

Interne notater

STATISTISK SENTRALBYRÅ

INO 86/28

26. mai 1986

DATASYSTEM FOR INTEGRERT PERSONSTATISTIKK (DIPS)

av
Halvard Skiri

INNHold

	Side
1. INNLEDNING	1
1.1. Prosjektet DIPS-en del av utredningen om FOB 90	1
1.2. Bakgrunn for prosjektet - prosjektets gang	1
1.3. DIPS - også et generelt programsystem	3
1.4. Forventet prosjektresultat	3
1.5. Situasjonen for prosjektet i dag	4
1.6. Prosjektorganisering - bemanning	5
2. DATASYSTEMET DIPS	7
2.1. Skisse av systemløsning	7
2.1.1. Bakgrunn	7
2.1.2. Produksjonssystemer generelt	7
2.1.3. Lagringsstruktur	8
2.1.4. Metadatabasesystemet (ISDS)	9
2.1.5. Sentral datakjerne	10
2.1.6. Interaktivt system for styring av produksjons- systemer etc.	11
2.2. Personvern - kryptering	11
2.2.1. Kryptering	11
2.2.2. Omfanget av bruken av kryptering	12
2.2.3. Datatilsynet - krypteringsnøkkel	13
2.3. Personregisterbasen (DSP) og Befolkningsstatistikksystemet (BSS)	13
2.4. Rasjonalisering ved DIPS	14
2.5. Sammendrag om datasytemet DIPS	15
3. REGISTERDATA TIL FOB 90	16
3.1. Identifikasjonskjennemerker	16
3.2. Geografisk kjennemerker knyttet til bosted	16
3.3. Kjennemerker knyttet til personer	16
3.3.1. Demografiske kjennemerker	16
3.3.2. Utdanning/skolegang	17
3.3.3. Yrkesaktivitet, arbeidstid mv.	17
3.4. Kjennemerker knyttet til familier	22
3.5. Kjennemerker knyttet til ektepar/samboerpar	23
3.6. Kjennemerker knyttet til husholdninger	23
3.7. Kjennemerker knyttet til boliger	24
3.8. Oversikt over aktuelle kjennemerker	24
4. KONKLUSJONER	28
Litteratur	29
Vedlegg	
1. Prinsippskisse for DIPS	31
2. Omfanget av bruken av kryptering	33
3. Befolkningsstatistikksystemet (BSS) og forholdet til DIPS og det sentrale personregisteret (DSP)	35

1. INNLEDNING

1.1. Prosjektet DIPS - en del av utredningen om FOB 90

DIPS står for Datasystem for integrert personstatistikk som er navnet på delprosjekt nr. 6 av prosjektet "Utredning av opplegget for Folke- og bolig telling 1990" i Statistisk Sentralbyrå (SSB).

Denne utredningen består av en generell og en spesiell del. Den generelle delen skal utrede:

- SSB's adgang til adm. registre (delprosjekt nr. 1).
- SSB's lagring av data (del 2).
- Regler for kopling og gjenbruk av data i SSB (del 3).

Den spesielle delen skal utrede alternative planer for FOB 90:

- Et opplegg omtrent som i 1980, dvs. en kombinasjon av spørreskjema og registre (del 4).
- En telling basert på opplysninger fra registre, der opplegget lages spesielt for FOB 90 (del 5).
- Oppbygging av en beredskap i SSB for å utarbeide folketellingsliknende statistikk ved behov, basert på registre (del 6, DIPS).

1.2. Bakgrunn for prosjektet - prosjektets gang

Prosjektet ble skissert første gang i 1982. I januar 1983 ble det vedtatt et forprosjekt med navnet "Databeredskap for persondata". Forprosjektet skulle skissere mulighetene og behovet for å opprette en database med plass til alle data som inngår i SSB's personstatistikker og som gir mulighet for uttak ved hjelp av standard tabellprogram. En skulle også si noe om kostnader, rasjonaliseringseffekter, og organisering av arbeidet.

Et utkast til rapport fra forprosjektet forelå i juli 1983 (WJo/-JKT/KKv/LHå, 20/7-83). 10. og 31. august 1983 ble det holdt to møter der det ble vedtatt at prosjektet skulle dekke følgende statistikkområder: Folketelling, regionalstatistikk, sysselsettingsstatistikk, og inntekts- og skattestatistikk. Utdanningsstatistikken skulle være med i den grad en kunne tilpasse seg reglene for utdanningsdataene. I en seinere fase kunne det bli aktuelt "å bygge ut beredskapen til også å omfatte data som det pr. i dag ikke er like stort behov for å ta med, f.eks., befolkningsstatistikken og Sosiodemografisk forskningsgruppes databehov".

Første fase i videreføringen skulle konkretisere innholdet i beredskapen ved filer og kjennemerker, for deretter å etablere en beredskap som dekket disse fire områdene.

Det ble ikke laget noe nytt prosjektskriv for prosjektet, ettersom møtene hadde vedtatt at det skulle fortsette. Utover høsten ble det arbeidet med prosjektet langs flere linjer. Den ene var personvernsspørsmål (Aurbakkens notat om "statistikknummer", møte med Datatilsynet). Den andre var kartlegging av metadata i beredskapen, med behovsanalyse og kravspesifikasjon. For dette formålet ble det laget et spørreskjema fra Systemkontoret som de aktuelle fagkontorene fylte ut. Bearbeidingen av disse skjemaene og forsøk på tilpasninger, konsentrerte seg i første omgang om inntekts- og formuesstatistikk og utdanningstatistikk.

På et møte 30. august 1984 ble det vedtatt at også den løpende befolkningsstatistikken i prinsippet skal betjenes av DIPS idet det tas sikte på å etablere en statistikkversjon av DSP, ved 12. kontor, Befolkningsstatistikksystemet (BSS). (Denne statistikkversjonen ble en tid omtalt som "Befolkningsstatistikkregisteret" (BSR)). Det ble klart at ettersom DSP-basen - som ble satt i drift ved Statens Datasentral (SDS) i juni 1985 - i hovedsak skal tjene administrative formål, vil BSS erstatte det tidligere sentrale personregisteret i SSB som datakilder for løpende befolkningsstatistikk. Data fra BSS ville derfor stå sentralt i DIPS-prosjektet.

Gjennom prosjektskrivet for Utredninga er databeredskapen for persondata ført videre som et delprosjekt under Utredning av opplegget for Folke- og bolig telling 1990 (nevnt som Personstatistikkberedskap), prosjektskriv datert 27. juli 1984.

Styringsgruppa for Folke- og bolig telling 1990 gikk i møte 26. oktober 1984 inn for å trekke inn i prosjektskrivet et notat fra Systemkontoret om produksjonssystemet for personstatistikk.

Det synes allerede da klart at en tradisjonell databaseløsning ikke ville være fornuftig. Mer aktuelt syntes det å benytte deler av databasesystemer (metadatabaser) kombinert med tradisjonelle metoder som satsvise kjøring på sekvensielle filer.

Et eget prosjektskriv for DIPS, datert 11. desember 1984 ble godkjent av kst.adm. direktør 21. januar 1985. Formålet med prosjektet ble beskrevet slik:

"Formålet med prosjektet blir derfor å planlegge og bygge opp et datasystem (en personstatistikkberedskap) for produksjon av statistikk på data hentet fra to eller flere av følgende statistikkområder:

- folketellinger
- løpende befolkningsstatikk
- sysselsettingsstatikk
- inntekts- og formuesstatikk
- utdanningsstatikk
- regionalstatistikk

Systemet skal inneholde en metadatabase.

Datasystemet må spesielt kunne gi folketellingsliknende statistikk i 1990 i den grad registerdata skal nyttes.

Systemet skal, sammen med de nye rutinene ved fagkontorene, utgjøre hensiktsmessige produksjonsrutinger for hvert enkelt statistikkområde."

1.3. DIPS - også et generelt programsystem

I februar 1985 ble formålet med DIPS utvidet til også å gjelde utvikling av et mer generelt datasystem for statistikkproduksjon der data fra flere statistikkområder må kobles. I tillegg skal det utvikles et programsystem for kryptering ("uleseliggjøring") av fødselsnummer til bruk blant annet i DIPS.

Hovedformålet med programsystemet vil være effektiv kopling og produksjon av statistikk på data fra ulike statistikkområder, samtidig som det tas nødvendige hensyn til datavern. Planene om kryptering av fødselsnummer til K-nummer (et 7-sifret nummer uten opplysninger om personen) ble skissert i samråd med Datatilsynet.

DIPS ble lansert som et tilbud også til 1. kontor og 9. kontor, uten at det nå vil bli tatt stilling til hvilke av disse kontorenes produksjonsrutiner som bør ligge under DIPS. Bakgrunnen for å utvide DIPS til flere statistikkområder var at planene om kryptering av fødselsnummer ventes å fjerne tidligere hindringer for å kople til helse- og sosialdata pga. deres sterke krav til personvern.

