

Interne notater

STATISTISK SENTRALBYRÅ

85/28

10. juli 1985

OMLEGGINGEN AV PERSONREGISTERET OG AV PRODUKSJONEN AV BEFOLKNINGSSTATISTIKK, OG KONSEKVENSER FOR GJENBRUK AV DATA.

Av Lars Østby

1. Notatets formål	2
2. Forholdet mellom BSR og annen personstatistikk	4
2.1. Metadatabase - DIPS	4
2.2. ISDS - programverktøy for metadatabasen	6
2.3. Kryptering	6
2.4. Sammendrag	8
3. Forholdet DSP-BSR	8
4. Beredskapsfil og historiske data	11
4.1. Dagens beredskapsfil	12
4.2. Beredskapsfile under det nye produksjonssystemet	13
4.3. Oppbyggingen av dagens beredskapsfil	14
4.4. Oppbyggingen under et nytt produksjonssystem	15
4.5. Organisering av arbeidet	18
Referanser.....	19

1 Notatets formål

I forbindelse med de endringene som nå er i ferd med å skje på flere ulike felter som berører befolkningsstatistikken, har Sosiodemografisk forskningsgruppe behov for å sikre seg en samlet oversikt over det som skjer. Først med en slik oversikt vil vi kunne vite hvilke konsekvenser endringene kan få for analyseformål. Sosiodemografisk forskningsgruppe regner det som en viktig oppgave å se til at data som er samlet inn blir organisert slik at det er praktisk og økonomisk mulig å bruke dem til forskningsformål, også slike formål som ikke var kjent eller prioritert ved etableringen av datagrunnlaget. (Se f.eks. Bakgrunnsnotat nr. 6 til Statistisk Sentralbyrås Langtidsprogram for 1980-årene. IN 02/27, av Per Sevaldson.)

Med så mange endringer som gjennomføres eller planlegges gjennomført samtidig, (vi kommer seinere tilbake til hvilke) er det en viss fare for at ingen på brukersiden har full oversikt over hvilke endringer som har betydning for datatilgjengeligheten for analyseformål. Dette notatet er ett skritt i retningen av å få oversikt over endringene. Notatet gir uttrykk for hvordan situasjonen i øyeblikket oppfattes fra brukersiden. Den videre utvikling vil gi avklaring på punkter der løsningen ennå ikke er fastlagt og kan også føre til endringer i forhold til den beskrivelsen som er gitt i dette notatet. For brukerne vil det da være en fordel at de nye utviklingstrekk kan refereres til og forstås i forhold til den situasjon som er beskrevet her. Notatet har vært lest av personer i alle tre avdelinger i Byrået, og det har kommet fra nyttige synspunkter som er innarbeidd. Særlig har kommentarer fra Egil Hahberstad, Eva Jemblie Monsen og Halvard Skiri vært til stor hjelp. De oppfatninger som legges fram i notatet er likevel bare forfatterens.

Det er minst tre viktige forandringer underveis som har stor betydning for befolkningsstatistikken:

- 1) Det sentrale personregister (DSP) vil fra 1. juni 1985 ligge på en database i Statens datasentral for administrativ databehandling (SDS). Denne basen skal omtales som DSP-basen. Bakgrunnen for dette står i Forprosjektnotatet NSH/VFr, 4/2-82. Innholdet i basen finnes i Feltnummerkatalog fra SDS, oppdatert 20.06.84. Planene for overføringen er beskrevet i DSP-prosjektet, Milepelplan for totalbasen, oppdatert 25.02.85.
- 2) Produksjonsrutinene for befolkningsstatistikken skal overføres til IBM (se prosjektskriv Ski/KWa, 5.8.83, godkjent 31.8.83, og notat JBy/KWa, 6.12.84). Som en følge av blant annet kostnadsstruktur og formål med DSP, skal befolkningsstatistikk produseres på grunnlag av et Befolkningsstatistikkregister (BSR) på IBM-anlegget på Kongsvinger og ikke mer eller mindre direkte fra Personregisterets filer som nå (se også møtereferat Ski/Rac, 17.9.84).
- 3) I forbindelse med Folke- og bolig tellingen 1990 er planlagt utredet et datasystem for integrert personstatistikk, DIPS. (Prosjektskriv JKT/Ski/HDr/11/12/84 og notat fra Systemkontoret KKv/SMS, 3/12-84. Se også Databeredskap for persondata. Rapport fra et forprosjekt WJo/JKT/KKv/LHå, 20/7-83. DIPS er dessuten beskrevet av Halvard Skiri i artikkelen "DIPS - et redskap i arbeidet med personstatistikken" i Byråpraten 2-1985.) Det viktigste formålet for utredningen er vel å vurdere om FOB-90 kan gjennomføres som en ren registertelling uten bruk av innsamlede personskjema. Som resultat av DIPS-prosjektet, er det sannsynlig at en vesentlig del av Sosiodemografisk forskningsgruppe's datagrunnlag blir beskrevet i en metadatabase - en base med opplysninger om definisjoner og lagring av data, men uten selve dataene.

Siden organiseringen av BSR vil få konsekvenser for det vi ønsker fra DSP, vil jeg her gå nærmere inn på denne organiseringen før jeg tar opp forholdet mellom BSR og DSP og mellom Beredskapsfilen og DIPS.

