Robbin, A. (1981). Strategies for improving utilization of computerized statistical data by the social science community. *International Journal of Social Science Information Studies, 1,* 89-109.

Social Science Information Studies (1981), *1* (89–109)

STRATEGIES FOR IMPROVING UTILIZATION OF COMPUTERIZED STATISTICAL DATA BY THE SOCIAL SCIENTIFIC COMMUNITY

ALICE ROBBIN*

Data and Program Library Service, University of Wisconsin-Madison, U.S.A.

ABSTRACT

In recent decades there has been a notable expansion of statistical data produced by the public and private sectors for administrative, research, policy and evaluation programmes. This is due to advances in relatively inexpensive and efficient data collection and management of computer-readable statistical data. Corresponding changes have not occurred in the management of data collection, preservation, description and dissemination. As a result, the process by which data become accessible to social researchers and others is frustrating, time consuming, and inefficient. This paper describes the reasons for this situation: the problem-solving workstyle of social data users, the nature of the data and their relationship to computer technology, and an inchoate social science information infrastructure.

Since statistical data play an increasingly important role in social research and policy decisions, social science information specialists must be prepared to meet the computer-readable statistical data user's needs. Four strategies are recommended for improving utilization of these data: improving the quality of statistical evidence, educating information professionals and end-users in numerical information, using the existing information infrastructure to preserve and disseminate data, and developing retrieval tools for improving access to information about social data.

PART I. INTRODUCTION

In recent decades, there has been a notable expansion of information recorded by government, primarily for administrative purposes, but also for research, policy, and evaluation activities.

Technological and intellectual advances have made possible the recording of these statistics in machine readable form, the collection of new types of data and increased utilization of governmental records (Rokkan, 1976; Robbin, 1979b;

* Alice Robbin is head of the Data and Program Library Service at the University of Wisconsin-Madison. She lectures in both the Department of Economics and School of Library Science, offering a three credit course entitled, 'Management of machine readable numeric data for the social sciences'. This course is designed for archivists, librarians, and social scientists. She is also President of IASSIST—the International Association for Social Science Information Services and Technology.

United Nations, 1979). Computer technology has changed the techniques of collection, processing, retrieval, and use of recorded statistical information. Computer technology now makes it possible to maintain information in its original form and to access, edit, and re-use it efficiently.

Since the 1930s, improvements in sampling techniques and survey methodology have made it possible to collect data more inexpensively and reliably (David, 1980; Lansing and Morgan, 1971; United Nations, 1979). Consequently, we have seen a great increase in the quality and quantity of available data.

Goals of the paper

As intermediaries in the communications system, information specialists are providers and users of statistical information. However, few have analysed the nature of the information with which they deal, and questions about content, quality and utility are seldom posed.

The information-seeking behaviour and needs of social scientists have not been the focus of a great deal of study (see American Psychological Association, 1969; Brittain, 1970; Garvey *et al.*, 1972; Garvey and Griffith, 1972; Johns Hopkins University, 1967; Line, 1969, 1971; Line *et al.*, 1971a,b; Paisley and Parker, 1967; Skelton, 1973; Vondran, 1976). More specifically, the information activities and needs of the users of computer-readable statistical data have been studied by only three researchers (Heim, 1980; Miller, 1979; White, 1974). Therefore we have limited information on the process of inquiry into social phenomena and lack sufficient understanding of the implications for information services of the great range of cognitive styles and interests, problems, approaches and values which make up the social scientific enterprise (Brittain, 1979).

Although the information system has been a subject of study for some time, the *social science* information system *in which statistical data are imbedded* has not been adequately examined. The dynamics of the interactions among the information base, the individual, and the technology employed by the individual to conduct his inquiry are extremely complex and rarely investigated. We typically confine our investigations to the discrete parts of the system and rarely recognize, much less analyse, the external conditions which affect these interactions.

This paper focuses upon social scientists and others who use quantitative data with modern computer and telecommunications to answer informational, research, instructional and public policy questions.[1] It examines how the social researcher uses information (Part II), gives an introduction to the nature of quantitative social data (Part III), and describes the infrastructure and retrieval tools which provide access to statistical data (Part IV).

Inherent in this approach is the belief that greater understanding will lead to better research products, more effective use of human resources and improved user services. Statistical data are an increasingly important information resource for research and decision making in all sectors of society. As such, inadequacies in the present scientific and technical information system must be rectified and strategies must be developed for improving utilization of computer-readable social data (Part V).

The description of the users, information base, information system, and recommended strategies for improving utilization offered here are derived from experiences at the University of Wisconsin Data and Program Library Service.

PART II. CHARACTERISTICS OF THE SOCIAL SCIENCE USER

A. The problem-solving mode of activity

Users of statistical data typically employ a problem-solving workstyle. Each identifies a problem and employs complex strategies for solving it, including developing testable hypotheses, locating appropriate data and statistical estimation techniques, retrieving and manipulating data, and deriving facts, a basis for new data inputs, or reports used in decision-making. These activities may be undertaken by different people or by one individual. A typology of the numerical data user is offered here for heuristic purposes.

B. A typology of data users

FACT-FINDER. This user needs specific numerical data, which represent facts and he is frequently encountered by the reference librarian. The fact-finder might ask for the population size of a county or the consumer price index for non-durable goods produced by an industry during 1978. He needs to locate a body of data (or a data base) which contains this fact or an easily understood series of numbers and to retrieve the numbers. This user has few analysis needs. Although the user could often turn to published sources for this information, he may try an automated data base first and discover that the information is indeed available, but that it costs money to retrieve. He discovers that although his needs are minimal the effort required to satisfy them is great.

BOTTOM-LINER AND TREND SEEKER. This user generally seeks a set of facts or numbers in order to make generalizations about certain conditions or processes. He requires discrete data items which can then be summarized, classified, or sorted to reduce them to manageable groups. He may perform such small calculations as addition or multiplication to generate useful results. For example, this user may need to classify all income tax returns by whether state government pays, refunds or makes no exchange of money, or a researcher examining historical census returns in Wisconsin might need to know what changes in crop production occurred over the ten years between two agricultural censuses.

NEGOTIATOR-TRANSFORMER. This user is a gatherer and transformer of numbers. He is often the middle-man in the process of problem-solving. He rarely initiates the process, but is responsible for searching and selecting data according to predetermined requirements of a project. Once the various discrete data are gathered, he may transform or reformat them or prepare them to be linked to other data. This individual typically makes few judgments about the appropriateness of the data, but presents the results of his search to the analyst. For example, this person might be charged with locating data for a sociologist examining family history records in Boston to test a variety of hypotheses about social mobility. These data will be contained in a variety of vital statistics, court, church or parish, school district, and city directory records. Once these records are located, they must be selected from the large body of data bases, identified for appropriate linkages, and merged to create a new data base. During this complicated process, a number of alternative strategies for selecting, retrieving, and linking relevant records will be discussed. The transformer will participate in this process, since during his search for appropriate data, he has become knowledgeable about their quality and condition.

