse samma tidpunkt/period. Både Aktivitets- och Kontrolluppgiftsregistret avser kalenderår och behöver kunna matchas mot en sådan årgångsversion av Befolkningsregistret.

11c. Kopplingsvariabler bör vara informationslösa så att t.ex. förändrad regional indelning inte påverkar kopplingsvariabeln.

11d. Som kopplingsvariabler mot Fastighetstaxeringsregistret används idag län, kommun, församling och fastighetsbeteckning. I stället för att matcha med dessa fyra variabler bör ett riksunikt fastighetsnummer införas.

Fastighetsregistret

Det nuvarande Fastighetstaxeringsregistret har inte samma struktur som de övriga tre basregistren. Registret är för innehållsrikt med 100-tals statistiska variabler för att vara ett renodlat basregister och uppdateras endast en gång per år. Kopplingen till Företagsregistret är dålig.

12a. Ett renodlat basregister avseende fastigheter och relaterade objekt (ägare, taxeringsenheter, värderingsenheter byggnader/mark) bör skapas. Samarbetet med RSV och LMV bör utvecklas så att SCB får regelbundna aviseringar till detta nya basregister.

12b. Kopplingen mellan Företags- och Fastighetsregistret bör förbättras. Geografidatabasen måste utvecklas så att arbetsställets belägenhetsadresser kan kopplas till rätt fastighet. Inom LMV är man medveten om betydelsen av verksamhetsadresser, då dessa behövs för studier av lokal och regional utveckling.

Företagsregistret

SCB får administrativa data avseende *juridiska enheter* från PRV och RSV. Ett omfattande arbete sker inom SCB för att samla in uppgifter från företag med verksamhet vid mer än ett arbetsställe för att kunna skapa *arbetsställeenheter*. 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 utvecklingsarbetet med Företagsregistret:

- 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. De som t.ex. samlar in data från industriföretag skall vidarebefordra den information de får om industripopulationen till Företagsregistret.

Inom Registerprojektet anser vi att dessa principer skall gälla hela SCB, inte bara inom avdelningen för ekonomisk statistik. Konkret innebär detta att de som arbetar med Lantbrukets företagsregister skall lämna underlag för lantbruksdelen av Företagsregistret och att de som arbetar med register avseende skolor och utbildningsställen lämnar underlag så att Företagsregistrets utbildningsdel blir så bra som möjligt. Motsvarande princip skall gälla för de övriga tre basregistren – inga parallella objektmängder får förekomma!

13a. Årgångsversioner och beståndsregister avseende årsskifte bör regelmässigt skapas. Både Aktivitets- och Kontrolluppgiftsregistret avser kalenderår och behöver kunna matchas mot en sådan årgångsversion av Företagsregistret. Bestånd vid olika årsskiften och förändringar under respektive år behövs för demografisk företagsstatistik.

13b. Samma objektsavgränsningar och kopplingsvariabler avseende företag, företagsenheter och arbetsställen bör finnas i Företagsregistret som i Aktivitetsregistret. Företagsregistret speglar aktuellt bestånd. Det kontrolluppgiftsbaserade Aktivitetsregistret ger efter lång tidsfördröjning delvis motstridig information om samma population.

13c. För vissa delar av företagspopulationen har dagens Företagsregister onödiga kvalitetsbrister. Aktuell information avseende jordbruksföretag och utbildningsställen finns inom SCB men används inte av Företagsregistret. Om organisations- och arbetsställesnummer införs inom jordbruks- respektive utbildningsstatistiken kan kvaliteten på Företagsregistret höjas.

13d. Inga parallella objektmängder skall förekomma. Detta innebär att alla register skall vara avstämda mot respektive basregister.

Aktivitetsregistret

Dagens *Jobbregister* ligger till grund för den registerbaserade sysselsättningsstatistiken, som är en av grundförutsättningarna för att kunna göra en registerbaserad FoB. Registret beskriver idag förvärvsarbetsaktiviteter som anställningar och eget företagande. Registret kan förbättras innehållsmässigt om alternativen till förvärvsarbete förs in i registret. Arbetet med att skapa ett sådant basregister för arbetsmarknadsstatistiken har inletts. Detta *Aktivitetsregister* kommer att innehålla aktiviteter som förvärvsarbete, studier, sökande efter arbete och även andra alternativ till förvärvsarbete som sjukskrivning, förtidspension etc. Det är viktigt att detta arbete fullföljs. Kopplingsvariabler i Aktivitetsregistret är personnummer, arbetsställesnummer och organisationsnummer.

Traditionell redovisning ger en för statisk bild av förändringarna i samhället. Aktivitetsregistret kan ge oss möjligheter till att få en bättre bild av dynamiken. Förstudier på Registerprojektets Gotlandsdatabas (som är en miniatyr av SCB:s registersystem) visar att ca en tredjedel av förvärvsaktiviteterna försvinner på ett år och ersätts med nya.

För att även studieaktiviteter skall kunna kopplas till Aktivitetsregistret fordras att utbildningsstället identifieras med arbetsställesnummer och utbildningsanordnaren med organisationsnummer. Detta kommer också att ge förbättrad pendlingsstatistik som beskriver både de förvärvsarbetandes och de studerandes pendling.

Sysselsättningsregistret, som idag bygger på Jobbregistret, störs av vissa klassifikationsfel. Enligt en evalveringsundersökning var 5 % av de sysselsatta felaktigt klassificerade som ej sysselsatta och 14 % av de ej sysselsatta felaktigt klassificerade som sysselsatta. Det är novembersysselsättning som skall beskrivas och då tidsangivelserna på kontrolluppgifterna har låg kvalitet medför detta att vi får dessa felklassificeringar.

Löneregistren däremot innehåller uppgifter om anställningar under november (för vissa september till oktober), dvs. tidsangivelserna har hög kvalitet. Om löneregistren kopplas till Aktivitetsregistret kan kvaliteten i Sysselsättningsregistret förbättras kraftigt. Detta kommer att vara viktigt för FoB2005.

14a. Jobbregistret byggs ut till ett Aktivitetsregister som även omfattar studier, sökande efter arbete, långtidssjukskrivning, förtids- och sjukpension mm.

