

Arbeidsnotater

Statistisk Sentralbyrå

I 769

S T A T I S T I S K S E N T R A L B Y R Å

Statistisk sentralbyrå

IO 69/2



Oslo, 2. mai 1969

DATAREGISTRERING, DATAARKIVER OG SAMFUNNSFORSKNING

Av

Odd Aukrust og Svein Nordbotten

311.3931

829 ar

Innhold

| | Side |
|--|------|
| 1. Innledning | 2 |
| 1.1 Arkivstatistikkens grunnidé | 2 |
| 1.2 Ansvar for data-arkivenes oppbygging og vedlikehold | 4 |
| 2. Utviklingen fram til i dag | 5 |
| 2.1 Innføring av permanente identifikasjonsnummersystem | 5 |
| 2.2 Det nåværende persondataarkiv | 6 |
| 2.3 Det nåværende bedrifts- og foretaksdataarkiv | 8 |
| 3. Nye informasjonsbehov | 9 |
| 3.1 Generelle og spesielle behov | 9 |
| 3.2 Behov for mer informasjon om befolkningen | 10 |
| 3.3 Behov for mer informasjon om økonomiske enheter | 12 |
| 4. Registerproblemer | 13 |
| 4.1 Utbygging av personregisteret | 13 |
| 4.2 Utbygging av bedrifts- og foretaksregisteret | 14 |
| 4.3 Adresse/eiendomsregister | 15 |
| 5. Datainnsamling | 17 |
| 5.1 Direkte og indirekte innsamling | 17 |
| 5.2 Persondata | 18 |
| 5.3 Bedrifts- og foretaksdata | 21 |
| 6. Dataarkivering | 23 |
| 6.1 Arkivsystem | 23 |
| 6.2 Tekniske spørsmål | 24 |
| 7. Utnytting av dataarkivene | 25 |
| 7.1 Publikasjoner | 25 |
| 7.2 Dataservice og spesialbearbeidinger | 25 |
| 8. Analytiske anvendelser av persondataarkivet | 26 |
| 8.1 Innholdet i persondataarkivet | 26 |
| 8.2 Persondataarkivet som grunnlag for sosialvitenskapelig forskning | 29 |
| 9. Konfidensiell behandling av data | 33 |
| 10. Kostnadsoverslag og ressursbehov | 34 |

Ikke for offentliggjøring. Dette notat er et arbeidsdokument og kan siteres eller refereres bare etter spesiell tillatelse i hvert enkelt tilfelle. Synspunkter og konklusjoner kan ikke uten videre tas som uttrykk for Statistisk Sentralbyrås oppfatning.

1. INNLEDNING

1.1 Arkivstatistikkens grunnidé

Den tradisjonelle måte å lage statistikk på, er ved undersøkelser i form av tellinger. Ved en telling blir opplysninger i de fleste tilfelle hentet inn på skjemaer fra individuelle oppgavegivere. Skjemaene undergis en bearbeiding en gang for alle. Resultatene publiseres i form av et tabellverk, som da vil være tellingens resultat. Opplysningene om den enkelte oppgavegiver vil derimot for alle praktiske formål være utilgjengelige i det øyeblikk skjemaene er pakket vekk fordi "søking" i store skjemabunker er en kostbar sak.

Det arkivstatistiske system bygger, i motsetning til tellingene, på den grunnidé at opplysninger fra individuelle oppgavegivere arkiveres gjennom lengre tid slik at de er tilgjengelige og kan hentes fram fra dataarkivet ved behov. De tekniske muligheter for dette er skapt ved datamaskinene, som har åpnet adgang til å ordne, lagre og gjennomføre store datamasser på kort tid og med små kostnader.

Et vilkår er at det er etablert et permanent identifikasjonsnummer-system for de statistiske enheter (oppgavegiverne), slik at vi vet hvilke enheter de lagrede opplysninger refererer seg til. Slike identifikasjonsnummersystem (fødselsnummer for personer, faste foretaksnummer for foretak osv.) spiller en sentral rolle i et arkivstatistisk system, fordi de gjør det mulig å hente fram og koble sammen opplysninger om de enkelte enheter hentet inn fra ulike kilder, for ulike formål, eller på forskjellige tidspunkter. Opplysningene kan kobles sammen "på tvers", dvs. fra forskjellige kilder, og vil da til sammen gi en bredere og mer omfattende karakteristikk av enheten enn de enkelte kilder kan gi hver for seg. Eller de kan kobles sammen "på langs over tiden", dvs. vi kan føre sammen opplysninger om en og samme enhet hentet inn på forskjellige tidspunkter, og vi vil få et bilde av enhetens livsløp.

Denne informasjonsmasse kan så tappes ved behov og brukes til å lage statistikk, eller den kan utnyttes direkte for analyseformål. Det kanskje viktigste fortrinn ved det arkivstatistiske system er derfor at det tar bedre vare på den informasjon som samles inn. Vi kan benytte oss av dette til å rasjonalisere statistikkproduksjonen og å gjøre den billigere. Enda viktigere: Vi vil kunne lage bedre og mer omfattende statistikk ved å utnytte de samkoblingsmuligheter som er omtalt ovenfor. For eksempel:

- a. Det er mulig å bearbeide under ett opplysninger som er hentet inn fra ulike kilder. Eksempelvis vil vi fra og med 1967 kunne gi en vesentlig

bedre inntekts- og formuesstatistikk ved å koble opplysninger om personers inntekter og formue tatt fra likningsansettelsene sammen med demografiske opplysninger om de samme personer tatt fra det sentrale personregister.

- b. Det er mulig å plukke ut en hvilken som helst gruppe enheter (et bestemt årskull av akademikere, barn som har vært under sosial omsorg e.l.) på et bestemt tidspunkt og å belyse denne gruppes senere livsløp ved å analysere den informasjon vi har om enhetene i gruppen på senere tidspunkter, eller å sammenlikne livsløpet for ulike grupper av enheter. Denne adgang til "longitudinelle" studier åpner store perspektiver for den samfunnsvitenskapelige, sosialmedisinske og sosialpsykologiske forskning (eksempler er gitt senere).
- c. Det statistiske beredskap vil øke ved at produktet av statistikkproduksjonen ikke lenger vil være bare de trykte publikasjoner, men like mye informasjonsmassen i selve dataarkivet og de muligheter denne gir for å betjene konsumentene med utkjøring av spesielle tabeller ved behov. (Mens "å bruke statistikk" hittil stort sett har betydd å lese tall ut av trykte publikasjoner, kan det i framtiden like ofte bli å hente tall ut fra dataarkivets magnetbånd.)

Hvilke muligheter et arkivstatistisk system konkret sett vil åpne, avhenger selvsagt av hvilke opplysninger det til hver tid inneholder. Generelt gjelder - og her skiller arkivstatistikken seg radikalt ut fra tradisjonell metode - at enhver ny opplysning som kommer til, ikke bare har verdi i seg selv, men samtidig øker verdien av andre opplysninger som arkivet allerede inneholder og som den nye opplysning kan kobles sammen med. Vi har allerede nevnt at likningsoppgaver koblet sammen med opplysninger i det sentrale personregister kan gi grunnlag for en bedre inntektsstatistikk. Men samtidig øker verdien av (for eksempel) personregisterets opplysninger om innenlandske flyttinger fordi det blir mulig å analysere flytterne ikke bare etter kjønn og alder som nå, men også etter deres inntekts- og formuesforhold. Hvis en som et neste skritt (for eksempel) tilføyer arkivet opplysninger om de enkelte personers utdanning, vil dette ikke bare gi grunnlag for å utarbeide en utdanningsstatistikk som i seg selv er verdifull. Det åpner samtidig muligheter for å analysere den personlige inntektsfordeling etter utdanning og flytterne simultant etter utdanning, kjønn og alder. Vi har altså å gjøre med kumulative virkninger som innebærer at den fulle nytte av dataarkivene først vil komme til syne etter hvert som nye opplysninger blir tilført dem. Nyttens av det sentrale personregister fra

statistisk synspunkt er i øyeblikket begrenset, fordi det inneholder bare et mindre antall opplysninger om den enkelte person (se senere). Nyttens ville mangedobles om det kunne bygges ut med noen få sentrale tilleggsopplysninger, først og fremst om hver enkelt persons utdanning, næring og yrke.

1.2 Ansvar for dataarkivenes oppbygging og vedlikehold

De opplysninger som skal inngå i de sentrale dataarkivene, kan innsamles direkte fra publikum av Statistisk Sentralbyrå. Men de kan også tenkes tilført arkivet fra andre offentlige etater. En må her velge praktiske løsninger. Det er på det rene at ansvaret for å skaffe fram og koordinere den statistikk som samfunnet trenger, er lagt til Byrået. Også ansvaret for innsamlingen og arkiveringen av de grunndata som trengs for å lage statistikken - herunder planleggingen av dataarkivenes innhold og oppbygging - må derfor i siste instans være Byråets. Men samtidig er det ofte slik at dataene allerede er tilgjengelige i andre etaters arkiver og registre (inntektsdata hos skattemyndighetene) eller lettest og sikrest kan samles inn av andre etater som har direkte publikumskontakt.

Det ville åpenbart være urasjonelt om Byrået henvendte seg direkte til publikum for å få opplysninger som allerede foreligger i andre offentlige registre og arkiver, eller opplysninger som andre etater kan skaffe billigere og bedre. Den praktiske konklusjon må bli følgende: Hvis det er på det rene at et samfunnsmessig anerkjent behov for en bestemt type statistikk foreligger, og de nødvendige rådata finnes hos eller best kan skaffes av andre etater, må disse kunne pålegges å bistå Byrået med å skaffe dataene til veie. Etatene må med andre ord så langt det er praktisk, oppfatte det som del av sine funksjoner å medvirke til at samfunnets informasjonsbehov blir dekket. Dette må også gjelde selv om etatene, av hensyn til statistikken, kan bli nødt til å samle inn informasjon som de strengt tatt ikke trenger for sitt arbeid, eller å nytte registreringssystemer, definisjoner o.l. som ikke faller helt sammen med etatens egne ønsker. Et visst samarbeid mellom Byrået og de fleste etater eksisterer, basert på disse retningslinjer. Overgang til EDB i stadig større deler av den offentlige forvaltning, som betyr at grunndata ofte kan stilles til disposisjon for Byrået i form av magnetbånd eller andre maskinmedia, har aksentuert nødvendigheten av koordinert innsats for at informasjonsmassen i samfunnet skal komme til størst mulig nytte.

Å bearbeide dataarkivenes innhold til statistikk, hva enten rådata er innsamlet på den ene eller andre måte, må derimot som hittil være Byråets spesielle oppgave.

