PRODUCTION NOTE

University of Illinois at Urbana-Champaign Library
Library Trends

Library Trends, a quarterly journal of librarianship, provides a medium for evaluative recapitulation of current thought and practice, searching for those ideas and procedures which hold the greatest potentialities for the future.

Each issue is concerned with one aspect of librarianship. Each is planned with the assistance of an invited advisory editor. All articles are by invitation. Suggestions for future issues are welcomed and should be sent to the Managing Editor.

Published four times a year, in summer, fall, winter, and spring. Office of publication: University of Illinois Graduate School of Library and Information Science, 249 Armory Bldg., 505 E. Armory St., Champaign, IL 61820. Entered as second-class matter under the act of August 24, 1912. Copyright 1982 by The Board of Trustees of The University of Illinois. All rights reserved; nonprofit organizations may, however, quote from or reproduce material copyrighted here by The Board of Trustees of The University of Illinois for noncommercial, educational purposes. Full credit should be given to both the author and Library Trends.

Subscription price is $16.00 a year (plus $1.00 postage for overseas subscribers). Individual issues are priced at $5.00. All foreign subscriptions and orders should be accompanied by payment. Address orders to Journals Department, University of Illinois Press, 54 E. Gregory Drive, Box 5081, Station A, Champaign, IL 61820. Editorial correspondence should be sent to Publications Office—Library Trends, 249 Armory Bldg., 505 E. Armory St., Champaign, IL 61820.

Indexed in Current Contents, Current Index to Journals in Education, Library and Information Science Abstracts, Library Literature, PAIS, and Social Sciences Citation Index.

PRINTED IN THE U.S.A.
Data Libraries for the Social Sciences

KATHLEEN M. HEIM

Issue Editor

CONTENTS

Kathleen M. Heim 321 INTRODUCTION
Judith S. Rowe 327 EXPANDING SOCIAL SCIENCE REFERENCE SERVICE TO MEET THE NEEDS OF PATRONS MORE ADEQUATELY
Sue A. Dodd 335 TOWARD INTEGRATION OF CATALOG RECORDS ON SOCIAL SCIENCE MACHINE-READABLE DATA FILES INTO EXISTING BIBLIOGRAPHIC UTILITIES: A COMMENTARY
Joseph W. Duncan 363 ACCESSING SOCIAL STATISTICS
Margaret O’Neill Adams 377 MODELS OF DATA LIBRARY DEVELOPMENT AND INFORMATION SYSTEMS SERVICES: AN OVERVIEW
Ray Jones 383 THE DATA LIBRARY IN THE UNIVERSITY OF FLORIDA LIBRARIES
Laine G.M. Ruus 397 THE UNIVERSITY OF BRITISH COLUMBIA DATA LIBRARY: AN OVERVIEW
Alice Robbin 407 THE DATA AND PROGRAM LIBRARY SERVICE: A CASE STUDY IN ORGANIZING SPECIAL LIBRARIES FOR COMPUTER-READABLE STATISTICAL DATA
Margaret O’Neill Adams 435 ONLINE NUMERIC DATA-BASE SYSTEMS: A RESOURCE FOR THE TRADITIONAL LIBRARY
CONTENTS—Continued

<table>
<thead>
<tr>
<th>Author</th>
<th>Page</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laine G.M. Ruus</td>
<td>455</td>
<td>TRAINING OF DATA SERVICES PROFESSIONALS: PAST, PRESENT, AND FUTURE</td>
</tr>
<tr>
<td>Howard D. White</td>
<td>467</td>
<td>CITATION ANALYSIS OF DATA FILE USE</td>
</tr>
<tr>
<td>Jeff Sobal</td>
<td>479</td>
<td>THE ROLE OF SECONDARY DATA ANALYSIS IN TEACHING THE SOCIAL SCIENCES</td>
</tr>
<tr>
<td>David F. Linowes</td>
<td>489</td>
<td>DATA CONFIDENTIALITY, SOCIAL RESEARCH AND THE GOVERNMENT</td>
</tr>
<tr>
<td>Michele M. Hoyman</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barton M. Clark</td>
<td>505</td>
<td>SOCIAL SCIENCE DATA ARCHIVES AND LIBRARIES: A VIEW TO THE FUTURE</td>
</tr>
</tbody>
</table>
Introduction

KATHLEEN M. HEIM

August Comte's grand vision of the final synthesis of the natural sciences and social physics focused on sociology as the universal link "completing the upward flight of our contemplation of reality." The end result would be an applied social science that would ameliorate the state of humanity. Comte's call for a general social theory that would consist of a set of general, testable, explanatory propositions applicable to the total area of collective human behavior was shared by later nineteenth-century statisticians, sociologists and anthropologists. These were the individuals who endeavored through the comparative method to establish an internationally and interculturally valid body of knowledge about variations and regularities in the functioning and development of human societies.