1.4. Forventet prosjektresultat

DIPS-prosjektet forventes å resultere i

- et programsystem for effektiv kopling og bearbeiding av registre (fortrinnsvis med kryptert fødselsnummer)
- en bedre databeredskap når det gjelder personstatistikk på grunnlag av data fra flere statistikkområder
- en felles metadatabase for fagområdene i DIPS
- et produksjonssystem for folketellingsliknende statistikk i 1990 i den grad registerdata skal nyttes
- et INO i mars 1986 skal tas med i Utredninga om FOB 1990

Under prosjektframdriften har avgrensinger av DIPS i forhold til de enkelte fagkontorers framtidige produksjonsrutiner vært tolket ulikt og derved medført uklarheter. Dette har særlig vært følbart overfor BSS.

I hvilken grad statistikk innenfor det enkelte fagområde kommer til å bli produsert "ved hjelp av" DIPS er ikke avklart. Men de mange produksjonsrutiner som etter hvert blir lagt til IBM vil bli samordnet med DIPS og metadatabasen i den grad dette synes hensiktsmessig. Resultatet ventes derfor å bli forbedrede produksjonsrutiner og bedre dokumentert data også innenfor hvert enkelt statistikkområde. Det har hele tida vært en forutsetning at opplegging av personstatistikkrutiner til IBM må skje parallelt med prosjektarbeidet.

1.5. Situasjonen for prosjektet i dag

Situasjonen for prosjektet pr. 1.12.85 kan beskrives i forhold til mandatet for Styringsgruppa for DIPS, der målet for programsystemdelen av prosjektet tenkes nådd som beskrevet til venstre nedenfor. En kort situasjonsbeskrivelse er gitt til høyre.

Situasjonsbeskrivelse

- | | |
|--|--|
| 1) Utvikle et programsystem for kryptering av fødselsnr. (krypteringsrutine) | Krypteringsrutine skissert, men ikke programmert |
| 2) Utvikle et programsystem som sørger for effektiv kobling og bearbeiding av registre, fortrinnsvis med kryptert fødselsnr. | Programsystem skissert, utviklingsarbeid konsentrert om befolkningsdata (DSP-data) |

Styringsgruppa skal spesielt sørge for:

- | | |
|--|--|
| 1) at behovet for og innholdet i en sentral datakjerne blir utredet | Noe er gjort, men ingen konklusjon pga. lite spesifiserte planer om koblinger til nå |
| 2) at krypteringsrutinene blir forelagt Datatilsynet med forslag til praktisk ordning av kontrollen av krypteringen | Utført |
| 3) å føre tilsyn med Systemkontorets arbeid med programsystemene | Utføres |
| 4) å holde fagkontorene orientert om de muligheter som ligger i det nye systemet slik at disse blir tatt i bruk i rett tid både til tilfeldige oppdrag, løpende statistikk og folketelling | Utføres. Informasjonsmøter i mars 1985 og oktober 1985 |

Situasjonsbeskrivelse

- | | | |
|----|--|--|
| 5) | å utarbeide langsiktige planer for arkivering og kobling av registre med tankte på, gjennom kryptering, å redusere lagringen av registre med fødselsnummer | Gjenstår idet resultatet av 2) ikke er klart |
| 6) | at hensynet til den løpende personstatistikk blir ivaretatt, spesielt de problemer for feilretting som følger av at dataene ikke kan identifiseres | Utført |
| 7) | å utarbeide kostnadskalkyler, organisasjonsplaner og arbeidsprogram for det videre arbeid | Gjenstår |

1.6. Prosjektorganisering - bemanning

Forprosjektet "Databeredskap for persondata" ble ledet av Willy Jordet fram til han sluttet i SSB i februar 1984. Fra april 1984 overtok Halvard Skiri.

Inntil februar fungerte Styringsgruppa for Folketellinger som styringsgruppe. Som følge av planen om aidentifiørning av data ved hjelp av kryptering av fødselsnummer ble det da utpekt en egen styringsgruppe for DIPS med Johan-Kristian Tønder som formann og Erik Aurbakken, Liv Bjørnland, Sverre Hove og Per Sevaldson som medlemmer.

Samtidig ble det utpekt en prosjektgruppe med Skiri som leder og Knut Kvisla, Sissel Solvin og Torgeir Vik som medlemmer. Det var ellers en forutsetning at representanter for de impliserte fagkontorene og Sosio-demografisk forskningsgruppe, skulle delta i arbeidet i prosjektgruppa etter behov.

Vik deltok i prosjektgruppa fra 2. kvartal som erstatning for Solvin som sluttet i SSB i juni 1985. På grunn av ressursproblemer, bl.a. som følge av skifte av personer, har ingen av medlemmene i prosjektgruppa kunnet bruke tilnærmet hele sin arbeidstid på prosjektet. I samband med uttak av data fra DSP-basen til BSS og gjenbruk av persondata har særlig Eva Jemblie Monssen, Halvor Strømme og Lars Østby deltatt aktivt.

Prinsippansvarskart, milepelplan og aktivitetansvarskart for prosjektet ble godkjent 12. februar 1985. Månedlige rapporter om framdriften i prosjektet er sendt styringsgruppa.

Skiri og Vik besøkte Danmarks Statistik 3.-4. juni 1985 for å studere danske erfaringer fra registerbaserte folketellinger (rapport SRJ/Ski/TVi/HDr, 27/6-85).

Skiri og Østby deltok på kurs i demografisk livsforløpsanalyse i Ystad 5. - 16. august 1985 (rapport (SKi/TNo/E1B, 24.09.85).

Tønder ble 30. august 1985 på møte i Statistiska Centralbyrån i Stockholm informert om svenske erfaringer med kryptering av personnummer (rapport JKT/SRT/Rac, 12.09.85).

2. DATASYSTEMET DIPS

Dette kapitlet omtaler de system- og produksjonstekniske sidene av prosjektet DIPS, foruten personvern og planene om kryptering av fødselsnummer.

2.1. Skisse av systemløsning

2.1.1. Bakgrunn

DIPS trenger et programsystem for effektiv kobling og håndtering av personregistre/datafiler.

Systemkontoret har bl.a. hentet inn opplysninger fra fagkontorene om hvilke kjennemerker som inngår i de forskjellige statistikkene, hvor kjennemerkene kan finnes og hva en ønsker å få ut av koblede data, foretatt beregninger på forbruk av CPU-tid, behov for plass på platelager, kanalbelastning etc.

Et såkalt 4. generasjonsverktøy har også vært prøvd som et mulig brukbart verktøy for å utvikle og drive basen.

Det har etter hvert blitt klart at en tradisjonell databaseløsning ikke vil være fornuftig i et system for produksjon av personstatistikk. Men ved å benytte deler fra databasesystemer (metadatabaser) og kombinere disse med mer tradisjonelle produksjonsmetoder (satsvise kjøring på sekvensielle og indekssekvensielle filer) mener en å kunne tilby brukerne et system som vil bli vesentlig lettere å håndtere både når det gjelder å produsere den løpende statistikken og foreta forskjellige koblinger i forbindelse med ad-hoc-oppdrag. Systemet ventes å gjøre det vesentlig enklere å hente inn kjennemerker fra andre statistikkområder, f.eks. inntektskomponenter som ønskes brukt i mange forskjellige sammenhenger. En gikk derfor tidlig bort fra betegnelsen database i denne sammenhengen. I stedet nyttes betegnelsen datasystem. Datasystemet er skissert i vedlegg 1.

2.1.2. Produksjonssystemer generelt

Det synes hensiktsmessig å produsere den løpende statistikken på hvert fagområde (fagkontorenes sektorstatistikk) ved hjelp av satsvise programmer og med de fleste dataene lagret på sekvensielle (flate) filer.

De kjennemerker som er felles for flere statistikkområder kan likevel lagres i en felles kjerne som blir liggende fast på platelager (disk). Individdataene for de enkelte statistikkområdene vil stort sett ligge på magnetbånd. Men individdata kan lastes inn på platelager og bli liggende der i kortere perioder når kjøring forekommer hyppig.

Dessuten vil en sentral metadatabase (metadata = data som beskriver data) bli opprettet. Den vil bli av svært stor betydning, spesielt ved kobling av kjennemerker fra forskjellige statistikkområder, men også ved utvikling og vedlikehold av de løpende statistikkrutinene på hvert fagområde. Beskrivelsen av dataene som finnes i metadatabasen kan hentes automatisk til ulike produksjonsprogrammer.

Det manuelle arbeidet rundt produksjonsrutinene ventes å bli rasjonalisert vesentlig ved å lage interaktive overbygninger som styrer de enkelte trinnene i produksjonsrutinene. Teknikken er brukt tidligere bl.a. i DIRIGENTEN (SCB) og vår egen TAB68GEN.

2.1.3. Lagringsstruktur

Størsteparten av databehandlingsoppgavene vil kreve gjennomløp av hele filer, enten det gjelder kobling, aggregering eller ekstrahering. Den mest ressursvennlige måten å få adgang til en hel fil på er å lese den sekvensielt (i rekkefølge) og på en slik måte at en får tak i flest mulige poster (records) i én fysisk lesing. Selv om en ved enkelte ekstraheringer ikke hadde trengt å lese hele filen, vil det likevel lønne seg å lese filen sekvensielt i stedet for å ha "nøkkel" på kjønn, og så søke data via denne. En må faktisk ned på en nokså lav prosent av ekstraherte poster før det går raskere å søke data via nøkler. På grunn av at det trolig vil være minst like mange ønskelige kombinasjoner av kjennemerker som nøkler, ville nøkkelfilene ha svulmet opp og tatt mye plass og lagrings-/oppdateringstiden for poster med mange nøkler knyttet til seg, ville blitt uakseptabelt.