2. UTVIKLING FRAM TIL I DAG

2.1 Innføring av permanente identifikasjonsnummersystem

Da datamaskinene kom i midten av 1950-årene, ble de nye muligheter dette åpnet for arkivering og direkte analytisk utnyttning av individualdata, drøftet i forskjellige sammenhenger i Byrået. I første omgang tenkte en seg data systematisk ordnet i hullkortarkiver, senere kom magnetbåndarkivene inn i bildet.

Før disse mulighetene forelå, var det en hovedoppgave for statistikkprodusenten å redusere de innsamlede datamassene til statistikk for grupper av individualenheter. Gruppetata - i motsetning til individualdata - var det eneste som en med den tids hjelpemidler maktet å arbeide med. Tallet på grupper måtte ikke være større enn at materialet kunne arkiveres i trykte og utrykte tabeller. For at statistikken fra de forskjellige felter skulle bli mest mulig sammenliknbar og anvendelig for videre analytiske studier, ble det satset meget på å nytte de samme standardgrupperinger i all statistikk.

En god del ble oppnådd på denne måten, men en har ikke helt kunnet unngå at ulike enhetsdefinisjoner på forskjellige områder, uoverensstemmelser i primæroppgavene o.l. har slått ut i uoverensstemmelser i statistikken.

Med de nye tekniske hjelpemidler har det åpnet seg muligheter for å bygge opp en integrert og konsistent statistikk ved å sørge for standardisering allerede ved datainnsamlingen. For å kunne føre data om den enkelte person eller økonomiske enhet fra forskjellige kilder sammen, og for å kunne bygge opp individuelle tidsserier i arkivene, er det etablert systemer for entydig og permanent identifikasjonsnummerering av de enkelte statistikk-enheter.

Et system med entydige og permanente fødselsnummer for hver enkelt person i landet ble innført 1.oktober 1964. Numrene tildeles av Byrået som et ledd i den løpende folkeregistrering og omfatter 11 siffer hvorav de seks første angir fødselsdag, -måned og -år innenfor et århundre. De følgende tre siffer angir kjønn, hvilket århundre vedkommende er født i og en nummerering av alle som er født på samme dag. De siste siffer er to kontrollsiffer som er beregnet som bestemte funksjoner av de foregående siffer, og hensikten er å oppdage feil i nummeridentifiserte data før de går inn i arkivene.

For bedrifter og foretak ble det innført et system med entydige identifikasjonsnummer i tilknytning til Byråets sentrale bedrifts- og

foretaksregister i 1956. I 1965 ble nummersystemet endret slik at også disse identifikasjonsnummer nå er permanente. De nummer som her nyttes er 7-sifret, hvorav de seks første siffer er en fortløpende nummerering av enhetene, mens det siste siffer er et kontrollsiffer.

De permanente identifikasjonsnummersystemer spiller for individualdata den samme rolle som de fast etablerte grupperingsstandardene for statistikk; i begge tilfelle er hensikten å gjøre det mulig å lage tids-serier.

Da en stor del av den egentlige dataregistrering foregår i administrative institusjoner, er det av vesentlig betydning å få disse til å innse sin fordel av slike identifikasjonsnummersystemer og ta dem i bruk. Dette vil gi Byrået data som allerede er identifisert på den måte et arkivstatistisk system krever for sammenkobling av data. Tallet på individualenheter og alle de kjennemerker som i form av livshistorier og beskrivelser kan knyttes til hver enkelt enhet, gir en datamasse som det ikke lenger er hensiktsmessig å lagre i utskrevne tabeller og lister. I stedet holdes dataene nå på magnetbånd og utgjør vårt fysiske dataarkiv.

2.2 Det nåværende persondataarkiv

Fram til i dag har oppbygging av dataarkiver foregått vesentlig på to felter:

1. Registrering og lagring av data om personer
2. Registrering og lagring av data om bedrifter og foretak

Disse to typer av enheter er ikke de eneste som kan tenkes. De utgjør imidlertid de enhetstyper som er mest sentrale i en detaljert beskrivelse og registrering av det som foregår i samfunnssystemet.

Registreringen av nummeridentifiserte persondata er kommet lenger i Norge enn i de fleste andre land. Det offisielle fødselsnummersystemet nyttes nå ved registrering i en rekke administrative prosesser, blant annet innen folkeregistreringen, liknings- og skattevesenet, trygdeadministrasjonen, helseadministrasjonen, militær administrasjon, skole- og undervisningsvesen og i valgmanntallene. Fra slike administrative prosesser får Byrået data som det ved hjelp av fødselsnummersystemet kan systematisere og lagre i sine persondataarkiver i form av individuelle tidsserier. Fordi utvalgsteknikk sjelden nyttes i administrative prosesser, og fordi en trekning av et utvalg fra de masser vi overtar fra de administrative organer vil bli nesten like kostbar som det å lagre massen, er persondataarkivet bygd opp på fullstendig registrering.

Innholdet i persondataarkivet i dag går fram av tabell 1, som viser at vi allerede har et omfattende og verdifullt arkiv. Med det nåværende omfang går det inn i arkivet flere millioner individuelle registreringer pr. år.

Oppbyggingen og vedlikehold av persondataregistreringen har hittil kostet om lag 5 mill.kr. Beløpet er belastet folkeregistreringen. Etableringskostnadene har vært store, men de er antakelig innspart allerede ved de fordeler som blant annet folketrygdadministrasjonen har hatt av å kunne overta et ferdig utbygd personnummersystem.

Tabell 1. Innhold og registreringshyppighet i persondataarkivet¹⁾

| Kjennemerke | Registreringshyppighet | | |
|----------------------------|------------------------|--------------|--------------------|
| | løpende | årlig | 10-årlig |
| 1. Fødselsdata og -sted | x | | |
| 2. Bostedskommune | x | | |
| 3. Navn | x | | |
| 4. Adresse | x | | |
| 5. Ekteskapeleg status | x | | |
| 6. Dødsdato og -sted | x | | |
| 7. Inn- og utvandringsdata | x | | |
| 8. Fødselsnummer for mor | x | | |
| 9. " " far | x | | |
| 10. " " ektefelle | x | | |
| 11. Inntektsdata | | x(f.o.m. 67) | |
| 12. Formuesdata | | x(" ") | |
| 13. Skattedata | | x(" ") | |
| 14. Examen artium | | x(" 68) | |
| 15. Familiedata | | | x(pr. 1/11-60) |
| 16. Utdanning | | | x(" ") |
| 17. Yrkesdata | | | x(" ") |
| 18. Næringsdata | | | x(" ") |
| 19. Boligdata | | | x(" ") |
| 20. Valgdeltakelse | | | x(valg f.o.m.1967) |

1) Det offisielle tidspunkt for oppstartning av denne registrering var 1. oktober 1964. Av den grunn mangler det enkelte kjennemerker for deler av befolkningen, f.eks. foreldrenes fødselsnummer for personer født før dette tidspunkt. Likeens vil vi heller ikke kjenne utdanning og yrke pr. 1. november 1960 for personer som har innvandret etter Folketellingen.

Hittil har storparten av de disponible ressursene gått med til å bygge opp registreringssystemet og arkivorganisasjonen. Det har vært små muligheter for analytiske anvendelser, og arkivene har bare vært nyttet i enkelte demografiske analyser og for bygging av Byråets regionale befolkningsprognosemodell. Med de data som allerede finnes i arkivene, er det lagt et grunnlag som med overkommelige kostnader kan utbygges videre og gi et datagrunnlag for samfunnsvitenskapelig forskning som få eller ingen land vil ha maken til.

For å kunne stimulere analytisk utnyttning av persondataarkivet, er det etablert en egen Gruppe for personmodeller i Byrået. Gruppen har som særlig arbeidsområde å utføre forsknings- og analysearbeid i forbindelse med befolkningens sammensetning, utvikling og atferd bygd på matematisk formulerte modeller. Arbeidet har fram til i dag vesentlig vært begrenset til demografiske aspekter, herunder flyttinger, særlig med sikte på praktiske anvendelser i form av bedre regionale befolkningsprognoser.

2.3 Det nåværende bedrifts- og foretaksdataarkiv

I motsetning til det sentrale personregister blir det sentrale bedrifts- og foretaksregisteret bare nyttet i Byråets egen datainnsamling. Bedriftsdataarkivet omfatter derfor, som tabell 2 viser, hovedsakelig data fra produksjon og omsetning. En del av disse data foreligger bare på utvalgsbasis. Det har derfor ennå ikke på langt nær samme omfang og betydning som persondataarkivet. For et mindre antall foretak og bedrifter, henholdsvis om lag 600 og 1 300, har en søkt å føre registreringen bakover til 1958.

Tabell 2. Innhold og registreringshyppighet i bedrifts- og foretaksdataarkivet

| | Registreringshyppighet | | |
|--------------------------------|------------------------|--------------|-------------|
| | løpende ¹⁾ | årlig | annen |
| 1. Kommune | x | | |
| 2. Navn | x | | |
| 3. Adresse | x | | |
| 4. Næringsgruppe | x | | |
| 5. Størrelsesmål | x | | |
| 6. Eierforhold | x | | |
| 7. Foretaksnummer | x | | |
| 8. Data for utvalgte foretak | | x(f.o.m. 58) | |
| 9. Produksjons-/omsetningsdata | | x(" 65) | |
| 10. Regnskapsdata | | x(" 67) | |
| 11. Bedriftstellingsdata | | | x(for 1963) |

1) Gjennomføring av det nåværende system med entydige og permanente identifikasjonsnummer ble foretatt på grunnlag av Bedriftstellingen 1963. Den "løpende" registrering betyr her at en foretar registrering av endringer så snart en får meldinger om dette.

Kostnadene ved å legge om registreringen i vårt sentrale bedrifts- og foretaksregister til permanente identifikasjonsnummer var ubetydelig, og de antas ikke å ha medført nevneverdige tilleggsutgifter i forhold til det system som tidligere ble nyttet.

Hittil har en verken hatt ressurser eller tilstrekkelig lange tidsserier til å ta opp longitudinelle analyser på grunnlag av arkivets individuelle data. Byråets forskningsavdeling har imidlertid tatt opp arbeidet med å utnytte tidsseriene for de utvalgte store foretak, bl.a. med sikte på å studere bedriftenes investeringsatferd.