2 Forholdet mellom BSR og annen personstatistikk

Befolkningsstatistikkregisteret må sees i sammenheng med produksjonen av annen personstatistikk som også skjer/vil skje ved IBM-anlegget på Kongsvinger. Dette produksjonssystemet (DIPS) har også vært omtalt som Personstatistikkdatabasen og Databeredskap for persondata. Produksjonssystemet planlegges nå med utgangspunkt i prosjektskriv JKT/Ski/HDr/11/12/84. I henhold til någjeldende planer vil det omfatte produksjon av individstatistikk på følgende felter. Det er mulig at flere kan tas inn seinere.

Sysselsettingsstatistikk,	}	Møtereferat WJo/JKT/HDr/15/8-83
Inntekts- og formuestatistikk		
Folketellingsstatistikk,		
Regionalstatistikk,		
Utdanningsstatistikk,		Møtereferat og prosjektskrivet
		JKT/Ski/HDr/11/12/84
Befolkningsstatistikk		Møtereferat Ski/Rac, 17.9.84.

2.1 Metadatabase-DIPS

Statistikkgrunnlaget innenfor disse områdene er tilsammen for omfattende til at det kan velges en tradisjonell databaseløsning (lagring på disk, og direkte aksess til hvert enkelt individ). Byrået vurderer istedet å opprette en felles metadatabase, mens dataene selv trolig vil ligge organisert på filer (oftest magnetbånd) som metadatabasens programutstyr ved behov vil kunne bruke som om det var data på en og samme fysiske fil. En metadatabase er en base som inneholder bare opplysninger (definisjoner, hvor de lagret osv.) om de data som skal brukes i produksjon og analyse av

statistikk, men den inneholder ikke selve dataene. Forskjellen fra den databaseløsningen som en opprinnelig tenkte seg (Personstatistikkdata-basen), ligger i produksjonsmetodene. Det er nå snakk om batchkjøringer på sekvensielle filer eller muligens oppretting av "partielle databaser" på temporær basis, ikke interaktiv kjøring mot en database hvor en har direkte adgang til hele basens informasjon. Forskjellen i produksjonsmetode har konsekvenser for hvor lenge en må vente på resultatene, og kanskje for fleksibiliteten i produksjonen, men disse konsekvensene er lette å akseptere.

Det kan bli aktuelt å la en sentral kjerne av koplede, ofte benyttede data ligge på disk. Innholdet i denne kjernen er ikke bestemt, og kan vel i noen grad bli regulert av etterspørsel. Det virker sannsynlig at denne kjernen vil inneholde også elementer av befolkningsstatistikk.

Det dokumentasjonssystemet som det legges opp til, kalles Data-system for Integrert Person Statistikk - DIPS. Uten at endelig innhold og omfang av DIPS er avklart, synes det rimelig å tro at data innenfor minst de seks områdene for individstatistikk som er listet opp foran vil bli dokumentert innenfor metadatabasens rammer, men ikke alle filer innenfor hvert av områdene. Utredningen av hvordan DIPS skal legges opp, skjer som et ledd i forberedelsene til Folke- og boligtellingsen 1990, og ledes av en egen styringsgruppe oppnevnt i februar 1985 med Tønder som formann (EAU/JKT/HaR, 19/2-85). Halvard Skiri er formann i den prosjektgruppa som skal gjennomføre arbeidet. Prosjektgruppa arbeider etter Milepelplan Ski, 7/2-85.

2.2 ISDS - programverktøy for metadatabasen

Til å ta data ut fra de filene som er knyttet sammen i DIPS, trengs et hjelpemiddel (programsystem) som gjør det mulig å behandle filene, ekstrahere data og bearbeide disse til tabeller eller til analysefiler som kan behandles videre f.eks. med SAS eller SPSS. Ekstraheringsopplegget som ble opprettet i forbindelse med den databasen som 3. kontor har for kommunaløkonomi, planlegges utbygget til et slikt hjelpemiddel som også er egnet for individdata. Det er skrevet for ekstrahering av aggregatdata. Systemet heter ISDS (Integrert Statistikk Data System, se KKv/SMS, 3/12-84). Den som skal bruke ISDS på data organisert og beskrevet i en metadatabase, trenger i prinsippet ikke vite noe om den rent fysiske organiseringen av data. Det at lagringen skjer på atskilte filer, skal ikke redusere bruksmulighetene sammenliknet med en ren databaseorganisering. Dersom en ønsker å knytte til muligheter for tabellproduksjon, kan dette skje ved at en "henger" på en versjon av TAB68 eller et annet tabellprogram. Det er fortsatt mulig at basen blir styrt med et ferdig innkjøpt 4. generasjons programmeringsspråk.

2.3 Kryptering

Ved å føre ulike statistikkområder med en felles metadatabase inn i ett stort system for statistikkproduksjon, åpnes mange muligheter for kopling av data. Byrået er svært opptatt av at slike koplinger skal skje innenfor gjeldende regler, og Byråets og Datatilsynets tolkning av disse reglene. Derfor foreslås det at det meste av individdata innenfor DIPS skal