14b. Skol- och högskolestatistikens register skall ha länkar till Aktivitetsregistret med kopplingsvariablerna personnummer och arbetsställesnummer.

14c. Löneregistren skall på samma sätt ha länkar till Aktivitetsregistret.

6. Metodarbete och kvalitetssäkring på registernivå

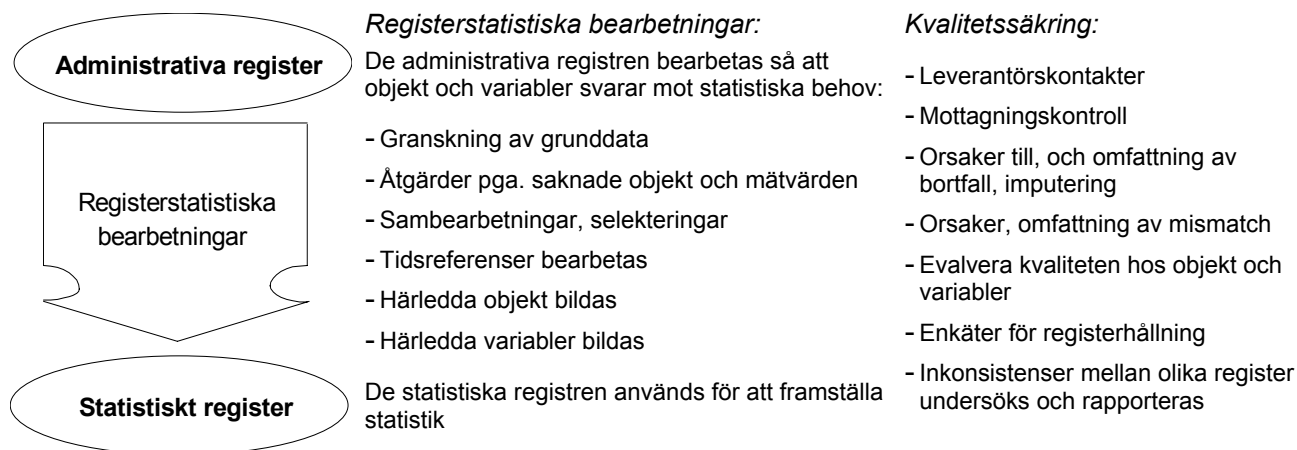
Urvalsundersökningarna inom SCB använder väldokumenterade metoder som diskuteras i rapporter och vid seminarier. Många kvalificerade statistiker arbetar med dessa urvalsundersökningar. För registerstatistiken är situationen annorlunda. Ad hoc metoder används, dessa diskuteras inte, metoderna har inte dokumenterats och endast ett fåtal statistiker har intresserat sig för området.

Man kan säga att halva SCB:s verksamhet är (har varit) osynlig. När metodfrågor diskuteras dominerar urvalsundersökningarna tänkandet. Detta är självklart när det gäller urvals- och estimationsfrågor, men samma dominans gäller även granskning, bortfall samt dokumentationen i SCBDOK. I SCB:s publikation *Att köpa en statistisk undersökning* nämns endast urvalsundersökningar trots att många uppdrag är registerbaserade undersökningar. Registerprojektets verksamhet har i någon mån ändrat detta synsätt, men mycket återstår att göra.

15. Det registerstatistiska arbetet måste få berättigad uppmärksamhet. Registerstatistiska metoder skall diskuteras och dokumenteras. Registerbearbetningar måste bygga på etablerade metoder med goda egenskaper.

I avsnitt 3 angavs de olika moment som ingår i det registerstatistiska arbetet.

Bild 2 (från avsnitt 3). Registerstatistiska arbetsmoment



Administrativa data skall bearbetas för att svara mot statistiska behov. Det är inte självklart att medvetandet om detta är så stort. Vid Registerprojektet utomnordiska kontakter har ”pilen” med Registerstatistiska bearbetningar väckt uppseende. Det finns skepsis mot att använda administrativa data för statistiska ändamål. Vi tror att denna skepsis beror på flera missuppfattningar:

- Att man använder administrativa data som de är utan bearbetningar och kvalitetssäkring. Bild 2 ovan illustrerar att detta är ett felaktigt synsätt.
- Endast en administrativ källa används för ett visst ändamål. Ingår källan i ett system av register kommer flera administrativa källor att jämföras och varje källa bidrar till att höja kvaliteten hos systemet.

Felaktig och okritisk användning av administrativa data förekommer men kan förhindras med metadata som ger information om jämförbarhetsproblem. Metadata har en mycket betydelsefull roll vid arbetet med registerstatistik. När man bildar integrationsregister där data från olika register och årgångar samkörs är risken stor för mycket grova fel om man inte känner till jämförbarhetsproblem och definitionsändringar.

Bearbetningsmetoder

På samma sätt som vi har inventerat vilka register som finns inom SCB, så behöver vi inventera vilka metoder som faktiskt används för att skapa statistiska register. En naturlig arbetsuppgift för en central registerfunktion blir att fortsätta denna inventering och fungera som metodstöd för registerprogrammen.

Kontroll av grunddata

SCB har genomfört ett stort granskningsprojekt. Rapporten från granskningsgruppen (mars 1997) redovisar metoder för granskning av i första hand urvalsundersökningar. Makrogranskning, som är den metod man rekommenderar, är troligen mindre lämplig för registerstatistik.

Granskningen skall redovisas, dels vilken metodik som används, dels omfattningen av olika feltyper och hur de har åtgärdats skall beskrivas. Den registerlämnande myndigheten bör få feedback när fel upptäcks.

Leverantörskontakter

Samarbete mellan SCB och de myndigheter som levererar administrativa register skall vara ett regelmässigt inslag i arbetet med registerstatistik. SCB måste ha full information om de regler och begrepp som används i de administrativa systemen, dvs. tillgång till administrativa metadata. I dessa kontakter bör även SCB ges tillfälle att diskutera brister och ge synpunkter på planerade förändringar av de administrativa systemen. Syftet är att i god tid få information om systemförändringar som kan påverka statistikens innehåll och kvalitet. Vissa förändringar kan kräva extra insatser från SCB:s sida för att säkerställa statistikens kvalitet och detta måste SCB skapa förståelse för.