BOTTOM-LINER/TREND SEEKER and NEGOTIATOR-TRANSFORMER

may both have had previous experience with computers. They are likely to interact with the support staff in some phase of their work.

HIGH PRIEST. In the complex process of problem-solving involving numerical data, there needs to be an individual who can make judgments about reliability, quality, and acceptable degree of error. This individual is the interpreter of the numbers and he may well be a statistician. In the social sciences, this individual is responsible for selecting appropriate statistical estimation techniques and interpreting the results of statistical manipulation of the data base. The statistician interacts infrequently with the support staff and may have limited interactions with the technological and technical issues underlying computer operating system hardware and software.

SCIENTIST-SAGE. This user knows how to evaluate the results of statistical analysis of numerical data. His goal is to collect and organize the 'bottom lines' or outputs systematically and to make qualitative judgments about the relevance of these outputs for answering the question he posed at the beginning of the research process. His overall aim is to gain not more data, but rather knowledge-producing information. Allen (1977) states that the scientists's principal goal is a published paper which is a systematic compilation of the inputs of the information processing system, then 'made available to other scientists to employ in their work' (p. 3). Also included in this category is the policy maker whose policy and programmatic actions are in part based on the selection of relevant evidence derived from statistical data.

PART III. CHARACTERISTICS OF COMPUTER-READABLE STATISTICAL DATA

This section describes the characteristics of computer-readable data which limit the social scientist's ability to conduct research. These characteristics are identified in the sub-headings below.

A. Quantity and types of data

Just as the literature of the social sciences has witnessed dramatic growth in the last decade, so too has the quantity of numerical data bases. Data in machine readable form are produced by every government in the world, by industry, by research institutes, and individuals. We have no information on the number of existing numerical data bases or the size of the increase over the last decade. Any estimate would provide a conservative number that would not reflect the multiple copies of each of the data bases held by individuals and libraries, the associated documentation in machine readable form, or the files which have been reproduced in internal formats consistent with particular computer configurations. Taking the Data and Program Library Service as an example, the last eleven years have seen an 1,166 per cent increase in the size of the data collection (600 to over 7,000 data files).

Social data are derived from specially designed inquiries like surveys and censuses or by-products of administrative processes, reflecting the interests and needs of their producers. For example, data have been collected on individuals, cities and industries, schools, political and economic events, voting behaviour and the like. Statistical data from the twelfth century have been encoded by

historians and predictive models of the world economy for the year 2000 by the World Bank.

At the microlevel statistical data exist as substantive disaggregated information in the form in which they were collected. At the macrolevel these data are 'derived by adding up or otherwise combining similar units' (United Nations, 1979: 6) and contain summaries or aggregates of the microlevel information.

Aggregate data can only be derived from microdata, and theoretically, it should be possible to relate 'any given aggregate construct to the underlying microdata from which it must conceptually be constructed' (ibid). In practice, this is not possible without significant conceptual and statistical adjustments. Microlevel data can always be aggregated (although not always usefully). Once aggregation and summary have occurred microdata usually cannot be reconstructed because data are not always available at the microlevel and macrolevel data may have significant distortions and biases. Empirical findings can sometimes be enriched by linking data from microlevel and macrolevel sources. Survey research data files, for example, often contain valuable information on certain types of income that might not otherwise be reported to the authorities. A better estimate of household income may be gained by linking the survey data to administrative sources such as tax records or social insurance.

B. Consistency of concepts, definitions, classifications and reporting units

Lack of consistency in the concepts which underlie the different data sources, the definitions of measurement procedures, the classifications of similar data subjects, and the treatment of reporting units has significantly affected the quality of investigations by statistical offices and researchers. In some cases, integration of statistical data sources has led to erroneous conclusions (when, for example, family level and household level data sources are combined and analysed at the household level). Without standardization of classification schemes it becomes impossible to make any useful comparisons among similar reporting units. For example, until standardization of industrial and occupational classifications occurred, regional policy making bodies such as the EEC were unable to examine the employment situation in the Common Market (La Cour, 1980).

C. Data and documentation quality

Most of the potentially rich sources of statistical data are limited in their use because they have not been subject to the traditional quality control procedures, nor have they had appropriate funding, preparation, and description to provide a better quality product. There are many reasons for the generally poor condition of data and documentation. Most notably, standards or guidelines are lacking for quality control and for data collection, processing and description. Limited value has been placed on these secondary data sources for scientific research and public policy planning. Only recently have national governments recognized that, as primary producers and sponsors of data collection, they have an important responsibility to monitor the quality of their collection, processing, and documenting procedures and to assist in developing standards.

Many data bases contain serious flaws because of poor sample design, missing responses and data, lack of verification of the data collection procedures, and nonstandard ways of collecting the information (Bailer and Lanphier, 1978) which seriously affect analysis, replication and comparison of results. Most data have not been properly cleaned, corrected and edited during the processing

stages, and therefore may contain many errors. While it is not always possible to obtain information, statistical estimation techniques are available to impute reasonable values for missing data. Because data files are generally in poor condition more computer and staff resources may be needed to prepare a data file for analysis than are required to perform the actual statistical analysis.

Data documentation, the descriptive text accompanying a file, is the key to understanding its quality. The text should describe the conceptual framework of the file's creators as well as the development and contents of the data base. A data base has no value without adequate documentation. If gaps in the statistical information are documented estimation techniques may be employed later to arrive at useful statistical results. Documentation should be prepared at the time of a file's creation and may contribute significantly to future use of the data and may enable future quality audits of the data collection and preparation operations to be conducted. The quality of the preparation also affects other aspects of data accessibility because it supplies the archivist with information on the provenance of the file, is an equivalent for the conventional archival inventory, and supplies information for other bibliographic access tools (Johnson, 1977; Robbin, 1975, 1980b; Roistacher, 1980).

D. Technology-related: identification and copyright problems

Some of the characteristics which increase the utility of machine readable data also create identification and copyright problems. Machine readable records are easily reformatted, updated, and otherwise changed, copied and transformed, making identification and definition of the original data base especially difficult. Data files are often copied and diffused through numerous agencies (and also within agencies) and the question of which file or version of a file is the record copy becomes obscured. The origin of reformatted files or of multiple source documents used to create a file is not usually identified, nor are many data bases created by government agencies from administrative records and subsequently distributed by private sector vendors.