På grunn av at statistikkene har forskjellige referansetidspunkter, vil det bli nødvendig å ha flere versjoner av samme datasett tilgjengelig. De samme kjennemerkene for en person må lagres i flere filer fordi de på grunn av ulike referansetidspunkter ikke inneholder samme verdi. Det ser heller ikke ut til å være mange kjennemerker i de enkelte statistikkområders filer som er av interesse for andre fagkontor.

I personstatistikk er dataomfanget så stort, særlig når også historiske data tas i betraktning, at det ikke vil være aktuelt med de nåværende lagringsmedia å ha en så stor datamasse "on line".

Oppdatering av datafilene i personberedskapen vil for noen kjennemerker foregå en gang i året, og for andre en gang i kvartalet. Da vil enten en helt ny datafil bli "innlemmet" i systemet, eller oppdateringer

vil bli gjort fra meldingsfiler. Med andre ord skulle ikke samtidig oppdatering, og oppdatering fra skjerm være et aktuelt problem.

Konklusjonen blir at en er best tjent med å lagre persondata fordelt på flere flate filer.

2.1.4. Metadatabasesystemet (ISDS)

Hjelpemidler for på en enkel måte å gjøre uttak (ekstrahering, kobling, aggregering) tenkes bygd rundt en sentral metadatabase, hvor hvert enkelt statistikkområde har dokumentert sine individdata. Denne metadatabasen vil bli opprettet og styrt ved hjelp av systemet ISDS (Integrert Statistikk Data System), som ble laget i samband med 3. kontors database for kommunaløkonomi.

Metadatabasen og ISDS vil kunne dekke flere behov:

- 1) Dokumentasjon av emner, enheter og kjennemerker på de enkelte statistikkområder.
- 2) Edb-teknisk definisjon av kjennemerker, knyttet til filer og poster.
- 3) Generering av datadefinisjoner for variable til bruk i andre systemer, f.eks. EASYTRIEVE, SAS, PRINTERTAB, TAB68. (Dette vil medføre at "recordlayout" beskrives bare ett sted, nemlig i metadatabasen).
- 4) Være et ekstraherings-, koblings-, aggregeringsverktøy hvor brukeren kan sitte ved skjermen og "plukke ut" de variable han/hun er interessert i - gjerne på tvers av statistikkområder - og få generert enten en fil for videre bearbeiding av andre systemer f.eks. SAS, eller ta ut dataene direkte som en tabell. Slike tabeller vil bli tekstet ved hjelp av kataloger og navn på variable.
- 5) Inneholde faste standarder og kataloger, f.eks. over nummer for kommuner og handelsdistriktinndelinger. Disse kan nyttes i alle sammenhenger og behøver derfor ikke dupliseres for hvert statistikkområde.

Forutsetningen for å benytte ISDS er at alle kjennemerker som skal brukes og filene de er lagret på, må være definert i metadatabasen. Mulig lagringsstruktur vil være sekvensielle filer og VSAM-filer. Lagringsmedium kan være platelager eller magnetbånd.

Brukere behøver i prinsippet ikke kjenne til lagringsstruktur eller lagringsmedium for de forskjellige filene. For å kunne spesifisere uttak, behøver brukeren ikke engang å vite navnet på filen(e) der kjennemerkene som skal behandles ligger.

Det gjenstår noe utbyggingsarbeid før ISDS dekker behovene til DIPS. Den vesentligste utvidelsen består i å sørge for at en vil kunne bruke ekstraheringsprogrammet på vanlige sekvensielle filer. Idag kan det kun brukes på indekssekvensielle filer (VSAM/KSDS - Virtual Storage Access Method/Key Sequenced Data Set).

2.1.5. Sentral datakjerne

Det tas sikte på å etablere en felles kjerne på platelager for data som er felles for flere statistikkområder. Kjernen kan bestå av flere delfiler. Kjennemerker som er aktuelle for en sentral kjerne er:

- Demografiske data (alder, kjønn, ekteskadelig status, statsborgerskap, bostedskommune, fødeland, familienummer).
- Adresser (arbeidsstedskommune)
- Utdanningsopplysninger (høyeste fullførte utdanning, klassetrinn og dato for avsluttet utdanning)
- Sysselsettingsdata (viktigste næring, arbeidstid)
- Inntektsdata (pensjonsgivende og disponibel inntekt)

I tillegg kommer datoer for kjennemerkene.

Størrelsen på kjernen vil være avhengig av hvilke data som det vil bli aktuelt å nytte i samme statistikk rutine og hyppigheten av slike kjøringer. Anmeldte behov til nå synes å peke i retning av at en kjerne vil bli forholdsvis liten.

En koblingsberedskap uten en sentral kjerne vil være et aktuelt alternativ dersom omfanget av koblinger mellom fagområdene blir lite. En tredje mulighet vil være et system med små kjerner (koblede filer) for de to og to evt. tre og tre fagområder som trenger koblinger hyppig.

2.1.6. Interaktivt system for styring av produksjonssystemer etc.

Det vil bli bygd ut et menystyrt system som styrer oppstartingen og utføringen av de enkelte programmene i en produksjonsprosess. Brukeren (fagkontoret) trenger ikke lage eller endre på kontrolloppsett i et slik system. Det er bare nødvendig å spesifisere hvilke data en skal bruke. Hvor dataene befinner seg rent fysisk er uvesentlig. På grunnlag av de opplysningene om dataene som blir gitt av brukeren, vil det automatisk bli generert nødvendige kontrollkort for selve kjøringen.

Ekstraheringsprogrammet som henter data fra ulike statistikkområder kan trolig styres fra samme overbygning (menyssystem).

Et slik system kan bygges opp med verktøy som SSB har allerede, ISPF - Interactive System Productivity Facility. Systemet kan bygges opp trinn for trinn, og det er egentlig ingen grense for hvor mange statistikkområder som kan legges inn. SSB har erfaring med å lage slike systemer (TAB68GEN, DOKS, FOPU), og den sier at det er relativt enkelt og raskt å bygge opp menystyrte systemer.

2.2. Personvern - kryptering

2.2.1. Kryptering

Byråets adgang til å koble identifiserbare data er regulert av en konsesjon gitt av Datatilsynet. Konsesjonen gir anledning til slik kobling når det er "strengt nødvendig". Dette fører til at kobling må gjentas hver gang det er behov for det. Resultatet blir en viss sløsing med ressursene og redusert databeredskap, selv om utviklingen mot billigere og raskere datamaskiner gjør at dette betyr stadig mindre. Utviklingen er likevel betenkelig med hensyn til datavern ved at misbruk av data ved kobling blir enklere.

På bakgrunn av utviklingen beskrevet foran har SSB arbeidet med en løsning for kobling av persondata som tillater økt databeredskap med samme datagrunnlag som tidligere, men der en samtidig kan oppnå et bedre vern mot misbruk av data som kobles (EAu/BjF, 5/2-85).

Løsningen går ut på at identifiserbare persondata gjøres uidentifiserbare ved at identifikasjonen (fødselsnummeret) krypteres, dvs. gjøres uforståelig. Slike data kan ikke gjøres identifiserbare igjen uten ved hjelp av en nøkkel som oppbevares av en kontrollør.

Det er blitt beskrevet nokså detaljert hvordan slike edb-rutiner for kryptering av fødselsnummer kan lages (KKV, 10/4-85). Et system med flere nummerserier er antydnet. Krypteringen tenkes utført i to trinn, først til et tilfeldig nummer på 7-siffer (T-nr) og deretter til en av flere mulige K-nr.-serier.

2.2.2. Omfanget av bruken av kryptering

En krypteringsløsning vil kunne være nyttig ikke bare ved kobling av data i SSB, men også ved utlevering av data fra SSB og ved levering av følsomme data til SSB.

Internt i SSB kan det være aktuelt med kryptering av data også i de tilfelle dataene ikke skal inngå i en kobling. Kryptering kan likevel ikke foretas før dataene er kontrollert og eventuelle feil rettet.

Omfanget av bruken av kryptering vil bl.a. være avhengig av hvor ressurskrevende prosedyrene blir, både mht. til tid og edb-kostnader og av hvor følsomme data det er tale om. En svensk metode som er blitt testet viste seg å bli for omstendelig og kostbar. Det gjelder derfor å forsøke å komme fram til relativt enkle rutiner som sammen med Top Secret-systemet kan ivareta behovet for personvern. Det kan ellers bli aktuelt å innskrenke kryptering av fødselsnr. til tilfeller med følsomme data, (jfr. dansk strategi og praksis), særlig dersom kostnadene skulle bli betydelige eller dersom DIPS skulle bli etablert som en koblingsberedskap uten permanent koblet kjerne.