Åtgärder på grund av saknade objekt och mätvärden

Även här dominerar tänkandet inom SCB av bortfallsproblemen för urvalsundersökningarna. I den årliga *Bortfallsbarometern* redovisas endast bortfallsutvecklingen inom olika urvalsundersökningar. De bortfallshandböcker som tagits fram är också helt inriktade på urvalsundersökningarnas problem.

För variablerna i systemet bör omfattningen av partiellt bortfall redovisas. Metoderna för bortfallskompensation bör redovisas.

Sambearbetningar och selekteringar

Sambearbetningar av register kräver kunskaper och erfarenheter för att undvika missvisande resultat pga. bristande jämförbarhet eller andra slag av kvalitetsproblem. Mismatch mellan basregistren skall finnas dokumenterat och mismatch mellan varje statistiskt register och dess basregister skall också redovisas. Orsakerna bakom mismatch bör utredas och resultaten dokumenteras som t.ex. objektsbortfall eller kvalitetsbrister hos kopplingsvariablerna hos något eller några av de inblandade registren.

Inkonsistenser mellan olika register rapporteras

Alla de som arbetar med att sambearbeta register gör värdefulla iakttagelser som skall dokumenteras och rapporteras till dem som ansvarar för de register som berörs. Man skall enkelt kunna rapportera in upptäckta brister och de som ansvarar för registren skall undersöka orsakerna till dessa konstigheter. Resultatet av detta skall sedan finnas i metadata-systemet.

Evalvera kvaliteten hos objekt och variabler

De urvalsundersökningar som regelmässigt görs bör också användas till att evalvera kvaliteten hos objekt och variabler i registersystemet. Det är särskilt viktigt att härledda objekt och variabler kontrolleras regelbundet. Speciella enkäter för registervård måste användas i vissa situationer.

Kvalitetssäkring¹

Den traditionella felmodellen kan uppfattas som att den följer en urvalsundersöknings olika steg. De fel som uppkommer under processen är: Relevansfel, ramfel, bortfallsfel, mätfel, bearbetningsfel och urvalsfel.

På motsvarande sätt vill vi beskriva kvaliteten i registerstatistik genom att följa de olika produktionsstegen i bild 2 ovan. Arbetet med kvalitetssäkring syftar till att undersöka och åtgärda kvalitetsbrister som kan uppstå i olika delmoment i arbetet med registerstatistik. I alla delmoment skall kvalitetsbrister dokumenteras och informationen skall återföras till de berörda.

Det finns minst samma behov av kvalitetsarbete, metodstudier och kvalitetssäkring för registerbaserade undersökningar som för urvalsundersökningar. Vid urvalsundersökningar kan producenten genom egna åtgärder och val av metoder i sin design ”sätta nivån” på kvaliteten och mäta kvaliteten på den publicerade statistiken kvantitativt genom att beräkna felmarginaler för de slumpmässiga urvalsfelen.

Vid registerbaserade undersökningar är situationen helt annorlunda, den registeransvarige kan inte styra datainsamlingen och vet inte hur registret kommer att användas och vilka skattningar som kommer att beräknas av olika användare inom och utom SCB. Dessutom är det svårt att beräkna kvantitativa mått på kvaliteten, någon parallell till den väletablerade metodiken för att beräkna felmarginaler vid urvalsundersökningar finns inte.

Ofta finns emellertid en kvalitativ kunskap då den registeransvarige genom sitt ämneskunnande, genom sina iakttagelser och erfarenheter väl känner till förekomsten av ett antal feltyper men kan inte mäta hur vanliga eller betydelsefulla de är. Även sådan vag kunskap bör dokumenteras, för att dels varna användare för okritisk användning och dels kunna användas för att peka ut behov av förbättringar. Kunskapen om tillförlitligheten kan klassificeras efter sin skärpa på någon av följande nivåer:

1. Ingen kunskap alls.
2. Vag kvalitativ information baserad på bedömningar.
3. Dokumentation av produktionsprocesserna och deras egenskaper.
4. Systematisk kvalitativ information baserad på undersökningar.
5. Kvantitativa indikatorer på tillförlitligheten, t.ex. omfattning av mismatch.
6. Kvantitativa mått på skattningens tillförlitlighet.

2.(Från avsnitt 3). Arbetet med att ta emot stora administrativa material samt granska dessa är viktiga moment som är likartade för många program. Arbetet bör effektiviseras genom erfarenhetsutbyte samt utveckling av gemensamma verktyg och granskningsmetoder. På motsvarande sätt bör metoder och verktyg för bortfallshantering, sambearbetningar och evalveringar utvecklas. Ansvar för detta bör ligga inom det gemensamma metodprogrammet och arbetet bör ske i nära kontakt med registerprogrammen.

16a. Metodik och verktyg för att bilda integrationsregister utvecklas gemensamt. De kvalitetsbrister som upptäcks vid sambearbetningar skall rapporteras till de register som berörs.

16b. Basregistren har gemensamma statistiska metodproblem och IT-problem. De bör därför samarbeta kring frågor som evalvering av objektskvalitet, täckningsproblem, a jourhållning, demografiska metoder, systemstruktur och IT-verktyg.

¹ Detta avsnitt bygger på Håkan Lindström: *Kvalitetssäkring i register för statistikproduktion med administrativt underlag*. Rapport från Registerprojektet november 1999.

7. Produktionsprocess och driftsmiljö

För att registersystemet skall kunna fungera som strategisk resurs vid SCB är det viktigt att produktionsprocessens utformning och driftsmiljön stödjer en sådan utveckling. Den nya IT-plattformen ger inte automatiskt detta stöd. Plattformprojektet konstaterar i sin slutrapport att det skett en stor kompetensutveckling, men att den tekniska plattformen måste vidareutvecklas och kompletteras för en mer effektiv produktionsprocess. Det saknas enhetlighet i systemlösningar, programvarumiljö och uppsättningen av IT-verktyg. Man konstaterar också att registerbaserad produktion och ett väl fungerande metadata-system kommer att få fortsatt stor betydelse.

