

## THE IMPACT OF COMPUTER NETWORKING ON THE SOCIAL SCIENCE DATA LIBRARY

Alice Robbin

University of Wisconsin - Madison

ABSTRACT

This paper explores the factors which have constrained the social science data library's participation in the use of computer networks as a vehicle for accessing information. It also suggests why changes in the situation can be expected and further suggests some of the ways that computer network resource sharing will affect the social science data library structure and services.

INTRODUCTION

Traditional libraries are experienced in meeting information needs and in coordinating activities among users. Computer technology has made it possible for the traditional library to service the user more quickly and (many enthusiasts of on-line data bases would add) more complicated than when the reference librarian relied on manual methods for searching and retrieving information. Computer technology has made it possible for traditional libraries to avoid some costly duplication of human and technical resources.

While the traditional library is a relative novice to computer technology the social science data service organization (data library or data archive) has been linked to the computer and modern technological developments by the very nature of the medium of its collection, machine readable data. The social science data library, while tied to advanced technology, has continued to use traditional means for locating, transferring and accessing machine readable data files (MRDF) and for communicating its needs and coordinating activities related to MRDF. Moreover, the social science data library has neither benefited from the set of experiences of the traditional libraries in resource sharing within a networking environment nor utilized network computers to share resources and expertise and to cooperate for more efficient allocation of resources.

CONSTRAINTS IN THE USE OF  
COMPUTER NETWORKING  
BY THE SOCIAL  
SCIENCE DATA LIBRARY

For several years now, as a growing number of articles, monographs and books attests, computer networking has become an "important mode for remotely gaining access to

data and computations, sharing resources, and providing information services" (EDUCOM, 1973). At the same time, there continue to be a number of factors which constrain the development of network resource sharing and information servicing. These factors include:

1. technical considerations, which revolve around processor configuration, software and communications (Davis, 1972);
2. financial considerations, which involve the nature, size and distribution of monetary support;
3. organizational and political considerations, which include the structuring of the network, provision and nature of services, monitoring of performance, source and distribution of authority and responsibility (Davis, 1972), degree of control at the network level and local level, and integration of the local effort into national network efforts (EDUCOM, 1973);
4. legal considerations, which involve federal and state legislative restrictions, which at the federal level prohibit monopolies and restraint of trade (Clayton Act of 1914) and protect the use of communications services as a public utility (Federal Communications Act of 1934, and Neumann, 1973), and which at the state level are designed to protect the outflow of state dollars for the buying of non-state services; and,
5. user considerations, which include knowledge of user needs and

characteristics, ease of system access, use and operation, a variety of services to assist in efficient and productive operations, education and training, and documentation.

Even with these constraints, the use of networking by libraries, researchers and students, particularly in the natural sciences, has increased during the last several years. This has not been the case for most professional social scientists, who have had little or no experience in network use, or for the social science data library, whose clientele are social scientists.

What accounts for the low level of use of computer networks and why has there been no network resource sharing by the social science data library? The reasons are structural (the result of political realities and historical accidents), economic, sociological, and experiential. Structural factors include the lack of data services, of institutionalization of the information service at the local level, and of professionalization.

Efforts at coordination and resource sharing have been made at national and international levels, but only minimal efforts have been made to encourage the development of infrastructures at the local level. Major archives have maintained their dominance and have, contrary to public expressions of support, done little to encourage the development of the local data center. But, networking depends on the creation of local "nodes" and without the local effort, national networking will not be successful.

Most organizations which provide data services are structurally weak, existing as appendages to one unit of a larger parent organization, rather than as an independent unit within the parent organization. With only tenuous funding support, the staff must dedicate its efforts to maintaining services with a continually eroding funding base. The staff therefore has few or no incentives to develop or employ networks to communicate with information services outside its local environment. In general, these data services are small and operate on the periphery of computing activities of their parent organization (or are perceived by others to be little involved in computing activities) and therefore have little or no control over how computing resources are allocated within the parent organization.