Det er klart at kryptering ikke er en fullstendig anonymisering. Dersom et datamateriale inneholder tilstrekkelig mange og detaljerte kjennemerker - særlig geografiske - vil en person med gode detaljkunnskaper kunne identifisere noen personer. Datatilsynet legger mindre vekt på akkurat dette forholdet, idet det ikke vil kunne medvirke til å avdekke sammenhengen mellom F-nr. og K-nr. for hele K-nr.serien.

Omfanget av bruken av kryptering i tilknytning til DIPS kan diskuteres. Det er på det rene at en tidlig aidentifisering (krypering) av dataene), før kontroller og nødvendig oppretting etter søking i primærmaterialet er foretatt, ikke er akseptabelt. En oppbevaringstid med fødselsnr. som identifikasjon på iallfall 1-2 år anses for nødvendig. Videre bør en koblet, kryptert datakjerne i DIPS ikke være for stor, idet bare data som trengs koblet relativt hyppig bør finnes der.

Virkningene for statistikkproduksjonen ved bruk av kryptering ved lagring av data har vært drøftet. Konklusjonen ble at det kan settes i gang tiltak alt etter hvor sterkt en vil sikre seg mot at noen i SSB skal kunne avdekke og misbruke krypterte data. Slike tiltak vil bety økte kostnader og gjøre SSB's arbeid mindre effektivt, men tiltakene synes gjennomførbare og til å leve med for statistikkprodusenten, brukeren og trolig også for Datatilsynet.

2.2.3. Datatilsynet - krypteringsnøkkel

Opplegget med kryptering av fødselsnummer er drøftet med Datatilsynet, som har stilt seg positivt til saken. Det arbeides med nødvendige endringer i SSB's konsesjon, uten at noe endelig vedtak er fattet. I øyeblikket ligger det an til at et kryptert register ikke vil bli betraktet som avidentifisert, dvs. at det fortsatt vil bli definert som et personregister. Likevel tar Datatilsynet sikte på å finne fram til en formulering der det godtas at krypterte data lagres i SSB, både koblet og ukoblet.

Spørsmålet om hvem som oppbevarer nøkkelen til kryptering er heller ikke endelig klarlagt, men trolig vil den bli oppbevart i SSB.

2.3. Personregisterbasen (DSP) og Befolkningsstatistikksystemet (BSS)

I et system for produksjon av personstatistikk vil demografiske data nødvendigvis måtte ha en sentral posisjon, nærmest utgjøre en slags basis eller totalramme. Innenfor DIPS-prosjektet har demografiske opplysninger til nå spilt en svært sentral rolle, også pga. to andre forhold:

- 1) Omleggingen av Det sentrale personregisteret til database (DSP)
- 2) Omleggingen av den løpende befolkningsstatistikken til IBM.

I tilknytning til disse prosjektene har også problemstillinger omkring gjenbruk av persondata vært tatt opp, jfr. INO 85/28 av Lars Østby.

Oppbyggingen av DSP-basen har vært fulgt nøye av prosjektgruppa for DIPS med sikte på å oppnå løsninger som vil sikre bedre tilgjengelighet til utnytting av data for statistikkformål. Endrede forutsetninger mht. utnytting av DSP-basens basestruktur har resultert i at ulike løsninger for datauttak og system for produksjon av løpende befolkningsstatistikk har vært skissert. Som en følge av dette har det vært problematisk å klarlegge forholdet mellom DIPS og produksjonen av løpende

befolkningsstatistikk ved 12. kontor. På grunn av det sterke behovet for å få i gang et nytt produksjonssystem for befolkningsstatistikk vil arbeidet med dette systemet, Befolkningsstatistikksystemet (BSS) bli prioritert fortsatt.

BSS vil i første rekke bestå av et situasjonsregister og flere meldingsregistre, ett for hvert av de demografiske faktorene fødsel, dødsfall, flytting osv. (se vedlegg 3). I tillegg kommer filer med historiske data, livsløpsdata (beredskapsfilen) mv. BSS vil hente data fra situasjonsuttak fra DSP-basen og fra DSP's journalfil. Sistnevnte fil er en samling av informasjon produsert i samband med oppdatering av DSP-basen via skjerm.

Ifølge systemløsningen for BSS vil data fra journalfilen være grunnlaget for meldingsfilene. Situasjonsfilene vil delvis bli basert på situasjonsuttak fra DSP-basen og delvis på oppdatering ved hjelp av endringsdata fra journalfilen. Hvor ofte situasjonsuttak fra DSP-basen bør nyttes må avklares med det første. Av betydning her er bl.a. erfaring mht. edb-kostnader og mulighetene for samordning med oppdrag for andre formål enn BSS. Dette samordningsansvaret vil bli tillagt 12. kontor. Ressursbehovet (systemressurser mv.) ved en samordning skal utredes med det aller første. Når det gjelder livsløpsdata, medregnet beredskapsfilen for personregisterdata, er framtidig ansvars- og ressursforhold ikke endelig avklart.

Det tas inntil videre ikke sikte på å oppdatere situasjonsregisteret i BSS mht. familiesammensetning. Av hensyn til system- og edb-ressurser antas uttak fra DSP-basen å være mest formålstjenelig.

Det foreligger detaljert dokumentasjon av journalfilen fra DSP-basen. Testdata er til kontroll. All dokumentasjon av DSP-data som planlegges tatt inn i BSS, er lagt inn i metadatabasen (ISDS).

2.4. Rasjonalisering ved DIPS

Kostnadene ved å etablere og holde i drift en databeredskap som beskrevet i dette kapitlet vil være avhengig av flere forhold som fortsatt er uavklart. Stikkord i denne sammenhengen er oppdateringshyppighet for dataene, størrelsen på en sentral kjerne, omfanget av historiske data mv.

Uavhengig av disse forholdene vil DIPS representere et produksjonssystem for personstatistikk, ikke bare for de tilfelle data hentes fra ulike statistikkområder, men også for fagkontorenes egne løpende statistikker. Det menystyrte systemet vil bety forenklete og mer tidsmessige produksjonsrutiner. Metadatabasesystemet vil komme til å spille en viktig rolle bl.a. mht. større muligheter for grupperinger, og avledninger, beskrivelse av data og overganger til andre 4. generasjonsspråk. Den sentrale metadatabasen vil også føre til mer rasjonell dokumentasjon

2.5. Sammen drag om datasystemet DIPS

For kobling og håndtering av personregistre og datafiler vil det bli satset på at programsystemet består av en sentral metadatabase kombinert med satsvise kjøringar på sekvensielle og indekssekvensielle filer. Det er derfor snakk om at datasystem, ikke en database.

Løpende statistikk på hvert fagområde (fagkontor) forutsettes produsert ved hjelp av satsvise programmer der de fleste data er lagret på sekvensielle filer.

Behovet for en sentral kjerne er ikke endelig avklart, vesentlig på grunn av få spesifiserte behov for koblinger av data fra ulike områder. En koblingsberedskap uten sentral kjerne er fortsatt en mulig løsning. En eventuell kjerne synes å ville bli heller liten.

Et menystyrt system som styrer oppstarting og utføring av de enkelte programmene i en produksjonsprosess vil kunne bygges opp nokså enkelt og raskt. Overføringen av de enkelte rutinene fra Honeywell til IBM kan foregå relativt uavhengig av arbeidet med databeredskapen.

Det lanseres en metode for kobling av persondata som tillater økt databeredskap samtidig som et bedre vern mot misbruk av data oppnås. Løsningen går ut på at persondata identifisert med fødselsnummer gjøres uidentifiserbare ved at fødselsnummeret krypteres, dvs. gjøres uforståelig. Mulige edb-rutiner for kryptering er beskrevet i detalj. Planene er drøftet med Datatilsynet, men det er foreløpig ikke full enighet om et opplegg som både Datatilsynet og SSB kan godta.

Nøkkelen til kryptering vil trolig bli oppbevart i SSB.

Oppbyggingen av datasiden ved DIPS-systemet har til nå vært konsentrert om data fra DSP-basen og Befolkningsstatistikksystemet (BSS) ved 12. kontor. Her har behovet for nye produksjonsrutiner vært størst. Alle data i BSS hentet fra DSP er dokumentert i metadatabasen (ISDS).

Systemløsningen for BSS er i store trekk klar. Situasjonsfiler vil delvis bli basert på situasjonsuttak fra DSP-basen og delvis på oppdatering ved hjelp av endringsdata fra journalfilen.

3. REGISTERDATA TIL FOB 90

Kapitlet beskriver hvilke data til en folketelling som en kan regne med vil være tilgjengelige i registre i 1990, på bakgrunn av situasjonen i november 1985 og opplysninger om utbyggingsplaner.

Det tas utgangspunkt i kjennemerkene i FOB 80 slik disse er gruppert i dokumentasjonen av tellingen, RAPP 85/15.

3.1. Identifikasjonskjennemerker

Fra den sentrale personregisterbasen (DSP-basen) hentes fødselsnummer og familienummer. Herfra eller fra Befolkningsstatistikksystemet (BSS) ved 12. kontor kan også hentes et planlagt sambofelleskapsnummer for samboende par (ugifte eller personer ikke gift med hverandre) som har felles barn. Sambofelleskapsnummer er forutsatt brukt til en utvidelse av den løpende familiestatistikken. Etter behov kan også avledes ev. nummer for ektepar eller markering av personer som inngår i et par (jfr. FOB 80).