De erfarenheter och iakttagelser om produktionsprocessen som redovisas i detta avsnitt stödjer vad som sagts ovan. Inom projektet har vi dels gjort egna praktiska erfarenheter av registerbearbetningar och dels diskuterat dessa frågor med Plattformprojektet och Metadatatprojektet. Redovisningarna i Kontaktgruppen har ofta kretsat kring problem i samband med produktionen. SCB måste fokusera på att lösa dessa problem om det skall vara möjligt att realisera intentionerna om ett effektivt och samordnat registersystem.

En väl fungerande driftsmiljö och produktionsprocess krävs för att registersystemet ska kunna vidareutvecklas. Det måste vara möjligt att sambearbeta flera register även om de är stora. Dessutom bör metadata göras tillgängliga redan i produktionsledet.

Exemplet Bamse

De problem som finns idag kan illustreras med situationen kring den s.k. Bamse-servern. Sedan slutet på 70-talet har SCB utvecklat en speciell service till kommuner i form av underlag på delområdesnivå. Systemet har gradvis byggts ut från att ha omfattat framför allt befolknings- och folkräkningsmaterial till att även omfatta bl.a. data från Fordons-, Sysselsättnings- och Inkomst och förmögenhetsregistret. I den nya miljön var det nödvändigt att samla dessa tunga bearbetningar och den tillhörande uppdragsverksamheten i *en* server. Denna har i takt med att plattformen bytts genomförts fått en mycket hög belastning, ofta pågår 80-100 sessioner samtidigt. Bearbetningarna kan ta lång tid, varför det är svårt att tidsplanera och lova kunder tidpunkt för leverans. Det är värt att notera att IT-enheten och berörda IT-grupper på avdelningarna sedan en tid tillbaka gör stora ansträngningar för att lösa problemen och därmed lägga grunden för en fungerande driftsmiljö.

Att problemen med den nya plattformen framför allt tagits upp av personalen kring Bamse sammanhänger förmodligen med att det där finns sambearbetningsbehov som baseras på tydliga kundkrav och det är fråga om många stora register som skall bearbetas. Mot den bakgrunden är det ganska naturligt att de mest kvalificerade kraven på processerna och driftsmiljön kommer från de som arbetar med Bamse-servern. Troligen kommer liknande problem i framtiden att uppstå i andra delar av registersystemet.

Metadata och dokumentation behövs också i processen

Metadata och dokumentation utgör inte ett särskilt moment i slutet av produktionsprocessen utan skapas idealt "under resans gång". Tyvärr kan vi konstatera att det tycks vara svårt att få till stånd en önskvärd samordning, inte bara mellan de olika delarna av metadata-systemet, utan dessutom mellan metadata-systemet och de olika produktionssystemen, trots goda teoretiska förutsättningar och att de olika delarna i stor utsträckning vilar på en gemensam begrepps- och beskrivningsmodell.

För att många användare skall kunna utnyttja registersystemet effektivt krävs att metadata byggs in. Tyvärr är det få verktyg som kan hantera metadata parallellt med data. Tanken är att metadata för de register som skall sam-bearbetas skall följa med vid matchning, selektering och aggregering. En matchning resulterar i ett nytt, tillfälligt register som innehåller just de variabler som man behöver i det aktuella uppdraget. Till detta register finns initialt inga metadata kopplade. Här borde man på ett enkelt sätt kunna hämta metadata från de register som matchas och knyta dem till det tillfälliga registret med hjälp av t.ex. PC-Axis. För detta krävs att man får möjlighet att importera redan befintliga metadata från Metadok.

Behovet av metadata blir också uppenbart när man skall generera många tabeller med omfattande indelningar och många variabelvärden som behöver skrivas in. Ett verktyg som skulle kunna uppfylla detta, med viss utveckling, är också PC-Axis. Tester har visat att programmets prestanda är i klass med rena SQL-frågor som kräver betydligt mer utbildning för att kunna hanteras. PC-Axis har valts som standardformat för leverans av pakettabeller till kommuner via SCB:s webbplats och har en struktur för hantering av metadata. Färdiga funktioner för gruppering av hierarkiska variabler finns och tabeller kan enkelt exporteras till andra format

Produktionsmetoder och verktyg

En naturlig målsättning för SCB har varit att bygga upp ett gemensamt datalager som skall vara tillgängligt för "alla". En vision av detta datalager har beskrivits av Plattformsprojektet och Registerprojektet och presenterats vid ett chefsmöte. Ett problem är fortfarande att det saknas klara strategier och metoder för hur denna vision skall realiseras. Som påpekas i en analys som gjorts av Bamsesituationen finns en stor variation i utformning av databaser, många designade utifrån det egna programmets behov och man har inte tagit någon större hänsyn till hur databaser skall samverka. Struktureringen av indata- och utdatabaserna brukar också anges som en viktig effektivitets- och kvalitetsfråga. Men hur denna strukturering skall göras verkar inte klart utrett.

17. Det är viktigt att de satsningar som nu planeras inom IT på en utveckling av SCB:s datalager och hantering av stora datamaterial får de resurser som krävs.

Det finns behov av ett generellt matchningsverktyg för den nya SQL-miljön. Personer i produktionen behöver lätt kunna matcha register så att det skapas dels en SQL-tabell med angivna variabler för de objekt som finns i samtliga ingående register och dels olika SQL-tabeller med vissa variabler för olika kategorier av icke-matchande objekt. Det finns exempel på menybaserade moduler där matchningen beskrivs grafiskt och som sedan genererar nödvändig SQL-kod.

Förutom att ett generellt matchningsverktyg skulle underlätta produktionen så skulle verktyget bidra till ökad kvalitet genom att risken för att göra misstag vid matchningen minskar och de tabeller som skapas vid mismatch kan användas för fortsatt analys av de objekt som inte matchat.