ha et kryptert nummer (K.nr.), ikke fødselsnummer. Dette krypterte nummer er slik at en kan gå fra fødselsnummer til K.nr. ved retting, oppdatering osv., men ikke tilbake. Veien fra K.nr. tilbake til fødselsnummer kan en i teorien holde åpen ved å la Datatilsynet oppbevare en nøkkel. Det er mulig at Byrået vil opprette egne K-nummerserier innenfor hvert statistikkområde, og bare få kople data mellom områdene (f.eks. inntekt og sysselsetting) dersom Datatilsynet gir oss overgangsnøkkelen til et slikt formål. Krypteringen er behandlet på møte hos Direktøren 6. mars 1985, på grunnlag av notat EAU/BjF, 5/2-85. Det er laget en revidert versjon av notatet (samme ident) som er sendt Datatilsynet. Retningslinjer for den praktiske gjennomføringen utarbeides av Kvisla, utkast KKv, 19/3-85. Datatilsynet har foreløpig ikke tatt standpunkt hverken til det praktiske eller prinsipielle i forslaget (midtsommer 1985). Det er viktig at de som har ansvaret for DIPS sørger for at koplingsmulighetene ikke blir unødig begrenset, hverken juridisk eller praktisk. Valget av praktisk opplegg for kryptering kan få konsekvenser for den videre utnyttelsen av DIPS til analyseformål. Det er derfor nødvendig at de som skal hente sine data via DIPS setter seg inn i forslaget til kryptering (KKv, 19/3-85). Av hensyn til notatets oversiktighet, kan ikke dette forslaget refereres her.

Det er foreløpig ikke klart hvilke konsekvenser en eventuell kryptering vil få for BSR. En kan f.eks. tenke seg to utgaver av BSR. Den ene er en fullstendig versjon med Fnr. og alle data som er nødvendig for statistikkproduksjon. Dette blir da å oppfatte som et lokalt register i den terminologien notatet EHA, 9/5-85 bruker. Statistikken lages uten bruk av krypterte data. Den andre utgaven vil inneholde det utvalg av variable fra det lokale registeret som en regner med blir etterspurt for bruk innenfor DIPS, både i en eventuell kjerne og som selvstendige filer. Krypteringen kan være en betingelse for å gjøre demografiske data tilgjengelige for kopling innenfor DIPS, men den bør ikke være til noen ulempe for produksjonen av den løpende statistikken.

2.4 Sammendrag

Det som er sagt hittil, kan sammendras slik: Befolkningsstatistikken skal produseres ved IBM-anlegget på Kongsvinger. Det vil skje innenfor, eller i nær-tilknytning til et opplegg (metadatabase) hvor mye personsstatistikk blir integrert, og som kanskje blir en viktig del av FOB-90. Dette opplegget vil basere seg på individdata, som i stor grad vil bli lagret på krypterte individnummer på grunnlag av fødselsnummeret.

3 Forholdet DSP-BSR

I møte hos Direktøren 30. august 1984 (referat Ski/Rac, 17.9.84) i anledning omlegging av den løpende befolkningsstatistikken til IBM, ble sagt at den løpende befolkningsstatistikken bør produseres med utgangspunkt i et eget personregister i Byrådet bygd opp for statistikkformål. Dette registeret omtales som BSR. Grunnen til at Byrådet må lage et nytt opplegg for produksjonen av befolkningsstatistikk, er behovet for et bedre opplegg enn det gamle. Dessuten skal produksjonsgrunnlaget, Personregisteret, legges på database i SDS og statistikkproduksjonen er vedtatt flyttet over til IBM, se prosjektskriv Ski/KWa, 5.8.83. Planene for omleggingen av befolkningsstatistikken er beskrevet i notat J8y/KWa, 6.12.84 og diskutert i ulike sammenhenger, se f.eks. møtoreferat EJM/Rac, 4.3.85.

Det sentrale personregister har alltid hatt administrative hensyn som sitt hovedformål. Det har vært lokalisert til Byrådet, og har også hatt til formål å gi grunnlaget for den løpende befolkningsstatistikken. Regist-

eret har i noen grad tilpasset seg de behov som statistikk og analyse har hatt. Ved at det nå organiseres som en database og lokaliseres utenfor Byrået, er det grunn til å tro at hensynet til produksjon og analyse av befolkningsstatistikk kommer mer i bakgrunnen. De to viktigste grunnene til å anta dette er:

1) Prisen på uttak fra basen er satt slik at ekstraheringer som angår hele befolkningen eller store deler av den blir svært dyre, mens uttak av et lite utvalg personer går meget raskt, er relativt billig og vil være fra en base med en meget høy datakvalitet.