These small centers remain invisible to policy makers and thus when decisions are made about computer use and about activities which will involve interaction with centers outside the home institutions, these centers are never informed.

One of the results of the lack of institutionalization is the perception of these services as non-legitimate and its staff as unprofessional. The staff members are not viewed as professionals by the user community which employs their services, nor do the staff members perceive themselves as professional data specialists, although many data center personnel are indeed experts in data processing and handling. Although a staff member may in fact be performing the work of a reference librarian or information specialist who searches and retrieves selected information upon request, as indeed most staff at local data centers do, that staff member usually does not recognize the role he performs--that is, cannot assign a name to the function being performed. He usually lacks a methodology for the tasks he is performing. Professional training provides tools and products (resources), an explanation for the activities performed by certain individuals, and a Methodology for task performance. But, in most cases, the data center staff member has not been trained as a professional.

Why should the lack of professionalization affect the use of computer networking and resource sharing? Reference work implies knowledge of and understanding of the nature and potential of available information resources. If people are unaware of resources, they cannot utilize them. The information specialist today is made aware of on-line data resources at the introductory course level and receives training in informational (bibliographic) data base creation and manipulation. The housekeeping and maintenance functions performed by libraries are facilitated by a networking environment. In other words, our data services personnel are uninformed of the potential use of networking facilities because they have not been trained as professional information specialists.

Economic reasons also explain why the social science data library has made little use of computer networking. In general, there has been enough money for the duplication of multiple copies of data and the duplication of very large data bases. There has been enough money to deal with incompatible data

structures, so that special purpose software could be written, and for inefficient use of computing resources. Staff time has been an inexpensive commodity, so that both the user and staff (and administrator) have had little incentive to revise documentation, provide better program support and better user assistance. Delays, as a result of an extended process of locating data, processing them and locating or creating software with certain analytic capabilities have been acceptable aspects of the data analysis process. In other words, the society has had enough money to support waste: duplication, incomparability, and delays.

The society has made an enormous commitment (investment) in computer technology for direct applications purposes, but not for information servicing purposes. This is probably because it is far easier to develop measurement tools to judge a tangible byproduct than to judge a process, which information servicing is. Products can be offered to the funding agent and potential user. But, information services, which rarely offer a "product" and whose "product" is rarely visible a short time after the service has been supplied, have been unable to generate sufficient funds to support its intended goal, the supply of particular data required to solve particular problems, required at a particular moment in time, selected from a very large set of information, most of which is irrelevant to the problem. It is very difficult to justify intangibles--they must be taken on faith alone. It is difficult to demonstrate that supplying information will reduce the total social cost of research, for example. Although networking presents a useful technique for communicating information efficiently, the resources to provide communications are not available.

MRDF collections range in size from several hundred observations to the population of the U.S., from a few to literally thousands of variables. Although technological problems have not proven to be the primary constraints in networking use, the social scientist does not realistically have the possibility of transmitting the information to his home site. Transmission costs are prohibitive. Utilizing data at the data's home site requires computer funds which may be unavailable to that researcher because his university prohibits the use of computing dollar expenditures outside the campus. When data are physically transferred from one organization to another, costs which

are incurred are fairly easy to calculate (staff and computer time for processing the request, documentation, parcelling and postage, and overhead). Thus, the buyer of data is billed for the capitalization investment incurred by the data producer, and the seller appears to have no problem justifying the cost of a data file when it is distributed in this fashion. But, if this physical transfer is not made, how do we protect the data producer's capitalization investment?

Theoretically, at least on the face of it, it appears quite easy to write an accounting algorithm for the computer billing system which adjusts costs of accessing selected files by the status of a user, such that the local user of the data would not be billed for accessing the data file, but the remote user would. A variety of problems need to be resolved. There are accounting ones which require reprogramming of the current accounting system of most computers. There are philosophical issues, about who should pay for the data and how much, which become more evident when networking and remote access are involved.

