DATA-ARKIVER

– Har vi behov for noget nyt? –

JØRGEN TVEVAD

De offentlige arkiver samler foruden retsstiftende dokumenter og administrative fonds i vidt omfang private papirer fra personer, hvis skriftlige efterladenskaber kan have interesse for fremtidens historieskrivere. På samme måde søger bibliotekerne i deres håndskriftsafdelinger at samle forfatteres og videnskabsmænds manuskripter. Med det stigende omfang af undersøgelser, som forvaltningen og forskningen må foretage, vokser interessen for de arbejdsfaser, der ligger mellem materialeindsamlingen og det færdige resultat. Ofte vil en embedsmand eller en videnskabsmand forsøge sig frem ad veje, som viser sig ikke at føre til et resultat, men hans foreløbige bearbejdelse kunne måske være til udmærket nytte i en anden forbindelse. Dette forhold gør sig især gældende inden for de dele af samfundsvidenskaberne, hvor man arbejder med kvantiteter eller kode-omsættelige kvaliteter, og derfor kan tage det moderne EDB-maskineri i brug. Da forberedelserne til en databehandling, programmeringen og kodningen, imidlertid er en tidkrævende proces, ville det være af værdi, om man kunne udnytte andres forberedende, om end upublicerede arbejder af statistisk karakter.

Brugen af statistiske metoder inden for den humanistiske forskning er ikke gammel, selv om historikere, teologer og filologer vel altid har udnyttet tallenes tale ved indholdsanalyser af tekster. I 1921 forsvarede bibliotekar Søren Hallar for Københavns universitets filosofiske doktorgrad en disputats om synselementerne i naturskildringen hos J. P. Jacobsen, sammenlignet med farvebeskrivelser i Blichers digtning. Hallar, der ikke var æstetisk, men naturvidenskabeligt uddannet, forsøgte i sin afhandling af benytte matematiske metoder på et sprogligt stof. Hans synspunkt var, at kvantitative bestemmelser måtte bære fortrinnet frem for alene kvalitative, at »tal er videnskabens versefødder«. Adgangen til korrelationsberegninger og til prøvning af diverse variable muligheder ville blive øget. Nutidens data-teknik, såvel ved hulkort som ad elektronisk vej, har åbnet yderligere muligheder og givet store lettelser for avancerede indholdsanalyser. Hvis man behersker eller kan få kodet til et problemorienteret programmerings-

sprog, kan det være muligt at udtrykke kvalitet ved kvantitet og derefter lade maskinen køre. Og medens gamle dages seddeludskrivning kun kunne tjene sit begrænsede formål, vil en nutidig hulkortbunke eller et magnetbånd kunne benyttes ved sekundære undersøgelser, såfremt problemangrebet er tilstrækkeligt bredt og programbeskrivelsen fyldestgørende.

Muligheden for praktisk udnyttelse af allerede foreliggende statistiske og andre data er imidlertid betinget af, at man ved, om og hvor de findes. Det er klart for enhver, at der ligget et kodet basismateriale til grund for en moderne statistisk publikation og for de fleste kvantitetsvurderende samfundsanalyser. Man det fremgår sjældent af beretninger og rapporter, om hele basismaterialet er blevet kodet på hulkort eller magnetbånd, og om det således er muligt at køre flere simple randfordelinger eller indviklede krydsfordelinger af de indsamlede fakta.

Medens f. eks. de statistiske bureauers folketællingsskemaer og vareimportindberetninger og de statslige survey-institutters interviewskemaer klart er destineret for en sluttelig aflevering til de offentlige arkiver, gælder dette næppe deres arbejdsmateriale til brug i tællingsanlæggene. Universitetsinstitutters hulkort osv. når næppe heller frem til de offentlige arkiver efter brugen og private markedsanalyse-firmaers givet ikke. Og det afhænger ganske af tilfældigheder, om et bibliotek får fat i en afdød personalhistorikers æske med kant- eller sigthulkort. Der har derfor, i første række i USA, rejst sig en bevægelse for oprettelse af såkaldte data-arkiver, som samler det tekniske materiale, der ikke hører hjemme i de offentlige arkiver og biblioteker.