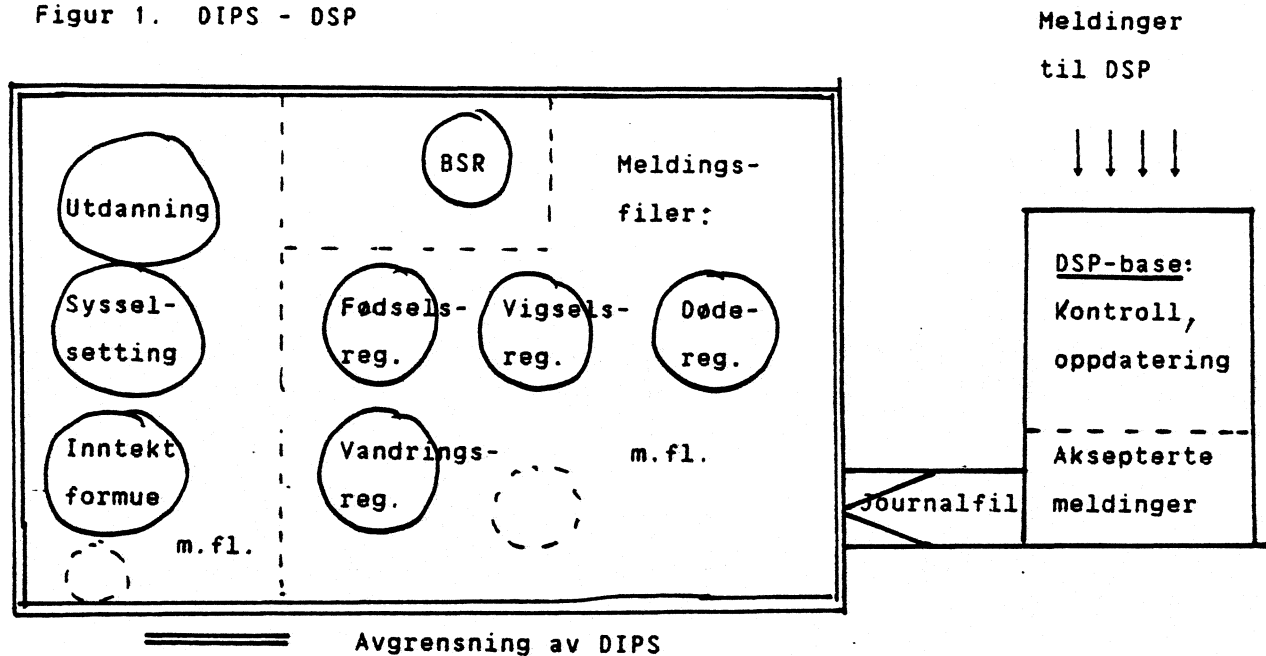
2) Gjennom hele arbeidet med omleggingen er registerets administrative formål understreket, og hverken teknisk eller økonomisk vil det være så lett å få SDS til å tilpasse seg også statistikkhensyn.

Såvel Fagavdelingen (12. ktr.) som Forskningsavdelingen (Sos.dem) har vært representert i den styringsgruppen som har ledet omleggingen til database, slik at det har vært mulig å få lagt fram de behov som statistikk og analyse har hatt ved omleggingen. Disse behovene er, slik det nå ligger an, tenkt ivaretatt som beskrevet videre i dette notatet.

Det opprettes et befolkningsstatistikkregister (BSR) på IRM-anlegget på Kongsvinger. Oppstartingsgrunnlaget for dette registeret vil være en situasjonsfil fra DSP, kanskje den som representerer starten på DSP-basen pr. 1. juni 1985. BSR skal oppdateres med de samme meldingene som brukes til å oppdatere DSP-basen, ved hjelp av den såkalte journalfilen. Alle meldinger som kommer til DSP-basen blir først kontrollert og eventuelt rettet. De som så blir brukt til å oppdatere DSP-basen, blir kopiert og lagt ut på journalfilen. 12. kontor har spesifisert innholdet de ønsker både i den situasjonsfil som starter BSR-registeret og i de enkelte meldingsfiler som skal brukes i oppdateringen. (En serie notater med tittel Innhold på melding fra personregisterkontoret til 12. kontor, EJM, januar, februar og mars 1985.)

Organiseringen av dette kan tenkes slik:

Figur 1. DIPS - DSP



===== Avgrensning av DIPS

----- Alternative avgrensninger av DIPS, mindre ønskelig fra Sosiodemografisk forskningsgruppes side.

Tegningen leses slik: I henhold til Lov om folkeregistreringen av 16. januar 1970 og Forskrifter om føringen og ordningen av folkeregistrene, kommer det meldinger til DSP-basen. Disse kontrolleres og rettes før de oppdaterer registeret. Samtidig med oppdateringen hentes eventuell relevant informasjon ut fra basen, og meldingene legges ut på journalfilen sammen med denne informasjonen. Journalfilen vil med regelmessige mellomrom, bestemt av behovet, brukes til å oppdatere BSR og et sett av meldingsfiler for hver meldingstype som inngår i BSR. Oppdateringen vil skje minst én gang i kvartalet. Utvalget av kjennemerker som inngår i BSR vil være langt mindre enn det som inngår i de enkelte meldingsfilene, men dersom såvel meldingsfilene som BSR blir tilgjengelige for DIPS eller et annet metadatabasesystem, vil en alltid kunne bruke et kjennemerke enten fra meldingsfilene eller BSR. Grunnen til oppdelingen er at BSR ikke bør bli så stort og uhåndterlig at det får for dårlige brukeregenskaper. Vi vil også legge til at det ikke er tatt noen endelig avgjørelse om hvor stor del av BSR og tilgrensende filer som skal inn i en eventuell sentral kjerne i DIPS eller gjøres direkte tilgjengelig for DIPS ved en koplingsberedskap. Fra Sosiodemografisk forskningsgruppes side synes det ønskelig at både BSR og alle meldingsfilene blir gjort lettest mulig tilgjengelig for DIPS. Men

begrensninger i bruken av et koplet materiale med mange identifiserende kjennemerker, kan medføre at det er mest hensiktsmessig at bare deler av BSR og meldingsfilene er i den eventuelle kjernen, og at bruken av resten av registrinnholdet er mulig bare dersom en får tillatelse til det. 12. kontor vil dokumentere hele BSR og alle meldingsfilene slik at de kan være tilgjengelige for koplinger i DIPS.

Slik 12. kontor har planlagt oppdateringen, synes det som om de data som brukes i ajourholdet dekker de behov Sosiodemografisk forskningsgruppe mener BSR bør dekke. Det synes som om at alle de behov som blir dekket av det eksisterende system for produksjon av befolkningsstatistikk, fortsatt blir tilgodesett. Videre blir kvaliteten bedre i den forstand at registeret til enhver tid vil være bedre ajour og at en vil trekke ut flere informasjoner fra DSP-basen til meldingsfiler og ajourhold enn det en får nå.

