The Data and Program Library Service: A Case Study in Organizing Special Libraries for Computer-Readable Statistical Data

ALICE ROBBIN

Organizational Structure and Activities

Historical Background

SEVENTEEN YEARS AGO, the importance of a library service for computer-readable statistical data for the social science community was formally recognized at the University of Wisconsin-Madison.[1] It is worth quoting directly from the document about this data library facility because the statement makes explicit the theoretical and functional bases which constitute what became known in September 1966 as the Social Science Data and Program Library Service (DPLS).

> Every successful science involves, in endless interaction, the following activities: hypothesis formulation, model building, data gathering for purposes of testing and estimation, testing, estimation, and prediction. All of these activities pose significant difficulties, but in general the most expensive to carry out is...gathering the data needed for...testing and estimation. In the physical and biological sciences the large accelerators, radio and optical telescopes, electron microscopes, instrumented rockets, atomic piles, high altitude balloons, and innumerable other expensive devices are all devices to collect desired data. In the social sciences collection of data is inadequately financed and frequently far too expensive to be engaged in by individual research workers or even by individual research centers or institutions. Except in certain experimental areas, the individual research worker must depend upon data gathered by others, including governmental units.

Alice Robbin is head, Data and Program Library Service, University of Wisconsin-Madison, and president of IASSIST.

> The result of this situation is that the individual social scientist who wants to engage in research must usually spend the major part of his time in discovering where data of interest are located, in obtaining permission to use desired data, in raising substantial funds to extract desired data from files, in transcribing the data, in coding the data, and finally in preparing magnetic tapes for use in a large computer. In general the Ph.D. candidate does not have the resources or the time to engage in the above process. Typically he must restrict his attention to bodies of data that his major professor has on hand or can readily obtain. His major professor, for his part, must typically make his plans [for using existing data] from one to three years in advance of when he can hope to have them...in usable form. Even then, after extensive effort and expense he may well be completely frustrated. It is an understatement to say that one of the major stumbling blocks to achievement of greater success by social scientists is the sheer inaccessibility within reasonable time limits of major bodies of data already in existence.
>
> The Computer Programming, Data Processing, and Data Library Facilities contained in the proposed Social Science Research Complex are designed, among other things, to [bring] together and [store] in fully indexed and...highly accessible form on magnetic tapes the major bodies of data that should be brought to bear on the study of man.[2]

The Data and Computation Center (DACC),[3] established in the mid-1960s, is the result of an articulated demand[4] for assistance in social science research problems related to large-scale data collection and computation. Within an instructional and research environment, DACC integrates services that are usually found in libraries, computer installations, survey research facilities, and instructional support services. A unit of DACC, the Data and Program Library Service (DPLS), was created because faculty members of the social sciences, especially in economics, political science and sociology, became convinced that the university needed a facility for managing the increasing quantity of available machine-readable social science data being produced on the campus and elsewhere. They recognized the importance of preserving data, collected often at considerable cost, which had significant subsequent value for other researchers and students. In addition, they believed that computer analysis and data management programs should be stored, documented and disseminated in conjunction with the data archive. In 1966, with assistance from the Graduate School and the Social Systems Research Institute, DPLS was established.

DPLS was designated as the local campus repository for quantitative social science machine-readable data. Its major functions concerning data were defined as acquisition, storage, maintenance, and

dissemination of data files from the social sciences, as these files became available from individual researchers, other local archives, national social science data repositories, and profit and nonprofit agencies. DPLS was also mandated to acquire studies created by faculty, students and researchers on this campus. Researchers and institutions elsewhere sometimes designated DPLS to maintain and distribute their data. Successive deans of the Graduate School (and later of the College of Letters and Science) and members of the DACC Faculty Policy Committee believed that a service to archive and distribute data should also be available to individuals outside the Madison campus, within the constraints imposed by DPLS's principal mandate as a local campus library service.

During the 1960s and early 1970s, the social science faculty felt that the most natural setting for DPLS was the traditional campus library.[5] Efforts were made either to incorporate DPLS as a special library within the university library system or to incorporate it within the university library. During this period, there was some enthusiasm by the university library, but it was evident that the library was not capable of absorbing DPLS, a computerized information facility, within its structure due to the library staff's inexperience with computers and data. There was also evidence that the library was unprepared to accept computerized data as a legitimate information resource, and was also unprepared to accept nonlibrarians (that is, people not holding a degree from a library school) who would be responsible for technical and public services. Nevertheless, the social science bibliographer served for some years on the DACC/DPLS Faculty Policy Committee, although he was completely frustrated in his attempts to create enthusiasm for DPLS as a legitimate library. By the mid-1970s, special libraries proliferated on the Madison campus, operating funds for the university library were significantly reduced (making it impossible to integrate DPLS into the library), and the social science community had become so used to easy access to machine-readable statistical data and information services that they had little interest in pursuing a formal organizational arrangement with the university library. Since the mid-1970s, DPLS's link with the professional library community has extended to membership in the Madison Campus Special Library Association (composed largely of special libraries which are not part of the university library system), to a Library School faculty appointment to the DACC Faculty Policy committee, and to a three-credit, graduate-level course, "The Management of Machine Readable Numeric Data for the Social Sciences," jointly offered by the Department of Economics and Library School.

The Collection

Beginning in 1966 with only a few data files contributed by the economics and sociology departments, DPLS's collection has grown to more than 2000 unique titles (and more than 8000 data files). The contents of the collection are varied and reflect the many disciplinary interests of its user community, including enumerative and vital statistics, communications and mass media; economic systems, structures, attitudes, and behavior; environmental and natural resources, international systems, legal systems; political and governmental systems, structures, attitudes, and behavior; social systems, structures, attitudes, and behavior; surveys conducted by commercial polling agencies containing many discrete topics; and reference materials in machine-readable form, some of which act as indexes to the contents and observations contained in other machine-readable data files (MRDF). (See appendix B for the index to the *Directory* of the data and program holdings at DPLS, which provides a detailed description of the types of data in the collection.) The size of data files (number of observations) varies from a few observations to the population of the United States. The number of variables contained in the files ranges from several hundred to well other three thousand. The time period ranges from the twelfth century to 1980.