The University of Wisconsin and the State Historical Society of Wisconsin have been carrying out a federally supported research project to determine the impact of computerization on State agency public records. During our investigations, we have found a number of data bases whose data elements are drawn from multiple source documents, data elements which appear to be redundant (but upon investigation are not because they have come from different sources), and whose data elements are said to contain all the information from the source documents. The reformatting of these multiple source documents in machine readable form, deletion of selected data elements, updating and transformation of the original elements, and lack of identification of original inputs and data products have made it difficult for the project archivist to identify the master data files and to link the source documents and data files. A systematic inventory of machine readable public records within one state agency thus becomes a frustrating task for the archivist, calling for extensive detective work and almost infinite patience (Hedstrom, 1980a,b).

E. Technology-related: hardware and software dependence

Computer-readable numerical data are inextricably bound to technology. The technology has changed so dramatically that various computers used less than

fifteen years ago are no longer available. Data files designed in the past may not be usable on contemporary machines and, consequently, data files sometimes cannot be transferred from one computer site to another. The degree of transportability is an important characteristic of access to numerical data.

Two basic problems are encountered with data management, processing, and statistical software for reformatting, retrieving, and manipulating data. Data base management systems have been developed which allow a user access to different files from different locations, bring these files together, select elements of the different files, and create entirely new files. While this is an exciting development, most of the present data base management systems (DBMS) do not facilitate transfer to other computers.

The growing complexity and scope of the internationally and nationally distributed statistical packages is also a problem. Although many more people have sufficient computing skills to use these software packages, they may not properly understand them or the consequences of making mistakes in using them.

F. Technology-related: storage and preservation

Current magnetic storage media and most of the mass storage devices now in various stages of development will not meet archival storage requirements of preserving digital data for a very long time. Problems associated with permanent preservation of data derive from the physical size of the data file, machine dependence and media standardization, reliability of the storage medium, the medium's sensitivity to environmental conditions, lifetime maintenance cost of the medium, ease of accessibility required of the data, and cost of duplication (Robbin, 1980a). Magnetic tape offers compatibility when written on different tape drives if the same density and character codes are used and utility software is available to translate the character codes (Dollar, 1977). However, all other magnetic storage media are machine dependent and non-standard, creating problems when equipment used for the initial creation and copying of data is removed. Only limited empirical evidence is available to ascertain how long the various storage media retain a reliable image of the data. Environmental and handling conditions affect the lifetime of the magnetically encoded data and necessitate expensive environmental controls. To ensure permanent accessibility to stored information, data tapes must be duplicated periodically. A regular programme of maintenance and preservation must be instituted and contributes to the hidden costs of maintaining a library of numerical data.

G. Confidentiality and privacy

Confidentiality and privacy constitute one of the most troubling problems of access to statistical information on data subjects. The private sector maintains control over collection of information deemed potentially harmful to individual human (and corporate) subjects and in a number of countries data protection and privacy laws have been established to limit both collection of and access to personal data. But, as funds for primary research become more difficult to obtain and greater demands for accountability are placed upon government, the research community becomes a more active consumer of both research information and

administratively-produced information of a confidential nature. The growing conflict between governments and the social research community (Mochmann and Muller, 1979) and the private sector (Walsh, 1978) on what constitutes risk to data subjects presents a problem which has important political, research and development, and economic facets for researchers, government, and industry. It is unlikely that the debate over access to data deemed 'personal' or confidential will abate over the next few years. It is imperative, of course, that safeguards be instituted to ensure adequate protection of subjects. Although various administrative and statistical techniques are available to protect individual identity (Boruch and Cecil, 1979; Robbin, 1979a) and to create disclosure-free or limited-risk, public-use files, ethical considerations, legal rulings, and illogical and *ad hoc* administrative rules have significantly affected research and statistical manipulation of data.

H. Costs

Numerical data represent a costly capital investment which usually must be recouped from a consumer public which is significantly smaller than for other information resources. Numerical data cost substantially more to produce than other types of information. There are data collections whose costs have been estimated at well over a million dollars and which are made available to users (including documentation) at a cost of three hundred dollars. The limited resources of the social scientist are a factor which must be considered when there is a need for numerical data.

Use of computer channels is also expensive. The social scientist typically 'passes' a great deal of data, making his input-output costs significant, even though in comparison to the natural or physical scientist, he does limited computation. The cost structure of most computational facilities reflects the needs of the natural scientist, at least in an academic environment, unless there is a machine dedicated for social science use.

Many social researchers typically study large populations and employ a large number of variables to test hypotheses. Technical developments make it possible to receive and transmit data from great distances, obviating the necessity for actually maintaining the data at a local repository. However, it is not yet cost-effective to transmit large quantities of numerical data for a user community which in general has limited computing funds, and at a high enough speed to match on-site computing capabilities.

We have already described the need for data independence from computer hardware and software. In some cases, the cost of making data independent may be high because special software must be written to extract the data base from the software and specialists' time is involved in writing the program instructions and executing the operation. Hedstrom (1980b) provides an example of this cost in her review of the State of Wisconsin Department of Revenue's data processing activities and their relationship to public records. She notes that data in the *Local Financial Assistance* data base can be extracted and reformatted into fixed format records which can then be accessed without special software only at considerable expense: the costs of reformatting one file (one data set for one year) have been estimated at fifty dollars for computer time and between fifty and several hundred dollars for analysts' time.

PART IV. CHARACTERISTICS OF THE SOCIAL SCIENCE INFORMATION SYSTEM FOR COMPUTER-READABLE STATISTICAL DATA

Social research and administrative statistical activities are becoming more dependent on numerical data which permit cross-national comparisons and replication when the investigations concern problems which transcend political and geographic boundaries. Examples might include the extent of assistance provided by the public sector for unemployment compensation and family planning, the role of the private sector and its influence on government programme planning, compensatory education for immigrants, or the levels of literacy required to cope with an automated society.

Information systems do not operate in isolation, but within political, social and economic contexts. Development of appropriate subnational, regional, national and international networks and tools is dependent on the relationship of the members of the information system to both national and international policy makers who provide political, economic and institutional support for the development of the information infrastucture. The following discussion examines the nature of this infrastructure and the formal and informal communications channels for statistical data. If the examination appears unduly pessimistic, it is because problems first observed more than twenty years ago have still not been resolved. However, the remaining part of this section provides some concrete examples of efforts to ameliorate the situation, which give hope.

A. Nature of the infrastructure for computer-readable statistical data

The infrastructure can be characterized as inchoate for numerical data. Most producers of machine readable data files have neither the resources nor expertise to preserve, maintain and distribute their data. They do not have established procedures for easy access to and inexpensive use of their data. Few data producers can supply adequate user support services related to teaching, research, coordination of multiple research projects, computation and general information services related to machine-readable data files, which are often by-products of large data-gathering projects.