Økologiske arkiver.

Data-arkiverne falder i to hovedgrupper, de kodede samlinger af almene baggrundsoplysninger til brug ved løbende og kommende undersøgelser og de kodede svar på spørgeskemaer og lignende fra tidligere eller løbende undersøgelser. Grænsen mellem de to slags data-arkiver er naturligvis flydende, fra den komplette samling statistiske fakta til de spredt indsamlede oplysninger.

Begrebet økologisk er lånt fra biologien, fra læren om dyrs og planters tilpasning til de naturforhold, der står til deres rådighed i forskellige geografiske områder. Springet herfra til socialgeografien, til en betegnelse for beskrivelsen af de fysiske, økonomiske og sociologiske milieuer, mennesker i bestemte geografiske områder lever i, er ikke stor.

Den topografiske litteratur har naturligvis altid meddelt en mængde økologiske fakta, men ikke altid i lige let tilgængelig og navnlig ikke i lige let behandlelig form. Nogle eksempler vil vise, hvorledes man i de senere år har søgt at imødekomme behovet for lokale data. I 1956 udgav det statistiske departement »Statistiske meddelelser. 4–162–2. Erhvervsgeogra-

fisk materialesamling«. I forordet hedder det: »Behovet for lokale oplysninger af denne art har været i fortsat stigning i de senere år, hvor de i Departementet beroende arbejdstabeller og grundmateriale har været stillet til rådighed for en lang række forskellige formål. Dette gælder ikke alene undersøgelser foretaget af offentlige institutioner, men også større erhvervsvirksomheder har efterspurgt oplysninger om befolkningens størrelse og sammensætning til brug ved planlæggende arbejde for produktion og afsætning og ved markedsanalyser på mange felter. Hertil kommer, at de nyere omkostningsbesparende metoder med undersøgelser baseret på udvalgte (»repræsentative«) landsdele eller befolkningsgrupper kræver mere detaillerede lokale oplysninger ved udvælgelsen.« Det samme generelle formål som »Erhvervsgeografisk materialesamling« forfølger socialforskningsinstituttets publikation nr. 3, »Kommunernes sociale struktur«, 1960, og, til eksempel, det kommuneregister, der i 1964 er udarbejdet som bilag til strukturundersøgelsen for Randers amt.

Til imødekommelse af marketing-arbejdet og planlægningsovervejelser har det statistiske departement indarbejdet visse af sine tal og oplysninger i de særlige handelsområder, jfr. »Statistiske undersøgelser. Byernes opland«, 1960. Om netop denne inddeling af landet kan være tilfredsstillende og formålstjenlig i det lange løb, er et spørgsmål. Konsulenttjenesten ved Fællesforeningen for Danmarks Brugsforeninger har således ud fra sin særlige forbrugsvareudvælgelse måttet instituere afsætningsområder, som ikke altid følger og i en del tilfælde går på tværs af grænserne for departementets handelsområder. Men så længe de grundlæggende tal fra sogne eller kommuner er tilgængelige, om end kun i arkivalsk form, vil sammenlignende undersøgelser dog altid kunne gennemføres. Gennemfører man derimod indsamlinger af data fra forskellige slags områder, som overlapper hinanden geografisk, vil det økologiske arbejdsmateriales værdi blive truet.

De foreliggende statistiske oplysninger er vel de første, hvortil man går, når en repræsentativ undersøgelse skal planlægges. Dette materiale giver om økonomiske og beslægtede forhold et brugbart grundlag for stratificeringen ved en sample-undersøgelse. Men trods alt ved man endnu langt fra nok om de økologiske data. Kun få af de mere seriøse danske undersøgelser har nærmet sig spørgsmålene om befolkningens holdning i forskellige, vanskeligere konkretiserbare relationer. Det er et faktum, at to geografiske områder med ganske samme økonomiske og befolkningsmæssige udseende kan reagere ganske forskelligt på ganske ens holdnings- og meningsspørgsmål. Dette gælder ikke blot for landdistrikterne, hvor den stigende urbanisering vel udjævner visse særpræg. Også de industrialiserede områder giver mange eksempler på ikke-konforme træk.