The DPLS collection is particularly strong in the areas of demography and family planning, policy, employment, and historical and current census materials, which reflect the historical importance of major research and policy contributions by the university's departments of history, sociology and economics. Somewhat more than one-third of its collection has been obtained through the university's membership in the Inter-University Consortium for Political and Social Research (ICPSR), an international research repository for quantitative data housed at the University of Michigan in Ann Arbor, to which Wisconsin has belonged since the Consortium's inception in 1962. Somewhat less than one-third has been obtained from other research archives in the United States and Western Europe and U.S. federal agencies, and about one-third has come from the University of Wisconsin's research community and researchers at other institutions. DPLS has, by default, become a repository for various federally produced data which have not been preserved by their respective agencies or the National Archives. With the Center for Demography and Ecology and the Institute for Research on Poverty, DPLS has reformatted and disseminates the *Current Population Surveys*, 1963-1980, in order to make these data more tractable for social science research and policy activities. All the

data for which DPLS serves as an archive of record are available free of charge to University of Wisconsin-Madison students and faculty and on a direct cost basis to individuals outside the Madison campus. For data files deemed of significant research value, DPLS has created public use files and fully documented the data.

As a library service, DPLS acquires data upon request, upon recommendation of faculty members who demonstrate the potential utility of a data file to a diverse set of researchers and students, and as part of collection building in the areas in which DPLS is particularly strong (i.e., for which there is a significant long-term institutionalized research commitment at the university). DPLS either supports fully or cost-shares with researchers and departments the purchase of data, obtains data free of charge or without direct charges (e.g., through its membership in ICPSR or exchanges with other archives), or receives data through a researcher's or department's donation.

Technical and Public Services

DPLS devotes a considerable effort to acquiring, accessioning, describing, preserving, and disseminating data. Determining the existence and availability of data is a time-consuming activity for which one staff person is primarily responsible for reading the published and unpublished literature of the social sciences, information profession, other archives, federal agencies, and the like. Because industry standards for the transfer of data are not utilized by all data producers and disseminators of data, obtaining a data file which can be immediately processed or analyzed can be a time-consuming activity. DPLS has invested some time in developing forms for accurate description of the physical structure of a data file on magnetic tape and in assisting researchers and students to obtain data without DPLS's intervention in the acquisition process.

Once the data arrive at DPLS, they are copied on magnetic tape(s); examined to determine whether their contents are fully described in the accompanying descriptive documentation on their physical and logical structure and contents; and a permanent historical record is created of their acquisition, evaluation of their quality, physical structure, medium of preservation, and physical location (on which magnetic tape the data are stored and in what storage facility). This process of accessioning is the most time-consuming technical activity that the DPLS staff performs on a daily basis. We have estimated that it may take as much as twenty to forty hours to evaluate a data file, particularly if, during the evaluation, errors are located in the original writing of the

data, the data do not correspond to their description in the documentation, delays are encountered locally with the computer, or if we must obtain answers to our questions from the data producer. Although the quality of data and documentation have improved in recent years, there are many problems with data. DPLS has committed itself to providing this technical service because few researchers and students have had sufficient experience in understanding data. To the extent possible, DPLS tries to reduce the difficulties in accessing and retrieving the data in its collection. This preliminary evaluation is one way of doing so. But the activity is very time-consuming and requires expertise not only in the substantive discipline of the data producer but also in data management and data processing.

After accessioning is completed and before the data file is publicly available, DPLS creates a catalog entry and an abstract summary describing the contents of the data file (see appendix C). Although librarians now have access to the *Anglo-American Cataloging Rules*, 2d edition,[6] the problem is not so much with the actual cataloging rules, but with the lack of bibliographic control over MRDF. In only a few cases have title pages been created with enough information to create a catalog entry (see appendix C for an example of a title page), and documentation has been so poor as to make it difficult, if not impossible, to augment the title page with the required elements for an entry. DPLS is sometimes forced to return to the data producer to obtain information about authors, producers, edition, date of production, and so forth. Writing an abstract, which contains a bibliographic citation, summary statement of the methodology employed to create the data, summary statement of the scope and contents, descriptors, technical information on the structure of the data, information on the file's availability, and relevant publications based on analysis of the data file, is similarly difficult because documentation is inadequate. Writing this abstract does take time because the abstractor must become familiar with the data. Yet, both the catalog entry and the data abstract, not including the title entries in the *Directory*,[7] are requirements for facilitated access because they are the first indications of the existence of an MRDF. Thus, DPLS devotes some amount of time to bibiographic access tool development.

In recent years, DPLS has been devoting more time to the problems of maintenance and preservation of data because its collection is rapidly aging. All its data are stored on magnetic tape, which is a fragile medium and must be regularly monitored to assure its reliability.[8] Between June and December 1979, DPLS carried out a complete

inventory of the magnetic tapes in its collection. Since January 1980, the staff has been involved in a long-term project for implementing a data preservation and tape maintenance program, which involves the complete conversion of all data files dating from a storage date of 1960 through 1979, examination of the quality of existing magnetic tape in the collection, purchase of new tape and its quality control, analysis of the present recordkeeping system, standardization of both bibliographic and technical descriptions of all data files in the collection, and development of a procedures manual for a data preservation program.[9]

All the experiences that DPLS has built up through its technical services have been useful for its archival program, the second component of its mandate. We have applied our experiences to assist the State Historical Society of Wisconsin Archives Division in determining whether data can be scheduled for retention, with maintenance and preservation to assist the State Archives in determining the long-term costs of administering an archival program for machine-readable public records, with documentation and bibliographic control to assist the State Archives in inventorying and describing the holdings of the state agencies. These same experiences have been applied to individual research projects carried out by the teaching and research faculty and students at the university and elsewhere, as part of the public services we provide. We help individuals locate, obtain, understand data and documentation, and plan research projects involving the gathering, coding, processing, and description of data. We assist project staffs in organizing a program for managing their own collections of data. We teach library management of machine-readable data to students in the Library School and the Archives Administration Program. DPLS disseminates data to students and faculty at the University of Wisconsin and throughout the world, as well as to public and private organizations. Perhaps most importantly, DPLS teaches students how to use the library service, so that novice users of the facility can become more independent of the library staff and then teach their colleagues and others how to enrich their learning and research programs through the use of statistical data to solve social, political and economic problems facing the society.

It is indeed a rich program that the DPLS support facility carries out. But it must not be forgotten that the quality of its technical and public services are quite dependent on the rich knowledge resources available to its staff, which are provided by the University of Wisconsin-Madison setting. The source of DPLS's strength and productivity lies in the university's long tradition of quantitative research and of providing appropriate and adequately funded support facilities for research.