3. NYE INFORMASJONSBEHOV

3.1 Generelle og spesielle behov

Byrådet registrerer med tiltakende styrke behov og etterspørsel for statistisk informasjon av følgende slag:

- a. Standardbearbeidinger for geografiske standardområder
- b. Spesielle bearbeidinger for geografiske standardområder
- c. Standardbearbeidinger for geografiske spesialområder
- d. Spesielle bearbeidinger for geografiske spesialområder

Bak etterspørselen i gruppe a. ligger det alminnelige behov for å kunne orientere seg i samfunnet. Å tilfredsstille dette behov betyr en videreføring av Byråets alminnelige publiseringsprogram.

I tillegg merkes en stadig økende etterspørsel etter bearbeidinger som kan tilfredsstille individuelle og spesifikke brukerbehov. Pågang etter spesielle bearbeidinger kommer fra offentlige organer i forbindelse med utrednings- og planleggingsoppdrag. De kommer også fra organisasjoner eller private som i sin virksomhet trenger kunnskap om detaljer som ikke er alminnelige etterspurte, og som derfor ikke kommer i våre publikasjoner som standardbearbeidinger. Brukerne kan også være samfunnsforskere med nye problemstillinger som tidligere ikke har vært gjenstand for statistisk belysninger og derfor krever spesialbearbeidinger.

Den spesielle bearbeiding som etterspørres, varierer sterkt i de krav som stilles, fra relativt enkle tabuleringer til tallfesting av strukturene i omfattende matematiske modeller. Ofte forutsettes det bruk av data som vanligvis ikke innhentes samlet, men i forskjellige prosesser eller/og på forskjellige tidspunkter. Den mest hensiktsmessige måte en kan tilfredsstille slike spesielle behov på, er å etablere et arkivstatistisk system hvor det er overkommelig å gå tilbake til grunnmaterialet, kople sammen data fra forskjellige enheter, og deretter gjennomføre spesielle bearbeidinger når behovene melder seg.

Etterspørselen har også i de senere år fått en annen dimensjon som blir stadig mer utpreget: Brukerne vil ha regional informasjon. Her kan man sondre mellom de brukere som har behov for mer informasjon for geografiske standardområder, som fylker, kommuner, handelsdistrikter etc., og de som har ønsker om data for spesielle geografiske områder avgrenset f.eks. ved koordinatbestemte geografiske polygoner. Det er særlig i forbindelse med distriktsutbygging, byplanlegging, utvikling av kommunikasjonsnett, etc. at den sistnevnte type av behov gjør seg sterkt gjeldende.

Å tilfredsstille behovet for regional statistikk på spesielle områder vil kreve at vi innfører og vedlikeholder et system som tillater en geografisk klassifisering av våre individuelle statistiske enheter i klasser representert ved et finmasket koordinatnett.

3.2 Behov for mer informasjon om befolkningen

Grunnlaget for all analyse av befolkningsutvikling og personatferd er selvsagt kjennskap til de demografiske forhold. Selv om demografiske studier har vært og er en oppgave det blir arbeidd med på løpende basis, er det neppe tvil om at det ennå er mange spørsmål som må besvares før vi har gitt en tilfredsstillende beskrivelse av fruktbarhet, ekteskapsinngåelse, flyttinger, dødelighet etc. Her vil det være av betydning eksplisitt å trekke inn som eksogene variable flere av de faktorer som kan antas å bidra til forklaring av de demografiske variasjoner og som kan betraktes som styringsinstrumenter i planleggingsprosesser.

Blant de hyppigst framhevede er behovet for økt kunnskap om befolkningens reaksjoner med hensyn til å skaffe seg utdanning under forskjellige forhold. Flere og flere peker på at vårt samfunns relative evne til å hevde sin plass i utviklingen internasjonalt, vil avhenge av våre kunnskaper. For planlegging med sikte på å heve befolkningens generelle kunnskapsnivå med en sammensetning som utnytter våre naturlige forutsetninger på den beste måte, kreves imidlertid statistisk innsikt i de faktorer som påvirker folks innstilling til å skaffe seg utdanning. Tradisjonelt har en måttet nøye seg med utdanningsfrekvenser i de forskjellige skoleslag etter kjønn og alder. Slike frekvenser kan være nyttige for preliminare betraktninger og prognoser. Skal en imidlertid ha håp om å nå fram til relasjoner som kan nyttes i planlegging og styring av utdanningssystemet, må en søke lenger ned i detaljene for å kunne avdekke de mulige strukturer mellom utdanningsvalg og de personlige og miljømessige

faktorer som f.eks. tidligere utdanning, ekteskapelig stilling og personlig inntekt, foreldrenes utdanning, yrke, inntekt, familiestørrelse, bostedets arbeidsforhold, dets utdanningstilbud og økonomi.

Et annet viktig område hvor samfunnsplanleggingen nå gjør seg sterkt gjeldende, er analyse av tilgangen på arbeidskraft. Også på dette felt er forholdene i stor utstrekning hittil bare beskrevet ved det relative antall arbeidstakere i forskjellige yrker etter alder, kjønn og næring på noen få tidspunkter. Disse frekvenser har begrenset verdi som utgangspunkt for prognoser som skal gi grunnlag for beslutninger fordi de ikke gir noen innsikt i sannsynlig virkning på arbeidskraften av endringer i de størrelser myndighetene kan påvirke. Det er nødvendig med mer informasjon om hvordan arbeidsstyrken reagerer under alternative forhold for å se om f.eks. behovet for arbeidskraft, slik det kommer til uttrykk i økonomiske planer, lar seg dekke gjennom den tilgang som sannsynlig vil genereres under de forhold planene medfører. De personlige utdanningsfaktorer vil uten tvil være av stor betydning. Likeens må en anta at tidligere yrkeserfaring må anses å være en sentral faktor.

Norsk helsestatistikk har hittil ikke kunnet gi viktige tall om sykkelighet av forskjellig art og varighet hos personer i forskjellige yrker, alder og regioner. "Vi vet hva folk dør av, men lite om helsetilstanden hos de levende." Dette er opplysninger av stor betydning som helseadministrasjonen trenger, både for å kunne vurdere effektiviteten av tiltak som er foretatt, og som grunnlag for videre rasjonell utbygging av helsevesenet. Helsevesenet kan i denne sammenheng oppfattes som en samfunnsvid prosess hvor innsatsfaktorene er helsepersonalets størrelse og sammensetning, og institusjoners karakter og kapasitet. Problemet er å dimensjonere disse faktorer slik at sykkelighet i befolkningen reduseres. Men da må de forskjellige relasjoner mellom innsatsfaktorer og resultater klarlegges. Sykkeligheten er også en viktig størrelse ved bedømming av landets arbeidsstyrke og i hvilken utstrekning denne kan påvirkes. Det kan her nevnes at resultater fra helseundersøkelsen høsten 1968 tyder på at 10-20 prosent av den norske befolkning til enhver tid har redusert arbeidsaktivitet på grunn av en eller annen skade eller sykdom. Informasjon om hvordan denne sykkelighet slår ut i spesielle yrker og alderstrinn, om tidligere yrkesaktivitet og miljø har betydning osv. synes å være av største betydning for å kunne konsentrere helsearbeidet om de viktigste oppgaver.

I inntektsanalyser er det behov for å få fram det som karakteriserer inntektstakerne i de forskjellige inntektsgrupper. Hittil har analyser av personlige inntektstakere måtte innskrenkes til det som framgår av de data

som registreres i liknings- og skattesektoren. Det er imidlertid ønskelig å få klarlagt sammenhenger mellom inntekt og inntektstyper, forbruk, utdanning, yrke, næring osv. for om mulig å kunne få bedre innsikt i de eventuelle konsekvenser endringer i nevnte faktorer kan tenkes å ha. Av spesiell betydning for den offentlige personaladministrasjon er behovet for mer informasjon om statsfunksjonærenes lønnsforhold. I tillegg til den ordinære lønnsstatistikk er det her behov for å belyse hvordan og med hvilken hastighet de forskjellige grupper av tjenestemenn beveger seg i lønssystemet. Videre synes det å være behov for å få innsikt i samspillet mellom de personmessige variable og lønnspolitikken i staten.

Også på andre områder som krimnologi, sosiologi, valgatferd, fritidsbruk, osv. er det et økende behov for å kunne gi informasjon som kan belyse konsekvenser og virkninger av forskjellige tiltak.

Felles for alle de ovenfor nevnte problemer er at de vil kreve data som en fortsatt utbygging av persondataregistrering og -arkivering vil gi.

For å sikre sammenliknbarhet og konsistens i den informasjon som er omtalt, vil det være behov for å etablere visse definisjonsmessige overbygninger. En oppgave er å etablere et befolkningsregnskapssystem med analogier til nasjonalregnskapet. Et slikt system vil imidlertid være begrenset til problemstillinger hvor det er tallet på personer etter forskjellige kjennetegn som er det sentrale. En annen og mer omfattende oppgave er å etablere et sett av konsistente definisjoner av personkjennemerker som sikrer konsistens også på individualnivå.

3.3 Behov for mer informasjon om økonomiske enheter

Bedre utnytting av våre økonomiske ressurser vil kreve detaljert innsikt i de relasjoner som bestemmer næringslivets struktur og utvikling. Relasjoner som beskriver produksjons- og omsetningsstrukturer og som er basert på data for vide næringsgrupper vil neppe være tilfredsstillende i det lange løp for dette formål. Faktorer som ligger skjult, f.eks. fordelinger innen den enkelte gruppe, kan vise seg å være vesentlige. En må forutse et behov for studier som i større utstrekning enn hittil vil kreve estimering på grunnlag av individualdata.

Ikke bare bedriftenes tekniske produksjonsstruktur, men deres tilpassingsmønstre med hensyn til produktutvalg, investeringer, etterspørsel og bruk av arbeidskraft, lokalisering, markedsføring, etc. er vesentlige sider ved problemstillingen. På grunn av treghet som ofte vil gjøre seg gjeldende ved reaksjoner og atferd av denne type, vil en ofte ha behov for

å studere atferd i relasjon til tidligere egenskaper, erfaringer og resultater ved bedriftene og det næringsmiljø de arbeider innenfor. Disse spørsmål lar seg neppe besvare uten gjennom longitudinelle studier med bruk av data som beskriver de individuelle økonomiske enheter utviklingen over tiden. De individuelle egenskaper som her vil være relevante, spenner fra de rent tekniske til de finansielle faktorer.

Et område hvor det er foretatt en lang rekke bearbeidinger og analyser, er utenrikshandelen. Analysene knytter seg her imidlertid ikke til de enheter som tar beslutninger om eksport og import, men til de varer som eksporteres og importeres. For å få dypere innsikt i de faktorer som påvirker importens og eksportens sammensetning og størrelse, kan det være ønskelig å knytte relasjoner mellom størrelsen på og arten av de varemengder som eksporteres og importeres, og data som karakteriserer de enheter som tar beslutninger, leverer og mottar varene.