Such an aim proved difficult to reconcile with other compelling objectives as the social sciences developed. The need to establish strict canons of evidence and inference in order to achieve a high level of analytical precision, as well as the need to gain academic recognition for these emerging disciplines, tended to force early social scientists to abandon universalistic theorizing and focus their inquiries on the local, the concrete and the specific. The social sciences achieved academic status by increasing attention to methodological rigor and deliberate concentration on well-delimited local or national inquiries. Though the gain in precision has been great, the original aim of the social sciences has tended to be sacrificed and the idea of an encompassing or

Kathleen M. Heim is Assistant Professor, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.
grand theory that synthesizes universal concepts is not a prominent one in the current panoply of social science philosophy. Nevertheless, the idea of the eventual achievement of a universal theory holds a persistent grasp on the imagination and presents to the layperson the most obvious rationale for the importance of the social sciences.

Before social scientists could begin to reconsider postulating theories at a highly general level, the raw information required to develop such ideas has had to undergo four revolutions. Karl Deutsch has described these as: (1) the collection of largely disjointed facts and figures chiefly for administrative, tax and military purposes in the seventeenth and eighteenth centuries; (2) the use of historical data dealing with successions of types of societies by such social scientists as Herbert Spencer or Karl Marx; (3) the rise of new methods in the mid-1930s for gathering partial and sectoral data, along with new quantitative techniques for organizing and interpreting them in order to put discrete, disjointed data in relationship (helped along by advances in survey research and sampling theory); and (4) the rise of multiple methods and complex data bases with the eventual aim of all-to-all comparison.

The fourth data revolution, with the possibilities of all-to-all comparisons, has come into being because of technological advances in the computer which allow ever more complicated statistical analysis of data and expanded capacity for storage of these data in facilities such as data archives or data libraries. The data library contains machine-readable collections of survey, census, polling, or legislative voting information and provides a laboratory for the social scientist to analyze data in order to make sometimes narrow and sometimes highly general observations about the nature of society.

Data archive and data library development was the focus of intense national and international debate in the late 1960s and early 1970s. The massive programs of international economic and political integration undertaken after World War II demanded much comparative research which was unavailable because the theoretical underpinnings of any efforts at cross-national comparison were poor and fragmentary. Unesco support to forward the state-of-the-art in the comparative social sciences often focused on the data library as a primary component in the information system required to enable researchers to understand and provide solutions which might alleviate disparities in the development of various nations. In spite of the extensive international and national efforts to develop not only individual data libraries throughout the world but networks of such information as well, the data library move-
Introduction

ment has persisted but not flourished. Although researchers at every university in the world could make use of data libraries, the high cost and poor understanding of their role in the social science information system has mitigated against full-scale development on the level of more traditional information facilities, such as libraries. Put most simply, the material in data libraries is a vital source of information for social research, yet because of the nontraditional format of these data, has been ignored by the information community at large. The failure of most models of the social science information system to include the data library in schemata and diagrams indicates that this resource is poorly understood. If the highly motivated librarian attempts to provide information resources to scholars wishing to study voting behavior and can only direct them to books, periodicals and government reports, major lacunae in the information infrastructure are evident, for machine-readable collections on the topic are probably more pertinent than anything published.

Failure of traditional libraries to consider machine-readable data files as within their purview has caused data library development to take place, in the main, outside of traditional library settings. Worse, most librarians in large research libraries do not even recognize the disservice they do their clientele in omitting information provision in this area. This issue of Library Trends is intended to place the data library and its holdings squarely in the forefront of vital information resources for social sciences research. Judith S. Rowe provides a general introduction to the importance of Machine-Readable Data Files (MRDF) in the social sciences reference exchange and demonstrates examples of their use in typical library situations.

Sue A. Dodd describes the arduous struggle to develop bibliographic control over MRDF and thus legitimatize them to the library community. She observes that “communicating the availability of usable data is an inseparable part of research and an integral part of librarianship. In the near future libraries will have no choice but to become more involved.”

The enormity of information collected in MRDF is characterized by Joseph W. Duncan, who notes the underutilization of federal social data due to a lack of adequate information. Duncan discusses the data access policies of the federal statistical system and selected source documents which aid researchers in accessing these data.

Margaret O’Neill Adams introduces four models of data library development and notes that “there is no clear administrative structure for services related to social science data files nor for numeric informa-
tion systems that is ideally suited for all institutional settings." The four models which Adams presents provide alternative functional strategies for meeting information needs for MRDF. These models include the integrated University of Florida facility outlined by Ray Jones, the University of British Columbia Data Library operated jointly by the University Library and Computing Centre described by Laine G.M. Ruus, the Data and Program Library Service at the University of Wisconsin characterized by Alice Robbin, and Adams's online numeric Kentucky Economic Information System at the University of Kentucky.