There are a variety of sociological reasons which explain why the social sciences data library user community has not made use of the computer networking and resource sharing. In general, the library's user community is not prepared to use electronic means to access information and communicate with each other. Computer networking still requires a moderate familiarity with computers. Computer use, while becoming more common in the social sciences, is usually limited to one or at most two courses a semester in each discipline, and use is largely for a class project and an exercise in data handling, not information or data management. Although the professional social scientist may from time to time be frustrated in his inability to locate MRDF sources, in general he relies on his colleagues and the "invisible college" for information on sources of data and the data themselves. This action perpetuates the lack of support for a data library, since it perpetuates the fact that the best data are not archived (but if one knows the right people, one gets access to the information) and reside in private hands, and presumes that data library staff can provide little assistance in information seeking for the social scientist.

Resource sharing implies that resources are known, and in this sense, emphasizes the impact of the invisible college, since it makes individuals less dependent on whom they know and more dependent on knowing information resources which are machine based and require the assistance of intermediaries who are specialists in information retrieval. Some years ago, at an EDUCOM conference, Richard Hofferbert (1973, p. 149), former director of the Inter-University Consortium for Political Research, and himself a political scientist, summarized social scientists' needs with regard to MRDF: He said they have three needs: (1) information about data, (2) access to the data themselves, and (3) available analytic capabilities. How could networking meet these social science needs? The discussion group of social scientists was confused by the concept of networking, and Hofferbert concluded that social scientists were not ready for networking. Thus, it has been perhaps the sociological aspects of the user community which provide the best explanation for the low-level of networking use: lack of experience and the lack of understanding about what networking is and why the social scientist could benefit from its use.

### CHANGES IN ORIENTATION

For data libraries to make use of computer networks and to engage in network resource sharing, there must be infrastructures for organizing information services and for creating and utilizing information products. To share resources, products describing these resources must be created. There must be standards for their production in order to retrieve the information and to enhance the product's utility. Although the structural, economic, sociological and experiential factors described in Section 1 will continue to constrain network use, recent developments suggest some optimism is warranted for future social science data library networking activities. These developments stem from the recognition that information about MRDF require organization and management. Thus, we are seeing (1) development of an infrastructure of social science data library services, (2) development of standards for describing and controlling the information, (3) recognition that professionals are necessary to manage the information, (4) recognition that costs of production and transfer of the information warrant

new methods of organizing and managing access to information, and (5) understanding of various technological advances which the social scientist can utilize to access information and transfer and retrieve data more efficiently and effectively.

Acceptance among researchers that secondary analysis of data is essential to realizing the full potential of expensive data collection has been increasing. The growing cost of creating complex data files to meet a variety of needs is resulting in a new philosophy that these data represent a national resource, publicly available, to be widely disseminated in order to realize their full potential. Inter-disciplinary research, involving a variety of data, is becoming of increasing importance for public policy planning and making and for sustained analysis of social, political and economic change. Complex data files produced by large-scale data gathering activities have induced a need for specialists who can organize, manage and document the information. Rapid reductions in the cost of remotely accessing these complex files is making it unnecessary to transfer to a local site for analysis purposes. More sophisticated analytic techniques are being required in order to realize the potential of these complex files, and the social scientist is readjusting his perspective on requiring that these techniques be locally available. The researcher and analyst are finding it necessary to support the creation of structures which will organize a collection of machine readable data, assist the researcher in locating data for his needs, and provide the analyst with sufficient documentation to make retrieval of statistics possible and to eliminate error in the process.

Failures in information retrieval have engendered a growing recognition that collections of data files need to be organized in the manner of other collections of information. There is a growing recognition that the enormous quantity of information is making it very difficult, indeed sometimes impossible, to locate and select relevant bits of information and to evaluate the quality of the information collected and the collection process itself, and that the society must develop rational strategies for collecting and retrieving information which is a byproduct of society's recordkeeping and evaluation of itself.