En spesiell type økonomiske enheter er husholdningen eller familien i situasjoner hvor den opptrer som inntekts- eller forbruksenhet. Det har her vært skissert problemstillinger som forutsetter simulering av framtidig utvikling for et utvalg av slike enheter. Estimering av strukturen i simuleringsmodellen vil imidlertid kreve et videre utbygd datagrunnlag.

4. REGISTERPROBLEMER

4.1 Utbygging av personregisteret

Med register mener vi en liste, på papir, hullkort, magnetbånd, etc., med navn, adresse, m.m. over en bestand av en bestemt type av enheter på et gitt tidspunkt. For at registeret skal kunne holdes ajour med hensyn til tilgang og avgang av enheter fram til et nytt tidspunkt, må hver enhet i listen være utstyrt med et entydig og permanent identifikasjonsnummer eller annen identifikasjon, og det må etableres rutiner som sender melding om tilgang og avgang til registeret.

Når det gjelder personregistreringen, må vi anta at det sentrale personregisteret vil tjene de fleste formål. Dette utelukker selvsagt ikke at det kan være praktisk med spesialiserte registre for forskjellige befolkningsgrupper.

Det er imidlertid et stort og foreløpig udekket behov for registrering av en beslektet type enheter, nemlig familier og hushold. I mange, viktige sammenhenger er det disse enheter, ikke de enkelte personer, som er

de naturlige atferdsenheter. Det er som eksempler nok å peke på flytteeatferd, boliggetterspørsel og etterspørsel etter forbruksvarer.

En kunne tenke seg etablert og vedlikeholdt et særskilt register over familier eller husholdninger som da ville utgjøre det naturlige utgangspunktet for innsamling og arkivering av data om disse enheter. Mer hensiktsmessig vil det antakelig være å bygge videre på de registreringsrutiner som allerede eksisterer, nemlig ved å bygge ut personregisteret til å omfatte en felles familieidentifikasjon for hver person i samme familie. En slik utbygging vil være mulig i forbindelse med Folketellingen 1970 og den løpende folkeregistrering og vil kunne gjennomføres med langt færre praktiske problemer og ressurser enn en selvstendig utbygging etter 1970.

En utbygging av personregisteret som ovenfor antydnet må også antas å være av betydelig interesse utover den forskning som den kan gi grunnlag for.

4.2 Utbygging av bedrifts- og foretaksregisteret

Vårt bedrifts- og foretaksregister omfatter langt fra alle bedrifter og foretak i vid forstand. Det arbeides med flere utvidelser, blant annet for enheter fra jordbruks-, helse- og utdanningssektorene hvor det synes å foreligge gode muligheter for at et register, når det først er etablert, vil kunne vedlikeholdes med rimelige omkostninger.

Et registers verdi øker som nevnt sterkt med antall brukere som nytter det som felles referanse for avgrensning og identifikasjon av enheter. Det må åpenbart være en viktig framtidig oppgave for Byrået å tilpasse og utbygge det sentrale bedrifts- og foretaksregister slik at det kan nyttes også av andre offentlige myndigheter, jfr. utviklingen når det gjelder personregisteret. Av særlig stor betydning vil det være å tilfredsstille liknings- og skatte-myndighetenes behov for et sentralt register over ikke-personlige skattytere og arbeidsgivere. Et felles register vil for statistikken blant annet bety at administrativt innsamlede skatte- og regnskapsopplysninger vil kunne kobles sammen med opplysninger som Byrået samler inn direkte. Selv om en slik data-registrering neppe vil kunne komme skikkelig i gang på mange år, er det viktig å utbygge og eventuelt omlegge Byråets nåværende bedrifts- og foretaksregister til et sentralt felles register så snart som mulig. Det vil være for sent å ta opp dette når administrativ dataregistrering eventuelt er i full gang på et annet registergrunnlag.

I et sentralt bedrifts- og foretaksregister vil det finnes mange enheter som også vil være enheter i personregisteret. Dette gjelder f.eks. eiere av personlige foretak. Det er ikke utenkelig at den mest rasjonelle løsning på lang sikt, både fra administrativ og statistisk synsvinkel, vil

være et universalregister som omfatter såvel personer som ikke-personlige enheter. Et slikt universalregister vil eliminere en del dobbelt vedlikeholdsarbeid som ellers vil være nødvendig. Det vil dessuten gi grunnlag for en enklere sammenkopling av opplysninger som er innhentet for enheten delvis som person og delvis som økonomisk enhet. Dette er imidlertid praktiske spørsmål som det vil føre for langt å diskutere videre her.

I bedrifts- og foretaksregisteret registreres det to typer av enheter, bedrifter og foretak, slik disse enheter er definert i Norsk Standard for Næringsgruppering. Vi må imidlertid være merksam på at det for framtiden kan bli behov for - i innsamling, bearbeiding og analyse - å ta vare på data også om andre beslektede enheter som f.eks. arbeidssted avdeling. Slike behov har Byrået allerede støtt på i arbeidet med å bygge ut registeret til å omfatte enheter innenfor helse- og undervisningsvesen. Det er viktig at en i arbeidet med disse nye enheter nyttiggjør seg forarbeider, systemopplegg og erfaring som er gjort i foretaks- og bedriftsregistreringen hittil. Om en velger formelt å bygge opp registrene for arbeidssteder, avdelinger o.l. som særregistre, eller om en utvider rammen for den nåværende registrering av bedrifter og foretak, slik at en får plass for dem, må være et spørsmål om hva som er mest hensiktsmessig.

4.3 Adresse/eiendomsregister

For å tilfredsstillere behovet for statistiske bearbeidinger for spesielle geografiske områder, er det nødvendig med geografisk klassifisering av personer og økonomiske enheter som er langt mer detaljert enn den en får fram ved å bruke kommunene som minste geografiske område. Etter som det er langt flere personer og bedrifter enn forskjellige adresser, er det mer hensiktsmessig å geografisk klassifisere hver adresse enn hver person og bedrift. Dette vil best kunne løses ved oppretting av et adresseregister som inneholder en beskrivelse som gjør det mulig å få en entydig kobling av personer og økonomiske enheter til adresseenhetene. Med adresse menes her enhver tilgjengelig stedsspesifikasjon, ikke bare adresse i postal betydning.

Den mest detaljerte klassifisering oppnår en ved å nytte de geografiske koordinater for meget små ruter, tilnærmet punkter som de statistiske enheter kan gis en eller annen spesifisert tilknytning til. Dette går under betegnelsen som "koordinatmetoden", og har vunnet stor tilslutning i f.eks. Sverige. Grunnlaget for gjennomføring av en mer detaljert geografisk klassifisering, har vi hos oss i et allerede vel utbygd system for meldinger om adresseendringer og i det eksisterende kartverk.

Hovedlinjene i etablering og vedlikehold av et slikt register, vil være følgende:

Fra våre to basisregistre, personregisteret og registeret over bedrifter og foretak, ekstraheres alle forskjellige adresser som forekommer. En regner med at det anslagsvis er registrert nærmere 1 million forskjellige adresser i våre registre. For hver adresse registreres sentralkoordinaten for den rute adressen tilhører innen et på forhånd opptrukket rutenett. Koordinatfesting må gjøres lokalt og ved hjelp av eksisterende kartverk hvor det fastsatte rutenett inntegnes. Regulerings- og utbyggingsmyndighetene er allerede sterkt engasjert i andre sider av samme sak, og Byrået kan derfor tenkes å etablere et slikt landsomfattende register uten vesentlige eksterne kostnader.

Enkelte steder vil adressene være mindre spesifisert enn andre steder. Dette gjelder de fleste landkommuner. Det ville her være ønskelig om en i forbindelse med Folketelling og Bedriftstelling kunne foreta en løpende nummerering av bosteder og arbeidssteder og få de lokale myndigheter til å nytte disse som pseudo-adresser ved koordinatbestemmelsen. En annen, men ikke fullt så tilfredsstillende framgangsmåte, kan være å nytte gårds- og bruksnummereringen fra Folketellingens huslister som en utfyllende adressekomponent.

Enkelte områder vil ha dårlige kartverk eller spredt bebyggelse. I disse vil vi kunne arbeide med større ruter i første omgang. Det kan være aktuelle alternativer å arbeide med store ruter på 1 000 m x 1 000 m, mellomstore ruter på 500 m x 500 m eller små ruter på 100 m x 100 m. Andre kommuner har eller er i ferd med å opprette registre over koordinatregistrerte eiendommer. Disse vil uten videre kunne nyttiggjøres ved eiendomsadressen eller ved gårds- og bruksnummer som nøkkel. Den nøyaktighetsgrad en regner med ved "punktregistrering" svarer til ruter på 10 m x 10 m.

Adresseregisteret holdes ajour med tilgang av nye adresser ved at adresser som innføres i basisregistrene, kontrolleres mot adresseregisteret. Dersom en adresse ikke finnes, sendes den nye adresse ut til den lokale kontakt for koordinatbestemmelse.

Adresseregister og adressenes geografiske klassifikasjon må også sees i sammenheng med det pågående arbeid med å få etablert et jordregister. Det har vært stor interesse for at den maskintekniske del av oppretting og vedlikehold av dette registeret skal legges til Byrået. I forbindelse med registeret skal også samles inn visse data Byrået vil være interessert

i, blant annet geografiske koordinater for de enkelte jordstykker, og Byrået bør samarbeide på dette område. Ellers vil vel dette registeret stå på linje med sentrale register over andre kapitalgjenstander som biler, skip, etc., og ha et noe annet perspektiv enn adresseregisteret.

Med den sentrale rolle adresseregisteret vil spille for dekning av det økende behov for regional statistikk av forskjellige slag, vil det være meget uheldig om Byrået ikke får høve til å utnytte de muligheter som de store tellinger i begynnelsen av 1970-årene kan gi i utbygging av et landsomfattende adresseregister.

5. DATAINNSAMLING

5.1 Direkte og indirekte innsamling

Med datainnsamling mener en innsamling av data identifisert til de enkelte enheter som inngår i registrene. Innsamlingen kan foregå enten ved direkte innhenting eller ved å overta data som er registrert i en eller annen administrativ prosess. Datainnsamlingen kan skje fortløpende eller i et periodisk mønster. Den fortløpende innsamling vil gi grunnlag for å få aktuelle informasjoner hyppigere og vanligvis mer fullstendig. Felles for all datainnsamling innenfor et arkivstatistisk system er at det entydige og permanente identifikasjonsnummer alltid må nyttes som identifikasjonsopplysning.