Once the need for data libraries has been made clear, the type of resources they might hold particularized, strategies of bibliographic access delineated, and model facilities described, the question arises: Who will staff and maintain such services? Laine G.M. Ruus explores the training of data services professionals and includes a suggestion that graduate schools of library education consider incorporating such training in their curricula.

The next three papers in this issue examine use of data files, the role of data files in social science teaching, and issues of confidentiality and privacy. Howard D. White explores the current state of affairs vis-à-vis citations to MRDF through an analysis of Social Sciences Citation Index detailing the complexities and vagaries of accurate bibliographic control over these files.

Jeff Sobal, a social scientist and a user of MRDF in his own research, discusses the role of secondary data analysis in teaching the social sciences, and suggests ways in which traditional libraries might assist in expanding awareness of MRDF as an information resource.

A pressing concern for data librarians is the confidentiality of the information retained in data files. David F. Linowes, former chairman of the U.S. Privacy Protection Commission, and Michele M. Hoyman of the Institute of Labor and Industrial Relations at the University of Illinois analyze the broad issue of information privacy in general as well as its relationship to the function of the librarian and archivist.

Observations by Barton Clark, head of the Social Sciences Libraries at the University of Illinois conclude this issue. Clark is not optimistic about the wholesale integration of traditional libraries and data libraries, but rather takes a middle stand in recommending that traditional libraries become bibliographic brokers for data libraries.

If long-range goals for the systematization of a social science infrastructure as outlined by the UNISIST International Committee for Social Science Information and Documentation are met, the data library will be fundamental to the aim of creating a world system of social
Introduction

science information and documentation. I hope that this issue of *Library Trends* contributes to an understanding and greater visibility for this vital social science resource.

References


3. For a detailed account of national and international efforts to develop a network of data archives, see Heim, Kathleen M. “Social Science Data Archives: A User Study.” Ph.D. diss., University of Wisconsin-Madison, 1980.

This Page Intentionally Left Blank
Expanding Social Science Reference Service to Meet the Needs of Patrons More Adequately

JUDITH S. ROWE

In the beginning all social science data were created in printed form: books, journals, pamphlets, documents, technical reports. But slowly the world changed, and additional social science data began to appear on microform and as Machine-Readable Data Files (MRDF). Librarians designed cataloging and access systems for microfilm and microfiche, and patrons learned to master the use of film and fiche readers. Thousands of feet of shelf space were saved in the process, acquisitions budgets went further, binding and maintenance costs were reduced, and additional materials became readily available to users of social science information.

At almost the same time, public opinion pollsters and social science researchers in academic institutions, private companies and government agencies began to utilize the computer to store and analyze data. As a byproduct of these efforts, social science data became available in machine-readable form. At first these files lacked bibliographic identity; and although most of them were used only by their individual or corporate creators, copies on cards or tape were sometimes passed on to friends or fellow researchers. The establishment of the Roper Center in 1946 as an archive for public opinion polls and of the Inter-University Consortium for Political Research in 1962 as a political data archive were the first organized efforts to formalize the distribution of MRDF.

Since MRDF are not eye-readable even with the aid of special devices, librarians have been initially diffident about regarding them as extensions of traditional information resources, in spite of the fact that

Judith S. Rowe is Associate Director for Academic Data and Program Services, Princeton University Computer Center.
computers, computer terminals and computerized information are becoming less foreign. Circulation librarians have been using automated systems such as CLSI, DataPhase and GEAC for many years. Technical services librarians are familiar with MARC, OCLC, UTLAS, WLN, and at the larger research libraries, with RLIN. Reference librarians are using these same tools for author, title or subject access to monographs, and have learned to search the growing number of bibliographic data bases available through Lockheed, SDC and BRS. Some libraries are also providing access to both bibliographic and nonbibliographic data available online from more specialized vendors. For most librarians, however, none of these activities involve dealing with the physical entity, usually a tape, on which data are stored. And so the job of acquiring, cataloging and providing access to machine-readable information typically falls outside the purview of the central library.

MRDF are normally housed at data libraries in academic departments, research institutes or computer centers. A very few libraries have incorporated MRDF into the library's main collection. Some have acquired only codebooks or technical documentation which accompany and describe the tapes, and some have included records for MRDF in their public catalogs or in their subject bibliographies, leaving the tapes themselves to be serviced elsewhere. Most libraries have done nothing.