3.2. Geografiske kjennemerker knyttet til bosted

Kommune/fylke hentes fra kommunenummer i DSP-basen. Herfra avledes også handelsdistrikt og kommunetype. Grunnkrets/delområde hentes fra numerisk adresse (adressekode) i DSP-basen etter kobling mot adressedelen i GAB-registret. Samtidig fås menighet, skolekrets, valgkrets og om bostedet ligger i tettbygd eller spredtbygd strøk, forutsatt at adressedelen er tilfredsstillende oppdatert. Inntil nå (1985) har koblinger mellom adressedelen og Skattedirektørens magnetbåndregister vært av akseptabel kvalitet bare for en del av kommunene som er koblet. Alt tyder likevel på at kvaliteten i 1990 vil være tilfredsstillende selv om en nok må regne med uoppgitt krets for inntil 1-2 prosent av de bosatte før en ev. ekstra ajourføring av adressedelen. Tettstedkode, som angir hvilket tettsted adressen tilhører, forutsettes skaffet som ledd i arbeidet med avgrensning av tettsteder.

3.3. Kjennemerker knyttet til personer

3.3.1. Demografiske kjennemerker

Kjønn og alder fås fra fødselsnummeret.

Ekteskapelig status og statsborgerskap fås fra DSP-basen

Fødeland vil i 1990 - for de fleste personer - trolig kunne hentes fra 12. kontors befolkningsstatistikksystem (BSS). For personer inn-

vandret etter 1970 må fødeland trolig tilnærmes med fraflyttingsland eller tidligere statsborgerskap. Jfr. godkjent prosjektskriv fra 12. kontor. Det tas sikte på å få inn et spørsmål om fødeland på blanketten nyttet ved innvandring.

Trossamfunn vil sansynligvis ikke være tilgjengelig i noe register i 1990, og vil derfor mangle helt.

Opplysning om hvilke andre personer en person bor sammen med (i samme bolig vil ikke kunne skaffes fra noe register så lenge massen av bebodde boliger ikke gis nummer, jfr. leilighetsnummeret i bygningsdelen i GAB. I DSP-basen er det lagt til rette for registrering av leilighetsnummer.

3.3.2. Utdanning/skolegang

Elev/student. Kjennermerket, som viser situasjonen pr. 1. okt. med koder for heltid og deltid, vil være tilgjengelig slik som i 1980.

Høyeste fullførte utdanning. I tillegg til det som var tilgjengelig i 1980 vil kjennermerket i 1990 sannsynligvis også omfatte utdanning som nordmenn tar i utlandet. Registreringer samles også i dag, men massen (omfang mv.) er foreløpig for uklar til at årlig statistikk kan publiseres.

Utdanning som utenlandske statsborgere bosatt i Norge har tatt i utlandet vil neppe foreligge registrert i 1990, da forsøk på å få etablert faste rutiner for registrering av utdanning ved innvandring til Norge ikke har lyktes til nå. I øyeblikket finnes ingen planer om nye forsøk.

3.3.3. Yrkesaktivitet, arbeidstid mv.

Sysselsettingsfilen

6. kontor forutsetter at det til bruk i en registerbasert folketelling blir laget en sysselsettingsfil med sysselsettingsforhold som enhet. Både sjølstendige og ansatte skal være med og tidspunktet for start og slutt i arbeidsforholdet registreres.

Pr. i dag foreligger flere beskrivelser av planene, jfr. notatene "Synspunkter på statistikk og beregningen av timeverkopplegget for slik statistikk ved 6. kontor" (SGå/EBr, 26/1-85), "Sysselsettingsstatistikk for sjølstendige. Fil til bruk av registerbasert folketelling" (HNN/kin, 14/1-82) og "Sysselsettingsfil til FOB 90 - 6. kontors bidrag". (SGå/EBr, 19/11-85). Med hensyn til mulighetene for å skaffe en tilstrekkelig god sysselsettingsfil til FOB 90 mener 6. kontor av dette i hovedsak er et spørsmål om ressurser til SSB.

Forutsatt at de planer det arbeides etter kan gjennomføres, vil en registerbasert sysselsettingsfil for kalender året 1990 være klar i 1. kvartal 1992.

Den første årgangen av sysselsettingfilen, for kalenderåret 1987, vil være klar i løpet av 1. halvår 1987.

De data som trengs finnes i administrative registre i dag, men kobling mellom arbeidsgiverregisteret og bedriftsregisteret er ikke etablert på en tilfredsstillende måte for alle næringer. Det er heller ikke etablert kobling mellom en registerbasert fil over sysselsatte sjølstendige og bedriftsregisteret.

Det som gjenstår for å kunne utarbeide en registerbasert sysselsettingsfil er for det første at koblingsrutinen mellom bedriftsregisteret og arbeidsgiverregisteret bygges ut til å omfatte alle næringer. Det er Byrået som må ha ansvaret for dette. For det andre må vi se til at de arbeidsgivere som er delt opp i flere bedrifter fører riktig undernummer på meldingsblanketten. Det er Byrået som må ha ansvaret for å kontrollere dette og trygdekontorene må følge opp de feil og mangler vi kan dokumentere. 6. kontor regner med at det vil være en forholdsvis grei sak å sikre kobling mellom sjølstendige næringsdrivende og bedriftsregisteret.

For den kvartalsvise arbeidstakerstatistikken er det et problem at arbeidsgiveren ikke sender meldingene i rett tid. Dette forholdet skaper derimot ikke problemer for en sysselsettingsfil som utarbeides så lang tid etter utgangen av kalenderåret at administrative data for pensjongivende inntekt og arbeidsgiverens lønnstrekksoppgaver foreligger.

Lover og forskrifter for arbeidstakerregisteret gir ikke grunnlag for å pålegge arbeidsgiveren å følge den oppdeling i undernummer som er fastlagt. Det er heller ikke høve til å pålegge arbeidsgiveren å utføre årskontroller av arbeidstakerregisteret.

En forutsetning for at planene om en sysselsettingsfil kan realiseres er at arbeidstakerregisteret fortsatt vil eksistere i 1990 og at kvaliteten på det vil være den samme eller bedre enn i dag.

Enhet i sysselsettingsfilen - kjennemerker

Enheten i sysselsettingsfilen er det enkelte sysselsettingsforhold

$$S_{ij}(t_1 - t_2, h, p, a)$$

Det enkelte sysselsettingsforhold er identifisert ved fødselsnummer (i) og bedriftsnummer (j). Kjennermerker knyttet til hvert enkelt sysselsettingsforhold er:

t_1 og t_2 = start- og stoppdato for forholdet

h = betalte timeverk i perioden $t_2 - t_1$, beregnet verdi

p = pensjonsgivende inntekt for perioden $t_2 - t_1$

a = type sysselsettingsforhold (arbeidstaker, sjølstendig, vernepliktig)

Datakilder for sysselsettingfilen

For arbeidstakere/lønnstakere:

- arbeidstakerregisteret
- arbeidsgivernes lønnstrekkoppgave
- pensjonsgivende lønnsinntekt

For vernepliktige/sivilarbeidere:

- filer fra de ansvarlige myndigheter

For sjølstendige:

- likningsregister (næringsoppgaver?)
- momsregister
- bedriftsregisteret
- eventuelle spesielle bransjeregister

Andre aktuelle data

Den sysselsettingsfilen 6. kontor utarbeider vil inneholde data i tillegg til de som er spesifisert foran. Dette er av omsyn til beregning av timeverk, komplettering av start- og stoppdato og for statistikkformål.

Aktuelle data er:

- registrerte arbeidsledige, permitterte
- sysselsatte i arbeidsmarkedsstiltak
- uføre- og alderspensjonister, andre trygder
- utdanningsaktivitet
- beregnede verdier av omfanget av husarbeid
- sysselsettingsregister for offentlig sektor

Data som er aktuelle, men som trolig kommer seinere enn 1990:

- registrert sykefravær
- individdata som hentes inn for lønnsstatistikk

Referanseperiode

Sysselsettingsfilen omfatter alle sysselsettingforhold i kalenderåret. Når start- eller stoppdato mangler, nyttes antatt dato etter maskinelle rutiner. Om vi ikke lykkes å få til en tilfredsstillende løsning på denne måten, er alternativet for situasjonen 1.11. 1990 beregnede verdier basert på kombinasjon av register og utvalg.

Årskontrollen av arbeidstakerregisteret gjennomføres i februar. Opplysningene om start- og stoppdato vil ha best kvalitet denne måneden. Årskontrollen går på faktiske datoer slik at opprettingen gir kvalitetsforbedringer også tilbake i tid, f.eks. for 1. november året før.

Ressurser

Når det gjelder framdriften i arbeidet med sikte på å etablere en registerbasert sysselsettingsfil er usikkerheten knyttet til om det er avsatt tilstrekkelig ressurser. Det svakeste punktet her er at bedriftsregisteret ikke er kommet i gang med å etablere et tilfredsstillende bedriftsregister for offentlig sektor.

Arbeidet med å utvide arbeidstakerstatistikken til flere næringer i 1986 går langsommere enn ønskelig på grunn av snaue ressurser.