This has led to a reevaluation of the importance of libraries and information services and the growing need for individuals who have expertise in information selection and retrieval, organization, management, dissemination, and documentation. The reference function is becoming recognized as a critical activity for the supply of information to the user community. The individual who serves as a reference librarian for MRDF will in the near future be viewed as a professional.

Support for the establishment of data libraries is the result of recognition of the cost of social research: Data libraries represent savings in scarce resources with their potential for collecting in one location studies which are of value to a variety of individuals; providing a systematic description of these studies to facilitate locating and utilizing data resources; reducing duplication of purchases; providing centralized expertise in file creation, processing, and description; and, providing a basis for a data services infrastructure to facilitate access to data at a reduced cost through memberships in consortia and through exchanges, enhance communications about data resources, and lay the foundation for the development of data information products to benefit the user community.

Part of the failure in information transfer and retrieval is due to the lack of standards for documentation. In general, MRDF documentation has been poor and bibliographic control undeveloped. Sue Dodd, of the Social Science Data Library at the University of North Carolina, has commented extensively on this through the the Classification Action Group of IASSIST (1977 a, b, c) as have others involved in the development of documentation standards (Nielsen, 1977; Mochmann, 1977; and Robbin, 1977). The social science data librarian has often been unable to locate files because no title or producer statement was provided. Analysts do not know how to acknowledge the use of secondary sources of data in their publications. Study description providing brief histories of a data file have never contained adequate descriptions of the data.

In the past year, however, we have seen growing support for documentation standards. It is significant that support for documentation standards comes not only from the social scientist who does not have easy or regular access to data or is not linked to a major information network, but also from the

experienced data handler who has never paid much attention to the quality of documentation and for whom the invisible college has operated effectively to make it possible to obtain the data he needed. This support comes at an appropriate time: During the last several years, data information specialists have been working on guidelines for documenting MRDF, ranging from various types of bibliographic descriptions, products such as catalog records, indexes and classification schemes and standardized study descriptions, to file and variable level descriptions. Software, such as the Interchange file, developed by Roistacher and Noble (1976), will obviate the necessity of rewriting codebooks formatted for different statistical software packages. Thus, there is support in the user community for the data professional's concern about documentation and an apparent willingness to accept the recommendations of these specialists so that better descriptions of MRDF will be available.

These developments encourage an atmosphere in which computer networking can be accepted as a viable and desirable means to (1) access information processing services outside the local environment, (2) share resources which will provide economy of scale or operation to a number of participants of the social research process, and (3) share intellectual resources and cooperate in joint programs (Roistacher, n.d.). Networks will provide a mechanism for more effective communication, cooperation and coordination of information through services such as a social science data library.

#### THE IMPACT OF NETWORKING ON THE SOCIAL SCIENCE DATA LIBRARY

The social science data library is a special purpose library or information service which has been created to respond most directly and immediately to its special clientele, a diverse community of users composed of social science researchers and students, policy planners and analysts. It is located in a variety of settings, within government, commercial organizations, foundations, and academic institutions. The data library may be part of a computer center, a larger library (facility), a (general) information center, a data archive (in the European sense), a social science department, a research institute,

market research organization, or may be an independent service organization. The organization in which the library is imbedded is (in most cases) the library's major funding source, except in those instances when part of its activities may be supported by external funding arrangements for the creation of special products (which may or may not be a byproduct of the library's regular activities) or the provision of special services (which may be the result of staff expertise).

The collection reflects as closely as possible the clientele's activities and is organized to provide (efficient) access to its contents and retrieval of selected elements within the collection. The library staff's substantive and technical training provides a clientele, who have differing levels of substantive, methodological and technical skills, with appropriate information on the contents of the library's collection and technical assistance in utilizing the information. The staff has a comprehensive knowledge of resource material described in its reference collection and potentially available elsewhere, the transfer of the materials from one location to another, the preparation of data for analysis, and the relationship of the data to the computer and management and analysis software required for the data in the collection.