The world of information is changing. A decade ago a reference librarian who provided a patron with assistance in searching the card catalog, who identified relevant bibliographies and other printed reference tools, who pointed the patron to the indexes of possibly useful serials, and who in some cases went right to an appropriate monograph or vertical file, had done a thorough literature search. If the information could not be located through these strategies, it probably did not exist. But the computer has changed reference service. Not only has it provided online access to bibliographic data bases but it has also made it possible for library patrons to access primary data and, in some instances, to create their own information. Some typical reference inquiries illustrate the relation of these new resources to the library's more traditional ones.

One library patron is a public opinion pollster. Month after month he has conducted polls in which representative samples of the population have been interviewed and asked questions ranging from which toothpaste they last bought to which candidate for president they voted for in a recent election; from their satisfaction with their local mayor to their satisfaction with their doctor, their lawyer or their accountant;
from their attitudes about a school redistricting plan to their knowledge about the issues involved in the increase in the U.S. military budget. In order for the pollster to be confident that his samples are of the proper size and distribution to represent accurately the population he seeks to describe, he must periodically look at the universe from which these samples are drawn. This means that every ten years he must look at the full count census data. If he were only concerned with, e.g., large areas, states or Standard Metropolitan Statistical Areas (SMSAs), he might consult the printed census reports or use one of the special online data services. If he needed small area income or housing data for a substate area such as a Minor Civil Division (MCD) or a census tract, he might copy these from a printed or microfiche source. However, if he needed enumeration, district or block group data for these areas, or census tract data for the whole country, these would be either unavailable or inefficient to retrieve from any source but the summary tape files provided by the U.S. Bureau of the Census. Although stored in machine-readable form these data are intellectually similar to the more familiar census printed reports. They are aggregate data stored in tabular form.

Another library patron comes to the reference librarian. She is interested in correlates of contemporary American voting behavior. Either through a catalog search using such subject headings as elections, political participation, public opinion or voting, or at the recommendation of the reference librarian, she may locate such printed sources as Campbell, Converse, Miller, and Stokes's *The American Voter,* their *Elections and the Political Order,* or the more recent *American National Election Studies Data Sourcebook 1952-1978.* Through the use of appropriate periodical indexes or computerized bibliographic searches using such data bases as *Social Sciences Citation Index, U.S. Political Science Documents* or the *New York Times Information Bank,* she may find additional references and additional information based on a variety of public opinion polls. For specific tabulations from the polls themselves, she might consult the monthly *Gallup Opinion Index: Political, Social and Economic Trends,* its earlier antecedents, press releases from other polling agencies or such publications as *Public Opinion Quarterly, Current Opinion* or *World Opinion Update.* But still the particular data she seeks elude her. With the aid of an integrated multi-media catalog, a knowledgeable reference librarian and/or an available data library, she might be directed to the *American National Election Studies,* the MRDF which contain the actual data collected from every interview taken in each of the continuing series of surveys conducted by the Institute for Social Research at the
University of Michigan in every even year between 1948 and 1980. Together these files contain almost 30,000 records—each one representing a separate interview in which as many as 900 questions were asked. Each question was coded and the data keypunched, or in recent years entered on a terminal, and then written on tape. These tapes, as well as those from the CBS/New York Times Surveys, are now archived with their accompanying codebooks at the Inter-University Consortium for Political and Social Research, and copies are made available for use at local computer facilities. Tapes containing data on voting behavior from many other surveys, including those conducted by Gallup, Roper, Yankelovitch and numerous state polls, may be obtained from the Roper Center located at the University of Connecticut at Storrs. Louis Harris data are archived by the Institute for Research in Social Science at the University of North Carolina, Chapel Hill. Using these tapes and special computer programs known as statistical packages, a user can relate any of the questions asked to any of the others. Are younger high school graduates less likely to get campaign information from newspapers than from other media? Is place of residence more important than education in determining the voting behavior of blue-collar workers? Unless a previous researcher has performed and published these tabulations, our patron must use a computer to analyze these basic data herself. This can only be done because the data have been preserved for such secondary analysis. Intellectually, these survey records more closely resemble the individual records stored in traditional archives or in manuscript collections than the statistical reports with which librarians are familiar. These records contain the original data used to create aggregations rather than the aggregations themselves.

A third patron, aware that the library collects government documents and concerned about the effect of time zone differences on voting choices, seeks to find information on the time of day at which various types of people vote. As a result, the patron consults the Current Population Reports in Series P-20 which deal with voting and registration. The librarian does a search using both the Government Printing Office Monthly Catalog and the American Statistics Index to learn which reports will be helpful and finds that they do indeed contain information on the time of day at which voters in various regions go to the polls, but they tell nothing about differences between young and old voters, men and women, employed and unemployed, etc., nor do they provide data for individual states. The solution is to gain access to the public-use microdata samples from the Current Population Surveys used to produce these reports. These samples consist of individual records divested
of identifying information and are available for every congressional
election year since 1972. With these it is possible to see, for example, if
the prototype “early voter” in New York is different from the one in
California, if employment status affects voting behavior, etc. No printed
source or online data base can provide these data.