De informasjonsbehov som er skissert foran vil kreve en langt mer omfattende datainnsamling enn den vi i dag foretar. Det synes imidlertid som om det meste av den videre datainnsamling som er ønskelig, vil kunne begrenses til at det sentrale dataarkiv systematisk blir tilført allerede eksisterende administrative data. Det er riktig nok slik at dette materiale foreløpig ofte registreres på en måte som ikke helt tilfredsstillende statistiske krav, og det må bli en oppgave for Byrået, i samarbeid med administrative organer å få dataregistreringen lagt om hvor dette er nødvendig. Målet må være at de administrative organer kan dra nytte av de sentrale registre som Byrået oppretter og holder vedlike, mens de til gjengjeld innretter sin datainnsamling slik at de foruten å imøtekomme egne behov, bidrar til å dekke samfunnets generelle informasjonsbehov så rasjonelt og billig som mulig.

Et av de første og viktigste forarbeider er å etablere en kjenne-merkekatalog med definisjoner av de individuelle kjennemerker. En slik

katalog vil ha to formål: Den skal sikre at det overalt blir arbeidet med konsistente begreper, og den skal bidra til standardisering og riktig bruk av de forskjellige data. På grunn av den omfattende mengde av begreper som registreres, er dette et stort standardiseringsarbeid.

Administrativt registrerte data vil vanligvis være fullstendige datasett i den forstand at de vil omfatte alle enheter i en masse. Bare unntaksvis vil en arbeide med utvalg. Foreløpig må en anta at fordelene ved å ha fullstendige data vil være så store at det berettiger fullstendig arkivering. I tillegg til disse data vil det også bli behov for at Byrået innhenter flere data ved direkte innsamling av data når det ikke er mulig å få disse fra administrative prosesser. Ved direkte innhenting vil utvalg spille en stor rolle fordi de faller så mye billigere. Men også ved utvalgsundersøkelser bør det sørges for at individuelle data blir identifisert ved de offisielle identifiseringsnummer.

5.2 Persondata

Innsamling av data om individuell utdanning foregår nå i tilknytning til examen artium og adademiske studier. Behovet krever imidlertid at en samler inn individuelle utdanningshistorier som beskriver de enkeltes vei gjennom utdanningssystemet, det høyeste utdanningsnivå de har nådd og helst et mål for den grad av kunnskap de representerer, f.eks. i form av oppnådde eksamensresultater.

En viktig målsetting ved Folketellingen 1970 er at den skal registrere en bestand som blant annet er karakterisert ved utdanningsnivåer. Dersom dette ikke fortsatt skal bli en ti-årig foreteelse, må vi allerede ved tellingstidspunktet ha innført et opplegg for løpende registrering av utdanningsdata. Grunntanken for et slikt løpende vedlikehold av utdanningsdata er at det for hvert enkelt kurs skal samles inn elevlister med fødselsnummeridentifisering såvel ved kursets begynnelse som ved dets avslutning. Avbrudd og liknende forutsettes rapportert løpende. En slik registrering skulle foruten å gi det informasjonsgrunnlag en søker, være av betydelig interesse for undervisningsmyndighetene med sikte på å forenkle og forbedre kommunikasjonssystemet i skolesektoren.

Et stort problem i denne sammenheng er hvordan en best kan få gjort fødselsnumrene lett tilgjengelige i systemet. På langt sikt vil løsningen etter all sannsynlighet ligge i at numrene påføres alle fødselsattester og følger elevene gjennom systemet ved hjelp av avgangsvitnemål fra de forskjellige skoler og kurs. På kort sikt må en imidlertid finne andre løsninger som må antas å bli mer kostbare.

Forutsetningen for en vellykket innsamling er at før den etableres, må de nødvendige klassifikasjonsstandarder og definisjoner utarbeides med sikte på å ivareta de viktigste trekk ved utdanningshistorien og gi grunnlaget for en ensartet og tilfredsstillende registrering. Disse forarbeider er allerede under utvikling.

Det er liten grunn til å anta at undervisningsanstaltene eller de sentrale undervisningsmyndighetene vil ha behov som gjør at de vil kode og punche dette materialet. En betydelig del av disse oppgavene vil falle på Byrået.

Befolkningens yrkesaktivitet, yrke og næring (arbeidssted) er kjennemerker som tradisjonelt blir registrert i folketellinger. For å tilfredsstille informasjonsbehovene er det imidlertid ønskelig med en hyppigere registrering. Alle data som er nødvendige for å fastsette verdien på disse kjennemerker, registreres allerede i en eller flere administrative sammenhenger om enn noe upresist. Det som trengs er først og fremst en rutine for innsamling av data fra disse kilder.

Yrkesaktivitet, yrke og arbeidsgiver blir registrert i forbindelse med sykestrygden. Her oppgis arbeidsgiverens beskrivelse av yrke som trolig mer er en stillingsbetegnelse enn en yrkeskarakteristikk. Om arbeidsgiverens rapporter om arbeidstakerne vil bli gitt med fødselsnummer er usikkert. Det er sannsynlig at endring av yrke ikke uten videre betinger noen melding fra arbeidsgiveren til trygdeadministrasjonen. En må derfor regne med at opplysningene om yrke kan være relativt upresise. På den annen side kan selvsagt dette materialet få stor verdi med hensyn til å belyse endringene i graden av yrkesaktivitet. Et spørsmål er hvilke behov trygdeadministrasjonen vil ha for å kode og punche disse data. Dersom den ikke har behov for dette, vil materialet måtte kopieres lokalt og kodes og punches i Byrået.

Av denne grunn er trolig følgende alternativ lettere realiserbart. Alle som inngir selvangivelse, skal på denne hvert år gi stilling/yrke. Når disse opplysningene er med på selvangivelseskjemaene er det fordi de likningsmessig anses for nyttige tilleggsopplysninger. I dag kodes og punches store deler av innholdet i selvangivelsen inkl. fødselsnummer og nyttes i automatiske rutiner. Målet for likningsmyndighetene må trolig være å innføre en mest mulig automatisk likning, og en viktig basisopplysning vil her være inntektstakerens yrke.

Forutsetter vi f.eks. at yrke fra 1970 preutfylles sammen med navn, adresse og fødselsnummer på selvangivelsene med oppfordring til utfylleren å korrigere eventuell feil yrkesangivelse, vil det som kreves bli koding og

punching av yrkesdata for dem som ifølge selvangivelsen oppgir å ha endret yrke i løpet av siste året. På lengre sikt må en vente at likningsmyndighetene vil finne det verdt å utføre denne koding og punching.

Grunnlag for koding av næring finnes også på selvangivelsen som skal gi arbeidsgiverens navn og adresse. Her er det imidlertid mer tvilsomt om likningsmyndighetene er interessert og vil makte en koding av arbeidsgivers næring.

En tredje mulighet som foreligger er å samle inn de data som registreres for skattetrekkblankettene. For dem som det gis melding om, vil dette omfatte fødselsnummer, arbeidsgivers angivelse av lønnstakerens stilling/yrke og eget arbeidsgivernummer. Her er det særlig det siste som vil være interessant. Dersom likningsvesenets arbeidsgivernummer og Byråets foretaksnummer var felles, ville dette gi grunnlag for å koble vår standard-spesifisering av foretak etter næring til arbeidstakernes inntektsdata. Da dette alternativ vanskelig kan utnytte de yrkesdata vi får gjennom Folketellingen som basis, må vi imidlertid anta at yrkesopplysningene vil bli dårligere og kreve vesentlig mer arbeid i form av koding og punching enn i selvangivelsesalternativet.

Innen helsevesen og medisinsk forskning er det for tiden stor interesse for bruk av fødselsnummer som identifikasjonsmiddel. Her journalføres som kjent en rekke individuelle medisinske data om pasientene, men på uensartet form. Dersom en fikk systematisk samlet inn disse opplysninger, ville en trolig kunne tilfredsstillende betydelige behov både direkte på det helsestatistiske område og indirekte på andre områder.

Det er her to sentrale problemer. For det første må helsesektoren få tilfredsstillende adgang til fødselsnummeret. Dette vil kreve en innsats fra Byrådet for å yte den registerservice dette vil medføre. Løsningen kan enten bli at Byrådet gir regelmessige registeropplysninger til et sentralt helseorgan som foretar en videre fordeling innen sektoren. Dette er uten tvil den for Byrådet gunstigste løsning. Et annet alternativ vil være at Byrådet også påtar seg fordelingsfunksjoner etter en avtalt fordelingsplan. Som på en rekke andre felter, er det vi virkelig savner, et opplegg som gjør det mulig for publikum selv å identifisere seg med fødselsnummer. En kan kanskje tenke seg muligheten av å innføre såkalte "helsekort" som må medbringes ved helsekonsultasjoner. Dersom dette lot seg gjennomføre, kunne en senere gjøre publikum merksam på at de samme helsekort også kunne nyttes i andre forbindelser når det gjaldt å identifisere seg ved fødselsnummer, f.eks. når en skulle oppgi skattnummer til arbeidsgiver.

Det andre problemet, som Byrådet bare delvis kan bidra til å løse, er den nødvendige standardisering i helsesektorens registreringer. Det er behov for definisjoner og klassifikasjoner m.v. på en lang rekke felter, fra de rent medisinske aspekter og den behandling som gis, til økonomiske og kostnadsmessige sider av helsetjenestene.

Her synes det å være mange analogier og likheter mellom utdannings- og helsesektorene, men sannsynligvis er det enklere å løse problemene i utdanningssektoren, fordi det i denne ikke synes å være behov for den samme grad av detaljert personbeskrivelse som i helsesektoren.

Når identifikasjons- og klassifiseringsproblemene er løst, vil en praktisk registreringsløsning være etablering av et sentralt utskrivningsarkiv. Alle helseinstitusjoner pålegges å gi visse minimumsdata om utskrevne/døde/friskmeldte pasienter. Hovedbegrunnelsen for dette arkiv vil ikke være den statistisk-vitenskapelige verdi det vil få, men den medisinske beredskap det vil kunne sette helsevesenet i.

I stor utstrekning synes en vellykket datainnsamling fra helsesektoren å avhenge av at det også tas i bruk EDB. Interessen er for tiden meget stor selv om de oppnådde praktiske resultater hittil ikke på noen måte står i relasjon til de mål det snakkes om. Foreløpig er det særlig i arbeid omkring kreft- og tuberkulosesykdommer hvor det nyttes fødselsnummer, standardbeskrivelseskoder og EDB.

Allerede i flere sammenhenger nyttes nå fødselsnummer på data fra politi, fengsel og rettsvesen, og det må antas at det innen denne sektor vil bli stadig mer nyttet. Det Byrådet her ville kunne bidra med, vil først og fremst være å tilby og yte mer effektiv registerservice som vederlag for data som mottas.