Critical Issues Facing DPLS: Now and the Next Five Years

In the previous section on technical and public services, I alluded to some of the problems that DPLS faces. The quantity of machine-readable data potentially of interest to the social science community is enormous, thus creating difficulties in determining their existence and availability. Reference services are an important part of any library, yet the lack of bibliographic access tools, standards and control make it difficult and time consuming to ascertain the existence of machine-readable data. Lack of quality control over data products, either in the design, collection or processing stages, makes it difficult for the user to access and retrieve data in an easy and efficient manner. User documentation is not adequately prepared during the data collection and processing stages, again impeding easy access to data products. The medium of storage is fragile, and although advances are being made in new storage technology, magnetic tape will remain the principal medium of storage for some years, thus requiring the data library to institute a regular (and perhaps expensive) program of tape maintenance and data preservation. Lastly, the requirements of computer technology, data structures and substantive knowledge of social science call for special expertise in the social sciences, library management, data management, and data processing, thus demanding that personnel have special skills not ordinarily obtained in the formal requisites for a degree in either library science, computer science or a social science subject area. All these problems have been discussed elsewhere in the literature[10] and will not be elaborated here. Rather, I will comment in this section on the implications of the DPLS organizational arrangements for long-term stability and support-building within the university, and management of information in a systematic and efficient manner.

Implications of Organizational Independence

There have been definite benefits to maintaining DPLS as an independent department, organizationally unrelated to any teaching department. But at the same time, its dependence has meant a tenuous existence, one highly dependent on a minority of scholars (those who do quantitative research), on the availability of sufficient funds for all departments within the college, and on the Wisconsin tradition of adequate research support facilities which have been built principally with extramural funding.

Organizational independence has meant:

1. a collection reflecting the many substantive discipline interests of the multidisciplinary social science community, rather than the parochial interests of one department or subdiscipline. This has made good sense, considering that most MRDF contain data items which are of potential use to a variety of people.[11] The result has been a very rich collection and, decreasingly, the need to obtain data which are useful only to a small subset of social scientists at the University of Wisconsin.
2. a library situated in a multidisciplinary social science research environment which provides access to a wide array of social scientists with substantive and methodological expertise in an enhanced environment for learning about diverse research applications of social science data. The result has been an enlarging of the vision of social research and an opportunity for enlightenment by the DPLS staff. It has meant that users from one department, usually not knowing about research interests and activities in another department, can be referred to appropriate individuals who are doing work in the individual's area of interest. Because some of the work of the social science faculty has been on the cutting edge of social research and policy, the collection building effort has been greatly facilitated for the DPLS library staff.
3. politically, the ability (and need) to be responsive to a variety of needs articulated by the many social science disciplines on the campus. Although this has sometimes meant a somewhat precarious balancing act, the result has been to educate the staff to be responsive to needs of a wide variety rather that a vocal minority of users. The result has also been a continuing staff effort to develop and foster a broad base of support for the data library during times when funds were becoming unavailable to all teaching departments. All departments have thus had a stake in DPLS's existence. Because not one department alone has the resources to maintain DPLS, all departments could agree that designating a small percentage of their (potential) funding support to DPLS would not endanger the quality of their teaching and research programs.
4. the possibility to create more easily informal organizational and daily working arrangements with other departments and support facilities, such as the university library, other campus special libraries, the computing center, Library School, State Archives, and Survey Research Laboratory. The results have been to give the staff far more access to a wide variety of expertise than is typically avail-

able to a special library staff and to increase the funding base for data acquisitions. Access to expertise bears directly on the quality of services that the staff provides, particularly for reference and referral services and for answering highly technical questions about which the staff has little knowledge. Because few individuals or single departments have the funds to acquire expensive data collections, DPLS has been able to acquire these collections by pooling the funds from its own budget and from individual departments. Its direct effect has been to enrich the data collection and stretch the funds available to individuals and single departments.

5. the ability to establish DPLS's unique identity, unrelated to any particular substantive research area, department, or faculty member(s). DPLS's identity has provided faculty and administrators with an opportunity to demonstrate that the University of Wisconsin supports unique facilities for social research that can be matched only by a few other institutions (and ones which have more disposable funds for social research than the University of Wisconsin). DPLS's unique identity has nurtured among the staff the desire to assist other universities to develop a similar support facility, to teach the library and information science community how to organize and manage a special library for quantitative data, and to participate in national and international activities where this expertise and experience can be shared to improve access to social science computer-readable data.

On the other hand, this organizational independence creates problems in a university structure where support facilities are dependent on available funds to teaching departments, where a minority of the teaching department's members are quantitative researchers, and where the support facility is almost wholly dependent on state (rather than extramural) funding to carry out its program. In an era of increasingly diminishing available dollars for universities, due principally to a deteriorating federal and state economy and to a somewhat unfavorable political climate for universities, support facilities which survive by the grace and wisdom of the deans are an easy target for administrators when there are not enough funds to support the principal mandate of the university, which is teaching. When budget cuts affect all teaching departments, organizations like DPLS come to be considered luxury items. It is difficult to build support for a facility like DPLS when teaching departments are asked to handle more students with fewer faculty and teaching assistants and with a 4 percent decrease in supplies and capital equipment budget items (and an inflation rate which effec-

tively decreases real dollars to 1969-70 levels). When the university library cannot buy books for an entire fiscal year, can the university support data purchases which may average between $350 and $6000? When the university cannot hire assistant professors who are paid $15,000, can the university hire highly skilled library and data management specialists whose starting salaries range between $18,000 and $25,000? When teaching departmentss have no books because the library book budget has been eliminated, can the university justify paying between $1000 and $10,000 for a new piece of computer hardware?

It is generally agreed that modern social science research and teaching depend on data and computer experience, and that students need to be exposed to computers and data during their undergraduate and graduate careers because society has become so dependent on computers and on a well-educated populace. However, in a period of declining funds to operate the university, organizations like DPLS have come under increasing scrutiny. Thus, the next five years of DPLS's existence will be critical ones for survival. Although the arguments for maintaining its organizational independence seem indisputable, the economic environment is sufficiently unfriendly as to raise questions about DPLS's continuing autonomy in a university structure in which it "fits between the cracks." Whether DPLS will continue to operate as it has during the last fifteen years is now open to question. There are hopes that during 1981-82, this issue will be resolved.