The data library engages in a number of activities related to the medium of its collection and its "mission." A data library may be a publisher and producer of information in machine readable form and/or may provide information services much like that of the traditional library. The data library may be responsible for study design development, data preparation and collection, data processing and data management, data analysis, data archiving (or accessioning) and records management, data dissemination, and data reference services. In general, however, the data library which responds as a special purpose library is characterized by three functions: It is responsible for locating information on behalf of its clientele, accessioning materials it acquires, and disseminating these materials upon request.

The three basic functions of reference, accessioning and dissemination require access to different types of information. The reference function requires access to information about the existence, availability and logical structure

of MRDF. The accessioning function requires access to information about the history of the data, physical structure, relationship of the data to management and analysis software and computer hardware, file storage, processing and maintenance, quality of the logical structure and documentation, and relationship of the logical structure to the physical structure. The dissemination function requires access to information about users' needs for efficient retrieval of relevant data from a larger collection of information.

In other words, the data library is involved in information seeking and management processes on a regular basis. The data library acts as an intermediary between information and the user, to retrieve and organize information to meet specialized present or potential needs for data resources which will be used to generate statistical products. To develop the data library's collection, the staff gathers appropriate information on potentially useful data resources. It classifies the gathered information to facilitate its retrieval upon request. The library records information on acquired materials which it then integrates into the collection. The library produces a variety of information products describing the collection, which represent access tools for information about data resources and for the data resources themselves. Successful information gathering and management requires a communications network among organizations which perform a similar function and can supply each other with information, organizations which produce information products, and staff trained in classifying and retrieving the information.

What then is the effect of computer networking and resource sharing on the social science data library's structure, services, collection, information seeking and management processes in which it engages, and relationship to other organizations? Networking has the potential for affecting the relationship of the data library to other information and computation services within the parent organization.

Networking appears to indicate a trend toward centralization of services. The experiences of the 1960's have shown that centralization of services was not cost effective and has not provided more effective and better user services; however, in the last two years, we have seen a move to centralize organizations which perform similar

services (e.g., the Carter Administration's efforts to reduce "inefficiencies" in government by reorganizing its bureaucracy). The central computing center appears to be making an effort to exert control, both directly and indirectly, over other organizations which perform functions related to computing. After several years of reassessing its role in the parent organization, the computing center appears to be moving toward efforts to centralize computing services and computation activities within the parent organization. It appears that the computing center may be successful in its efforts because of its size (and budget) and the mythology of expertise which the computing center perpetuates. In an era of continuous inflation, there are increasing pressures on the administrators of the parent organization to maintain existing facilities which service the largest number of users and require the largest budgets.

The social science data library, specializing in assistance to social scientist who have traditionally not been big users of computing services, is not able to generate large-scale support. Thus, I would predict that as networking becomes an integral part of the activities of a computing facility or large information service, data libraries which do not have strong, independent constituencies within the parent organization, will be subtly and not so subtly pressured by the parent organization either to be absorbed by a computing center or to have its functions provided by another organization which is responsible for information services. It will be hard to counter the trend toward centralization of services, especially because the computing center provides experience and know how which are strong arguments for extending services to a wider market, one which is covered by a social science facility like the data library. But, the trend should be resisted because it will mean that special needs of a specialized user community will most likely go unmet.