Primary and secondary information services either do not exist or provide limited information and data services in local environments. Few libraries have integrated the numerical data base into their collections, although it can logically be considered another information resource. As a result, the traditional abstracting and indexing services and other tools and services which serve a document-based community do not exist for numerical data. With few exceptions, data libraries have grown up outside the traditional library, without the foundation of generations of experience in providing technical and public services. Data libraries typically exist as adjuncts of social science departments and have not employed trained specialists in information management. There are historical reasons for this.

In the late 1950s, Lucci and Rokkan (1957), an American library school faculty member and a Norwegian sociologist, proposed establishing data libraries within the library framework. During the 1960s, Adams and Dennis (1969) and others continued to argue that the natural setting for providing data was through institutionalized, local libraries. Responsibility for the lack of integration of the data resource into the library setting falls on the library profession. Until recently, it has been slow to respond to computer technology, intimidated by automation

and by numerical information, and highly resistant to introducing new methodologies and concepts into the curricula in library schools and new information resources into the traditional library. Library and information science students have shown an amazing ignorance and lack of interest in anything which suggests an interdisciplinary approach, particularly outside the library school. This has made it difficult to introduce innovations and flexible approaches to problems of information organization, transfer and management. Because the requirements for adequate data servicing involve skills in social and behavioural theory building, the principles of instrument design and measurement, statistics and computer science (David, 1980), few librarians have been qualified to work in a social science data library.

In recent years the information industry has begun to offer seminars on non-bibliographic data bases to libraries and information specialists. But training in these data bases is limited to learning the access and retrieval protocols. Librarians are left with the impression that little information is required to understand the data which they extract for the end-user. No evaluation of statistical estimation techniques and computer algorithms is provided. Lack of real understanding leads to general acceptance of the veracity of the information.

Few information specialists can assist in locating data, so the social scientist must spend a great deal of time searching for data. In only a few places, primarily in the academic environment, are there information specialists prepared to provide public and technical data services. Most data services are not formally linked to one another (although the International Federation of Data Organizations is designed in part to respond to this situation) and data services professionals have only one professional arena in which to interact (the International Association for Social Science Information Services and Technology—IASSIST). Numerical data base services are not usually institutionalized in the United States, in contrast to certain countries in Western and Eastern Europe, where data services now receive support from the national governments.

The conjunction of lack of integration into the traditional library framework, of formal training in information management by data library staffs, of expertise by library students, and of institutionalized professional structures have contributed to an inefficient and inadequate information system for the quantitative social scientist.

B. Nature of the formal communication channels

Serendipity or accidental discovery of information is a common phenomenon for the quantitative social scientist and data services professional. Mention of use of a data file on which the analysis was based may be made in a journal article and reference to ongoing research involving data collection may be made in a publication or report. But typically, published articles and reports contain no mention of the machine readable data file. Without proper citation, machine readable data files cannot enter the formal information system. Lack of formal channels is also costly: many students and researchers conduct primary data collection because they do not know about existing collections. In some cases, several years have been devoted to coding data from published sources when the data have long been available in machine readable form. Current information on work done in foreign countries is always difficult to obtain. Efforts to report on

ongoing or completed research are being made, but information on these efforts is not widely known.

One reason why numerical data bases are not abstracted or indexed in journals is that most have not carried formal bibliographic information and thus it has been impossible to define the data base's existence. Its producers and users have not viewed the numerical data base as another medium of information, probably because it is the product of data base use which is considered useful (the published report or article) and because the numerical data base is considered part of the process by which one solves problems and answers questions. Until recently (Dodd, 1977, 1979a,b; Robbin, 1980b; Roistacher, 1980), data producers and consumers and the library profession (American Library Association, 1979; Dodd, 1980a) have had no guidance in describing and documenting machine readable data.

Another reason why data bases are not indexed is that few catalogues or inventories of numerical data exist and most of those have inadequate entries. The reasons for these inadequacies include:

(1) the terminological difficulties inherent in the social sciences;
(2) the difficulty of describing the contents of a data file to individuals of different disciplines having different analytical frameworks and methodologies, but who could potentially make use of the same data file;
(3) since data services are designed for a local user community, inventories do not need to fully describe the contents of a data collection because the user can easily examine the data documentation;
(4) most data users search for specific variables or subsets of variables and need to know the conceptual linkages between and among the variables (Heim, 1980). With the growing number of data files containing more than a thousand variables (items of information), it becomes increasingly difficult to describe the contents of a data file;
(5) data staffs lack experience in developing bibliographic tools for their clientele, no matter how knowledgeable they are about their collections; and
(6) there has been limited agreement on what constitutes an adequate description that will lead the potential user to a data file.

Staffs have tried to respond to the inadequacy of catalogues and inventories by producing abstracts and detailed study descriptions of the contents of the data files in their collection and to automate the detailed study descriptions. In general, these descriptions are non-standard and their contents vary in quality depending upon the resources allocated to this problem and the expertise of the staff (knowledge of content analysis, of retrieval languages and of the quality of information storage and retrieval software). These abstracts and study descriptions have had limited distribution within a small community of data libraries and have not entered the wider information system. There is still no agreement on the desirable minimal levels of information which must be provided.

There are no formal channels for supplying data to the user. Again, the local nature of data services and lack of institutionalized and permanent structures for numerical data transfer make it very difficult for the user to obtain data. Data are supplied by data processing departments, individuals or computing centres, once the user has located a file and specified the configuration of his site computer. Most of these organizations have little experience with servicing the user outside the centre. Few end-users are experienced in using computer networking facilities

to access available data at other installations and the costs of computer networking are still too high for the average social scientist.

C. Nature of the informal communication channels

It has already been noted that few information specialists are knowledgeable about numerical data bases. But even these few are not fully utilized to locate and acquire data and use the data base. The quantitative social scientist prefers to rely on colleagues and on the periodical literature which may mention a potentially interesting data file. Although consultants on the computing aspects of his problem are queried more often than other information specialists, the social scientist prefers to investigate the problem himself.

Results of a small user study of the Data and Program Library Service graduate student population confirmed the importance of informal communication channels (Heim, 1980). More than 60 per cent of the students learned of the existence of numerical data from a faculty member, colleague or friend; less than 30 per cent learned about data files from reports, journal articles, monographs, or by coming to the Data Library to search its directory and abstracts. The faculty member is usually a specialist in a subject area and has a well developed network of colleagues who are working on similar problems. He has more extensive resources for locating information about numerical data, and is therefore less dependent on formal channels of communication. He only uses the data specialist for 'filling in the gaps' in case he might have missed something and for communicating the technical language required for obtaining the data base from another colleague.