Det må derfor være en opgave til brug for såvel generelle som repræsentative undersøgelser at indsamle og gøre tilgængelige så mange data som

muligt, også om de lokale enheders kulturelle, religiøse og politiske særtræk. Dyberegående folketællingsspørgsmål om skole- og faguddannelse ville give værdifulde oplysninger. Foreningsaktiviteten må kunne belyses bedre. Mange traditionelle kirkelige forskelligheder kan i et vist omfang beskrives som af Paul Nedergaard i »Personalhistoriske, sognehistoriske og statistiske bidrag til en dansk præste- og sognehistorie. Kirkelig geografi«, 1949–66. Og navnlig er den politiske geografi moden til den dybtgående forskning, som instituttet for statskundskab ved Århus universitet har påbegyndt med oprettelsen af et økologisk data-arkiv. Ligeså har det sociologiske institut ved Københavns universitet et økologisk data-arkiv for hovedstadsområdet under opbygning.

Undersøgelses-arkiver.

Data om enkeltpersoner vil, bortset fra selektive fortegnelser som »Dansk Biografisk Leksikon« og »Kraks Blå Bog«, næsten altid foreligge generelt, enten for hele befolkningen i folkeregistre og lignende eller for særlige grupper med samme uddannelse, samme erhverv eller lignende. Også de økologiske data vil som regel være generelle, kun sjældent færre for nogle end for andre egne. Det vil sige, at anvendelige personelle og topografiske oplysninger altid vil kunne findes hos offentlige myndigheder eller på bibliotekernes hylder.

Anderledes forholder det sig for mange repræsentative undersøgelsers vedkommende. Socialforskningsinstituttets og andres publicerede rapporter vil vel kunne fremfindes ved hjælp af bibliografier og dokumentationsservice. Men mangfoldige undersøgelser, der har dannet basis for administrationens beslutninger eller for videregående forskningsresultater, foreligger, selv om ikke fortrolige, så dog kun i stencileret eller skrevet form, og dermed i praksis utilgængelige. Såfremt der med tiden kunne udgives registre over planlagte og påbegyndte projekter, således som det f. eks. sker i Holland, ville der åbne sig store muligheder for blandt andet orientering om upublicerede undersøgelser.

Også råmaterialet til mange undersøgelser, f. eks. spørgeskemaer og kodede IBM-kort, kan have den allerstørste betydning for sekundære analyser, parallel forskning og senere undersøgelser. Til sikring af baggrunds- og sammenligningsmateriale stiller man vel ofte flere spørgsmål og indhenter flere oplysninger, end man får anvendelse for, og der køres vel som regel flere tabeller, end der indgår i den endelige undersøgelsesrapport. Hele dette »ubrugte« eller i det mindste upublicerede materiale vil kunne have den allerstørste betydning for andre undersøgelser. Således kunne socialforskningsinstituttet som et »spildprodukt« fra sine undersøgelser over handicappede og ældre i 1966 udgive nogle tabeller om uddannelse, erhverv og helbred, som giver et mere detailleret billede af visse befolkningsforhold end

de af det statistiske departement offentliggjorte tal. I det hele gælder det, at man ved at gå tilbage i basis- og kortmaterialet for mange undersøgelser vil kunne finde særdeles interessante data, som imidlertid har været mindre relevante for de pågældende undersøgelser og derfor ikke er blevet anvendt. I det omfang, man kan gennemføre repræsentative undersøgelser, hvor de samme baggrundsspørgsmål indgår, vil man få mulighed for at opbygge vigtige statistiske data-serier.

Nationale data-arkiver.