In principle, networking creates the potential for independence for computing centers and other facilities providing computer-related information services. Networking frees the data library to operate in a free market situation: The data library can look for a seller which provides the best products and services at the lowest rate. Networking, however, poses a threat to a computing center, an organization which has operated like a

monopoly, because it affects its utilization and revenue. Thus, the computing center will make every effort to control the outflow of computing dollars elsewhere. Rather than entering into a free market environment and upgrading the quality of its services, the computing center retrenches and begins to exert pressure on the parent organization's policy makers to stem the flow of money elsewhere. While it may be impossible to prevent the user community from taking its money elsewhere, the computing center may convince administrators to make it difficult or at least inconvenient through a variety of bureaucratic measures to buy services outside. I would predict that this is a short term response and in the long run it will be desirable to develop competitive capabilities which will favor the user by helping to reduce prices for comparable services and to improve service quality and service availability (Neumann, 1973, p. 28). Thus, I think that in the long run networking will free the data library from overwhelming dependence on its local computing center and at the same time provide the data library with better services from its local computing facility. Protectionism has never worked to the long term benefit of the protected--as economic history has shown in the last century.

Networking will also affect the relationship between staff and user. Most data libraries have worked on a one-to-one, face-to-face basis with their users. The staff has provided extensive and time consuming user support services, training, tutorial information, documentation and human consultation, because its user community has preferred to use the system on an "as needed basis" rather than to spend time preparing for future system use (Neumann, 1973a, p. 3) and because different methods of user assistance have been necessary to meet different levels of expertise. When networking is introduced as a means of acquiring new information and statistical products, the library will have to extend its resources to provide services for users unfamiliar with other operating systems. These services will probably be carried out without additional funding support. Networking will mean the development of and greater use of automated interactive tools for user support services. Data library staff will probably become more involved in computers as a result, and in the development of automated and interactive support capabilities (and ironically, the expertise required for their

development will propel the data library toward the computing center which has more experience in systems development).

Networking will make the user community even more aware of the necessity for good documentation for data, software and operating systems. While documentation has been primarily hard copy, networking will probably increase the trend toward automating instruction, updating through the terminal, and development of systems directories and catalogs of services and products.

Probably the greatest opportunity that networking presents for the data library is in the area of retrieval of information about the existence and availability of MRDF. As documentation becomes automated, networking will make it possible to search for information contained in "on-line data bases" of directories of data holdings, contents of data files, and codebooks and other documentation for data files located at institutions far away from the local data library.

Data libraries will have an opportunity to participate directly in cooperative ventures to create data bases containing information on the contents of data files which are located at local data centers. There will of course be non-trivial administrative and financial problems to be resolved. But, networking presents an unprecedented opportunity to create resources of utility to a wider user community than now served by small local data centers and to enhance the quantity and quality of information now available.

As networking becomes a more accepted activity in the generation of statistical products, we can expect to see an increasing amount of Statistical analysis done remotely, although it is doubtful that there will ever be more remote than local access and analysis of data. Remote access to data files will affect the data library: its collection, how much time it allocates to the accessioning process, and how it (and the user community) pay(s) for data which are physically not in the data library's collection.

At present, a data library's collection grows by acquisition of a copy of data archived and maintained by other data libraries and repositories. It has never really been feasible for the data library to engage in an interlibrary loan type of data exchange because the library has had to bear the cost of

duplicating a copy of a data file each time there has been a request from outside its local environment. (This is in contrast to a book loan, where one copy of a book circulates and there is no need to duplicate a copy of the book each time a request is made for its use.) Other reasons which explain why data have not been distributed on an interlibrary loan basis include the extensive capitalization investment that a data library makes in developing its collection, difficulties in the physical transfer of data and repeated use of the medium on which the data are stored (typically magnetic tape); need to prepare a data file in a physical structure compatible with the host environment to avoid additional processing; and, far more human time required to prepare data for an external environment. Furthermore, it is not yet economically feasible to transmit data remotely from one site to another because transmission speeds are too slow for the quantity of data typically analyzed by the social scientist, and because transmission costs are still too high.