Informal channels for obtaining data contain hidden costs of which the social researcher may be unaware. By obtaining a copy of the data from a colleague, the researcher makes it impossible to capitalize on the initial investment in data preparation, typically lengthens the time it takes to access the data because neither of them understands how to communicate technical information on the physical structure of the data, and contributes to the problem of maintaining records on the diffusion of data.

D. Nature of the infrastructure and communication channels and statistical computing

This discussion has emphasized information needs in terms of the data themselves, but some additional comments on the nature of communications channels and statistical computing needs are warranted. Formal information channels for acquiring software and information on statistical computing are few. There is little formal course instruction offering students laboratory experience in solving statistical problems. Few academic institutions provide self-teaching materials and remedial assistance for the social scientist. Informal channels provide most of the information about software and the computer environment. Formal statistical consulting activities are limited, in comparison to consulting provided for other types of computation. Most computational facilities are not staffed with people who are interested in providing social science data management consulting services. Unless the social science community represents a significant constituency at an institution, the computer centre generally remains uninterested in providing computational services. Lack of formal channels for communicating information about available statistical software results in unnecessary duplication of effort.

E. Efforts to improve formal and informal communication of information about numerical data

Several data repositories (such as the University of Iowa and the University of Bergen) have responded to the absence of a formal system by publishing newsletters and inventories describing available data resources. The abstracting services are underwritten by the respective institutions and subscription fees are either minimal or non-existent. While the newsletters are highly dependent on the vagaries of interest by data services and individuals who provide the information (Kolp, 1980), they do appear in the formal guides to the literature.

The formal journals of many social science professional associations are beginning to provide information on data gathering and numerical data bases available for public distribution. For example, the Data and Program Library Service is the archive of records for a series of American national fertility surveys conducted since 1955. When the last survey (1975) was ready for public distribution, DPLS announced its availability in every journal and newsletter which dealt with family planning and population studies. Within two months, twenty requests for the data files had been received.

Individuals, such as Sessions (1974) and Wasserman (1971) have devoted considerable time to developing descriptions of non-traditional information products. Recently, Dodd (1980b) discussed machine readable data files in a major journal for information scientists.

Efforts to standardize descriptions of machine readable data files in a computerized data base are being made by a group of data archives, within the International Federation of Data Organizations and the International Association for Social Science Information Services and Technology. The Zentralarchiv für Empirische Sozialforschung (ZA) at the University of Cologne has led the way in developing specialized information storage and retrieval software for an automated data base describing in great detail its survey data collection. The ZA, along with the Danish Data Archives at the University of Odense, the Steinmetz Archives (SWIDOC) in Amsterdam, and the Leisure Studies Data Bank at the University of Waterloo (Canada) have been cooperating in developing bibliographic data bases which describe the holdings of their respective collections. Efforts to integrate the data bases have begun and there is sharing of experiences during all the professional meetings. The next steps will be to enlarge the number of data archives participating in the cooperative effort, solicit active participation from the library and information science community, and work toward developing standards for retrieval terminology.

Other institutions have created automated bibliographic data bases describing the contents of their machine readable data files, such as Stanford University, University of Alberta, Northwestern University and the University of British Columbia. The British Columbia Data Library compiles a detailed automated abstract entry which is then transferred to a professional cataloguer in the main library, who in turn converts the information into the required UTLAS format (University of Toronto Library Automated Systems).

Two other institutional activities deserve mention, because they combine different and complementary areas of computer and information storage and retrieval software technology, classification and indexing.

The Institute for Research in Social Science at the University of North Carolina at Chapel Hill (Louis Harris Data Archive) has created a prototype of a multi-purpose information system of entries describing a subset of their statistical data holdings. Products such as catalogue entries, directories, inventories,

abstracts and indexes can be generated by this versatile system. Searches can be made on item level questions from surveys conducted by Louis Harris, Inc. The automated data base can be accessed through the Triangle Universities Computing Center via the on-line Telenet communications facility in an on-line mode.

The Inter-university Consortium for Political and Social Research (ICPSR), the international academic data repository for social research data, is now developing a multi-purpose data base to describe their holdings along the same lines as the University of North Carolina. The information system will be accessible to more than 240 member institutions through the MERIT network and the Telenet communications facility. In both cases, the system implements many of the data elements provided in the MRDF/MARC format (Dodd, 1980a). Because both information systems are available through Telenet, information on statistical data bases will be available internationally.

Through the efforts of Dodd and members of the Classification Action Group of IASSIST, the machine readable data file can now be properly cited in a publication. The bibliographic citation element has been incorporated in the study description record by the several university data services mentioned earlier and in a data abstract description and title page verso of a 'User's guide to the machine readable data file' (see Roistacher, 1980). The bibliographic citation incorporates all the required and optional fields of a main entry as recommended by the *Anglo-American Cataloguing Rules* (American Library Association, 1979). A cataloguing-in-source entry now appears in the 'User's guide' prepared by the Data and Program Library Service, National Opinion Research Center, U.S. Bureau of the Census, and the Inter-university Consortium for Political and Social Research. The International Standard Book Number (ISBN) is applied to the user's guide by several data libraries and the Data and Program Library Service applies it to the machine readable data file as well.

PART V. STRATEGIES FOR IMPROVING UTILIZATION OF COMPUTER-READABLE STATISTICAL DATA

The process by which data become accessible to researchers is frustrating, time consuming and inefficient (see David, 1980). Data and data information services, computer centres and statistical consulting services are not integrated into the larger social science information system, but are compartmentalized by carefully defined functions without recognition that the process of problem-solving is multi-dimensional and requires an information support structure integrated along the same dimensions as the intellectual problem-solving activity. Information about data is obtained haphazardly and often too late to contribute to the empirical investigation. Thus, the present system requires that the social scientist be highly articulate about his problem, motivated and self-reliant. To rectify this situation, I recommend four general strategies for improving utilization of computer-readable statistical data. These are set out below.

(1) Improving the quality of statistical evidence

With increasing dependence on statistical data for public policy decisions, it becomes critical that standards for data quality control be established to ensure

their utility. These standards reflect the philosophy that statistical data must represent objective and verifiable evidence that is unambiguously described, so that analysis or evaluation based upon this evidence can be effectively reviewed, criticized and replicated (Schmandt, 1978). Developing criteria for what constitutes good data is essential to providing access to both primary and secondary data sources. The primary goal of these standards should be to ensure that data producer and analyst can make maximum use of machine readable data files (Robbin, 1980b).

This does not imply that high quality statistical data production and description are inexpensive; their initial high cost is acknowledged. Nor does it imply that all data files are worthy of this investment. But there are potentially higher gains for the knowledge flow system in terms of 'efficiency, quality, and capacity to reach a [larger] audience' (Havelock, 1968: II-5) than the initial data collection effort intended. Provision of standards offers a strategy for maximizing the returns on capital investment in data resources and contributes toward better quality innovation, dissemination and information transfer.