Still another patron is concerned about obtaining information
about how attitudes toward social welfare have changed over time, both
in the United States and abroad. The card catalog is searched using such
subject headings as insurance, social; poor; poverty; public welfare;
social security; and welfare state. Publications dealing with the history
of social welfare provide some indications. General periodical and
dissertation indexes, economics, politics and sociology indexes and
abstracts identify other references of other types. But what did people
really think? Questions on this issue have been asked since the earliest
days of polling. The Roper Center can provide either individual data
files or specified tabulations from these data files going back to the
Gallup polls of 1936, in the midst of the Great Depression. Almost fifty
years of public attitudes on issues ranging from social welfare to gun
permits, from income tax to pornography, are available to the contem-
porary chronicler of changing social and political attitudes.

Three recent publications document changes in some important
political and social attitudes using MRDF produced during the past
twenty years. Throughout the world academicians and policy-makers
are performing secondary analysis of existing data files to address ques-
tions which had never occurred to the original investigators.

Why aren’t all these data just printed out, bound and shelved for
library users? There are three main reasons: their specialized nature, the
quantity of paper (or even of fiche) involved, and perhaps most impor-
tantly, the need to analyze or process these data in machine-readable
form. To illustrate, let us take a small survey containing 150 variables
for each respondent. Each user of these data has a somewhat different
research interest, i.e., each is interested in a different subset of these
variables. From the same survey one researcher may look only at ques-
tions dealing with employment status, another may be concerned pri-
marily with family composition, and still another with attitudes on
social issues. One research design may involve the creation of multilevel
cross-tabulations such as age by sex, by employment status for each
state, or SMSA. Another may involve correlations or regressions
designed to identify the demographic variables most likely to determine
given social attitudes, and so on. Any attempt to print out in advance all
of the possible relationships would require reams of paper and hours of
programming, and would almost certainly omit the specific computa-
tion which some researcher would require. In the ten years during
which Princeton has provided access to 1970 census data, we have never
had the same request more than once.

Librarians should be concerned with improving access to MRDF,
but even more importantly, with providing access to information about
MRDF. There is, at this point, no single printed or computer-readable
reference tool which can serve as a comprehensive finding aid for
MRDF. Publications of federal agencies (see Duncan's article in this
issue), the ICPSR Guide to Resources and Services, similar smaller
catalogs or directories published in the United States and abroad by
individual data libraries, and a variety of journals and newsletters cover
most of the easily available public data files, but many useful files
remain elusive. Efforts by professional associations, such as the Interna-
tional Association for Social Science Information Service and Technol-
ogy (IASSIST) and the Association of Public Data Users (APDU), and
by funding agencies to encourage the deposit of MRDF in data archives,
and by journal editors to require standard citation of this material in the
literature will go a long way toward increasing access to bibliographic
information describing these files and their availability and use. The
cataloging of individual MRDF and the entry of these catalog records
into printed or computerized single- or multimedia catalogs is now
being addressed. Both the cataloging rules and the MARC format have
been developed. Current recommendations call for more extensive
catalog records than are usual for other materials.

A next step which has been suggested is the creation of a variable-
or question-level file which could be searched by reference librarians.
Although there has been some exploration of this activity both in the
United States and abroad, it has first been necessary to deal with ques-
tions of bibliographic identity. The development of variable- or
question-level indexes for MRDF is akin to the development of chapter
and paragraph-headings indexes for monographs. It is a costly and
complicated project. Adequate financial support for the development of
question-level indexes has not been forthcoming, and techniques for
such development are still inadequate. It has also been recommended,
although seldom by commercial vendors, that all of these MRDF be
available online. Although the cost of online storage is decreasing, it
seems unlikely that any unsubsidized service could afford to keep all
public MRDF available online or even to acquire and maintain all of
these files. Such a recommendation is akin to supplementing each
bibliographic file with a file containing the full contents of each cited
Social Science Reference Service

article. Several online statistical data bases, currently available through commercial vendors, are heavily used by businesses. However, they do not meet the needs of potential users of small area data or of public use microdata. They do not, for example, meet the need for tabulations from the census containing information regarding marital status by income for census tracts in Omaha, Nebraska, or any other city; nor, for example, do they meet the need for joint analysis of complex income and employment indexes which must be created from individual records. In the future, we will no doubt see increased online access to both aggregate data and microdata, but it is unlikely that it will ever be economically feasible to store all available statistical data in this manner.

As a result, although major data archives and even local data libraries attempt to acquire comprehensive collections of MRDF to meet the needs of their users, like even the most diligent subject bibliographer, they often find their collections incomplete. Social science reference librarians can increase the quality of their service to patrons by becoming aware of machine-readable data products and of their place in modern research, policy-making and classroom instruction. With this awareness another step will be taken in expanding the librarian’s role as gatekeeper to the ever-growing wealth of social science data, and toward the recognition that physical form should not be a barrier to information access.