Obtaining a copy of a data file has been the only way that a data library increases its collection (and thereby justifies the number of personnel required to maintain the collection, since quantity is always a more tangible measure of service than quality). It has always been assumed that when a data file is needed that it must be acquired, even when, as in most cases, what the user community does is prepare a statistical overview of the population in the data file and prepare some inexpensive, preliminary statistical results. (This is called, "getting a feel for the data" and most researchers begin their projects in just this way.) In many cases, the data which are acquired by the data library on behalf of a user and are reviewed in this manner, are rejected as not meeting the user's needs; and the researcher never completes a detailed analysis of the file.

For every data file which is acquired, scarce resources of time and money must be allocated to integrate the file into the collection. This accessioning process is expensive because the data must be checked to verify that what is described as its physical structure actually is and that the descriptive materials (documentation) accompanying the data allow the user to understand the logical structure of the data and to carry out statistical analysis of the file, and

because duplicate copies must be made to protect the data for future use.

Thus, because of the need to duplicate copies of a data file, costs associated with capitalization investment and with moving data from one location to another, and the infeasibility of transmitting data remotely from one site to another, data libraries have been discouraged from utilizing resource sharing methods used by more traditional libraries through regional cooperative networks. There is extensive investment by the data library in acquiring and maintaining a data file which will be accessed at most a few times.

In a networking environment data would be acquired only after preliminary analysis of the data had taken place of when acquisition and accessioning could be justified because extended and detailed analysis was intended. Cost for remote access of the data could be transferred to the user. A fee would be attached to accessing the data and for any consultation that the data distributor would provide. A fee could be paid either for first time access of the file by an individual user or for each time the data file is accessed. If the remote user were to pay a one time fee which covered the cost of accessing the data and consulting assistance, the distributor could calculate access costs to all its files and apply charges across all data files available for network remote access.

Remote access via networking is one way of reducing the cost of data transfer from one library to another: by eliminating the physical problems of data transfer which inevitably occur because people are unfamiliar with producing copies of data to be accessed within a different computing environment; reducing staff and computer time involved in processing the data request; reducing time lags which inevitably occur due to processing individual copies of a data file; eliminating the burden of data review which takes place during the accessioning process; and, reducing the data library's allocation of resources for accessioning information which may be very infrequently accessed in the future.

The data library has justified its information gathering, acquisition of new materials, and retrieval and dissemination of selected information on the basis of the information's potential use for a variety of individuals in its user

community, although only a very small part of its collection is utilized on a regular basis. Networking makes it possible to shift the cost of acquiring information from the information service to the individual who required the specialized information and to justify a reassessment of the allocation of resources for information management tasks which the data library performs on a regular basis. If fewer resources are allocated to records management (which accessioning basically is) of the collection, resources can be reallocated to gathering information, retrieving selected information upon request, and to developing automated user support capabilities which create information products which could potentially benefit a wide variety of individuals in diverse user communities, while continuing to provide specialized services oriented to the social scientist.

Successful information gathering and management requires a set of relationships to communicate information among organizations. Computer networking is an excellent technological development which will assist the data library in communicating its and its user community's needs efficiently and rapidly. It should be easier to search for information by placing a message in the network system which will reach everyone at the same time. Request and response time should therefore be shortened.

I think that networking will force a reassessment of the relationship between the data distributor and local service data library, not just in the techniques for "costing out" data and passing on overhead and direct costs for data production and processing to the library (and thus to the user), but in the social relationships of the two organizations which respond to individual users quite differently and have quite different responsibilities to their different user communities. It may well be that data libraries will begin to respond more like individuals; that is, information will be acquired on a need-only basis. This means that services may be paid for only when an operation is performed, rather than in advance with a "blanket order" for anticipated future services. On the other hand, it would seem that more effective and rational planning can only be accomplished if resource sharing is done on the basis that "everything will get evened out in the end"; that is, some organizations will benefit more than others, but over time,

all organizations will benefit about the same amount. Information sharing would argue for its cost being borne by the largest number of potential users possible. Certainly, computer networking offers this possibility.