1999 genomfördes fullskaleförsök med Multi-server-join och stora datamängder och rekommendationer för användning togs fram. Slutsatserna var att det är möjligt att göra stora multiserverbearbetningar, att servrar bör struktureras efter typ av bearbetning och att det är värt att använda sig av Multi-server-join i vissa fall i stället för att dubbellagra data. Det bör dock betonas att det visade sig att en Multi-server-join gav bearbetningstider som är mycket längre än om motsvarande join görs inom en och samma server. Detta är

förmodligen skälet till att ingen, eller väldigt få personer, använder tekniken i praktiken.

SCB:s produktionsmiljö uppvisar en mycket splittrad bild av vilka verktyg man använder, vilket leder till att onödigt många systemlösningar förekommer (SQL, SuperCross, PC-Axis, Excel, SAS, SPSS). PC-Axis som nämnts mer ovan, har visast sig kunna användas för aggregering mm. av mikrodata och inte bara för makrodata, men detta är i praktiken mer eller mindre okänt inom SCB, vilket konstateras i en nyligen gjord utvärdering. AGDA, som utvecklats av SCB för aggregering, har endast fyra användare. Det SCB-utvecklade verktyget GRETA, för granskning, används av en enda produkt.

18a. Frågan om framtidens produktionsprocesser och krav på driftsmiljö bör studeras och utvecklas med särskild hänsyn till de problem som tar sig uttryck i Bamse. Det finns en risk att problemen med Bamse betraktas som undantag från en i övrigt ganska välfungerande miljö istället för ett exempel på frågor som måste lösas om och när hela registersystemet skall fungera i full skala.

18b. Produktionsmiljön är personberoende och i ganska liten grad standardiserad. Generella verktyg saknas (eller är alltför många för att kunna underhållas) och det är olyckligt att så mycket programmeringskompetens krävs för att utföra de vanligaste uppgifterna. Den nya IT-organisationen behöver ta ett större helhetsgrepp, genomlys processen och minska floran av verktyg genom tydliga satsningar.

Erfarenhetsutbyte och kompetensutveckling

Genom den nya IT-organisationen har det skapats förutsättningar för en ökad samverkan mellan olika delar av organisationen. Fler personer får därigenom en gemensam grund för systemarbetet och en praktisk kompetensutveckling. Det är angeläget att detta får genomslag även för det nödvändiga samarbetet i fråga om arbete med registerbaserad produktion. Vi har funnit en stor variation i fråga om praktiska metoder för registerprocessen. Goda lösningar behöver spridas och användas och gemensamma bedömningar kan läggas till grund för utveckling av generella verktyg som stöd för arbetet. Det är viktigt att detta inte bara involverar specifik IT-personal utan vidgas till att gälla även personer som är ansvariga för register och som med mer standardiserade hjälpmedel skulle kunna själva svara för en större del av produktionen än nu.

19. Inom plattformprojektet utformades ett förslag till erfarenhetsseminarier baserade på olika produktionsmoment och produkter. Detta förslag bör förverkligas av den nya IT-organisationen.

8. Dokumentation och metadata

Förenklat kan man säga att metadata är data om data. Metadata beskriver innehållet och den tekniska strukturen hos ett datamaterial och är viktigt för att tolka mikrodatas innebörd, möjligheter och begränsningar. Det har därför stor betydelse för såväl interna som externa användare. Registerstatistiken hämmas om tillgången till mikrometadata är låg och projektet har därför i stor utsträckning verkat för att få igång ett mer aktivt dokumenterande. Detta har bl.a. skett genom att ett antal viktiga register samlades i ett särskilt projekt, Dokumentationsprojektet, för att hjälpa personalen vid dessa register att komma igång med dokumentationsarbetet.

Det har i detta sammanhang varit naturligt att i första hand tänka på de *interna* användarna som behöver metadata när de arbetar med att framställa registerstatistik från egna register och från andra register inom SCB. Metadata är ett stöd för statistikproducenten vid underhåll och utveckling av produktionssystem. God tillgång på metadata gör produktionen mer effektiv och mindre sårbar för förändringar i personalen och underlättar utbildningen av ny personal.

Metadata har också stor betydelse för våra *externa* användare för att de ska kunna söka bland de register som finns och göra riktiga tolkningar och analyser av statistiken från SCB. Även i samband med beställning och produktion av uppdrag underlättar metadata kommunikationen mellan beställare och producent. Externa användares behov ställer därför också krav på ett fungerande metadata-system.

Situationen idag

På SCB finns ett stort antal register som är totalräknade. En förutsättning för att dessa ska fungera i ett registersystem enligt vår vision är att tillräcklig metadata finns om vart och ett av dessa register. Detta ställer krav på att ett omfattande dokumentationsarbete utförs på ämnesavdelningarna. Ansvar för att registren dokumenteras ligger hos respektive program. Trots att detta har funnits med i avdelningarnas styrdokument under flera år har arbetet gått långsamt.

Genom en inventering av dokumentationsläget för registren i registersystemet som gjordes vid årsskiftet 2000/2001 kan vi glädjande nog konstatera att andelen register som anger att de har minst en årgång dokumenterad i Metadok uppgår till nästan 70 % och många register har flera årgångar färdiga. När det gäller SCBDOK är läget sämre. Andelen register som har färdigställt minst en sådan dokumentation är ungefär 30 % och i de flesta fall finns enbart en årgång dokumenterad. Tyvärr har många av de här dokumentationerna inte skickats in för granskning och utläggning på Internet.

Det finns flera orsaker till att arbetet inte har genomförts i önskvärd takt. Dokumentationsarbetet fick stå tillbaka när plattformbytet krävde stora arbetsinsatser. Ett flertal har vid projektets kontaktgruppsträffar också påpekat att det ställs krav på registren att man ska dokumentera, men att en otillräcklig satsning har gjorts inom metadataområdet. I brist på resurser, både vad gäller personer och pengar, har t.ex. programutvecklingen av Metadok gått alldeles för sakta. Många har dessutom drabbats av tekniska svårigheter när de har använt programmet. Ett skäl till att antalet färdiga SCBDOK - dokumentationer är lågt verkar vara att många tycker det är svårt att förstå hur de olika avsnitten ska fyllas i. Denna synpunkt har ofta framförts på kontaktgruppsträffarna.