3.3.3.1. Yrkesaktivitet i kalenderåret

Yrkesaktive i kalenderåret. Kjennemerket vil måtte beregnes ut fra opplysninger fra de sysselsettingsforhold som er registrert på sysselsettingsfilen. Gruppen familiararbeider er bare delvis registrert. 6. kontor mener det er lett å finne gode metoder for å beregne tall for små grupper i dette tilfellet. Dette er ikke noen viktige grupper i FOB-sammenheng.

Svart arbeid er ikke registrert. Den viktigste gruppen her er dagmammaer. Det er noe usikkert hvor godt denne gruppen vil bli fanget opp ved en tradisjonell telling.

Betalte timeverk på registerfilen er basert på anslag ved arbeidsforholdets start. Anslaget korrigeres bare ved årskontrollen. Med dette forbeholdet kan tallet på yrkesaktive beregnes som tallet på personer med arbeidstid \geq 100 timer siste år, slik som i FOB 80.

Næring. En må velge et sysselsettingsforhold, omkring tellings-
tidspunktet. Erfaringer fra Tilleggsundersøkelsen til Folke- og boligtel-
ling 1990 (TU) tyder på at det vil bli problematisk å få tak i næring. 6.
kontor mener at det likevel er bra muligheter for en løsning og mener at
registeralternativet vil gi best kvalitet.

Yrke. Kjennermerket vil måtte utgå med sitt nåværende begrepsinn-
hold. Metodegruppas planer om mulig tilnærming er ikke blitt fulgt opp.

Yrkesstatus. Kjennermerket vil bare kunne ha verdiene selvstendig
og ansatt. Gruppen familiearbeidskraft uten fast lønn vil ikke kunne
skilles ut.

3.3.3.2. Yrkesaktivitet mv. siste uke

Arbeidstid siste uke. Beregnede tall vil være gjennomsnitt av
estimerte verdier for arbeidstid siste år. I motsetning til FOB 80 får en
derfor ingen opplysning om faktisk arbeidstid siste uke.

Arbeidsstedskommune. Stedfesting til kommunene synes mulig.
Dette er et prioritert kjennemerke i arbeidet med sysselsettingsstatistik-
ken.

Fast oppmøteplass

Det er vanskelig å få gjennomført noe skille mellom fast og ikke
fast oppmøteplass, unntatt for spesielle næringer. Planen er at arbeids-
takere som ikke har fast oppmøteplass skal registreres på egne undernum-
mer, men det er uklart når et slikt opplegg vil være gjennomført. Endel
fiktiv pendling vil derfor forekomme.

Oppmøteplassens beliggenhet (krets)

Dette må begrenses til bedrifter med fast beliggenhet. Det er
ikke tidfestede planer om å etablere dette kjennemerket. Bruk i 1990 er
derfor lite realistisk. Offentlig sektor er mest problematisk. Det kan
være et tilbud til kommuner som kan skaffe opplysningene. Eventuelt må
dette begrenses til næringer der dette er av spesiell interesse.

Tallet på arbeidsreiser i uka, reisetid for arbeidsreiser og
reisemåte/transportmidler på arbeidsreiser. Ingen registerbaserte opplys-
ninger vil finnes, men i tilfelle oppmøteplassens beliggenhet kan henføres
til krets, vil det være visse muligheter for å beregne reisetid for grup-
per av personer ved hjelp av metoder under utvikling ved 12. kontor
(koordinatfesting av grunnkretser mv.)

Pendling. Registreringen av arbeidssted er i dag ikke god nok for pendlingsstatistikk som er mer detaljert enn mellom kommuner. En mer nøyaktig registrering av oppmøteplass kan likevel tenkes utført i samarbeid med kommunale myndigheter, på tilsvarende måte som i betalte oppdrag ved Folketellingskontoret utført for en del kommuner. En vil likevel ikke kunne avgrense statistikken til de yrkesaktive siste uke med fast oppmøteplass, slik som i FOB 80.

Dersom en ikke kan få til mer nøyaktig registrering av oppmøteplass (arbeidssted) enn kommune, er en derfor mht. pendlingsstatistikk tilbake i en situasjon som er minst like lite tilfredsstillende som før 1980.

3.3.3.3. Viktigste kilde til livsopphold/inntekt

Viktigste kilde til livsopphold. Kjennemerket i 1980 bygde på personens egen vurdering. Et eventuelt forsøk på å avlede tilsvarende kjennemerke på grunnlag av registeropplysninger, standard for sosioøkonomisk status mv. vil i beste fall resultere i data som er svært lite sammenliknbare med FOB 80. Trolig vil registerdataene by på tolkningsproblemer i denne sammenhengen og medføre muligheter for betydelige feil på individnivå.

Arbeidsinntekt. De samme inntektsdata som ble nyttet ved FOB 80 ventes å være tilgjengelige også i 1990.

En har forventninger om at innføring av EDB i Skatteetaten vil kunne resultere i at flere kjennemerker kan bli aktuelle for FOB. Det er likevel atskillig usikkerhet mht. hvilke av disse kjennemerkene som vil bli knyttet til alle skatteyttere og til når de nye kjennemerkene vil kunne være tilgjengelige.

3.4. Kjennemerker knyttet til familier

Gruppering av personer til familier vil kunne foretas på grunnlag av data i personregisterbasen, på tilsvarende måte som i 1980. En arbeidsgruppe har lagt fram forslag til nye standard kjennemerker knyttet til familier og husholdninger, INO 85/31. Forslaget er ikke behandlet ennå.

Anbefalte kjennemerker:

Familietype. Foreslått endret siden 1980. Tre alternativer

Tallet på ugifte, hjemmeboende barn i familien (som i 1980?)

Tallet på personer i familien (som i 1980?)

Alder for yngste, ugifte hjemmeboende barn (som i 1980?)

Disponibel inntekt for familien. Kjennemerket vi kunne avledes slik som i 1980, da tilgangen til inntektsdata vil være den samme eller bedre.

3.5. Kjennemerker knyttet til ektepar/samboerpar

Kjennemerker for ektepar basert på data for personer kan avledes som i 1980 i den grad data for personer vil foreligge. (Ektefellens yrkesdeltaking, ektefellens arbeidstid i kalenderåret, ekteparets antall hjemmeværende barn). Dessuten vil det sannsynligvis bli mulig å avlede kjennemerker for samboende (ikke gift med hverandre) med felles barn, da slikt sambofelleskap er planlagt registrert ved 12. kontor ut fra data i Personregisterbasen. Grupperingen bygger på forutsetningen om at personer som har barn sammen og som er registrert bosatt på samme adresse (numerisk adresse) er samboende. Sambofelleskap uten felles barn vil sannsynligvis ikke kunne foreligge registrert i 1990, fordi det i øyeblikket ikke eksisterer noen planer om å tildele leilighetsnummer på bestanden av bebodde boliger.

3.6. Kjennemerker knyttet til husholdninger

Slike kjennemerker vil ikke kunne fås eller avledes fra registre som ventes å eksistere i 1990.

Kjennemerker som husholdningstype, tallet på personer i husholdningen, tallet på familier i husholdningen, tallet på barn i husholdningen og husholdningens disponible inntekt kan derfor ikke skaffes i 1990 uten informasjon via blankett. Det samme gjelder 1980-tellingens blankettopplysning "disponering av personbil".

Et alternativ vil være å nytte de motsvarende kjennemerkene knyttet til familier (se foran), med de faglige svakheter det kan innebære.

3.7. Kjennemerker knyttet til boliger

Uten en blankett med boligspørsmål vil en i 1990 stå uten boligopplysninger. I øyeblikket synes verken bygningsdelen i GAB eller noe annet register å kunne gi nødvendig datagrunnlag.

3.8. Oversikt over aktuelle kjennemerker

Aktuelle kjennemerker for FOB 90 gruppert etter muligheten for at de er tilgjengelig på en fullstendig FOB-fil i 1990. Vurdering pr. november 1985

Kjenne- merke- gruppe	Finnes som register- kjennemerke i dag, evt. helt uproble- matisk	Vil med stor sannsynlighet eksistere som kjenne- merke i 1990	Vil neppe være til- gjengelig som kjennemerke i 1990	Kjennemerker som bare vil være til- gjengelig for et utvalg i 1990
-----------------------------	--	---	---	---

<u>Identifi- kasjons- kjenne- merker</u>	{ fødselsnummer familienummer	sambofelleskaps- nr. (ekteparnr., markering av personer som inngår i par og har felles barn)	bolignr./ husholdn.nr.	markering av personer i samboerpar uten felles barn
<u>Person- kjenne- merker</u>	{ kjønn alder	fødeland		trossamfunn
Demografi	{ ekteskapelig status statsborger- skap			
Utdanning	{ elev/student høyeste full- førte utdanning			
Yrkes- aktivitet i kalender- året	{		yrkesaktive i kalender- året yrkesstatus arbeidstid i kalender- året næring	yrke

Aktuelle kjennemerker for FOB 90 gruppert etter muligheten for at de er tilgjengelig på en fullstendig FOB-fil i 1990. Vurdering pr. november 1985 (forts.)