(2) Educating the information specialist and end-user

To prepare the information transfer professional to deal adequately with future developments, there must be an

> 'intellectualization, expansion and deepening of library school goals, objectives, and curricula to encompass all facets of document organization and information service activities applicable for the organizing and servicing of recorded information in traditional as well as nontraditional information service environments.' (Kent, 1978: 12)

The library school student must develop a more 'penetrating understanding of the information transfer process' in order to create and participate in 'information systems to support the new intellectual technology' (Kent, op. cit). It seems certain that information has increasing potential uses by a diverse group. More emphasis in library school curricula will have to be placed on the interdisciplinary nature of statistical information, that is, the intellectual and technical means by which the information is produced, processed, transferred, reduced and interpreted. This emphasis calls for a new curriculum model where *process* (the communication model) becomes the guiding principle underlying the intellectual and technical skills which the library or information science professional may apply in any place of work (see Saracevic, 1978).

The general level of numeracy must be raised not only at higher institutions of learning, but also in elementary and secondary schools. At the university level, students and faculty must be trained to evaluate and use numerical information. Doubtless this suggestion is fraught with major educational, political and economic consequences. It must be recognized that the technologies of data processing, storage and transmission have changed the relationship between the individual and information and altered the economic structure of society. This technology has the potential for transforming society so that discrimination will be based less on the storage of information than on the ability to research and use it (Nora and Minc, 1980). Without a reorientation in the structure of learning and communication, we will be unable to create an equitable distribution of and access to information.

(3) Use of the existing information infrastructure

Libraries and archives preserve data for future use and give access to information about data. Librarians should consider integrating numerical data bases in their collections to encourage use of quantitative data and enrich the possibilities for research and instruction. Heim's findings show that the quantitative social scientist would prefer to have numerical data integrated into library collections, but with the proviso that the full panoply of services be provided. While integration of a full range of data services is not yet feasible, better access to information about data bases could be provided through catalogues, better documentation and union lists. Links could be created to other services, such as the computer centre. Linkages would reduce overall information service costs to an institution (many of which are invisible).

Networks appear to work quite well for the highly sophisticated, motivated and well-funded user, but for almost everyone else, assistance needs to be provided in close proximity to the user. There may be some change in the next decade, but the psychological, economic and political barriers to computer networking will be difficult to overcome and economic, political and ethical considerations have tended to constrain efficient use of the modern computer technology. These realities suggest that the information centre will continue to play a major role in the transmission of information (Robbin, 1980a).

The data archive is the best vehicle for preserving data and documentation. Data repositories act as vehicles for magnifying the return on the original capital investment by providing established distribution channels for data and information. The data repository provides easy and relatively inexpensive access to data and documentation. It can provide user support services related to the data file, such as copying data and documentation, showing users how to understand the contents and structure of a data file, and consulting on data use and computation. Further, a data archive can serve as a critical feedback mechanism for users of a data file because the centre circulates reports and publications related to research using the data as well as information on modifications, updates and errors in the data file. The archive offers the best mechanism for improving data access when the information dissemination and transfer channels are neither integrated nor coordinated (see Miller, 1976; Neswold, 1976; Rokkan, 1976; Traugott and Clubb, 1976; for additional arguments on the importance of the data archive).

(4) Developing retrieval tools for facilitating access to information about machine readable data files

Facilitating access to machine readable data files involves developing methods of publicizing their existence and improving standards for description and control.

Better bibliographic control over machine-readable data files is needed. The data producer should follow accepted bibliographic practices for describing his data collection. Data producers must be encouraged to prepare 'title pages' with appropriate information. They must be convinced that this effort will encourage more use of a significant capital investment and will enhance public access to data resources. This will only happen if information specialists play a larger role in numerical data base transfer. The cataloguing expertise of the librarian needs to be applied to machine readable data files (cataloguing-in-source).

Abstracting and indexing services could supply information about data bases

and also identify the data base as part of its description of the information resource. The medium designator, MRDF, is now an optional element in the title for a catalogue entry (American Library Association, 1979) and could be included in the title and abstracts of the on-line bibliographic data base record to indicate a machine readable data file resource. The MRDF designator in bibliographic descriptions would provide us with a useful crude tool for measuring the use of machine readable data files, just as is done for other information resources. Since data production is increasingly expensive it seems that a better case for its exploitation could be made if actual use could be demonstrated.

The quantitative social scientist, policy maker and practitioner require multi-purpose bibliographic data bases capable of generating abstracts, directories, indexes and card catalogues. An information data-base on machine-readable data must describe both the contents of the machine-readable data file, and the published and unpublished reports on methodological and substantive problems with the data and uses of the specific file (see David, 1980). It must provide direct access via software interfaces to the data themselves (the processes of search and statistical analysis are tied together to increase accessibility to the data [University of Wisconsin Data and Computation Center, 1978]). The interface procedures call for high-level programming and software development and probably a significant capital outlay; but, in principle, the set of procedures is possible with existing computer and telecommunications technology.

Where a major investment must also be made, however, is in the development of internationally standardized descriptions of the numerical data base. This effort calls for a funding commitment for a feasibility study and eventually may demonstrate the need for a viable bibliographic data base system supported by government, industry and the consumer. This investment requires cooperation by government, industry, the information societies, data producers, data libraries and archives professionals, and the social scientific community.

CONCLUDING REMARKS

If society is really committed to information dissemination to increase and make better use of knowledge, then formal channels for access to numerical data must be developed and provided at low cost and without difficulties to the end user. At present, the cost of access to and retrieval of numerical data bases places this information resource beyond the reach of a large potential-audience of students, researchers, practitioners, policy makers and most certainly the mass public. The high cost of numerical data suggests that they are a resource that needs to be shared in order to be supported.

The variety of data needs for problem-solving require different data structures and means for accessing data. While the research process is open-ended, practitioners and policy makers have very specific needs for data; these needs must be met in efficient, timely and inexpensive ways. The large data bases being developed to respond to administrative reporting requirements and used for research and policy questions must be organized in tractable ways.

While it is recognized that statistical information is a powerful tool and that safeguards must be instituted to protect the disclosure of information which places an individual at risk, there must also be a commitment to timely release of data. The rationale for supporting infrastructure development, tools for information and data retrieval, and changes in the curriculum of the library and

information professional is based on the principle that a society has a commitment to information as a national resource. Statistical data represent one of our valuable information resources.