An Automated Information Management System

During the fifteen years that DPLS has operated, manual library and administrative recordkeeping systems have been implemented to organize information about and access to the core collection. Computerization of selected technical and public service procedures has not been considered until recently for a variety of reasons, including lack of adequate software, peripheral computer and data processing equipment, programming support, and a reasonably stable staff with a low level of staff turnover which reduced the need for well-documented procedures and records. However, changes have occurred during the last several years. DPLS has begun to experience demands for improved access to the contents of the collection, retrieval of selected parts of the MRDF and reference collections, and reference services for MRDF located elsewhere. A review of the present recordkeeping system has indicated incomplete records documenting the history of each data file and its changes. The nature of the collection has changed, reflecting the more complex needs of the social research community at the university,

and the growth of the collection continues unabated. Increased emphasis on teaching has led to more requests for user-oriented, instructional data packages and better-documented data. Offline data entry equipment has reduced the cost of computing at the university and small computers are available at minimal or no cost to the department. At the same time, increased budgetary constraints on the College of Letters and Science have reduced available resources to DPLS at a point when manual tasks have become increasingly labor-intensive and time-consuming, but a larger staff is unavailable to carry out public and technical services designed to improve access to the collection. Future maintenance and development of the archives and library clearly depend on improving public (user) and technical services to avoid degrading the quality of current services, and on providing more cost-effective mechanisms to access the collection.[12]

On the basis of a preliminary systems analysis of DPLS, the staff believes that efficient use of available resources requires an integrated systems approach to archives/library and records management at DPLS. In order to cope with the size of the present collection, maintenance and dissemination activities, and recordkeeping requirements in a no-growth and inflationary economy, maintenance and improvement of current services and development of new services depend on automating the various activities/operations of the library and integrating these operations in a systematic way. The user community will become the principal beneficiary. Efficient recordkeeping systems will reduce redundancy of information and operations, improve user access and retrieval of information about data and to the data themselves, provide better file security and current information about a file's status, provide comprehensive statistical analysis of library operations for administrative and research purposes, improve non-University of Wisconsin access to the DPLS collection, decrease the extensive reliance by users on the DPLS staff for information related to file contents and structure, and release staff for more productive reference work, collection building, and preservation.

Optimum use of computing and data processing resources available to DPLS would permit a systems design which fulfills four requirements:

1. immediate response to user requests and file updates by staff in an online mode;
2. report generation and statistical analysis in an online, but also batch mode;

3. selective dissemination of information (SDI) capabilities to supply users with information directly relevant to their subject area (as related to secondary analysis of machine-readable data, and generally applicable to other archives and library administrative records management projects); and
4. access to the collection from users outside the University of Wisconsin environment, both in a time-sharing and batch mode. (See appendix D for an overview of the data-base management system.)

Developing a data-base management system to automate the recordkeeping activities for the data collection requires resources which are currently unavailable to DPLS. These resources include a systems programmer to carry out a detailed analysis of the present structure of the manual recordkeeping system and to develop the automated management system; someone knowledgeable about abstracting and information storage and retrieval technology to develop a bibliographic data base of descriptions of the MRDF in our collection; sufficient funds to acquire and modify or write software for the interfaces to the existing data-base management systems available at the University of Wisconsin; and data entry personnel to carry out the arduous task of entering all the information about the collection into the data bases. How much progress is made over the next few years will depend on support from researchers and library staffs in other departments, who could also make use of this information management system for their own data collection projects.

The Next Five Years: Where are Data Libraries Going?

Current Funding Strategies

For the last two decades, attention has been focused at the national level and funding priorities have dictated allocation of resources to national centers as principal sources of valuable archival data to serve a wide array of scholarly research and teaching activities. National funding priorities have promoted uses of data repositories which are too centralized, too structured and too hierarchical. This policy risks paralysis of the larger system, denies the pluralistic nature of information needs and services, and demonstrates ignorance of the intellectual and social processes of information exchange.

Local data libraries are important contributors in a pluralistic information system, and their importance for preserving, disseminating and describing statistical, textual and other types of data should not be underestimated. Although these local, campus-based libraries and

archives have not been integrated into the larger information system, they play a critical role in the knowledge-flow process and information transfer system for the social scientific community within and outside their institutions.[13] Their efforts in the areas of disseminating data and information about data, and of maintaining valuable archival data resources need to be fostered. Even as ever-accelerating technological developments in telecommunications, information processing and transmission, and computers make it possible to receive and transmit data from great distances (thereby obviating in principle the necessity for being a local repository), current economic and political realities and intellectual or problem-solving modes of behavior have acted to constrain efficient use of the modern computer technology. These realities suggest that distributed data libraries in one form or another, whether independent of or integrated in the traditional library, will continue to play a major role in the transmission of information and will continue to be critical to the process of intellectual inquiry.[14]

Without denying the importance of the national data repositories whose historical and future role and contributions to social science research and teaching are indisputable, we must at the same time develop new strategies for making data more easily accessible to individuals and institutions which do not have the infrastructure to support large-scale data services for teaching and research. Hierarchical and centralized data repositories do not promote infrastructure development at institutions which are not part of the small communications and information network in which these data repositories operate. Public policy must foster development of small-scale data services within an institutional setting where it is not economically feasible to develop large-scale computational facilities or to rely upon highly trained support specialists in data and computation. The long-term knowledge benefits seem clear. Although current economic conditions may not permit reversals in funding priorities, other factors will play a role in fostering the development of small data libraries in the public and private sectors. These factors are the subject of the concluding remarks.