Også på dette felt er det rimelig å tro at EDB vil gjøre sitt inntog i de nærmeste årene, og vi bør trolig avvende denne utvikling.

5.3 Bedrifts- og foretaksdata

Når det gjelder bedrifter og foretak, har hovedtyngden av datainnsamlingen tradisjonelt vært direkte innhenting. Med økende behov for felles styring av næringsliv og økonomi, vil uten tvil en standardisering av regnskapsførselen tvinge seg gjennom før eller senere. En standardiserende regnskapslov vil gi to datainnsamlingalternativer: Byrådet kan innhente regnskapsdata direkte gjennom spesielle regnskapsundersøkelser og bygge sin regnskapsstatistikk på denne. Med en standardiserende og bindende regnskapslov, er det imidlertid lite rimelig at spørsmål fra Byrådet som går ut over rammen for eller på kryss av de begreper som loven fastlegger, vil

bli særlig godt besvart. Det synes både å være rimeligere og trolig like tilfredsstillende å inngå avtaler med de administrative organer som vil innhente regnskapsdata for administrative formål om å få overta disse data.

Dersom vi på forhånd oppnår å samordne vårt eget foretaksregister med de registre som vil bli benyttet i forbindelse med registrering av regnskapsdata for administrative formål, vil vi her kunne kombinere slike regnskapsdata med våre egne innhentede produksjonsdata. De uoverensstemmelser som vi nå etter hvert oppdager i våre produksjonsdata sammenliknet med regnskapsdata, skyldes vel delvis at vi nytter andre bedrifts- og foretaksdefinisjoner enn de administrative organer og delvis at regnskapsførselen bygger på uensartede begreper.

Med den utvikling vi nå kan vente innen den administrative sektor i forvaltningen med hensyn til bruk av elektronisk databehandling, vil innsamlede regnskapsdata etter all sannsynlighet bli overført til hullkort og magnetbånd for å kunne tjene de administrative formål. Vi må derfor kunne regne med at når vi får en regnskapslov, vil vi relativt snart også kunne overta regnskapsdata på magnetbånd fra administrative organer som liknings-, skattemyndigheter, prismyndigheter, etc.

Fra begynnelsen av 1970-årene vil vi få et merverdiavgiftssystem. Ifølge de foreløpige forslag forutsetter dette en relativt hyppig rapportering fra de forskjellige foretak om omsetning i løpet av rapporteringsperioden.

Dersom administrative organer overhode skal kunne nyttiggjøre denne informasjon til kontroll, synes det ikke å være noen vei utenom automatisk databehandling av materialet. Dette vil bety at en for store deler av vårt næringsliv hver annen måned vil få omsetningsoppgaver på magnetbånd. Slike oppgaver må, selv om de er lite spesifiserte, være en vesentlig forbedring av grunnlaget for korttidsstatistikk og -analyser, dersom de kan kobles sammen med data for foretakene innhentet fra andre kilder. Byråets oppgave må derfor være å søke å påvirke opplegget for innhenting av disse oppgaver i retning av det benyttes registreringsenheter som svarer til de enheter vi finner det riktig å arbeide med, og at de definisjoner m.m. som nyttes, er i samsvar med våre ønskemål.

En oppgave som det dessverre ikke er så enkelt å se noen løsning på, er en registrering av import og eksport med en foretaksidentifikasjon som svarer til det som ellers nyttes. En kan kanskje tenke seg at skattemyndighetene i sin tid av kontrollhensyn vil være interessert i å få registrere handelen med utlandet for hver skatteenhet. Den praktiske løsning ville da være at eksport-/importskjemaene ble påført eksportørens/

importørens identifikasjonsnummer. Dersom dette var identisk med Byråets foretaks-/bedriftsnummer, vil vi også her få data som lå vel til rette for analyse av utenrikshandelen etter næring, produksjonsstruktur, størrelse, etc. hos importør og eksportør.

6. DATAARKIVERING

6.1 Arkivsystem

I hvilken utstrekning en økt dataregistrering etter de linjer som er skissert foran, kan bidra til å bedre informasjonsservice fra Byrået, vil avhenge av at vi organiserer data i arkivet slik at vi lett kan finne fram til og trekke ønskede data ut av arkivet. I de siste årene har det vært arbeidd med systemer for såkalte databanker. De arkiver Byrået søker å etablere er både mer omfattende og tildels forskjellige fra de vanlige databanksystemer.

Det som gjør våre dataarkiver forskjellige fra databankene er:

1. Til et meget stort antall enheter, nærmere 4 millioner personer, 100 000 - 200 000 bedrifter og foretak, osv., knytter vi mange kjennemerker over hele livshistorien slik at hver enhet kan være beskrevet av mange tusen tegn.
2. Vi får data ordnet i puljer fra forskjellige kilder, f.eks. demografiske data fra folkeregistreringen og inntektsdata fra likningsvesenet.
3. Vi søker uttak ikke av enkeltopplysninger, men av hele datasett.
4. Behovet for å arkivere og trekke ut data umiddelbart når den registrerte begivenhet skjer eller når en problemstilling er formulert, er svakt og berettiger neppe kostnadene ved såkalte "sann tid" opplegg for hele arkivet slik som de fleste databankopplegg forutsetter.

Byrået har arbeidet med arkiveringsspørsmål med sikte på å kunne komme fram til en løsning som tilfredsstiller våre spesielle krav. I denne sammenheng kan det her være grunn til å trekke fram 3 av de mest sentrale spørsmål som melder seg:

- 1) For å kunne orientere seg i arkivet på en forsvarlig måte er det behov for å opprette og vedlikeholde en kjennemerke-katalog for alle de kjennemerker som er registrert og arkivert i arkivet. Katalogen må inneholde hvert kjennemerkes offisielle betegnelse, eventuelle synonyme

betegnelser, kjennemerkets definisjon operasjonelle definisjoner som angir hvordan kjennemerket skal registreres, og referanse til det eller de delarkiver som inneholder data om kjennemerket. Slik katalogisering av kjennemerker eksisterer i dag bare delvis.

2) For å redusere letetid må arkivet ordnes i delarkiver, lagt opp på en slik måte at de kan kobles sammen etter behov. Den ideelle oppdeling er den som har minst forventet letetid/kostnad i forbindelse med arkivering og uttak fra arkivet. Den ideelle løsning vil trolig variere over tiden, avhengig av arkivets innhold og bruk. Hittil er vårt arkiv ordnet i delarkiver etter registreringskilde og årgang. Innenfor de enkelte delarkiv er enhetene vanligvis ordnet etter stigende identifikasjonsnummer.

3) For å holde rede på hva de forskjellige felter i delarkivet, inneholder, må vi også ha et format- eller beskrivelseskartotek. Det holdes i dag ajour på maskinteknisk form.

6.2 Tekniske spørsmål

Lagring av data kan teknisk sett skje i to forskjellige typer av lager, oppslagslager og serielager. Oppslagslager er kjennetegnet ved at lagerplassene er tilgjengelige like raskt i hvilken som helst rekkefølge, og at lagerkostnadene er høye. Serielageret, først og fremst representert ved magnetbånd, gir en lavere lagerkostnad, men lagerplassene er bare tilgjengelige i en bestemt rekkefølge.

Med de gjeldende kostnadsforhold mellom de to lagertyper, vil magnetbåndet være det som vil være mest effektive for et arkiv som skal betjene statistiske behov som i store trekk kan forutsees på arkiveringstidspunktet. For betjening av individualoppslag i ukjent rekkefølge og for ukjente framtidige uttak, kan imidlertid direkte oppslagslager være hensiktsmessige. Byrået må trolig innstille seg på å gjøre bruk av begge typer lagre i framtiden.

Med de relativt små forhold vi har i Norge, må en også vente at det vil skje en stadig økende koordinering av det offentliges EDB-anlegg. Det kan føre til at Byråets dataarkiver og registre om noen år, kan tenkes å bli samordnet med andre institusjoners arkiver. Dette vil kreve at vi overveier de tekniske løsninger på spørsmålet om konfidensiell behandling.

7. UTNYTTING AV DATAARKIVER

7.1 Publikasjoner

En utbygd arkivering av individualdata vil etter alt å dømme måtte få konsekvenser for vårt publikasjonsprogram. Vi vil sitte inne med mye mer nyttbar informasjon enn tidligere. En stor del av denne vil bli så allment etterspurt at den kan forsvare å gjøre den offentlig tilgjengelig ved publisering i et offisielt tabellverk. Hvor langt en skal gå i så måte, eller alternativt nøye seg med et beredskap for spesial-utkjøringer etter bestilling, lar seg ikke avgjøre på forhånd. Men et par utviklingstendenser kan forutses med forholdsvis stor sikkerhet.

Arkivene vil gjøre det mulig å lage tabeller basert på kryssklassifisering av enheter etter kjennemerker som hittil ikke har latt seg knytte sammen, fordi opplysningene har vært hentet inn fra forskjellige kilder. Slike kryssklassifiseringer vil ofte være like interessante som allerede etablerte tabuleringer, og vil bli like sterkt etterspurt. Vi har tidligere nevnt som eksempel hvordan inntektsstatistikken kan bygges ut ved at likningsansettelsene kobles sammen med andre opplysninger i persondataarkivet. Tilsvarende eksempler vil kunne hentes fra nær sagt alle felter av individualstatistikken. Det er derfor sannsynlig at et stort antall av våre publikasjoner vil vokse i tykkelse ved at nye tabeller kommer til.

Helt nye publikasjonsserier kan også tenkes. Fullt utbygd vil persondataarkivet på løpende basis inneholde det meste av den informasjon som hittil bare har vært innhentet ved folketellingene. En del av denne informasjon vil det antakelig vise seg hensiktsmessig å offentliggjøre årlig; en begynnelse er for så vidt allerede gjort ved de nye publikasjonsserier, Flyttestatistikk, Folkemengden i kommunene o.a.

Andre muligheter for regelmessige (årlige eller periodiske) publikasjonsserier vil åpne seg ved at hittil ikke geografisk stedfestede data vil kunne kobles sammen med stedfestede data. Det vil kunne gi grunnlag for mer allsidig beskrivelse enn hittil av befolkning og næringsliv i de enkelte geografiske områder. En del av denne informasjon vil offentlige myndigheter og andre ha behov for å få gjort tilgjengelig på løpende basis.