Ordningen og opbevaringen af datamaterialet stiller store krav. Spørgeskemaer og tabel-»tapeter« kan vel arkiveres som andre papirer, men computer-kort og magneto-bånd kræver særlige klimatiske forhold, som nuværende arkivlokaler sjældent byder på. Dertil kommer, at skemaer, kort og bånd altid må være let tilgængelige for genoptaget brug. Det ideelle samarbejde, hvor forskellige institutioner frit hos hinanden kan låne de data, man mener at kunne udnytte, vil kræve en central registrering og måske et fælles arkiv for ikke-dagsaktuelt materiale. Hertil kommer, at universitetsundervisningen i samfundsbeskrivelse kan stille de databehandlende institutioner over for ønsker om en bredere praktikantundervisning og om udlån af ældre materiale til orienteringsbrug.

Det norske Institutt for Samfunnsforskning har taget fat på de nævnte problemer og har blandt andet indsamlet økologiske data for en række områder til supplering af det almindelige statistiske materiale, således at der gives bedre muligheder for undersøgelser med lokale variable, herunder de sproglige og religiøse modsætningers betydning. Et af de første resultater af dette arbejde er »Henry Valen & Daniel Katz: Political Parties in Norway. A Community Study«, Oslo 1964.

I Köln og Amsterdam har man oprettet egentlige data-arkiver, som registrerer og samler materiale fra offentlige og akademiske institutioner samt eventuelt private markedsanalyse-firmaer m. v. og stiller disse data til rådighed for forskning og undervisning. Andre samlinger af maskinlæselige data fra samfundsvidenskabelige undersøgelser er under opbygning. Zentralarchiv für Empirische Sozialforschung an der Universität Köln blev grundlagt i 1960 til indsamling af survey-data, især fra Vesttyskland. I sommeren 1965 forvaltede arkivet materialet fra mere end 80 undersøgelser, især af forbrugsøkonomiske, politiske og sociologiske spørgsmål. Mere end 50 kommercielle og akademiske forskningsinstitutter er gået ind i samarbejdet og vil aflevere data, beskrivende materialer og anden information. Foruden sample-udvælgelser, interviewerinstruktioner, spørgeskemaer og kodnings-beskrivelser samler man forskningsrapporter og metodologisk litteratur.

Interessen for adgang til sample-survey-data til brug ved sekundære

analyser og ved tvær-nationale undersøgelser har dog hidtil været størst i Amerika. I 1946 oprettedes The Roper Public Opinion Research Center ved Williams College i Williamstown, Massachusetts. Dette forskningscenter rummer allerede over 8 millioner IBM-kort og giver samfundsforskere adgang til at få kørt tabeller eller hjemlåne duplikatkort. Andre større amerikanske arkiver er The International Data Library and Reference Service ved Survey Research Center, University of California, Berkeley, og The Inter-University Consortium for Political Research ved University of Michigan i Ann Arbor.

Internationalt samarbejde.

De amerikanske forskningsinstitutioner enedes i 1962 om at oprette et National Council on Social Science Data Archives med det formål at finde veje og midler til at lette adgangen til data og at inspirere til oprettelsen af arkiver. Spørgsmålet blev samme år taget op af International Social Science Council, (ISSC), der har indledt et samarbejde, som forestås af forskningsleder Stein Rokkan fra Chr. Michelsens Institutt i Bergen.

Ved et møde i La Napoule, Frankrig, i 1962 enedes forskellige europæiske og amerikanske socialvidenskabsmænd om at ville underkaste arkivspørgsmålene nærmere prøvelse, navnligt med henblik på hjælp og bistand landene imellem. I sommeren 1963 diskuterede europæiske eksperter ved et møde i Köln det nærmere samarbejde, og i efteråret 1963 afholdtes ved Yale Universitetet internationale drøftelser om brugen af kvantitative politiske, sociale og kulturelle data ved tvær-nationale undersøgelser. Beretninger fra den »første« konference om data-arkiver i Köln og fra Yale-mødet er offentliggjort i »Social Sciences Information«, vol. II, dec. 1963.