Factors which Will Influence the Development of Data Libraries

In recent decades, there has been a notable expansion of information recorded by government and the private sector. Increased statistical data collection and utilization of these records have been possible because of the technological and intellectual advances in computer technology and the development of survey methodology.[15] Much of this recorded information is not in readily usable form, but a great deal has

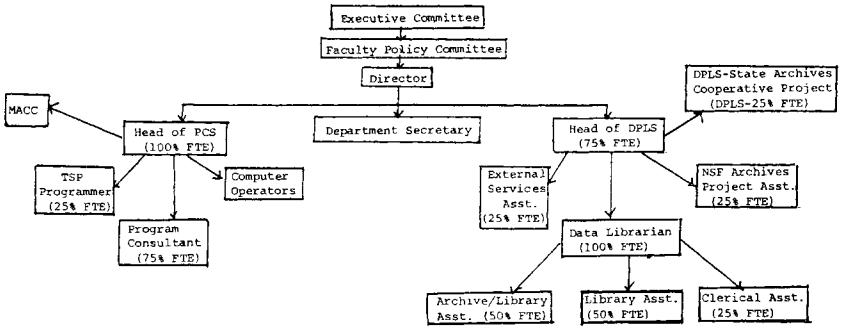
utility for teaching and secondary analysis. With significant improvements in the quality of data and development of standards for data documentation, we can expect that data will be more easily accessible to students and researchers. Although data structures are becoming more complex, data management software is becoming more "user friendly" and is making it easier to manipulate the large-scale data collections which are being produced. Statistical software is being written so that novice users can manipulate data more easily. Data management software is being developed to facilitate creating instructional data files from the large and complex data bases. Software interfaces are being developed to make access to computer operating systems less dependent on the expertise of computer specialists. Although we will see increasing quantities of statistical and other types of data produced by a myriad of individuals and organizations, international standardization of descriptions, information data bases, and computer technology and telecommunicaitons will make it easier to locate and obtain information about machine-readable data.

Computer hardware and peripheral devices are rapidly decreasing in cost and are changing the nature of people's relationship to computational devices. New organizational structures are being developed to take advantage of this new technology, and these structures are reorganizing our professional lives and relationships with other professionals and our clients. The technological and intellectual advances will surely facilitate the organizing and managing of data and allow small organizations to provide inexpensive services for their users.

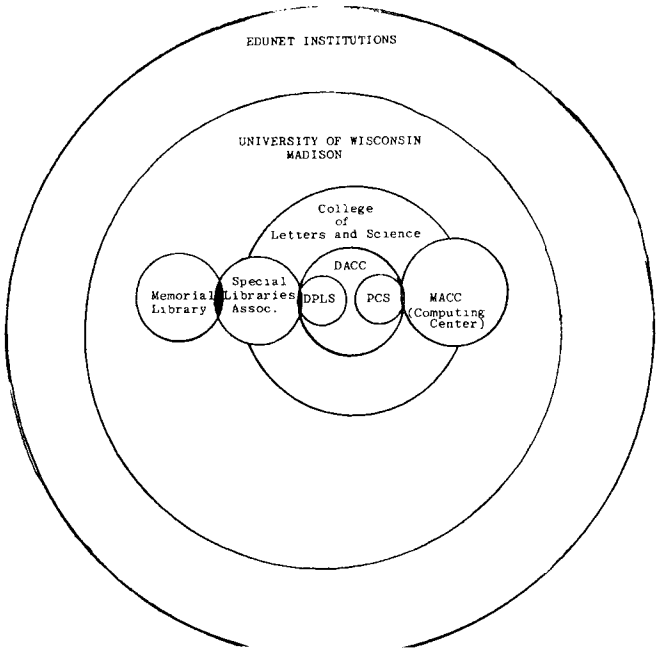
Recognizing and accepting these developments and acquiring the expertise to deal with these new resources will encourage the library community to improve the training of specialists who supply information to people, to use existing information structures such as traditional archives and libraries to preserve and disseminate data, and to develop new retrieval tools to improve access to information about data. After all, the rationale for supporting infrastructure development, tools for information and data retrieval, and changes in the curriculum of the library and information professions is based on the principle that a society has a commitment to information as a national resource. This commitment has always been fundamental to the principles of librarianship.

Appendix A

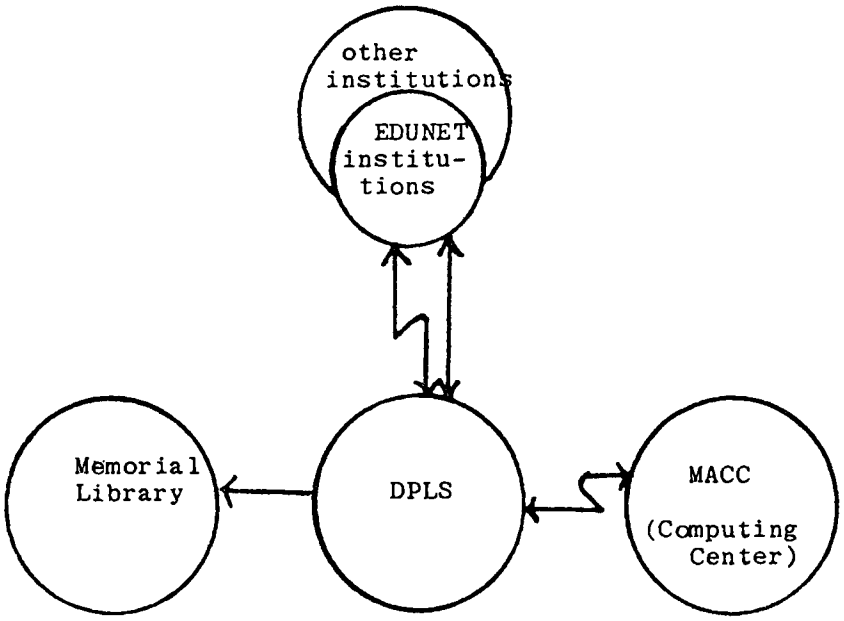
System Perspective and Organizational Chart of the Data and Program Library Service



Organizational Chart of the Data and Computation Center, September 1980



System Perspective of the Data and Program Library Service



Manual and Electronic Communications Links of the System

Appendix B

Classification Scheme and Table of Contents from the *Directory* describing the DPLS Data Collection

I. ENUMERATIVE STATISTICS
- AA Nation
- AB State or Province
- AC County
- AD Legislative District
- AE Standard Metropolitan Statistical Area
- AF City or Other Local Unit
- AG Household
- AH Individual
- AJ Vital Statistics

II. COMMUNICATIONS AND MASS MEDIA
- BA Communications and Mass Media

III. ECONOMIC SYSTEMS, STRUCTURES AND BEHAVIOR
- CA Economic Attitudes and Behavior
- CB Economic Processes and Indicators

IV. ENVIRONMENT AND NATURAL RESOURCES
- DA Environment and Natural Resources

V. INSTRUCTIONAL DATA SETS AND COMPUTER PROGRAMS
- EA Instructional Data Sets
- EB Computer Programs

VI. INTERNATIONAL SYSTEMS
- FA Dyadic and small Group Interaction
- FB Organizations
- FC Structural Characteristics of the International system
- FD Alliances and Military Affairs
- FE Conflict, Violence and Wars