Acknowledgements: I gratefully acknowledge Kathleen Heim's contribution to my education in library and information science and for sharing the preliminary findings of the use-study conducted at the University of Wisconsin-Madison Data and Program Library Service; Christopher Perry for assistance in developing the typology of the social science numerical data user; and Sue Dodd, Nancy McManus, Tefco Saracevic and James W. Duncan for their most valuable comments on an earlier version of this paper. I alone bear responsibility for the opinions registered here.

This paper was edited for *SSIS* by David R. Streatfield.

[1] Lindblom and Cohen (1979) describe these individuals as those involved in 'Professional social inquiry'. The diverse activities of this group include evaluation studies in profit-making firms; opinion research in commercial survey organizations; academic social research (broadly defined to include anthropology, history, urban and regional planning, social work, journalism and communication arts, and other areas where the study of human behaviour occurs); government policy analysis; journalism focused on aspects of societal behaviour; and research and analysis aimed at solving social problems carried out in specialized R & D 'laboratories' in such fields as education or mental health.

REFERENCES

ADAMS, M. and DENNIS, J. (1969). Creating local social science data archives. *Social Science Information, 9*, 51–60.

ALLEN, T. J. (1977). *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the Research and Development Organization.* Boston: MIT Press.

AMERICAN LIBRARY ASSOCIATION (1979). *Anglo-American Cataloguing Rules, 2nd edition.* Chicago: American Library Association.

AMERICAN PSYCHOLOGICAL ASSOCIATION (1969). *The American Psychological Association's Project on Scientific Information Exchange in Psychology.* Washington, D.C.: American Psychological Association.

BAILER, B. and LANPHIER, C. M. (1978). *Development of Survey Methods to Assess Survey Practices.* Washington, D.C.: American Statistical Association.

BORUCH, R. F. and CECIL, J. S. (1979). *Assuring the Confidentiality of Social Research Data.* Philadelphia, Penn.: University of Pennsylvania Press.

BRITTAIN, J. M. (1970). *Information and Its Users: a Review with Special Reference to the Social Sciences.* New York: Wiley-Interscience.

BRITTAIN, J. M. (1979). Information services and the structure of knowledge in the social sciences. *International Social Science Journal, 31*, 711–728.

DAVID, M. (1980). *Access to data: the frustration and Utopia of the researcher.* Presented at the International Association for Social Science Information Services and Technology Annual Meeting, 2–4 May 1980, Washington, D.C.

DODD, S. A. (1977). Titles: the emerging priority in bringing bibliographic control to social science machine-readable data files (MRDF). *IASSIST Newsletter, 1* (No. 4), 11–18.

DODD, S. A. (1979a). Bibliographic references for numeric social science data files: suggested guidelines. *Journal of the American Society of Information Science, 30*, 77–82.

DODD, S. A. (1979b). Building an on-line bibliographic/MARC resource data data-base for machine-readable data files. *Journal of Library Automation, 12*, 4–21.

DODD, S. A. (1980a). *Cataloguing Machine-Readable Data Files: an Interpretive Manual.* Chapel Hill, N. Caro.: University of North Carolina Institute for Research in Social Science (forthcoming).

DODD, S. A. (1980b). Data bases. *Bulletin of the American Society for Information Science, 6* (3), 16–17.

DOLLAR, C. M. (1977). Problems of magnetic recording in archival storage. *Digest of Papers.* Spring COMPCON 1977, 14th IEEEC Society International Conference, San Francisco, 28 February–3 March, 28–30.

GARVEY, W. D. and GRIFFITH, B. C. (1972). Communication and information processing within scientific disciplines: empirical findings for psychology. *Information Storage and Retrieval, 8,* 123–136.

GARVEY, W. D., LIN, N., NELSON, C. E. and TOMITA, K. (1972). Research studies in patterns of scientific communication: 1. General description of research program. *Information Storage and Retrieval, 8,* 111–122.

GARVEY, W. D., LIN, N., NELSON, C. E. and TOMITA, K. (1972). Research studies in patterns of scientific communication: 11. Information-exchange processes associated with the production of journal articles. *Information Storage and Retrieval, 8,* 207–221.

GARVEY, W. D., LIN, N. and TOMITA, K. (1972). Research studies in patterns of scientific communication: IV. The continuity of dissemination of information by 'productive scientists'. *Information Storage and Retrieval, 8,* 265–276.

HAVELOCK, R. G. (1968). *Planning for Innovation through Dissemination and Utilization of Knowledge.* Ann Arbor, Mich.: University of Michigan Institute for Social Research Center for Research in Utilization of Scientific Knowledge.

HEDSTROM, M. (1980a). *A report on the investigation of data processing activities and the machine readable data files in the Department of Public Instruction (DPI).* Madison, Wisc.: University of Wisconsin Data and Program Library Service and State Historical Society of Wisconsin. (Survey of machine readable public records project report #1.)

HEDSTROM, M. (1980b). *A report on data processing and machine readable records in the Wisconsin Department of Revenue.* Madison, Wisc.: University of Wisconsin Data and Program Library Service and State Historical Society of Wisconsin. (Survey of machine readable public records project report #2.)

HEIM, K. (1980). *The social science data archive: a user study.* Madison, Wisc.: University of Wisconsin School of Library Science. Unpublished dissertation (forthcoming).

JOHNS HOPKINS UNIVERSITY CENTER FOR RESEARCH IN SCIENTIFIC COMMUNICATION (1967). *A comparison of the dissemination of scientific and technical information associated with two meetings of the American Institute of Aeronautics and Astronautics.* Baltimore: Johns Hopkins University. (Report #1.)

JOHNS HOPKINS UNIVERSITY CENTER FOR RESEARCH IN SCIENTIFIC COMMUNICATION. (1967). *Scientific information-exchange behavior at the 1966 Annual Meeting of the American Sociological Association.* Baltimore: Johns Hopkins University. (Report #4.)

JOHNSON, S. (1977). *File documentation as a key to the appraisal of machine readable data sets.* Paper prepared for F. Gerald Ham, Madison, Wisconsin.

KENT, A. (1978). The information science-library science interface—curriculum design and change. *Information Reports and Bibliographies, 7* (No. 4), 10–14. (Conference on education for information science-strategy for change in library school programs, 24–25 April 1977, Albany, New York.)

KOLP, J. G. (1980). *Reference Tools for Machine Readable Data Files.* Presented at the International Association for Social Science Information Services and Technology Annual Meeting, 2–4 May 1980, Washington, D.C.

LA COUR, A. 'Policy and recent developments in EUROSTAT—methods of disseminating statistics' *Statistisk Tidskrift/Statistical Review. 15,* 333–347. Reprinted in *European Political Data Newsletter, 35* (June 1980), 7–16.

LANSING, J. B. and MORGAN, J. N. (1971). *Economic Survey Methods.* Ann Arbor, Mich.: University of Michigan Institute for Social Research.