References

7. A statistical package is an integrated collection of computer programs used to manipulate, transform and analyze data. The most widely used of these packages are BMDP, OSIRIS, P-STAT, SPSS, and SAS. All of these packages are designed for use by nonprogrammers. Although each has some report generation capabilities, their main strengths are in allowing the user to invoke a variety of statistical procedures with a small number of simple commands.
JUDITH ROWE


Toward Integration of Catalog Records on Social Science Machine-Readable Data Files Into Existing Bibliographic Utilities: A Commentary

SUE A. DODD

Formal recognition of the need for bibliographic control over computerized information has slowly been evolving within the library and information science profession over the past several years. A major landmark that helped to focus increased interest in the cataloging of social science data files was the inclusion of chapter nine on "Machine-Readable Data Files (MRDF)" in the second edition of the Anglo-American Cataloging Rules (AACR2). Publication of these rules in 1978, coupled with a number of other events, including the compilation of a machine-readable catalog (MARC) format for machine-readable data files, provided the important links that would facilitate the integration of bibliographic records into local automated systems and eventually into national information systems.

The most recent cataloging code (AACR2) and the MARC format for MRDF provide the standards required for describing and creating automated records, which in turn can be applied to many different purposes, such as shared cataloging, acquisition systems, and the building of a union list on all available MRDF. The primary purpose of this paper is to provide a commentary on the significant steps that have contributed to this current level of bibliographic control and to outline some of the remaining problems still to be considered before MRDF bibliographic records can successfully be integrated into existing bibliographic utilities. (A bibliographic utility as referenced in this paper is an organization that maintains a large bibliographic data base in an online mode via communications lines enabling it to offer computer

Sue A. Dodd is Associate Research Librarian, Institute for Research in Social Science, University of North Carolina, Chapel Hill.

WINTER 1982 335
base support to any interested users, including designated network participants.)

Overview and Definitions

If it is true that advances in modern information technology have far exceeded our ability to control data products generated by a computer, then it is equally true that the proliferation of the various types of data files has impeded our ability to apply a consistent vocabulary to describe and distinguish one from the other. Differing vocabularies have emerged depending on which segment of the information community is speaking. To the bibliographic-oriented portion of the information profession, "data bases" are machine-readable bibliographic files, whether produced by a library or by the American Chemical Society. When information other than that represented by the bibliographic journals and indexes was marketed online, then the terms bibliographic and nonbibliographic emerged. However, the use of this terminology encouraged others to offer another approach. Sessions suggested: "Although the terms bibliographic and non-bibliographic data seem clear enough, the negative term can be eliminated and a clearer relationship between the two kinds of information established by referring to primary and secondary data files." To Sessions, primary would be equated with original (or primary source of) information. The computerized version of the census data, for example, would be considered primary, while the resulting printed census volumes plus the bibliographic references to census documents would be considered secondary. Primary data sources in the social sciences predate the "online revolution" and bibliographic data bases. In fact, census data were the first to be represented in punched card form and the first to be computerized by means of UNIVAC I in 1951. Public opinion data represented by the established pollsters were another early source of computerized data, and as the collections of public opinion data increased, data archives and libraries were established to maintain these collections. The earliest and largest collection of public opinion data in the world today, which dates to 1935, is that of the Roper Public Opinion Research Center, founded in 1946.

To social scientists, a "data file" or "data set" will most often refer to a set of numeric values that can be manipulated by a predesigned statistical routine. Characteristically, data from numeric files are statistically manipulable and subject to quantitative analyses. Such files are manipulated using different forms of statistical software, such as tabu-
Integration of Catalog Records

 Numeric data can result from surveys of households and/or individuals, from scheduled censuses, from administrative records or economic reports, from test scores, and from other sources of statistical information. Data obtained from surveys or censuses can be classed into two groupings: summary data and microdata.

 Summary data are aggregations of individual record data. Totals and frequency distributions show numbers of persons, families, housing units, corporations, vehicles—whatever the unit of enumeration is—distributed by their various characteristics for different geographic areas. A subset of summary data are the so-called time series data files. Time series are observations of discrete variables—such as the price of wheat or grain, the GNP, or the employment totals of an industry—for periodic intervals, such as months or years. Microdata are unaggregated data, produced from basic household and person unit record data (i.e., the actual responses of each person who completes a questionnaire).