4 Beredskapsfil og historiske data

Sosiodemografisk forskningsgruppe har sterke interesser også i analyse av tilbakegående data, ikke bare i den løpende befolkningsstatistikk. Denne interessen kommer til uttrykk gjennom gruppas analyser, engasjement i beredskapsfilen og i et stort antall notater om gjenbruk av data. Se f.eks. IN 82/27, notatet Bruk av registerdata til demografisk forskning (PSe/MFo, 17/8-76), JaM/GHu, 10/8-76 og Løs/SiW, 11.1.83.

4.1 Dagens beredskapsfil

For å gjøre det noe lettere å analysere tilbakegående tall fra Personregistersystemet, er det årlig fra 1965 laget en såkalt beredskapsfil. Bare de kjennemerkene som en har antatt har interesse for analyseformål, er tatt med. Det kommer årlig 1.1-1.2 millioner meldinger til Personregisteret. De fleste av disse (noe under 1 mill.) blir tatt med i beredskapsfilen. Den er derfor blitt temmelig stor etterhvert. Uansett omleggingen av Personregisteret ville det være aktuelt med en omorganisering av beredskapsfilen som kunne gjøre den lettere å bruke.

Beredskapsfilen kan brukes både til prosjekter som trenger tilbakegående tall som en tidsserie av tverrsnittsstatistikker og til prosjekter som setter begivenheter hos enkeltindivider i relasjon til foregående begivenheter blant de samme individene. Eksempler på dette siste er å se flyttinger i relasjon til tid siden foregående flytting og tidligere bosteder ("tilbakeflytting", Alstad, 1982) å se fødsler i relasjon til tid siden foregående fødsel eller tid siden endring av ekteskapeleg status, eller som i RAPP 82/2 å analysere gjengifte etter tid siden skilsmissen, skilsmisse etter ekteskapsvarighet o.l. (Mønnesland et al 1982). Analyser som dette omtales ofte som livsløpsanalyser fordi de bygger på muligheten til å se begivenheter på ulike stadier i livsløpet i relasjon til hverandre.

4.2 Beredskapsfil under det nye produksjonssystemet

Slike analyser har vært sett på som en spesialinteresse som Sosiodemografisk forskningsgruppe har, og ikke som et ledd i den løpende befolkningsstatistikken. Vi håper imidlertid at livsløpstilnærmingen etterhvert kommer sterkere inn i såvel analyser som i den løpende befolkningsstatistikken. Det er de siste åra utviklet statistiske metoder som egner seg godt til analyse av livsløpsdata. Blant problemstillinger som er aktuelle for statistikkproduksjon og analyse, nevnes eksempler som fødsler og fruktbarhet i relasjon til tidligere statsborgerskap eller i relasjon til oppholdstid i landet for utenlandske statsborgere; omfang av tilbakeflytting til de ulike regioner; gjengifte i relasjon til tid siden oppløsning av forrige ekteskap/ektefelles død; forholdet mellom dannelse av egen familie og utflytting fra oppvekstfamilien. Statistikk av denne typen kan det nok være mulig å produsere innenfor dagens system, men DSP-basen organiserer personregisteret slik at det blir lettere, særlig når en må trekke inn relasjoner til andre familiemedlemmer.

Det er av stor betydning at de tidsserier som beredskapsfilen gir oss fra personregisterets start og fram til 1984, kan fortsettes også etter at systemet for produksjon av befolkningsstatistikk er omlagt. I og med omleggingen til IBM, vil det gamle produksjonsopplegget opphøre, og vi må finne et opplegg som kan føre tidsseriene videre. Vi bør dessuten ta med oss de forbedringene som omleggingen åpner for og vi bør bedre anvendligheten og tilgjengeligheten av beredskapsfilen. Beredskapsfilen er idag lite brukt i forhold til det analysepotensial den inneholder. I omleggingen til nytt produksjonssystem ligger det en mulighet for at beredskapsfilen kan bli bedre kjent og lettere tilgjengelig, og dermed mer brukt.

4.3 Oppbyggingen av dagens beredskapsfil

Før vi går inn på hvordan historiske data bør bevares i det nye BSR-DSP-systemet, skal vi kort gå gjennom oppbyggingen av beredskapsfilen. Gjennomgangen bygger på NOTAT Beredskapsfilen for analyseprosjekter. Nåværende organisasjon, begrensninger og muligheter, EHa/HaR, 29/9-80. Innholdet er listet opp i revidert vedlegg 1 til det, notatet EHa/HaG, 23. januar 1) 1981. Filen består egentlig av to fysiske filer, kalt historieekstrakt og statistikk ekstrakt. Disse to produseres årlig, og brukes sammen med den siste generasjonen av situasjonsfilen. Dermed har en mulighet til å lage situasjonsfiler for ethvert tidspunkt i det tidsrom filene dekker, samt til å rekonstruere individuelle livshistorier.