### SUMMARY

This paper presents a cursory review of the factors which have constrained the use of computer networking and resource sharing by the social science data library and its user community. These factors are structural (the result of political realities and historical accidents), economic, sociological and experiential. These factors will continue to constrain network use, but recent developments suggest that we will see greater use of networking by the data library and social scientists. There is growing acceptance of secondary data, growing costs of creating complex data files, need for specialists who can organize, manage and document these data, rapid reductions in the cost of remotely accessing these complex files, recognition that collections of data files need to be organized and well documented, reevaluation of the importance of libraries and information services, and support for the establishment of data libraries. These developments would seem to encourage an atmosphere in which networking would be accepted and seen as a viable and desirable means to access information processing services outside the local

environment, share intellectual resources, share resources to provide economy of operation, and provide a mechanism for more effective communications, cooperation and coordination of information.

We should not expect to see networking affect radical changes in the structure, services, staffing, collection and relationships with other data centers. But, we can expect that networking will produce a push toward centralization of computer-related services within the parent organization in which the social science data library is located, in the future a freer market in which to buy computer services, more user support services which are computer-based, greater demands for library staff expertise about computing by the user community, better documentation of information materials, interactive information products which describe the contents of MRDF, remote access for preliminary statistical analysis of data, the shifting of resource support from the library to the user, fees for services rendered rather than for services anticipated (as memberships in consortia are), and a reevaluation of the relationship of data distributor-archive to the local data library.

The conclusion is that in general computer network and resource sharing should provide better services, although the social science data library may be integrated into a larger information services organization.

Roistacher, Richard C. The Data  
Interchange File: A First  
Report. CAC Document No. 207.  
Urbana, IL: The University of  
Illinois, 21 June 1976.

REFERENCES

Davis, Ruth M. Practicalities of  
Network Use. Networks for  
Higher Education. (Princeton,  
NJ: EDUCOM, The Interuniversity  
Communications Council, Inc.,  
1972).

Dodd, Sue. Report of the Joint  
United States-Canadian Action  
Groups on Classification.  
IASSIST Newsletter. 1,2 (1977),  
9-11.

----- Report of the Joint  
Canadian-United States Action  
Groups on Classification.  
IASSIST Newsletter. 1,3 (1977),  
8-10.

----- The Emerging Priority in  
Bringing Bibliographic Control  
to Social Science Machine  
Readable Data Files (MRDF).  
IASSIST Newsletter. 1,4 (1977),  
11-13.

EDUCOM. Planning for National  
Networking. Proceedings of the  
EDUCOM Spring Conference, 1973.  
(Princeton, NJ: EDUCOM, The  
Interuniversity Communications  
Council, Inc., 1973).

Hofferbert, Richard. Networks and  
Disciplines. Common Themes and  
Consensus: Report and  
Discussion of Workshops.  
(Princeton, NJ: EDUCOM, The  
Interuniversity Communications  
Council, Inc., 1973).

Mochmann, Ekkehard. Information  
Access at the Data Item Level.  
SIGSOC bulletin. 6 (2,3).

Neumann, A. J. Review of Network  
Management Problems and Issues.  
NBS Technical Note 795.  
(Washington, DC: U. S.  
Department of Commerce, October  
1973), 22-23.

----- Network User Information  
Support. U. S. Department of  
Commerce, December 1973.

Nielsen, Per. Information Access  
at the Data File Level:  
Documentation Prerequisites on  
the File-Level Data Base Inquiry  
Process. SIGSOC Bulletin. 6  
(2,3).

Robbin, Alice. Managing  
Information Access through  
Documentation of the Data Base.  
SIGSOC Bulletin. 6 (2,3).

----- and Noble, Barbara B.  
Computer Network Support of  
Social Research Communities.  
Unpublished paper. Urbana, IL:  
University of Illinois, n.d.