VII. LEGAL SYSTEMS
- GA Legal Systems

VIII. POLITICAL AND GOVERNMENTAL SYSTEMS, STRUCTURES AND BEHAVIOR

Elites and Leadership
HA Elites and leadership

Governmental Structures, Policies and Capabilities
JA Behavior and Attitudes of Bureaucrats
JB Historical and Contemporary Public Policy Indicators
JC Statistics on Government Operations

Mass Political Behavior and Attitudes
Historical and Contemporary Electoral Processes
KA Primaries, Conventions and Candidate Evaluations
KB Election Studies
KC Election Returns
KO Merged Electoral and Ecological Data
Public Opinion on Political Matters and Political Participation
LA Public Opinion
Political Parties
LH Political Parties

Legislative and Deliberative Bodies
MA Roll Call voting Records
MB Decision making in Deliberative Bodies
MC Apportionment

IX. SOCIAL SYSTEMS, STRUCTURES AND BEHAVIOR

Community and Urban Studies
QA Citizen Attitudes Towards the Local Community
QB Community Structure

Education
QD Education Attitudes and Behavior
QE Education processes and Indicators

Medical and Health
QG Medical and Health

Organizational Behavior
QH Organizational Behavior

Religion
QJ Religion

Socialization, Students and Youth
QK Socialization, Students and Youth

Aging
QM Aging

Family
QN Family and Child in Society
QP Family Planning and Fertility

Social Indicators
- SA Attitudes Toward Self and Society
- SB Social and Occupational Mobility
- SC Social Processes and Indicators

Conflict and Aggression
- SJ Anomic Behavior
- SK Attitudes Toward Violence
- SL Domestic Conflict Indicators
- SM Minorities and Race Relations

Transportation
- TA Transportation

X. UNCLASSIFIED SURVEYS CONDUCTED BY COMMERCIAL POLLING AGENCIES
- YA Surveys Conducted by Commercial Polling Agencies

XI. REFERENCE MATERIALS IN MACHINE READABLE FORM
- ZA Reference Materials

Appendix C

Bibliographic Access Tools Prepared by DPLS

AMERICAN FAMILY GROWTH, 1957-1967
A User's Guide to the Machine Readable Data File

Principal Investigators

Charles F. Westoff
Robert G. Potter, Jr.
Philip C. Sagi
Elliot G. Mishler

Office of Population Research
Princeton University
Princeton, New Jersey

Edition Produced By

Mary Ann Hanson
under the direction of
Larry Bumpass
Center for Demography and Ecology
University of Wisconsin-Madison
Madison, Wisconsin
1978

DPLS ed. 1979

Distributed By
Data and Program Library Service
4452 Social Science Building
University of Wisconsin-Madison
Madison, Wisconsin

Data and Program Library Service University of Wisconsin-Madison
1979

Suggested Bibliographic Citation

American Family Growth, 1957-1967 [machine readable data file]. Principal investigators, Charles F. Westoff et. al. DPLS ed. Edition prepared by Mary Ann Hanson, under the direction of Larry Bumpass, Center for Demography and Ecology, University of Wisconsin-Madison. Madison, WI: University of Wisconsin Center for Demography and Ecology [producer], 1978. Madison, WI: University of Wisconsin Data and Program Library Service [distributor]. 1 data file (1,165 logical records), plus accompanying documentation. This edition was produced under grants PD-238416-7 to Larry Bumpass and HD 05876 to the Center for Demography and Ecology.

American family growth, 1957-1967 [machine readable data file] / principal investigators, Charles F. Westoff et. al. -- DPLS ed. / edition prepared by Mary Ann Hanson, under the direction of Larry Bumpass, Center for Demography and Ecology, University of Wisconsin-Madison. -- Madison, Wis. : University of Wisconsin Center for Demography and Ecology [producer], 1978 ; Madison, Wis. : University of Wisconsin Data and Program Library Service [distributor].

1 data file (1,165 logical records) + accompanying documentation.

The public use version of this study was reprocessed, edited, and multiple punches removed.

SUMMARY: This study examines the fertility history of American couples in the United States and the motivational connections between the environment and fertility decisions and behavior. Data describe fertility-planning, status, family composition, socioeconomic status, residential history, religiosity, level of status satisfaction of the husband, achievement of life goals of the wife, commitment to work (husbands), and a wide variety of background information, social and psychological attitudes, and behavior.

This study is also known as The princeton fertility study.

ISBN 0-89605-052-1

1. Westoff, Charles F. II. Bumpass, Larry.

ABSTRACT

Unique identification number(s): Accession number QP-003-004-USA-1957.

Citation: American Family Growth, 1957-1967. [Machine-readable data file]. Principal investigators, Charles F. Westoff et. al. DPLS ed. Edition prepared by Mary Ann Hanson, under the direction of Larry Bumpass, Center for Demography and Ecology, University of Wisconsin-Madison. Madison, Wis.: University of Wisconsin Center for Demography and Ecology [producer], 1978. Madison, Wis.: University of Wisconsin Data and Program Library Service [distributor].

Methodology. The target population was urban, native-born white couples with two children, couples whose marriages so far had been uncomplicated by death, divorce, separation, or extensive pregnancy wastage, with the second birth to have occurred during September 1956 for every couple. A probability sample, stratified by metropolitan area, was drawn from 7 SMSAs with population over 2 million (exclusive of Boston). Couples were interviewed three times in February-March 1957, 1960, and between 1963 and 1967 to determine eligibility and to complete questionnaires. Data checks and full-scale processing were run on the public use version. The final sample size is 1,165 couples; 814 couples completed all three interviews.

Summary of contents. American Family Growth, 1957-1967 is a longitudinal study which examines the fertility history of American couples in metropolitan America and the motivational connections between the environment and fertility decisions and behavior. Phase I looks at the social and psychological factors thought to relate to differences in fertility. Phase II focuses on why some couples stopped at two children while others had a third or fourth child during the first and second phase. Phase III examines how well attitudes and events of the early marriage determined the record of the later years of childbearing. The data file contains over 1000 variables.