 A major producer of social science data is the federal government. A significant proportion of federal social and statistical data is disseminated by five general-purpose collection agencies: the Statistical Reporting Service of the Department of Agriculture, the Bureau of Labor Statistics of the Department of Labor, the Bureau of the Census of the Department of Commerce, the National Center for Education Statistics (NCES) of the Department of Education, and the National Center for Health Statistics (NCHS) of the Department of Health and Human Services.

 With the appearance of cataloging rules for computerized data and the integration of such data into traditional library collections, still another vocabulary emerges. The generic term for computerized information that has been offered by the library community is "machine-readable data files." According to AACR2, a machine-readable data file is defined as any information encoded by methods that require the use of a machine (typically, but not always, a computer) for translation. The justification for the selection of this term by the ALA Subcommittee for Cataloging Rules for Machine-Readable Data Files is documented in their final report:

 Frequently-heard designations are those introduced by the word "data"; "data record," "data set," "data file," "data base," "data bank," etc. To many these terms convey a sense of size, a "data item" being the smallest unit, and "data base" or "bank" implying the largest accumulations. Between these extremes, "data set" and "data file" are sometimes used interchangeably, but "data file" is more unambiguously defined as a collection of related records to be treated
as a unit, while definitions for “data set” vary according to computer languages, glossaries, and individual usage. However, any designators which do not take into account the means of access to the information do a disservice to the catalog user, as any of the terms introduced by “data” could conceivably apply to information in another medium.

AACR2 further defined the generic term by stating that “machine-readable data file embraces both the data stored in machine-readable form and the programs used to process that data.” Consequently, the term machine-readable data file or its acronym MRDF as used throughout this paper will stand both for data files and program files.

The term machine-readable is easily understood, especially when it can be equated with computer-readable, but the term data file still warrants more explanation. A data file is defined here as any organized collection of automated records that are related in some way and treated as a unit, e.g., a payroll file with one record for each employee, showing his rate of pay, annual leave, deductions, etc. In most cases, the reader should conceptualize a singular MRDF to be an “inert file”—that is, existing alone as a separate entity on any number of data carriers such as a magnetic tape. It is this “inert file” of computerized information that conceptually becomes the “item in hand” to be described.

The opposite of an inert file may arbitrarily be defined as a “dynamic file” or a “dynamic data base.” A dynamic data base is one that is characterized by its fluid and constantly changing nature. It may be represented by economic time series, or bibliographic data bases, and may be corrected, revised retrospectively, updated, merged, partitioned, and blocked into subfiles without changing its bibliographic identity. Even though these data files are associated with online systems, many are also available on a magnetic tape subscription basis and could conceivably become part of a library’s collection of informational resources represented by a serial catalog entry.

Events Contributing to Bibliographic Control of Social Science MRDF

Although the early abortive attempts in 1957 to involve traditional libraries in the acquisition and management of social science data files have been well documented, it was not until the early 1970s that the library profession began to take a bibliographic interest in MRDF. In January 1970, the Executive Committee of the ALA’s Cataloging and Classification Section instructed the Descriptive Cataloging Committee to form a Subcommittee on Rules for Machine-Readable Data Files.
Their mandate was to recommend methods of describing data files that would be compatible with existing cataloging procedures for other media. This effort grew out of the perceived need to apply some type of bibliographic control to data files that were actively being collected by academic and research institutions. In the absence of any local data archive or center, these materials after their initial collection and application were often brought to the library for processing. Faculty members were asking librarians to acquire MRDF for them, just as they would request an important book or reference work. According to Byrum, the establishment of the subcommittee marked the formal recognition of the need for standards by which libraries could assist in the control and access of data files which academic and other institutions had already begun to collect as an additional and increasingly important resource of educational and research value.7

First under the direction of John D. Byrum, Jr., chief of the Descriptive Cataloging Division, Library of Congress, and later under the direction of Elizabeth Herman, Technical Services Department, University of California at Los Angeles, the subcommittee met twice a year for five years, drafting position papers and making recommendations on every component of the catalog bibliographic record. Their final report was filed in January 1976, and it was this document8 that laid the groundwork for chapter nine in the second edition of AACR, which in turn introduced rules for cataloging MRDF for the first time.

National Bureau of Economic Research (NBER) Workshop

The subcommittee made an effort to gather feedback from nonlibrary audiences who represented the data processing or data producing community. An important forum for an exchange of ideas and information on bibliographic aspects of MRDF took place in 1974 with the National Bureau of Economic Research’s conference on “The Computer in Economic and Social Research,” and its workshop on “Documentation of Large Machine-Readable Statistical Data Sets.” The focus of this workshop included an evaluation of the recommendations of the subcommittee’s work to date. Early cataloging examples were presented at the workshop, along with a checklist of descriptive bibliographic elements.