Med disse forudgående drøftelser som grundlag fandt ISSC det betimeligt at tage initiativ til den »anden« internationale konference om data-arkiver. Arrangør af Paris-konferencen 1964 var ISSC og International Committee on Social Sciences Documentation, der bl. a. har forbindelse med The European Coordination Centre for Research and Documentation in Social Sciences i Wien. Foruden af UNESCO blev Paris-konferencen understøttet ved bidrag fra Ecole Pratique des Hautes Etudes, Paris, Maison des Sciences de l'Homme, Paris, samt Social Science Research Council, New York. I møderne deltog ialt 41 personer fra Finland, Frankrig, Nederlandene, Norge, Polen, Storbritannien, Tyskland, Ungarn og Danmark samt internationale organisationer. Fra dansk side deltog direktør Henning Friis, socialforskningsinstituttet, og cand. mag. Jørgen Tvevad. Konferencens rapporteur général var Stein Rokkan, Norge, dens chairmen professeur Jean Stoetzel, Centre d'Etudes Sociologiques, Centre National de la Recherche Scientifique, Paris, Rokkan, professor Erwin Scheuch, Zentralarkiv für Empirische Sozialforschung, Universität Köln, samt dr. Marten Brouwer,

Seminarium voor Massapsychologie, Openbare Mening en Propaganda, Universiteit van Amsterdam.

Forhandlingerne 1964.

Konferencen er refereret i »Social Sciences Information«, vol. IV, sept. 1965, og af dens forhandlingsemner kan følgende fremhæves.

Interessen for økologiske arkiver synes stor i de nordiske lande, hvor, bortset fra Danmark, naturforholdene spiller så dominerende en rolle. Professor Erik Allardt, Helsingfors universitet, fremlagde en redegørelse for samlinger af data vedrørende territorielle enheder i Finland og Stein Rokkan og magister Henry Valen gjorde rede for statistiske studier af indenlandske forskelligheder i Norge. De nordiske deltagere drøftede et nærmere samarbejde på disse områder.

Redegørelser for antallet og karakteren af sample-undersøgelser var udarbejdet for det kontinentale Europas vedkommende, for Amerika og for Storbritannien. Antallet af surveys, i alt over 6000 i 1963, viste nødvendigheden af en bedre dokumentation samt, at centralarkiver til brug ved sekundære analyser og tvær-nationale undersøgelser ville betyde en arbejdslettelse for alle forskere.

Om arkiveringen af survey-data blev der forhandlet på grundlag af reglerne for The Roper Public Opinion Center samt redegørelser fra bl. a. Köln-arkivet om de tekniske problemer ved oplagringen.

Retrieval-teknikken og det øvrige dokumentationsarbejde i forbindelse med samlinger af undersøgelsesmateriale samt forskning ved indholdsanalyser blev drøftet, ligesom der fra amerikansk side blev gjort rede for dataarkivernes betydning for universitetsundervisningen.

Konferencens formelle resultater blev: en anerkendelse af bestræbelserne fra Zentralarchiv i Köln og Steinmetz–Stichting voor het Opslaan en Toegankelijk Maken van Bestaand Materiaal von Sociaal Onderzoek, Amsterdam, for at etablere en europæisk federation af samfundsvidenskabelige data-arkiver; en opfordring til International Social Science Council om nedsættelse af en stående komité for data-arkiver med direkte eller indirekte repræsentation for arkiverne samt for International Committee for Social Sciences Documentation, for International Statistical Institute, for World Association for Public Opinion Research og for International Council of Archives; en anbefaling af med støtte fra UNESCO at planlægge fortegnelser over eksisterende og kommende undersøgelsesdata, for den europæiske regions vedkommende efter en drøftelse ved det europæiske dokumentationscenter i Wien; en henstilling til de nationale samfundsvidenskabelige forskningsråd om at udvikle data-arkiver og give dem mulighed for internationalt samarbejde; en opfordring til UNESCO, ISSC og de internationale forskningsråd om at foranstalte træningskurser og metodologiske se-

minarier, især med tvær-nationale og kulturelle undersøgelser for øje, ved forskningsinstitutter og data-arkiver; et ønske om en senere international konference om økologiske arkiver og analysemetoder; en ansøgning til UNESCO og ISSC om økonomisk støtte til den foreslåede stående komité's informatoriske virksomhed.