Geographic coverage. United States SMSAs (New York, Indianapolis, Chicago, Los Angeles, Milwaukee, Cleveland, Minneapolis)

Descriptors. Fertility, family planning, family composition, socioeconomic status, work satisfaction, contraceptive practices, religiosity.

Technical notes. Rectangular file with 1,165 observations

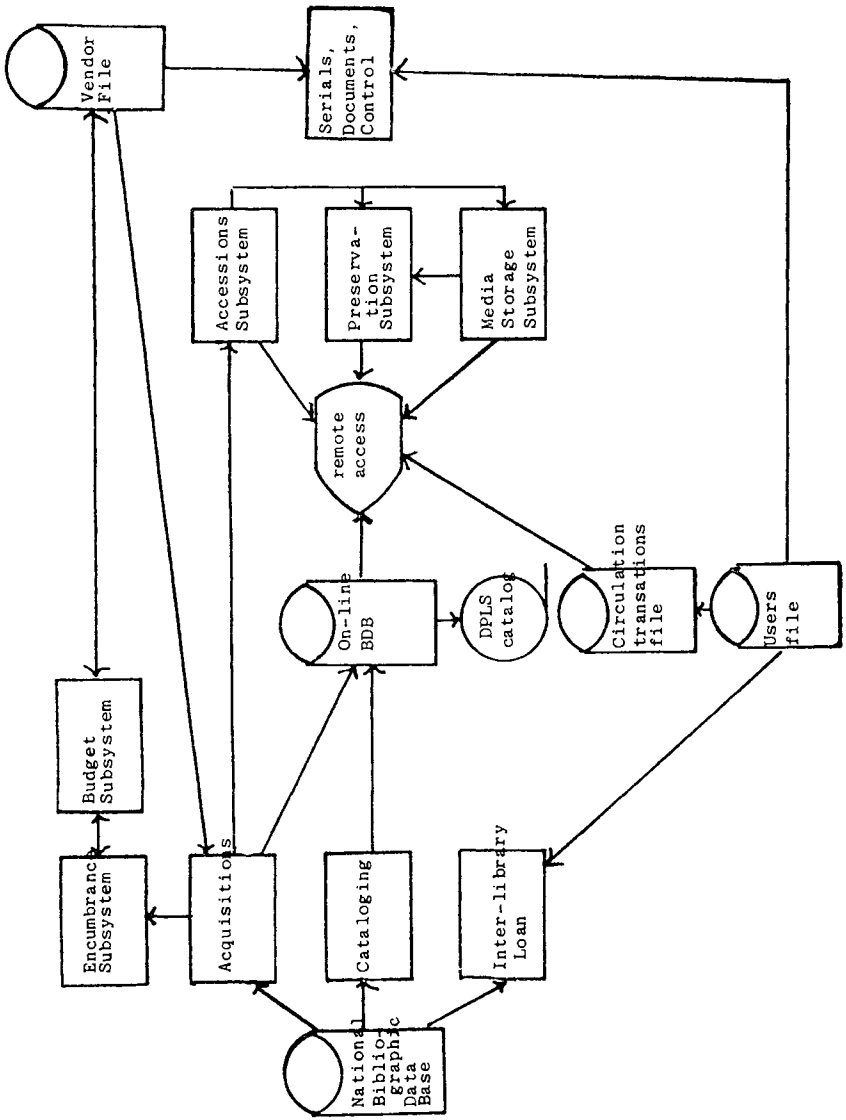
Terms of availability. Data checks and full scale processing have been performed on the public use file. There

are no restrictions on access to the public use file. Copies of the data and documentation can be obtained by writing to the Data and Program Library Service, 4452 Social Science Building, University of Wisconsin-Madison, Madison, Wisconsin 53706 USA; telephone number: (608) 262-7962.

Cited references. Principal monographs include Family Growth in Metropolitan America by Charles F. Westoff, Robert G. Potter, Jr., Philip C. Sagi, and Elliot G. Mishler (Princeton, NJ: Princeton University Press, 1961); The Third Child: A Study in the Prediction of Fertility by Charles F. Westoff, Robert G. Potter, Jr., and Philip C. Sagi (Princeton, NJ: Princeton University Press, 1963); and The Later Years of Childbearing by Larry Bumpass and Charles F. Westoff (Princeton, NJ: Princeton University Press, 1970).

Appendix D

Overview of an Automated Information Management System for the Data Library



References

1. University of Wisconsin-Madison. *A Proposal for a Social Science Research Complex*. Madison: University of Wisconsin, 1963. (Application for National Science Foundation Graduate-Level Research Facilities Development Grant.)

2. Ibid., pp. H-3, 4.

3. From 1966 until June 1974, the Data and Computation Center (DACC) was composed of two divisions: the Programming and Computation Service (PACS), a contract programming and software development service for the social sciences, supported by the College of Letters and Science; and the Data and Program Library Service (DPLS), supported by the Graduate School and the Social Systems Research Institute within the Department of Economics. In March 1974, the dean of the College of Letters and Science (L&S) recommended PACS be dissolved and functions originally performed by PACS be transferred to other organizations. The Madison Academic Computing Center (MACC) assumed management of the remote job-entry terminal and offers contract programming support. Software development and duplication services were reinstated via MACC and the Photo and Duplicating Services, respectively. At the same time, the deans of L&S and the Graduate School recommended that DPLS be reorganized. L&S funds originally allocated to PACS for software development, maintenance and operation of the unit record data processing and interactive terminal equipment, and the miscellaneous activities performed by PACS personnel were transferred to DPLS. A formal liaison with MACC was established to permit increased social science input into MACC policy-making and better communication and coordination of data processing activities among social scientists within the university system, and to administer the batch remote entry terminal. In June 1974, funding support for DPLS was transferred to L&S. In December 1976, DPLS was restructured so that the program consulting activities for which it had had responsibility since 1966 were transferred to a Program Consulting Service (PCS) within DACC. Since then, PCS provides consulting on software, generates some special-purpose software for file handling, is responsible for acquiring and maintaining equipment and administering the computer site in the Social Science Building, and with DPLS, provides support for and disseminates computer software originally developed by PACS. DPLS's responsibilities are solely library and archive management of computer-readable statistical data. DACC is an independent department within L&S. MACC provides half-time support for the head of PCS, who also serves as the DACC remote job entry site coordinator. L&S provides continuing support for PCS and DPLS, supplemented by extramural funds to DPLS for special research projects. (See appendix A for the DACC and DPLS organizational chart, Sept. 1980.)