7.2 Dataservice og spesialbearbeidinger

Selv om Byrået er en statistisk institusjon, kan det ikke bli spørsmål om å forbeholde data i arkivene utelukkende for egne statistiske bearbeidinger. Byråets målsetting må være å bidra mest mulig til å dekke

samfunnets informasjonsbehov i forsvarlige former. Parallelt med videreutbygging av analyse- og publikasjonsprogram må Byrået være innstilt på også å tilby informasjon på andre media som er mer hensiktsmessige for brukerne. Både statistikk og arkivdata må tilbys på hullkort, magnetbånd, etc. slik at brukere som ønsker å viderebearbeide data selv kan gjøre det uten unødig punching av noe som allerede finnes på maskinmedia.

De tekniske registrerings- og kommunikasjonsmidler vil etter hvert føre til at informasjon flyter lettere. Teknisk sett er det i dag f.eks. mulig for en statistikkbruker å få overført hele innholdet av månedsheftet for utenrikshandelen fra Byrået over en vanlig telefonforbindelse til egen datamaskin i løpet av den samme tid det tar å spørre og få svar om noen få varer over telefonen.

Et viktig hjelpemiddel for styring av informasjon til de relevante mottakere vil være registrene. Det er grunn til å tro at også andre informasjonsformidlere enn Byrået vil være interessert i å betale stadig mer for å skaffe seg tilsvarende registre. Vi kan derfor regne med stor etterspørsel fra næringslivet etter utdrag av Byråets registre. Det vil være urasjonelt å nekte å stille slike utdrag til disposisjon, fordi det vil føre til samfunnsmessig dobbeltarbeid. Det er derfor meget som taler for at Byrået bør selge slik registerinformasjon, selvsagt avgrenset til slike data som etter gjeldende lovregler til hver tid er offentlige tilgjengelige.

8. ANALYTISKE ANVENDELSER AV PERSONDATAARKIV

De største vinninger ved overgang til et arkivstatistisk system vil antakelig først og fremst vise seg på personstatistikkens område. Her kan vi allerede nå se muligheter som om få år kan gi administrasjonen, og de vitenskapsgrener som har individene og familiene (husholdningene) som studieobjekt, et helt nytt datagrunnlag å bygge på.

8.1 Innholdet av persondataarkivet

Dataarkivet for personopplysninger vil komme til å bestå av:

- (i) et generelt persondataarkiv som inneholder et utvalg av kjennemerker for alle personer fra forskjellige kilder, og
- (ii) forskjellige delarkiv, som for avgrensede grupper av befolkningen inneholder spesielle kjennemerker som er av stor betydning og som registreres for de respektive grupper.

Det generelle persondataarkiv vil inneholde bakgrunnsopplysninger om personene som trengs i mange sammenhenger. Først og fremst vil dette si opplysninger som kan gi en demografisk og sosio-økonomisk karakteristikk av personene. Det er innenfor rekkevidde, som vist foran, i løpet av noen år å bygge opp og vedlikeholde et arkiv som foruten identifikasjonsnummer inneholder følgende opplysninger om alle som er eller har vært bosatt i Norge:

- a. Demografiske data (fødselsår, kjønn, ekteskapelig status før og nå, antall levende og døde barn født med fødselsår for hvert barn og opplysning om legitimitet). Allerede i dag er de fleste av disse data i arkivet, og Folketellingen 1970 vil gi oss grunnlaget for å komplettere disse data.
- b. Familierelasjon til andre personer. Det tas sikte på i tilknytning til Folketellingen å registrere farens og/eller morens og/eller adoptivforeldres og ektefelles identifikasjonsnummer, oppfattet som kjennemerke ved personene, også for begivenheter som inntraff før fødselsnummeret ble innført. Opplysningene vil være nyttige (i) ved at persondataarkivet via foreldres nummer vil kunne gi viktige opplysninger om en persons miljø som barn (ii) ved at vi vil kunne hente fram opplysninger om grupper av personer som er knyttet sammen ved blodsbånd.
- c. Bosted (nå og tidligere).
- d. Utdanning (eksamen, år og kanskje resultat).
- e. Inntekt og formue (skattelikningsdata siste år og tidligere).
- f. Næring nå og tidligere (helst koordinering med bedrifts- og foretaksregisteret, dvs. bedriftens/foretakets nummer oppfattet som kjennemerke ved personen).
- g. Yrkesstatus, nå og tidligere.

I tillegg til ovenstående opplysninger, som vil være tilgjengelige for alle i det generelle dataarkiv, vil spesielle delarkiv inneholde ytterligere opplysninger om deler av befolkningen. Opplysningene i disse delarkiv kan, via identifikasjonsnummrene, kobles sammen innbyrdes og med opplysningene i det sentrale dataarkiv. Det blir et hensiktsmessighets-spørsmål og et kostnadsspørsmål hvor langt en ønsker å gå i retning av å bygge opp slike delarkiver og i hvilken utstrekning de spesielle kjennemerker skal introduseres i de generelle arkiv. Mulighetene for å etablere spesielle delarkiv er meget store som noen eksempler, hentet fra ulike felter, tydelig vil vise:

- a. Ved Jordbrukstellingen i 1969 vil fødselsnummer bli hentet inn for alle som eier jord eller holder husdyr. Opplysningene som jordbrukstellingen gir vil derfor være tilgjengelige og kunne oppfattes som kjennemerker ved disse personer.
- b. Til sosialstatistikken innhentes hvert år opplysninger på individualbasis om barn under sosial omsorg. Her kan fødselsnummer føyes til. Det vil gi et delregister for barn og ungdom i denne kategori som, sammenkoblet med det sentrale register, vil gjøre det mulig etter hvert å gjennomføre longitudinelle studier av hvordan denne gruppen tilpasser seg økonomisk og yrkesmessig.
- c. For spesielle tabelleringer av den samiske befolkning tenkes opprettet et register over alle samer i Norge i tilknytning til folketellingen i 1970. Personene vil være identifisert ved blant annet fødselsnummer. Når dette register først er etablert, er mulighetene gjennom fødselsnummer åpne for å utnytte opplysningene i det sentrale personregister til spesialstudier av den samiske befolknings yrkesstruktur, inntektsforhold osv. uten innhenting av opplysninger spesielt for dette formål.
- d. Fødselsnummer er delvis tatt i bruk i tilknytning til kriminalstatistiske data. Dette betyr, som ovenfor, at en også for grupper av lovbrytere kan studere deres tidligere og senere sosio-økonomiske historie slik denne kan beskrives ved data tilgjengelig i det sentrale persondataarkiv.

Andre eksempler på delregistre med nyttig informasjon for samfunnsvitenskapene har vi i valgmanntall, førerkortregister, sykehusregistre, kreftregister, bankenes innskudds- og utlånsregister, register over låntakere i offentlige banker, register over livsforsikringstakere etc. I alle slike registre kan personenes fødselsnummer registreres, og innhentede opplysninger kan arkiveres på en slik måte at de kan kobles sammen med opplysninger i det sentrale persondataarkiv. Det er neppe nødvendig allerede i starten å legge noen detaljert plan for hvilke spesialiserte delarkiv som skal opprettes. Men Byrået bør gjøre en innsats for å holde greie på de delarkiv som finnes eller opprettes og prøve å utvirke at de blir lagt opp slik at en samkobling med det sentrale persondataarkiv blir mulig ved behov. For offentlige registre er det antakelig ønskelig med en ordning slik at oppleggene skal godkjennes av Byrået, analogt med at Byrået i dag skal få seg forelagt alle offentlige skjemaer til statistiske formål.

8.2 Persondataarkivet som grunnlag for sosialvitenskapelig forskning

Den forskning som vil vokse opp omkring persondataarkivet, eller omkring statistikk som persondataarkivet vil muliggjøre, vil ha som mål å estimere strukturen i atferdsrelasjoner som kan inngå i personmodeller av ulike slag. En statistikk bygd på et persondataarkiv med innhold som angitt ovenfor, kan med sikkerhet sies å finne konsumenter blant forskere fra vidt forskjellige fagområder: demografi, sosialøkonomi, utdanningsøkonomi, sosialmedisin, sosiologi, kriminologi, psykologi, valgforskning, bedriftsøkonomi, næringsgeografi og anvendt forskning på feltene transportøkonomi, regionalplanlegging etc.

Det er umulig i dag å angi alle de analytiske anvendelser som et slikt personstatistisk system vil få, jfr. hvordan det statistiske system som nasjonalregnskapet representerer har fått anvendelser som ingen kunne forutse for tjue år siden. Mulighetene er imidlertid store. For å illustrere dette skal vi nedenfor nevne en del eksempler, hentet fra ulike fagområder, på forskningsoppgaver hvor det er åpenbart at et persondataarkiv vil gi mye av det datagrunnlag som trengs.

Demografi

Et viktig forskningsfelt, hvor Byrået allerede har engasjert seg gjennom sin Gruppe for personmodeller, er oppstilling av befolkningsmodeller til bruk for blant annet befolkningsprognoser, totalt for riket og for enkelte distrikter. Som forarbeider for slike modeller vil det bl.a. være nødvendig med:

- a. studier av brukbarheten som funksjon av blant annet alder, ekteskapelig status, antall barn, yrke, bosted, inntekt og yrkesmiljøfaktorer,
- b. studier av dødeligheten som funksjon av kjønn, alder, yrke, næring, bosted,
- c. studier av innenlandske vandringar som funksjon av kjønn, alder, ekteskapelig status, utdanning, yrke, inntekt, familieforhold, bosted, boligforhold og en rekke miljøfaktorer.

Studier som nevnt, krever opplysninger om demografiske variable og om miljø-variable. En del av de viktigste personlige variable - kjønn, alder, ekteskapelig status, inntekt m.v. - finnes allerede i persondataarkivet, mens en rekke miljøfaktorer knyttet til kommunene er tilgjengelig som statistikk.

Sosialøkonomi

- a. Kartlegging av inntekts- og formuesstrukturen og studier av årsakene til den observerte fordeling. Persondataarkivet vil muliggjøre studier av inntektens variasjon simultant med blant annet alder, utdanning, yrke, næring, bosted og eksamensresultat. Såvel personlige inntekter som familieinntekter vil være vesentlige studieobjekt. Utvalgte grupper av personer (f.eks. utvandrere fra jordbruket, innvandrere til byene, personer som har gjennomgått omskoleringskurser, ungdom som har vært gjenstand for sosial omsorg) kan følges over tiden med hensyn til inntekts- og formuesstruktur.
- b. Skatteforskning, blant annet bygging av personskattmodeller til bruk ved studiet av hvordan skattesystemet, eller planlagte endringer i dette, rammer ulike grupper av hushold. En vesentlig del av databehovet for slike undersøkelser - demografiske, økonomiske og sosiale kjennemerker som har betydning for skatteplikten - vil kunne møtes ved et utbygd persondataarkiv.