Når en melding kommer til personregisteret, fører den til at det individ den angår, får forandret en eller flere kjennemerkeverdier. Den tidligere verdien for disse kjennemerkene føres over i historiefilen og det produseres en melding om endringen til statistikkfilen. Den løpende befolkningsstatistikken kjøres på statistikkfilen og noen supplerende filer. Fra statistikk- og historiefilen lages det ekstrakter som består av de kjennemerker en har antatt er av interesse for analyseformål. Når det kommer en korreksjonsmelding til registeret, en melding som retter en tidligere registrert kjennemerkeverdi, produseres samtidig en melding til historiefilen. Neste gang denne verdien forandres gjennom en melding til registeret, inneholder det den korrigerede kjennemerkeverdier, og det er denne som flyttes over i historiefilen. Dermed vil historiefilen inneholde bare riktige kjennemerkeverdier om en ser bort fra de historierecords som

1) Vi har hatt stor nytte av samtaler med Habberstad under utarbeidelsen av denne delen av notatet.

er laget på grunnlag av korreksjonsmeldinger.

Historieekstraktet inneholder:

- ekteskapelig status
- familienummer
- oppgitt flyttedato
- ektefellens fødselsnummer
- tidligere bostedskommune
- bostedskommune
- fra-/tilflyttingsland
- statsborgerskap (fra 1. juli 1973)
- ekt.status dato (fra 1. september 1980)

Statistikketraktet inneholder:

- levende-/dødfødsel
- fødsel i/utenfor ekteskap
- fødselens multiplisitet
- enslig/familieflytting
- mors f.n. (bare dersom barnet er dødfødt)

I tillegg til kjennemerkeverdiene er også registreringsdato for endringer tatt med. Kjennemerkekoder og filredigeringer er beskrevet i AFd/VFr, 23.1.81 som er en revidert utgave av AAH/eh, 24.9.73 og AEd/bft, 13.2.78. Merk at Historieekstraktet er slik at kjennemerkeverdien er gyldig inntil registreringsdato, mens registreringsdato på Statistikketraktet viser den dato kjennemerkeverdien gjelder fra.

4.4 Oppbyggingen under et nytt produksjonssystem

Filene er organisert sekvensielt. Situasjonsfilen er sortert på fødselsnummer, de andre på fødselsnummer x registreringsdato x meldingstype. Meldinger fra noen årganger er sortert sammen for å spare lagringsplass. Om dette er den mest hensiktsmessige sortering også når ajourholdet er kommet over på IBM, er uklart. De som har kjennskap til planene for DIPS og de tekniske egenskaper i ISDS bør uttale seg om denne organiseringen. Det er godt mulig at det innenfor det nye systemet vil være hen-

siktsmessig å sortere etter meldingstype, på samme måte som de ulike meldingsfilene i figur 1. Til hvert meldingstyperegister i figur 1 kan en tenke seg en tilsvarende historiefil som lagrer nødvendige opplysninger om de endringer som skjer, og som brukt sammen med BSR skulle dekke den funksjon som beredskapsfilen har nå. Den kvinnefil som er under planlegging i Sosiodemografisk forskningsgruppe, og som inneholder alle fødsler etter 1953 sortert etter moras fødselsnummer, er eksempel på en slik historiefil som baserer seg på bare en meldingstype. Den vil kunne bli en viktig del av den nye beredskapsfilen.

Vi går så tilbake til oppbevaringen av historiske data i DSP-BSR-systemet. Historiske opplysninger blir, med unntak for navnehistorie, ikke lastet inn i DSP-basen. Så lenge det er plass vil det bli bevart en god del historie i basen, for "forhold" vil alt bli bevart (ThG, 28/8-84). På starttidspunktet er det relativt god plass i basen, som tida går blir en stadig økende del belagt, og historiedata må vike plass. Dette, sammen med kostnadsstrukturen, gjør at vi ikke kan regne med at DSP-basen kan ta over noen av de funksjonene beredskapsfilen idag skal ivareta. Derimot vil vi i meldingsgangen fra DSP til BSR få med de opplysningene som trengs for statistikk og analyse. Derfor synes det naturlig at disse meldingene også blir brukt i ajourholdet av beredskapsfilen siden de skal brukes til å oppdatere de ulike meldingsfilene. Det er viktig at opplegget av beredskapsfilen under det nye systemet sikrer like god datakvalitet som dagens system med bruk av historieekstraktet gir.