Europæisk samarbejde.

Paris-konferencens første resultat blev altså dannelsen af en Federation of European Social Science Data Archives, hvis principper, i deres foreløbige udformning, skal være følgende: For at lette forløbsundersøgelser og tvær-kulturel forskning bør der oprettes samfundsvidenskabelige arkiver med data fra forskellige kilder og for forskellige enheder; for at sikre den bedste udnyttelse af sådanne data bør der oprettes arkiver i hvert land eller for samarbejdende grupper af lande; for at muliggøre systematiske sammenligninger af de forskellige landes data bør de europæiske arkiver indgå samarbejdsaftaler både sig selv imellem og med oversøiske arkiver. De foreløbige arbejdsregler er følgende: Hvert arkiv påtager sig at erhverve, behandle og gøre tilgængelige data fra sit regionale område; der tilbydes UNESCO og ISSC bistand ved seminarer på data-arkiv-området.

Fra dansk side blev det betonet, at socialforskningsinstituttet ikke har bemyndigelse til at forhandle på andre institutters eller arkivers, endsige private markedsanalysefirmaers vegne. Men man måtte håbe, at et eventuelt dansk samfundsvidenskabeligt forskningsråd ville kunne overtage forbindelsen til de europæiske og internationale data-arkivbestræbelser. Lige som fra fransk måtte man også fra dansk side gøre opmærksom på, at adskillige undersøgelsesdata hidrører fra statslige institutter, som har pligt til at aflevere ældre materiale til statslige, administrative arkiver.

Et dansk data-arkiv?

Kort og båndmateriale med kodede oplysninger, der eventuelt kunne benyttes ved tabelkørsler til sekundære, tvær-nationale og didaktiske analyser, findes for tiden hovedsageligt i Data-Centralen, det statistiske departement, socialforskningsinstituttet, institut for sociologi ved Københavns universitet, institut for organisation og arbejdssociologi, handelshøjskolen i København, institut for statsvidenskab ved Århus universitet samt i Dansk Gallup Institut og andre markedsanalysefirmaer. Et fælles data-arkiv for disse statslige, akademiske og merkantile institutioner kan sikkert kun gennemføres ved et samfundsfagligt forskningsråds autoritet.

Videnskaben har i dag utallige lagerproblemer. Bibliotekarer og arkivarer kæmper med pladsproblemer, og dokumentalisterne kan knapt følge med i informations-eksplosionens tempo. Hvis man vil oprette et dansk survey-data-arkiv, må det derfor kræves, at det kan blive virkeligt anvende-

ligt. Og en ufravigelig forudsætning herfor må være, at programbeskrivelserne for de primære undersøgelser er så omhyggelige og grundige, at de kodede data uden videre kan bruges, også til sekundære undersøgelser. Det er derfor nødvendigt, at de forskellige institutioner enes om en nøje afstemt standardisering af deres spørgsmål og af svarenes kodning.

Inden man kan nå så langt, kræves imidlertid en fortsat kortlægning af de danske samfundsvidenskabelige bestræbelser, en bibliografering af foreliggende historiske økologiske data og en registrering af løbende undersøgelsesprojekter.

Spørgsmålet, om vi har brug for data-arkiver, vil derfor næppe kunne besvares, før man får en bedre oversigt over, hvad der måtte ligge af brugbart materiale rundt i forskellige institutioner. Den påbegyndte kulegravning af de statistiske undersøgelsesområder vil møde problemet. Den sidste menneskealders fordyrelse af trykningspriserne har medført, at flere og flere rapporter og beretninger aldrig når frem til bibliotekernes katalogisatorer, og langt fra alt undersøgelsesmateriale når nogen sinde frem til arkivernes registratorer. Den sammenlignende samfundsforskning og den fremtidige historieskrivning må imidlertid ikke blive glemt.