Arbeidskraft

Oppstilling av modeller til bruk for prognoser over tilbudet av arbeidskraft, geografisk og etter næring og yrke:

- a. Yrkefrekvenser som funksjon av blant annet alder, kjønn, sivilstand, utdanningsbakgrunn, bosted.
- b. Valg av næring og yrke som funksjon av blant annet utdanning, foreldres næring og yrke, bosted og lønnsforhold.
- c. Valg av geografisk arbeidssted som funksjon av eget (og ektefelles?) fødested, sted for egen utdanning, utdanningsmuligheter for barna etc.

Utdanningsvalg og utdanningsøkonomi

- a. Studier av årsakene til variasjoner i søkingen til ulike skoleslag som ledd i oppstillingen av utdanningsmodeller. Mange av de variable som vil måtte inngå som forklaringsvariable, vil finnes i persondataarkivet, blant annet foreldres utdanningsbakgrunn, næring og yrke, inntekt og bosted.
- b. Utdanning som økonomisk vekstfaktor.
 - (i) En måte å analysere utdanningens økonomiske betydning på, er å studere inntektens variasjon med utdannelsesnivået for grupper som i andre henseender er like (idet konstaterter

inntektsdifferenser kan tas som uttrykk for utdannelsens betydning både for den enkelte og for samfunnet). Persondataarkivet vil automatisk gi datagrunnlaget for slike undersøkelser.

- (ii) Problemet kan studeres ved hjelp av makroproduktfunksjoner - for samfunnet som helhet eller for enkelte næringer - hvor "kunnskapskapitalen" inngår som produksjonsfaktor ved siden av andre faktorer. Persondataarkivet vil inneholde data som trengs for å måle kunnskapskapitalen.

Sosialmedisin

- a. Forekomsten av ulike sykdommer i forskjellige miljøer. Studiet vil være mulig ved å koble sammen sykdomsdata fra forskjellige kilder (kreftregisteret, sykehusprotokoller) med generelle bakgrunnsdata fra persondataarkivet om pasientenes livshistorier karakterisert ved opplysningene om alder, næring, yrke, bosted, inntekt m.v.
- b. Analyse av dødeligheten etter årsaker i forskjellige miljøer. Studier vil være mulig ved å koble sammen opplysninger om dødsårsaker på de individuelle dødsmeldinger med generelle bakgrunnsopplysninger som ovenfor.
- c. Forbruksstruktur og dødelighet. Ved et forskningsprosjekt som er i gang, kobles dødelighetsopplysninger fra siste 10-årsperiode sammen med materiale fra forbruksundersøkelsen i 1958 med sikte på å påvise eventuelle sammenhenger mellom forbruk og senere dødelighet.
- d. Attføringsproblemer o.l. Ved at persondataarkivet tillater å følge utvalgte personer "over tiden", åpnes det muligheter for studier av hvordan grupper av personer som har vært under medisinsk behandling, senere evner å tilpasse seg i arbeidslivet. Det kan gjelde en hvilken som helst medisinsk interessant gruppe hvor medlemmene kan identifiseres ved fødselsnumre, f.eks. utskrevne fra psykiatriske sykehus, barn med medfødte handikap, et klientell av narkomane, personer behandlet etter trafikkulykker osv. Undersøkelser av denne art bør kunne gi holdepunkter for vurdering av de resultater som oppnås ved ulike behandlingsmetoder.
- e. Arv som sykdomsfaktor. Sykdoms- og dødelighetsdata koblet sammen med persondataarkivets opplysninger om foreldre-barn-relasjon kan på noe lengre sikt åpne muligheter for studiet av arvelige sykdomsfaktorer. Dette vil dog forutsette en systematisk registrering av sykdomsdata som det kan ta tid å få til.

Sosiologi

Dette er den vitenskapsgren som på lang sikt antakelig vil komme til å gjøre seg størst nytte av persondataarkivet. En gjennomlesning av Tidsskrift for samfunnsvitenskap fra senere år gir en lang rekke eksempler på undersøkelser som er utført på grunnlag av spesielt innsamlet materiale som i framtiden helt eller delvis vil foreligge tilgjengelig i persondataarkivet. Dette gjelder kanskje særlig:

- a. Studier av sosial mobilitet, dvs. bevegeligheten ut og inn av bestemte sosiale miljøer. Studier av denne type vil kunne gjennomføres kontinuerlig i den utstrekning miljøene lar seg karakterisere ved demografiske og sosiale kjennetegn som alder, kjønn, sivilstand, næring, yrke, inntekt, formue, familiebakgrunn, utdanning.
- b. Studier av størrelse, sammensetning og bevegelse i sosialt interessante grupper som "de gamle", "de skilte kvinner", "de ufullstendige familier", "de ugifte mødre", "den samiske befolkning". En ikke uvanlig teknikk kan bli å utnytte arkivdataene til en generell beskrivelse av slike grupper, samtidig som en ved intervjuer på utvalgsbasis innhenter supplerende opplysninger for å belyse bestemte problemstillinger.
- c. Studier hvor det er viktig å kunne følge en avgrenset gruppe individer over tiden, f.eks. ved å trekke inn som bakgrunnsfaktorer i en undersøkelse individenes tidligere livssyklus beskrevet ved demografiske og sosiale kjennemerker.

Kriminologi

- a. Forekomsten av lovbrutt i ulike miljøer. Studier vil være mulig ved å koble sammen lovbrotsdata fra ulike kilder (fangeregister, strafferegister o.l.) med generelle bakgrunnsdata fra persondataarkivet om tidligere livssyklus karakterisert ved demografiske og sosiale kjennemerker. ("Kriminalitetens årsaker".)
- b. Tilbakefallsundersøkelser. Grupper av personer som har vært i konflikt med loven kan følges med sikte på å studere deres senere tilpasning i samfunnet, blant annet for å klarlegge virkningene av ulike reaksjonsformer. Analyseteknikken vil likne mye på den som kan brukes for studiet av medisinske attføringsproblemer.

Sosialpsykologi

Forekomsten av bestemte psykologiske reaksjoner i ulike miljøer. Studier vil være mulig ved å koble sammen iakttagelser av reaksjoner som skal studeres med generelle bakgrunnsdata fra persondataarkivet.

Valgforskning

- a. De politiske partiers relative styrke i ulike miljøer. Studier vil være mulig ved å koble spesielt innsamlede opplysninger om stemmegiving (eller annet uttrykk for partipreferanse) sammen med data fra persondataarkivet til karakteristikker av relevante miljøer (alder, næring, yrke, bosted, inntekt, familiebakgrunn osv.).
- b. Studier av valgdeltakelse. Iakttagelser av opptreden ved valg, - f.eks. hentet fra valgmanntallene - koblet sammen med vanlige bakgrunnsdata fra persondataarkivet.

Bedriftsøkonomi

- a. Markedsundersøkelser:
Spesielle undersøkelser av forbruksstrukturen i ulike miljøer kan kobles sammen med data fra persondataarkivet om miljøenes størrelse og veksthastighet.
- b. Lytter- og seerundersøkelser. Mer generelt:
"Markedsundersøkelser" for tjenesteytende offentlig virksomhet:
Hvem (hvilke sosiale miljøer) gjør bruk av tjenestene?
- c. Lederrekruttering:
Data over næring, yrke og inntekt fra persondataarkivet, eventuelt kombinert med spesielt innsamlede data, kan brukes til å klarlegge fra hvilke miljøer lederne i næringslivet rekrutteres.

9. KONFIDENSIELL BEHANDLING AV DATA

De data som inngår i persondataarkiv og bedrifts- og foretaksdataarkiv er innhentet delvis med hjemmel i lov som pålegger publikum en plikt til å gi de etterspurte opplysninger og delvis ved oppgavegiverens frivillige svar på Byråets forespørsler. Lovverket fastsetter i det første tilfelle ikke bare en oppgaveplikt, men inneholder også bestemmelser som skal sikre publikum mot at de oppgaver som pålegges gitt ikke skal kunne

nyttes til skade for oppgavegiveren. Byrået har sett det som selvsagt at det er like viktig å gi publikum denne sikkerhet også når oppgavene gis Byrået på fortrolig, frivillig grunnlag. En har derfor gjennom alle år vært meget streng med hensyn til behandling av data om de enkelte oppgavegivere, også når disse har vært gitt på frivillig grunnlag.

Fra en rekke statistikk-brukere ytres det imidlertid ønske om å få større adgang til data i Byråets arkiver. Særlig har samfunnsforskere utenfor Byrået pekt på det etter deres mening urimelige i at de ikke får utnytte eksisterende data til løsning av viktige oppgaver i samfunnsforskning og -planlegging.

Løsningen på dette dilemma synes å være at det må forsøkes trukket opp en grense mellom (i) data som kan tenkes å føre til skade for oppgavegiveren dersom de ble ført ut av Byrået igjen og som derfor må være konfidensielle og (ii) data som ikke kan antas å ville påføre oppgavegiveren skade om de blir tilgjengelige for andre. I den første gruppen vil identifiserbare, individuelle data om private forhold utgjøre hovedmassen, mens den andre gruppen vil - foruten statistikk - omfatte anonyme, individuelle data og data om forhold som må antas å være offentlig tilgjengelig.

Når det gjelder den andre gruppen av data, bør det være grunnlag til å gi større adgang enn hva som hittil har vært tilfelle. For den første gruppen av data synes det rimelig å sikre utnyttelsen ved å bygge ut samfunnsforskningen innenfor Byrået. En absolutt forutsetning må under enhver omstendighet være at en opprettholder et godt tillitsforhold mellom publikum og den data-innsamlende institusjon.

10. KOSTNADSOVERSLAG OG RESSURSBEHOV

En gjennomføring av de prosjekter som er omtalt ovenfor vil ta mange år, fordi det vil kreve arbeidskraft som ikke kan skaffes på kort varsel. Kostnadene vil bli forholdsvis store. Grove overslag viser at selve etableringen av et dataarkiv som skissert vil kreve beløp av størrelsesordenen 10-20 mill.kr. fordelt over en årrekke, og at det senere vil kreve 6-8 mill.kr. pr. år å holde datamassen løpende vedlike. Beløp av denne størrelsesorden vil representere en øking av Byråets budsjett med en faktor på 0,25 eller mer. I tillegg kommer utgifter - men etter hvert også store besparelser - hos andre etater som Byrået vil samarbeide med for innhenting av data.

Byrået ville forsømme sin plikt hvis det ikke pekte på de muligheter som nå foreligger for å oppnå en radikal forbedring av data-grunnlaget for norsk samfunnsforskning og planlegging. Om planene skal realiseres, avhenger av om administrasjonen og de bevilgende myndigheter vil finne at pengene vil være vel anvendt.