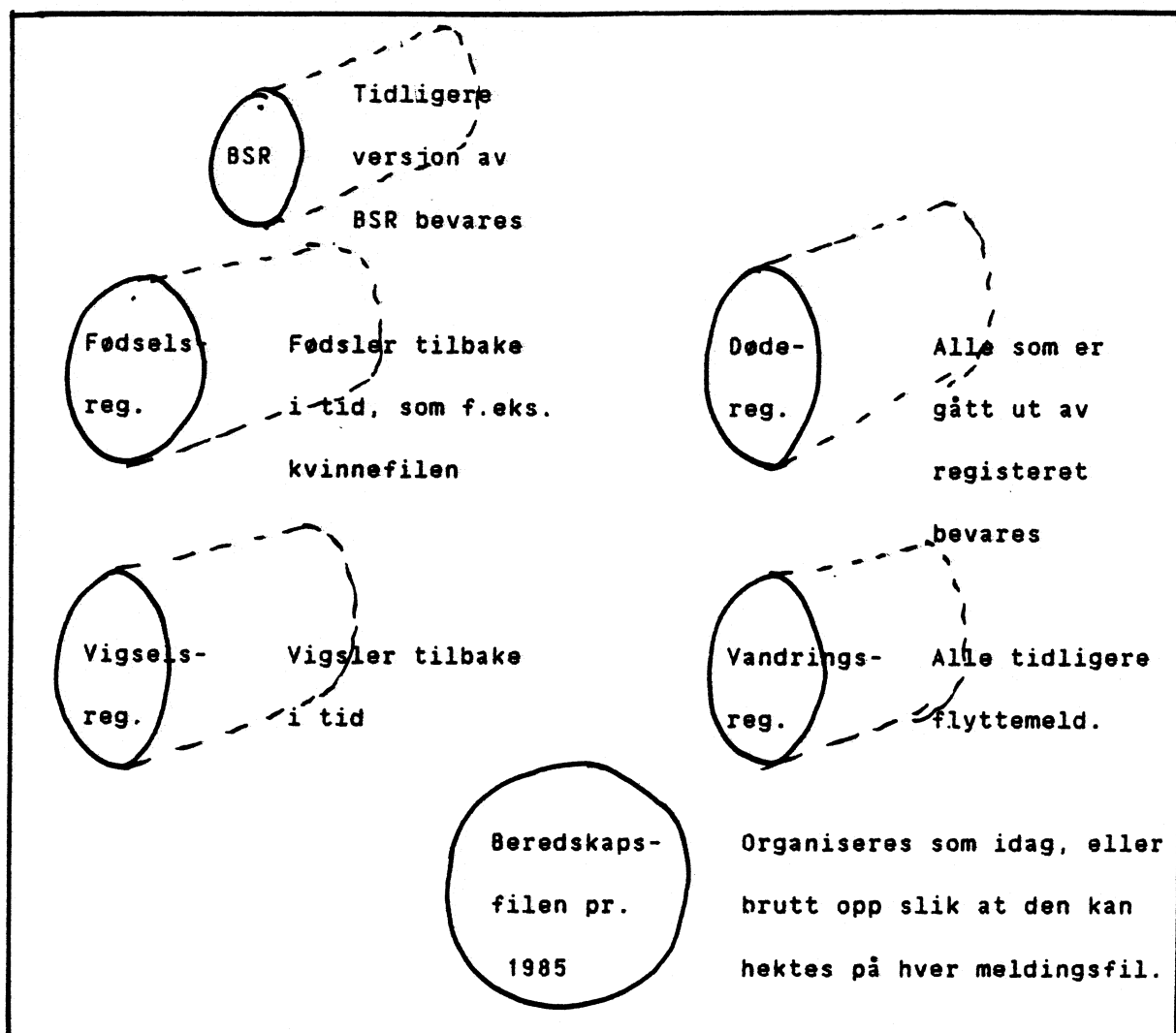
I samtaler med 12. kontor og i møtereferat EJM/Rac, 4.3.85 er det kommet fram en viss usikkerhet om hvor mange utgaver av BSR som skal bevares, hvor lenge journalfilene skal lagres og om meldingsfilene bare skal inneholde noen få års meldinger. I starten vil alle versjoner og årganger av journalfil, meldingsfiler og BSR bli lagret. Sosiodemografisk forskningsgruppe og 12. kontor må sammen med eventuelle andre interesserte parter planlegge hvordan data mest hensiktsmessig skal lagres i framtida.

Dette også for å begrense omfanget av lagrede data og hindre inkonsistens i det lagrede materialet.

Dersom Beredskapsfilen blir en integrert del av BSR (på samme måte som meldingsfilene), kan v oppnå en slik rasjnell lagring. Dessuten vil hensynet til beredskapen for tilbakegående tall kunne ivaretas bedre, og det vil også være et system som gjør det mulig å bruke disse dataene. Den delen av figur 1 som angår BSR gjentas derfor her, med beredskapsfilen som et tillegg.

Figur 2. Befolkningsstatistikksystem.

System for tverrsnittstatistikk (aktuell og tilbakegående), og for livsløpsstatistikk på individnivå



I tilknytning til figur 2 ligger det også nær å tenke seg at de

Øvrige statistikkområdene som inngår i DIPS bruker rutiner som bevarer tidligere kjennemerkeverdier hver gang det kommer en ny verdi. På den måten etableres livsløpshistorier for inntekt, sysselsetting, utdanning osv. som kan gi grunnlag for analyser av utviklingsbaner innenfor hvert felt og for analyser av hvordan endringene henger sammen mellom de enkelte statistikkområdene.

4.5 Organisering av arbeidet

Til nå er det Driftskontoret som har produsert de årlige oppdateringene av Beredskapsfilen, etter Systemkontorets opplegg. Når nå grunnlaget for dette opphører, og det samtidig oppstår et nytt grunnlag i BSR, kan det synes rasjonelt å legge ansvaret for beredskapsfilen samme sted som ansvaret for BSR, nemlig til 12. kontor. Dette forutsetter selvfølgelig at 12. kontor får de ressuser denne tilleggsoppgaven krever. Det løpende driftsarbeid bør ikke bli særlig stort, men det kan nok være noe arbeid forbundet med omorganisering av den gamle beredskapsfilen. Konverteringen til IBM av beredskapsfilen bør skje så tidlig at det ikke blir problemer med kontinuiteten. Opplegget av ny beredskapsfil bør Sosiodemografisk forskningsgruppe være med på sammen med Folketellingskontoret og 12. kontor, slik det er skissert i Ansvarskartet for DIPS - Arbeid mot milepelene 3-4, 9-10 og 12 (aktivitet 11-13) Ski, 7/2-85. Alle disse tre kontorer/grupper er interessert i at arbeidet blir gjort, og må sammen kunne finne fram til en fordeling av ansvar og arbeidsoppgaver som sikrer at tilbakegående data fortsatt blir tilgjengelige. I prosjektet DIPS har en allerede det organ som har til oppgave å koordinere slike samarbeidsprosjekter.