LINDBLOM, C. E. and COHEN, D. K. (1979). *Usable Knowledge: Social Science and Social Problem Solving*. New Haven, Conn.: Yale University Press.

LINE, M. B. (1969). Information requirements in the social sciences: some preliminary considerations. *Journal of Librarianship*, *1*, 1–19.

LINE, M. B. (1971). The information uses and needs of social scientists: an overview of INFROSS. *Aslib Proceedings*, *23*, 412–434.

LINE, M. B., BRITTAIN, J. M. and CRANMER, F. A. (1971b). *Information requirements of social scientists in government departments*. Bath, England: Bath University. (Report #2.)

LUCCI, Y. and ROKKAN, S. with MEYERHOOD, E. (1957). *A library center of survey research data: a report of an inquiry and a proposal*. New York: Columbia University School of Library Science.

MILLER, B. A. (1974). *Characteristics of access to machine-readable statistical data files*. Chicago: University of Chicago Graduate Library School (unpublished dissertation).

MILLER, W. E. (1976). The less obvious functions of archiving survey research data. *American Behavioral Scientist*, *19*, 409–418.

MOCHMANN, E. and MÜLLER, P. M. (eds) (1979). *Data Protection and Social Science Research: Perspectives for Ten Countries*. Frankfurt: Campus Verlag.

NESVOLD, B. A. (1976). Instructional applications of data archive resources. *American Behavioral Scientist*, *19*, 455–466.

NORA, S. and MINC, A. (1980). *The computerization of society: a report to the President of France*. Cambridge, Mass.: The MIT Press.

PAISLEY, W. M. and PARKER, E. B. (1967). *Scientific information exchange at an interdisciplinary behavioral science convention*. Palo Alto, Cal.: Stanford University Institute for Communication Research.

ROBBIN, A. (1975). Managing information access through documentation of the data base: characterizing social science data base text documentation as a minimal information system. *SIGSOC Newsletter*, *6* (Nos. 3 and 4), 56–68.

ROBBIN, A. (1979a). Ethical standards and data archives. *In: Secondary Analysis* (R. F. Boruch, ed.). San Francisco: Jossey-Bass, Inc., 7–18. Also published in *Data Protection and Social Science Research: Perspectives for Ten Countries* (1979) (E. Mochmann and P. M. Müller, eds). Frankfurt: Campus Verlag.

ROBBIN, A. (1979b). Understanding the machine readable numeric record: archival challenges, with some comments on appraisal guidelines. *The Midwestern Archivist*, *4* (No. 1), 5–24.

ROBBIN, A. (1980a). *A profile of data preservation: activities in university data libraries and archives, January 1980*. Madison, Wisc.: University of Wisconsin Data and Program Library Service.

ROBBIN, A. (1980b). Technical guidelines for preparing and documenting statistical data for secondary analysis. *In: Secondary Analysis: Policy and Practice in Applied Social Research* (R. F. Boruch, P. M. Wortman, and D. Cordray, eds). San Francisco: Jossey-Bass, Inc. (forthcoming).

ROISTACHER, R. C., with DODD, S. A., NOBLE, B. and ROBBIN, A. (1980). *A Style Manual for Machine Readable Data Files and their Documentation*. Washington, D.C.: Government Printing Office (forthcoming).

ROKKAN, S. (1976). Data services in Western Europe: reflections on variations in the conditions of academic institution-building. *American Behavioral Scientist*, *19*, 443–454.

ROKKAN, S. (1978). National primary socio-economic data structures: III. Norway. *International Social Science Journal*, *3*, 621–652.

SARACEVIC, T. (1978). A model of knowledge communication as a unifying framework for education in librarianship and information science. *Information Reports and Bibliographies*, *7* (No. 4), 26–34. (Conference on education for information science—strategies for change in library school programs, 24–25 April 1977, Albany, New York.)

SCHMANDT, J. (1978). Scientific research and policy analysis (editorial). *Science*, *201* (4359), 869.

SESSIONS, V. S. (ed) (1974). *Directory of Data Bases in the Social and Behavioral Sciences.* New York: Science Associates/International, Inc.

SKELTON, B. (1973). Scientists and social scientists as information users: a comparison of results of science user studies with investigation into information requirements of the social sciences. *Journal of Librarianship, 5*, 138–156.

TRAUGOTT, M. W. and CLUBB, J. M. (1976). Machine-readable data production by the Federal Government: access to and utility for social research. *American Behavioral Scientist, 19*, 387–408.

UNITED NATIONS. DEPARTMENT OF INTERNATIONAL ECONOMIC AND SOCIAL AFFAIRS. STATISTICAL OFFICE (1979). *The development of integrated data bases for social, economic and demographic statistics.* (Studies in methods series F (No. 27).) New York: United Nations.

UNIVERSITY OF BERGEN. NORWEGIAN SOCIAL SCIENCES DATA SERVICES AND THE EUROPEAN CONSORTIUM FOR POLITICAL RESEARCH (1971–). *Euorpean Political Data.* Bergen, Norway: University of Bergen Norwegian Social Science Data Services and the European Consortium for Political Research.

UNIVERSITY OF COLOGNE. ZENTRALARCHIV FÜR EMPIRISCHE SOZIALFORSCHUNG (1968–). *Empirische Sozialforschung.* München-Pullach: Verlag Dokumentation Saur K.G.

UNIVERSITY OF IOWA. THE LABORATORY OF POLITICAL RESEARCH (1971–). *S S Data: newsletter of social science archival acquisitions.* Iowa City, Iowa: University of Iowa Laboratory for Political Research.

UNIVERSITY OF WISCONSIN-MADISON. ACADEMIC COMPUTER ADVISORY COMMITTEE SUB-COMMITTEE ON STATISTICAL CONSULTATION (1979). *Report.* Madison, Wisc.: University of Wisconsin Academic Computer Advisory Committee.

UNIVERSITY OF WISCONSIN-MADISON. DATA AND COMPUTATION CENTER (1978). *A proposal to define and demonstrate capabilities for rapid retrieval of statistical data contained in machine-readable data files.* Madison, Wisc.: University of Wisconsin Data and Computation Center.

VONDRAN, R. F., JR. (1976). *The effect of method of research on the information seeking behavior of academic historians.* Madison, Wisc.: University of Wisconsin School of Library Science (unpublished dissertation).

WALSH, J. (1978). There's trouble in the air over transborder data flow. *Science, 202*, 29, 31–32.

WASSERMAN, P. (ed) (1971–). *LIST 1971–. Library and Information Science Today: an International Registry of Research and Innovation.* New York: Science Associates/International, Inc.

WHITE, H. D. (1974). *Social science data sets: a study for librarians.* Berkeley, Calif.: University of California (unpublished dissertation).

(Received 9 July 1980)