Kjenne- merke- gruppe	Finnes som register- kjennemerke i dag, evt. helt uproble- matisk	Vil med stor sannsynlighet eksistere som kjenne- merke i 1990	Vil neppe være til- gjengelig som kjennemerke i 1990	Kjennemerker som bare vil være til- gjengelig for et utvalg i 1990
Yrkes- aktivitet mv. siste uke			arbeidstid siste uke fast opp- møteplass arbeidssteds- kommune oppmøte- plassens beliggenhet (krets)	tallet på arbeids- reiser i uka reisemåte/ transport- midler på arbeids- reisen reisetid for arbeidsreiser
Viktigste kilde til livsopp- hold/inn- tekt		arbeids- inntekt		viktigste kilde til livsopphold
<u>Familie- kjenne- merker</u>		familietype tallet på ugifte hjemme- boende barn i familien tallet på per- soner i familien alder for yngste ugifte hjemmeboende barn disponibel inntekt for familien div. andre kj.merker		

Aktuelle kjennemerker for FOB 90 gruppert etter muligheten for at de er tilgjengelig på en fullstendig FOB-fil i 1990. Vurdering pr. november 1985 (forts.)

Kjenne- merke- gruppe	Finnes som register- kjennemerke i dag, evt. helt uproble- matisk	Vil med stor sannsynlighet eksistere som kjenne- merke i 1990	Vil neppe være til- gjengelig som kjennemerke i 1990	Kjennemerker som bare vil være til- gjengelig for et utvalg i 1990
-----------------------------	--	---	---	---

<u>Kjenne- merker for ektepar/ samboerpar med felles barn</u>	{	parets antall hjemneværende barn		
		alder for yngste barn		
<u>Kjenne- merker for ektefelle/ samboer (dersom felles barn)</u>	{		ektefellens/ samboerens yrkesdel- taking	
			ektefellens/ samboerens arbeidstid i kalenderåret	
<u>Kjenne- merker for samboerpar uten felles barn</u>	{			samboerens yrkesdel- taking
				samboerens arbeidstid i kalender- året
<u>Kjenne- merker for hushold- ninger</u>	{		hushold- ningstype	
			tallet på personer i hushold- ningen	
			tallet på familier i hushold- ningen	
			hushold- ningens disponible inntekt	

Aktuelle kjennemerker for FOB 90 gruppert etter muligheten for at de er tilgjengelig på en fullstendig FOB-fil i 1990. Vurdering pr. november 1985 (forts.)

Kjenne- merke- gruppe	Finnes som register- kjennemerke i dag, evt. helt uproble- matisk	Vil med stor sannsynlighet eksistere som kjenne- merke i 1990	Vil neppe være til- gjengelig som kjennemerke i 1990	Kjennemerker som bare vil være til- gjengelig for et utvalg i 1990
-----------------------------	--	---	---	---

Dispo-
nering av
personbil

disponering
av person-
bil

Bolig-
kjenne-
merker

tallet på rom
i boligen

bosatte
pr. rom

boligens
størrelse
(areal)

eier/leier-
forhold til
boligen

kjøkken

wc/annet
avtrede

boligens
oppvar-
mingsmåte

bad/dusj

hustype
byggeår

pipe

tallet på
leilig-
heter i
huset

brenselovn/
peis

telefon

4. KONKLUSJONER

Spørsmålet om det i SSB er mulig å bygge opp en statistikkberedskap for utarbeiding av folketellingsliknende statistikk basert på registerdata kan trygt besvares med ja, forutsatt at de data en trenger finnes i registre og at kvaliteten er akseptabel.

Målet kan nås på ulike måter. Det gjelder derfor å finne den mest hensiktsmessige løsningen. Denne vil i noen grad bestå i en avveining mellom motstridende ønsker, f.eks. mht. beredskap, datavern, effektivitet og aktualitet. Det synes også slik at et valg mellom sentral datakjerne eller koblingsberedskap neppe vil være avgjørende for om målet kan nås eller ikke, heller ikke valget mellom kryptering eller ikke kryptering av fødselsnummer.

Et menystyrt datasystem av typen DIPS vil imidlertid representere et forenklet og tidsmessig produksjonssystem, ikke bare for personstatistikk der det hentes data fra flere fagområder, men også for fagkontorenes egne statistikker.

Hovedproblemet ved å basere seg på registerdata ved en folketelling i 1990 er knyttet til hvorvidt ønskede data finnes i registrene eller det i registrene finnes data som sammen med annet materiale evt. ved hjelp av statistiske metoder kan omformes til ønskede data. Siden erfaring har vist at registre trenger en nokså lang innkjøringsperiode, vil det være problematisk og trolig uklokt å basere sentrale data for FOB-90-formål på registre som i dag ikke er etablert eller i rimelig bra gjenge. Et produksjonsopplegg som bygger på forutsetninger om en viss framtidig kvalitet på ett eller flere registre, evt. kvalitetsforbedring, vil være svært sårbart. I så fall bør muligheten for alternativ datainnhenting foreligge.

LITTERATUR

Prosjektskriv mv.

JKT/HDR, 2/2-1983. Prosjektskriv: "Databeredskap for persondata, forprosjekt."

WJo/JKT/HDR, 15/8-83. Referat fra møte om databeredskap for persondata.

JKT/LHå, 27/7-84. Prosjektskriv: "Utredning av opplegget for Folke- og bolig telling 1990."

WJo/JKT/KKv/LHå, 20/7-83. Første utkast. Databeredskap for persondata. (Rapport fra forprosjektet.)

JKT/Ski/HDR, 11/12-84. Prosjektskriv: "Datasytem for integrert personstatistikk (DIPS) (Sentralt statistisk befolkningsarkiv)." (Del av prosjektet Utredning av opplegget for Folke- og bolig-telling 1990.)

Milepelplan DIPS. Ski 7/2-85, revidert 15/3-85.

E Au/JKT/HaR, 19/2-85. Styringsgruppe for datasytem for integrert personstatistikk (DIPS).

Skiri, Halvard. DIPS - et redskap i arbeidet med personstatistikken. Byråpraten nr. 2 - 1985, side 25-27.

Personvern - kryptering

Konsesjon for personregistre i Statistisk Sentralbyrå, gitt av Datatilsynet i medhold av Lov om personregistre mm. § 9, første ledd, jfr. kgl.res. av 21.12.79, pkt. I 2. ledd (brev fra Datatilsynet av 2. april 1981)

Konsesjon for personregistre til bruk i statistikkproduksjon, gitt av Datatilsynet, datert 15. april 1986

E Au/BjF, 12/12-83. Datavern i Statistisk Sentralbyrå (om bruk av "statistikknummer").

E Au/BjF, 5/2-85. Kobling av registre uten bruk av kjent identifikasjon.

KKv, 10/4-85. Rutine for oversetting av fødselsnumre til et kryptert nummer via et tilfeldig rekkefølgenummer.

Hvor god er SSB's kryptering? (Udatert notat av Johannes Hansen, Datatilsynet.

JFu/Rac, 3/4-85. Datasytem for integrert personstatistikk (DIPS) (kommentarer til innhold og omfang av DIPS, utdanningsdata og personvern).

JKT/Rac, 10/6-85. Verknadene for statistikkproduksjonen ved innføring av kryptert nummer i lagring av data.

JKT/SRJ/Rac, 12/9-85. Rapport fra møte i Stockholm 30. august 1985. (Om kryptering av personnummer.)

Personregisterbasen og Befolkningsstatistikksystemet

Feltnummerkatalog for Dataelementer i DSP's database. SDS. Oppdatert 1/11-85.

Dokumentasjon av objektklasser og termer. Del av beskrivelse av DSP-basen, SDS. Oppdatert 20/9-84.

ThG, 28/8-84. Prioritering av historiske opplysninger i DSP's database (om hvilke tilbakegående opplysninger som vil bli lagret).

DHS/DHS, 09/10-85. Journal. (Beskrivelse av journaliserte meldinger fra DSP-basen.)

EHa, 9/5-85. Prinsippskisse for uttak fra Det sentrale personregister (DSP) og eksterne registre og produksjon av statistikk.

Ski/KWa, 5/8-83. Prosjektskriv: Omlegging av den løpende befolkningsstatistikken til IBM.

Ski/Rac, 17/9-84. Referat fra møte om omlegging av den løpende befolkningsstatistikken til IBM 30. august 1984.

JBy/KWa, 6/12-84. Notat. Omlegging av befolkningsstatistikken til IBM. Prinsipløsning.

EJM/Rac, 4/3-85. Referat fra møte om omleggingen av befolkningsstatistikken 18. februar 1985. Kongsvinger.

JKT/UDy, 26/11-85. Referat fra møte om behandling og organisering av data fra DSP-basen.

FOB på registergrunnlag

Børke, Sindre (1984). Folke- og bolig telling 1980. Dokumentasjon.Rapporter 84/15.

Børke, Sindre (1984). Tilleggsundersøkelsen til Folke- og bolig telling 1980. Om muligheter for å erstatte skjema med registeropplysninger i seinere folke- og bolig tellinger. Rapporter 84/22.

SGå/EBr, 19.11.85. Sysselsettingsfil til FOB 90 - 6. kontors bidrag.