REFERANSER

liste over notater det er referert til i dette notatet. For lettere å finne igjen et notat, er referansene, slik de er henvist i teksten, ordnet alfabetisk. Dewnotatene som er nevnt her, er samlet i en egen mappe hos Østby.

- 1) AAll/eh, 24/9-73. Omfanget av og kjennemerker i Beredskapsfilen. Revidert i 2)
- 2) AFd/bft, 12.2.78. Omfanget av og kjennemerker i beredskapsfilen for analysefilen. Revidert i 3)
- 3) AFd/VFr, 23.1.81. Kjennemerkekoder og fileredigering i beredskapsfilen. Sammen med 7 og 8 beskrivelse av beredskapsfilen opp til 1985.
- 4) Alstad, Johan 1982. Oppfølging av Flyttemotivundersøkelsen 1972. En studie av tilbakeflytting. Rapport 82/15.
- 5) EAu/BjF, 5/2-85. Kopling av registre uten bruk av kjent identifikasjon. Utkast. Kryptering i forbindelse med DIPS.
- 6) EAu/JKT/HaR, 19/2-85. Styringsgruppe for datasystemer for integrert personstatistikk (DIPS)
- 7) EHa/HaG 23. januar 1981 Kjennemerkeliste for "Beredskapsfile persondata. Revidert vedlegg 1 til 8)
- 8) EHa/HaR, 29/9-80 Notat Beredskapsfilen for analyseprosjekter. Nåværende organisasjon, begrensninger og muligheter.
- 9) EJM/Rac, Is. januar, februar og mars 1985. Innhold på melding fra Personregisteret til 12. kontor. En serie notater som beskriver hvilke kjennemerker som skal inn i

BSR, og hvordan de skal oppdateres.

- 10) EJM/Rac 4/3-85. Referat fra møte om omleggingen av befolkningsstatistikken, 18. februar 1985, Kongsvinger.
- 11) Filnummerkatalog for Dataelementer i DSP's database. SDS. Oppdatert 20.06.84.
- 12) Forskrifter om føringen og ordningen av folkeregistrene fastsatt av Statistisk Sentralbyrå den 26. november 1979. Utgitt av Sentralkontoret for Folkeregistrering.
- 13) IN 82/27 Datatilgjengelighet for forskningsformål. Statistisk Sentralbyrås langtidsprogram for 1980-årene. Bakgrunnsnotat nr. 6 Skrevet av Per Sevaldson.
- 14) JaM/GHu, 10/8-76 Spesifikasjonsmuligheter i Personregisteret.
- 15) JBy/KWa, 6.12.84. Notat Omlegging av befolkningsstatistikken til IBM. Prinsipløsning.
- 16) JKT/SKi/HDr/11/12-84 Prosjektskriv: "Datasytem for intergrert personstatistikk (DIPS) (SENTRALT STATISTISK BEFOLKNINGSARKIV)" (Del av prosjektet Utredning av opplegget for Folke- og bolig telling 1990).
- 17) KKv/SMS 3/12-84. Utkast Datasytem for integrert personstatistikk (DIPS). Skisse av systemløsning.
- 18) KKv, 19/3-85. Rutine for oversetting av fødselsnumre til et kryptert nummer via et tilfeldig rekkefølgenummer. Den praktiske løsning av problemene i notat 5).
- 19) Lov om folkeregistrering av 16. januar 1970. Utgitt av Sentralkontoret for folkeregistrering.
- 20) LØs/SIW, 11.1.83 Databaseorganiseringen av Det sentrale personregister. Notat fra Sosiodemografisk forskningsgruppe til Prosjektgruppen for omlegging av DSP til database.
- 21) Milepelplan for totalbasen - DSP prosjekt. SDS. Oppdatert

25.02.85.

- 22) Milepelplan DIPS. Ski 7/2-85, revidert 15/3-85.
- 23) Mønnesland, Brunborg og Selmer 1982: Inngåelse og oppløsning av ekteskap etter alder og varighet. Rapport 82/2.
- 24) NSH/VFr, 4/2-82 Forprosjektnotat fra prosjektgruppa for omlegging av Det sentrale personregister til database.
- 25) PSe/MFo, 17/8-76 Bruk av registerdata til sosiodemografisk forskning.
- 26) Ski/KWa, 5.8.83 Prosjektskriv: Omlegging av den løpende befolkningsstatistikken til IBM.
- 27) SKi/Rac 17.9.84 Referat fra møte om omlegging av den løpende befolkningsstatistikken til IBM 30. august 1984.
- 28) ThG, 28/8-84 Prioriteringen av historiske opplysninger i DSP's database.
Om hvilke tilbakegående opplysninger basen vil lagre.
- 29) WJo/JKT/HDr/15/8-83 Referat fra møte om databeredskap for persondata.
- 30) WJo/JKT/KKv/LHå, 20/7-83. Første utkast. Databeredskap for persondata. Rapport fra forprosjekt.

Notater kommet til etter at notat var ferdigskrevet:

- a) EHA- 9/5-85. Prinsippskisse for uttak fra Det sentrale personregister (DSP) og eksterne registre og produksjon av statistikk.
- b) Halvard Skiri: DIPS - et redskap i arbeidet med personstatistikken. Byråpraten 2-1985, side 25-27.