4. *See The SSRI Data Library—Its Functions and Role within the SSRI Computation Division: Discussion Paper #1*. Madison: University of Wisconsin, n.d.; *SSRI Data Library: Tentatively Proposed Scope of Functions and Services, A Discussion Document*. Madison: University of Wisconsin, 1965; Kelley, Allen C. *The Data Library and the Electronic Data Processing Requirements of the Social Scientist* (Data Library Paper 6501). Madison: University of Wisconsin, 1965; _____. *SSRI Survey of Data Holdings and Requirements of Selected Wisconsin Social Scientists* (Data Library and Computation Paper 6503). Madison: University of Wisconsin, 1965; and Day, Richard H., and Kelley, Allen C. *Proposed Development of Data Library and Computation Services for Social Science Research* (Data Library and Computation Paper 6505). Madison: Social Systems Research Institute, University of Wisconsin, 1966.

5. Adams, Margaret O., and Dennis, Jack. "Creating Local Social Science Data Archives." *Social Science Information* 9(April 1970):51-60. *See also* Lucci, York, et al. *A Library Center of Survey Research Data; A Report of an Inquiry and a Proposal*. New York: Columbia University, 1957; and Adams, Margaret O., et al., eds. *The Proceedings of the Workshop on the Management of a Data and Program Library*. Madison: University of Wisconsin, 1970.

6. *See* American Library Association. *Anglo-American Cataloging Rules*, 2d ed. Chicago: ALA, 1978, pp. 201-16.

7. University of Wisconsin Data and Program Library Service. *Directory of the Machine-Readable Data and Program Holdings of the Data and Program Library Service*, 2d ed. Madison: University of Wisconsin, 1980.

8. *See* Robbin, Alice. "A Profile of Data Preservation Activities in University Data Libraries and Archives, January 1980." *IASSIST Newsletter*, vol. 5, no. 1, pp. 3-22.

9. This activity has been supported in part by special grants from the College of Letters and Science and the Graduate School.

10. For discussions about the problem of bibliographic access tools, *see* Dodd, Sue A. "Titles: The Emerging Priority in Bringing Bibliographic Control to Social Science Machine-Readable Data Files (MRDF)." *IASSIST Newsletter* 1(Fall 1977):11-18; _____. "Bibliographic References for Numeric Social Science Data Files: Suggested Guidelines." *Journal of the ASIS* 30(March 1979):77-82; _____. "Building a Bibliographic/MARC Data Base for Social Science Data Files in a Network Environment." *IASSIST Newsletter* 2(Spring 1978):34-37; _____. "Building an On-Line Bibliographic/MARC Resource Data Base for Machine-Readable Data Files." *Journal of Library Automation* 12(March 1979):6-21; _____. *Cataloging Machine-Readable Data Files*. Chapel Hill: University of North Carolina, forthcoming; and Peters, Paul E. "Notes on the Distribution of Labor in a Social Sciences Data Information Network." *IASSIST Newsletter* 2(Summer 1978):69-76. For problems of data quality, documentation and needs for standards and control, *see* Robbin, Alice. "Managing Information Access through Documentation of the Data Base." *SIGSOC Bulletin* 6(Fall/Winter 1974-75):56-68; _____. "Understanding the Machine-Readable Numeric Record." *The Midwestern Archivist* 4(1979):5-24; _____. "Technical Guidelines for Preparing and Documenting Data for Secondary Analysis." In *Reanalyzing Program Evaluations*, edited by Robert F. Boruch, et al. San Francisco: Jossey-Bass, 1980; and Roistacher, Richard C., et al. *A Style Manual for Machine Readable Data Files and Their Documentation*. Washington, D.C.: USGPO, 1980. For problems on training and recruitment of data library personnel, *see* Robbin, Alice. "Toward Creating the Professional Data Librarian." *IASSIST Newsletter* 2(Fall 1978):95-100; and _____. *Strategies for Improving Utilization of Computerized Statistical Data by the Social Scientific Community*. Madison: University of Wisconsin, 1980. For problems on data perservation, *see* Dollar, Charles M. "Problems of Magnetic Recording in Archival Storage." *Digest of Papers* (COMPCON 1977, 14th IEEEC Society International Conference, San Francisco) pp. 28-30; and Robbin, "Profile of Data Preservation Activities."

11. Heim, Kathleen M. "The Social Science Data Archive: A User Study." Ph.D. diss., University of Wisconsin, 1980. (Reports on the multidisciplinary use of statistical data files from the DPLS collection.)

12. Robbin, Alice. "Enhancing User Services through Optimum Use of Computing and Data Processing Resources: A Progress Report on an Integrated Systems Approach to Library and Records Management at the Data and Program Library Service" (paper prepared for the May 1979 IASSIST Conference, Ottawa, Canada, May 1979). *See also* David, Martin. "Access to Data: The Frustration and Utopia of the Researcher." *Review of Public Data Use*, 8(June 1981):13-26.

13. For example, decentralizing maintenance activities ensures broad-scale efforts to preserve data and reduces the costly burden and responsibility of maintenance for those few archival institutions with a specific mandate to preserve machine-readable data. Although there are economies of scale that come with centralization of services, typically—in contrast to a small organization—centralized institutions allocate a greater portion of their budgets to administrative operations than to technical services. In the long run, this allocation becomes more costly for the total information system. In contrast, if each small data library keeps its own house in order, there are significant savings for all members of an information system. *See* Robbin, Alice. "By the Seat of My Pants." *IASSIST Newsletter* 4(Spring 1980):23-24.

14. For some excellent discussions on the benefits of data archives, *see* Boruch, Robert F., and Wortman, Paul M. "An Illustrative Project on Secondary Analysis. In *Secondary Analysis*, edited by Robert F. Boruch, pp. 89-110. San Francisco: Jossey-Bass, 1978; Hofferbert, Richard I., and Clubb, Jerome M. "Introduction." *American Behavioral Scientist* 19(March/April 1976):409-18; Nesvold, Betty A. "Instructional Applications of Data Archive Resources." *American Behavioral Scientist* 19(March/April 1976):455-66; and Rokkan, Stein. "Data Services in Western Europe." *American Behavioral Scientist* 19(March/April 1976):443-54.

15. David, "Access to Data"; and Rokkan, "Data Services in Western Europe."