SGå/EBr, 4/12-85. Sysselsettingsfilen i FOB 80 og Tu til FOB 80 (Notat til seminar om FOB 90 5. og 6.12.1985).

HNN/EBr, 4/12-85. Kommentarer til "Utredning og Folke- og bustadteljing 1990, delprosjekt 4".
Danmarks Statistik (1982). Personstatistik på registergrunnlag.
SRJ/Ski/TVi/HDr, 27/6-85. Rapport fra besøk i Danmarks Statistik 3.-4. juni 1985. (Om det person-
statistiske system og folke- og bolig telling på registergrunnlag.)

Datadokumentasjon - gjenbruk av data

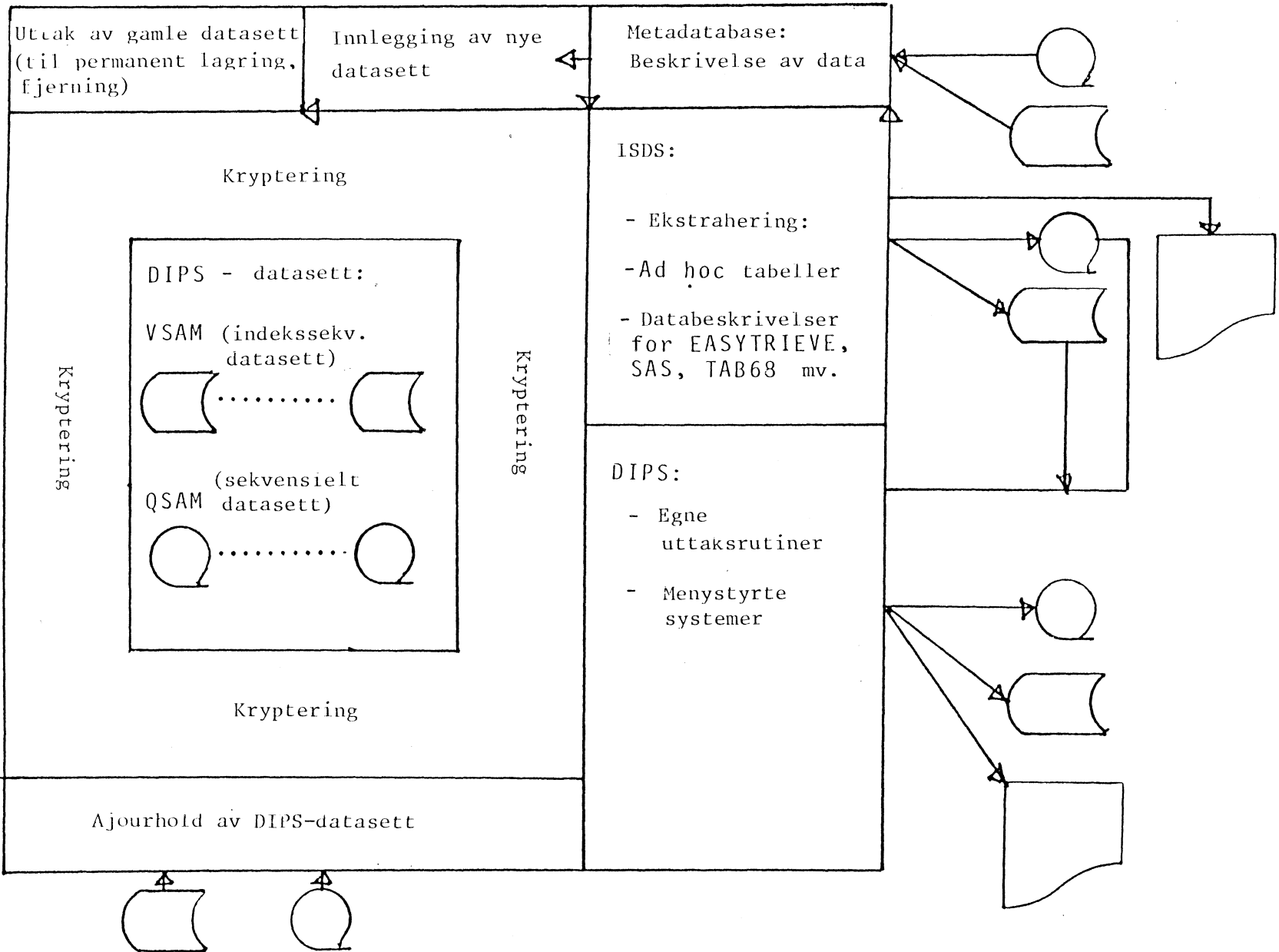
System for datadokumentasjon i Statistisk Sentralbyrå. Innstilling fra styringsgruppen. INO 85/18.
Forslag til standarder for kjennemerker knyttet til familier og husholdninger (fra en arbeidsgruppe).
INO 85/31.

AEd/VFr, 23/1-81. Kjennemerkekoder og fileredigering i beredskapsfilen.

Brunborg, Helge (1985). Omlegging av beredskapsfilen. INO 85/41.

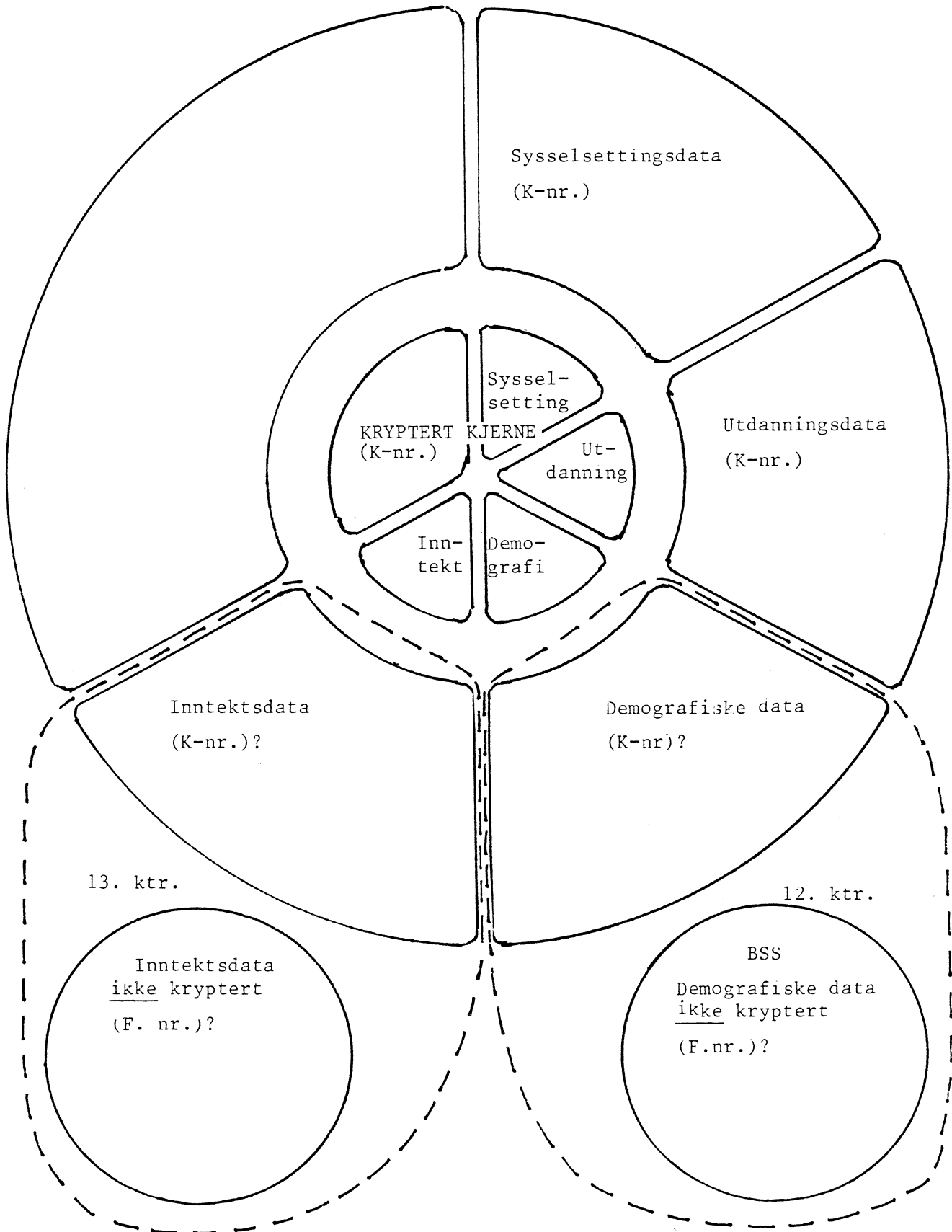
Østby, Lars (1985). Omleggingen av personregisteret og av produksjonen av befolkningsstatistikk, og
konsekvenser for gjenbruk av data. INO 85/28.

Ski/TNo/E1B, 24/9-85. Rapport fra kurs i demografisk livsforløpsanalyse, Ystad 5. - 16. august 1985.





OMFANGET AV BRUKEN AV KRYPTERING





Befolkningsstatistikksystemet (BSS) og forholdet til DIPS og det sentrale personregisteret (DSP)