Metadata och dokumentationssystem har under senare år främst utvecklats genom kraven på offentliggörande av den officiella statistiken i databaserna.

Arbetet har präglats av detta perspektiv som inneburit att man har dokumenterat färdiga undersökningar eller mikrodataregister för att göra statistiken mera lättillgänglig för de externa användarna. Projektet vill dock betona vikten av den interna nyttan med dokumentation. En ny fas måste nu inledas som innebär att metadata och dokumentation skapas och används successivt under själva produktionsförloppet och som därefter resulterar i den nödvändiga slutdokumentationen. Detta ställer nya krav på metadatamodellen och verktygsutvecklingen. Ett exempel är att värdemängder som har lagts in i Metadok även borde kunna användas vid framställningen av statistiktabeller istället för att särskilda hjälptabeller läggs in i den egna databasen.

En viktig komponent i metadatasystemet är KDB. Denna har under en längre tid inte varit tillgänglig på SCB:s nät p.g.a. omläggning, vilket många har sett som ett hinder i arbetet. Det finns önskemål om att fler klassifikationer läggs in och att uppdateringar görs snabbare. Dessutom efterlyses grupperingar i KDB. Dessa skulle bidra till ökad samordning och underlätta arbetet på programmen så att snarlika grupperingar inte skapas på olika håll inom SCB.

Det har under projektets arbete även framgått att det ibland brister i samordningen mellan olika projekt som berör metadata. Trots att de har anknytning till varandra har ingen haft ett övergripande ansvar för att koordinera arbetet.

Förslag till insatser

Dokumentation får ofta stå tillbaka för andra arbetsuppgifter som prioriteras högre på programmen. För att få igång detta arbete på allvar är det viktigt att uppföljningen av signalerna från verksledningen blir tydligare. Ett sätt vore att det sker en regelbunden avrapportering av dokumentationsläget för *mikrodata*.

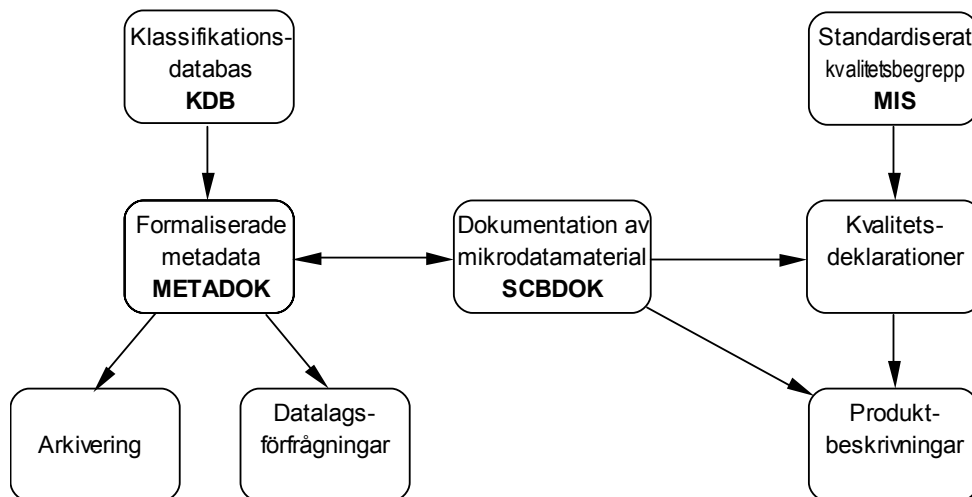
Om det ska vara rimligt att ställa krav på ämnesprogrammen när det gäller dokumentation måste mer resurser satsas på detta område. Metadok måste fungera stabilt och nya funktioner som efterfrågas av många bör utvecklas i en snabbare takt. I Metadataprojektet som haft ansvar för vidareutveckling av programmet fanns fram tills nyligen endast en person som arbetat med detta. Sedan hösten 2000 finns också en back-up vilket har minskat sårbarheten, men för att få en snabbare vidareutveckling behövs dock större programmerarinsatser. Önskelistan på nyutveckling är redan lång. Det som efterfrågas är t.ex. möjligheter för att *dokumentera longitudinellt* och söka variabelvis istället för register- eller årgångsvis. Även en koppling till Power Designer och SAS har efterlysts för att redan befintliga metadata ska kunna utnyttjas. Önskemål från Registerprojektets sida är att kvalitetsinformation ges mer utrymme i Metadok samt att ett gemensamt *kalendarium* utvecklas, där interna och externa händelser som påverkar statistiken kan läggas in.

Det är naturligtvis också viktigt att det finns resurser till användarstöd och granskning. Färdiga dokumentationer ska läggas ut på Internet och granskas dessförinnan av den centrala databasadministrationen på IP. Brist på personresurser till så viktiga projekt som metadata-, databas- och registerprojektet har medfört att granskningsarbetet inte har kunnat genomföras i den takt som varit önskvärd. Det är angeläget att denna flaskhals försvinner så att de som dokumenterar ser att materialet kommer till användning. Om det tar allt för lång tid innan de inlämnade dokumentationerna läggs ut är det svårt att motivera personalen att avsätta tid för sådant arbete.

20. Det behövs ökade resurser för att underlätta dokumentationsarbetet för ämnesprogrammen som har det yttersta ansvaret för dokumentationen. Även en satsning på användarstöd och granskning av färdiga dokumentationer krävs. Samtidigt behövs en central uppföljning av dokumentationsläget.

Den grundläggande idén i metadatasystemet är att information som en gång har lagts in i systemet ska kunna utnyttjas på flera olika sätt. I de fall samma information efterfrågas på flera ställen, som i "Beskrivning av statistiken", kvalitetsbeskrivningarna, SCBDOK och Metadok, ska man bara behöva skriva en och samma sak på ett ställe i systemet. Denna inläggning ska sedan kunna utnyttjas i de övriga delarna. Hur allt hänger ihop kan vara svårt att överblicka. Därför bör en översiktlig handledning tas fram som gör användaren uppmärksam på hur de olika delarna i systemet hör samman. Till exempel kan metadata som lagts in i SCBDOK och Metadok utnyttjas vid arkivering och för att hantera datalagsförfrågningar behövs Metadokdokumentation.

Bild 8. Schematisk bild över SCB:s metadatasystem



Registerprojektets kontaktgruppsträffar har varit ett viktigt och uppskattat tillfälle att diskutera metadatafrågor. Under träffarna har information getts om planer och pågående arbete inom Metadataprojektet och användarna har kunnat framföra sina synpunkter. I samband med att Registerprojektet avslutas försvinner denna möjlighet till kommunikation. Behovet av en sådan kommunikationskanal kvarstår dock.

Ett flertal har framfört att SCBDOK inte är anpassat för register då vissa delar av mallen bara är relevanta för surveyundersökningar. Därför behövs särskilda instruktioner för register tas fram. För att förenkla arbetet kan man också göra exempel på bra dokumentationer för både SCBDOK och Metadok tillgängliga i anslutning till befintliga handledningar. Det är också väsentligt att en kurs erbjuds i SCB:s kursutbud som innefattar metadatasystemet i stort och hur man använder dokumentationsverktygen. Tidigare har det funnits en återkommande kurs i SCBDOK, men den har ställts in p.g.a. lärarbrist. För att få en god kvalitet på de dokumentationer som görs behövs utbildning, inte minst för att SCBDOK upplevs som svårt av många.

21. Utbildning och handledningar om såväl metadatasystemet som dess delar måste finnas. Vi föreslår att det nätverk för registerfrågor som diskuteras i avsnitt 10, även fungerar som ett discussionsforum där synpunkter på metadatasystemet kan tas upp och vidareförmedlas. Dessutom vill vi särskilt uppmärksamma behovet av särskilda anvisningar till stöd för dem som ska dokumentera register i SCBDOK.

9. Sekretess och integritetsskydd

Samtidigt som registersystemet förbättras för att underlätta sambearbetningar och höja statistikens kvalitet, bör även sekretess- och integritetsskyddet inom systemet förstärkas. Det är angeläget att sekretesskyddet förbättras snarast.

Ökat skydd för personer med spärrmarkering

Det faktum att nazister fick känsliga data från en anställd på Försäkringskassan aktualiserar frågan om hur SCB kan förbättra säkerhetsrutinerna avseende personer med skyddad identitet. Dessa har så kallad spärrmarkering i en särskild variabel i Befolkningsregistret.

Personer med spärrmarkering (idag finns ca 8 000 personer som har skyddad identitet genom spärrmarkering) bör behandlas på något av följande sätt:

- Uppgift om adress, församling och fastighet tas bort ur Befolkningsregistret och alla övriga register. Variabeln med uppgift om spärrmarkering tas bort helt. Eventuellt kan dessa uppgifter sparas i en fil med begränsad behörighet.
- De spärrmarkerade personerna *tas bort helt* ur register med personnummer i det ”vanliga” registersystemet, men finns i ett särskilt väl skyddat register.

Idag ingår inte de spärrmarkerade i våra urval från Befolkningsregistret.

Ökat integritetsskydd

FoB90-debatten var negativ för SCB. Den aktuella debatten under hösten 2000 kring abortregistret visar att frågan fortfarande är mycket känslig. Detta aktualiserar frågan om SCB skall förbättra rutinerna för integritetsskydd innan en ny debatt om FoB2005 uppstår.

Namn, adress etc. i klartext bör endast finnas i basregistren. Det är helt onödigt att all SCB-personal som kommer i kontakt med övriga register skall kunna läsa dessa uppgifter. Även om detta inte är ett stort problem bör det åtgärdas så att *alla* övriga register är helt anonymiserade genom sifferkoder.

Kan vi minska användningen av personnummer? Den nya statistiklagstiftningen gör det möjligt att lämna ut avidentifierade uppgifter försedda med löpnummer för senare uppdatering. Vi kanske även skulle kunna använda avidentifierade register med löpnummer inom SCB. En förutsättning för detta är att nycklarna mellan person- och löpnummer är väl skyddade. Löpnummer kan användas för känsliga register inom SCB som idag innehåller person- eller organisationsnummer (som ibland är personnummer). Om FoB2005 endast kommer att innehålla löpnummer bör SCB kunna hävda att vi gjort en stor satsning för att öka integritetsskyddet.

I samband med detta vill vi påpeka att det har framförts önskemål om att reglerna för utlämnande av mikrodata bör preciseras. Dessutom finns det risker med att utlämna detaljerade metadata av teknisk natur.

22a. Det bör snarast utredas hur spärrmarkerade personer kan få ökat skydd inom SCB:s registersystem. Denna fråga gäller inte enbart Befolkningsregistret utan även Företags- och Fastighetsregistret. Namn, adress etc. skall inte förekomma i några av systemets register förutom i dessa tre basregister.

22b. Att ersätta personnummer med löpnummer skulle innebära en stor förändring och konsekvenserna måste undersökas noga.

22c. De nuvarande reglerna för utlämnande av mikrodata bör preciseras: Användare och personal måste göras uppmärksamma på att samma regler gäller för personnummer inom både företags- och individregister. Reglerna bör göras mer konkreta: Vilka variabler är känsliga och vilka är mindre känsliga?

10. Registerfunktion och nätverk

Registerprojektet har under några år haft till uppgift att hantera viktiga metod- och samordningsfrågor inom registerområdet. Nu upphör projektet och det blir då nödvändigt att ta ställning till behovet av och formerna för det fortsatta arbetet inom området. Därför måste dessa frågor få en mer permanent förankring i SCB:s organisation. Samma gäller frågorna om metadata som hanterats i projektform under senare år. Det finns starka kopplingar till centrala standardfrågor, som redan har en plats inom Metodenhetens basverksamhet, men med små resurser till förfogande.

Förslag och initiativ

Projektet har vid några tillfällen tagit upp dessa frågor i Vetenskapliga Rådet och i samtal med verksamheten. Det finns också ett förslag av Bo Sundgren (SOU 1999:96) som förordar att en registerfunktion inrättas inom den centrala databasverksamheten för bearbetningar och sambearbetningar av mikrodata för statistiska ändamål. Utredningen om en ny metodorganisation vid SCB, som ännu inte avslutats, pekar på behovet av en upprustning av den centrala funktionen för standardfrågorna. Det nya Metadataprojektet kommer att ha viktiga uppgifter avseende dokumentation och utveckling av metadatasystemet.

Slutligen kan noteras att U-avdelningen i sitt styrdokument för år 2001 fått i uppgift att tillsammans med IP och ämnesavdelningarna utreda hur en central registerservicefunktion skulle kunna utformas. Erfarenheter finns både från Danmarks statistik och från centralbyrån i Norge. Där finns enheter för kundservice, kvalificerade registeruppdrag och förmedling av mikrodata.

23. En central registerfunktion med ansvar för utvecklings- och samordningsfrågor bör finnas inom U-avdelningen. Till denna knyts ett nätverk av personal från övriga verket med metod- och produktionserfarenheter. De organisationsfrågor som tas upp i denna rapport behandlas i samband med U-avdelningens pågående utredning angående en central registerservicefunktion.

Arbetsfördelning

Vid SCB har den nya organisationen (1998) inneburit förändringar när det gäller basregistrens placering och det har blivit en bättre överensstämmelse mellan den formella organisationen och registerstrukturen. Tidigare fanns tre av basregistren inom en och samma avdelning. I den nuvarande organisationen kan vi se hur ämnesavdelningarna sammanfaller ganska väl med de fyra olika basregistren. Det här innebär att samtliga avdelningar är djupt involverade i registersystemet, men betyder samtidigt att samordningsbehovet ökat.

Registerprojektets kontaktgrupp har under de senaste åren fungerat som ett betydelsefullt forum för erfarenhetsutbyte och praktiska samordningsfrågor. Det finns i dag ingen permanent lösning och här räcker det inte med en central funktion. Ett nätverk föreslås, som kommer att överta kontaktgruppens roll. Dessutom kan arbetsgrupper inom detta nätverk utveckla nya typer av sambearbetningar.

SCB skall ha kapacitet att analysera stora mikrodatamaterial. Även denna analyskompetens bör kunna utvecklas inom nätverkets ram. En sådan kapacitet är viktig för att kunna utveckla ny registerstatistik. Dessutom kan forskare som arbetar med utlämnat SCB-material räkna fel eller komma till kontroversiella resultat. Då måste vi inom SCB kunna göra egna analyser.

Produktionen av registerstatistik och utvecklingen av registersystemet kräver samverkan över program- och avdelningsgränser. Rollfördelningen mellan program, avdelningar och den centrala registerfunktionen måste vara tydlig.

En grov arbetsfördelning mellan avdelningar och program samt den föreslagna centrala registerfunktionen skulle kunna se ut på följande sätt:

24a. Varje ämnesavdelning har ansvar för sitt basregister och övriga register inom sitt ansvarsområde och har särskilt ansvar för ett antal standardiserade variabler.

24b. Varje ämnesavdelning svarar för länkarna mellan sina statistiska register och respektive basregister.

24c. Utvecklingen av registermetodfrågor sker inom SCB:s gemensamma metodprogram.

Den centrala registerfunktionen föreslås ligga inom Utvecklingsavdelningen med följande ansvarsområden:

25. Den centrala registerfunktionen har följande uppgifter:

- Övergripande samordnings- och standardiseringsfrågor som avser definitioner och namnsättning av objekt och variabler.*
- Ansvar för metadatasystemet och utveckling av dokumentationsstöd.*
- Följer upp att allmänna riktlinjer efterlevs.*
- Håller i ett nätverk för erfarenhetsutbyte.*
- Ansvarar för att utveckling av nya sambearbetningar och analyskompetens sker inom nätverkets ram.*

Förteckning över utkomna R&D Reports

R&D Reports är en för IT-enheten och Metodenheten gemensam publikationsserie, som 1988-01-01 ersatte de tidigare "gula" och "gröna" serierna. I serien ingick fram till årsskiftet 1992-93 även **Abstracts** (sammanfattning av metodrapporter från SCB).

Reports published during 1999 and onwards:

- 1999:1 Täckningsproblem i Registret över totalbefolkning RTB. Skattning av övertäckning med en indirekt metod (*Jan Qvist*)
- 1999:2 Bortfallsbarometer nr 14 (*Per Nilsson, Antti Ahtainen, Mats Bergdahl, Tomas Garås, Jan Qvist och Charlotte Strömstedt*)
- 1999:3 Att mäta statistikens kvalitet (*Claes Andersson, Håkan L. Lindström och Thomas Polfeldt*)
- 2000:1 Kalibrering av vikter – beskrivning av tekniken och de SCB-fall den prövats i (*Sixten Lundström et al*)
- 2000:2 On Inclusion Probabilities and Estimator Bias for Pareto π ps Sampling (*Nibia Aires and Bengt Rosén*)
- 2000:3 Bortfallsbarometer nr 15 (*Per Nilsson, Ann-Louise Engstrand, Sara Tångdahl, Stefan Berg, Tomas Garås och Arne Holmqvist*)
- 2000:4 Bortfallsanalys av SCB-undersökningarna HINK och ULF (*Jan Qvist*)
- 2000:5 Generalized Regression Estimation and Pareto π ps (*Bengt Rosén*)
- 2000:6 A User's Guide to Pareto π ps Sampling (*Bengt Rosén*)
- 2001:1 Det statistiska registersystemet. Utvecklingsmöjligheter och förslag (*Anders Wallgren och Britt Wallgren*)

CATARINA ELFFORS
MR/BY-S

ISSN 0283-8680

Tidigare utgivna **R&D Reports** kan beställas genom Katarina Klingberg, SCB, MET, Box 24 300, 104 51 STOCKHOLM (telefon 08-506 942 82, fax 08-506 945 99, e-post katarina.klingberg@scb.se). **R&D Reports** from 1988-1998 can - in case they are still in stock - be ordered from Statistics Sweden, attn. Katarina Klingberg, MET, Box 24 300, SE-104 51 STOCKHOLM (telephone +46 8 506 942 82, fax +46 8 506 945 99, e-mail katarina.klingberg@scb.se).