An additional focus of the workshop was a discussion on the content and format of literature citations for social science data files. It was recognized that an accurate and complete literature citation for social science data files would benefit the researcher and potential user alike and would pave the way for social science data files to be included
in printed bibliographies, end-of-work references, and indexing and abstracting works such as the Social Sciences Citation Index. Today, guidelines on how to create a bibliographic citation are available to the reader, and a major journal in the social sciences—Social Forces—now carries in its “authors’ guide” section instructions with examples on how to cite a MRDF in the literature.

_**Early Cataloging Efforts**_

During the five years that the subcommittee met, the members had written and verbal contacts with data librarians who were beginning to catalog data files. Even as the subcommittee was meeting and debating on the bibliographic elements of MRDF, several research-oriented libraries and data centers were beginning to compile catalog records on data files held by their respective institutions. In Canada, the University of British Columbia and the Public Archives of Canada (Ottawa) took the initiative; and in the United States, it was Yale University (Social Science Data Library) and the University of North Carolina at Chapel Hill (Social Science Library). In other cases, like the University of California at Los Angeles, it was the automated 1970 census records that became the first MRDF to be included as part of a library's collection.

_IASSIST_

The ALA subcommittee was not the only group working to define standards for MRDF. Others included the Computer Media Working Party of the Library Association's Media Cataloging Rules Committee, the American Society for Information Science (ASIS) Special Interest Group for Non-Print Media, and the Association of Educational Communications and Technology (AECT) Cataloging Committee. Another newly formed group—the International Association for Social Science Information Services and Technology (IASSIST)—was established with a special subunit devoted to promoting cataloging and classification procedures for social science data files. As chairperson of this organization's Classification Action Group, I directed a special project aimed at testing the feasibility of cataloging MRDF. The participants were members of IASSIST who had expressed an interest in classification, although many were neither librarians nor catalogers. A brief manual based on the position papers of the ALA subcommittee was given to the participants, and each was asked to select six MRDF of numerical, text or program files and proceed to catalog these files, keeping records on time spent, problems encountered, etc. The rationale for this project was based on two major assumptions: (1) that data
files and programs are underutilized, and the lack of knowledge on the availability of existing data files has hampered the academic community and other interested parties in the ongoing process of scholarly research; and (2) that there is no standard format for providing information on the availability of data files which would make possible one central source of information. The library profession's historical commitment to standards and to providing bibliographic information on a variety of media provides a natural background to study the feasibility of applying library cataloging and classification procedures to MRDF.

By testing the ALA subcommittee's rules and providing the initial cataloging experience, it was expected that the most immediate outcome of this committee's work would be to help pave the way for the inclusion of MRDF catalog records into the local or most appropriate public facility. The effort yielded over forty individual catalog entries from nine participants representing the following institutions: National Opinion Research Center, Data Use and Access Laboratories (DUALabs), Yale University Social Science Data Archive, Drexel University Graduate School of Library Science, University of Pittsburgh Social Science Computer Research Institute, Rutgers University, and the National Archives Machine-Readable Archives Division. The variety of MRDF represented in the project was significant as well as interesting. The types of MRDF included text files, bibliographic data bases, census and census-related files, survey data, panel studies, time series, aggregate data banks, longitudinal files, serials, computer software programs, mathematical models, online program lessons, educational data packages, and simulation games. All of these different kinds of MRDF with their unique characteristics were successfully cataloged within the scope of the subcommittee's recommendations and with the guidance provided in the manual. While this project helped to establish the feasibility of cataloging MRDF by many different parties with varying degrees of cataloging experience, it also helped bring to the surface some of the problems that have come to be associated with the overall cataloging effort.

National Cataloging Conference

In March 1978, a national Conference on Cataloging and Information Services for Machine-Readable Data Files was held at Airlie House, Warrenton, Virginia. It was funded by the National Science Foundation and was organized by DUALabs. This was the first concerted effort at a national level to develop standards and to suggest cooperative efforts in establishing bibliographic control for MRDF. This meeting brought...
together key persons and organizations having an active interest in establishing a framework within which a national program of cataloging and information services could be developed. A heavy emphasis was placed on the problems of federal data producers and publicly available data.

While the conference did not attempt to provide solutions to the problems associated with applying standardized bibliographic control procedures for MRDF, there was general consensus that such procedures and related information services are urgently needed to improve user access to machine-readable data resources. The conference concluded with a "call for action" and with several recommendations, including the following:

— that "the AACR2 rules should be tested on a broad range of MRDF to determine the feasibility of using these rules as a standard for cataloging."
— that any resulting procedures should be directed toward an automated system of bibliographic records for MRDF;
— that the Library of Congress should be encouraged to design and establish a MARC format for MRDF;
— that products and services which could be derived from such a cataloging effort be defined; and
— that the feasibility of integrating the resulting catalog records into existing network systems be investigated.12

**Federal Task Force**

Immediately following the Airlie House cataloging conference, the Office of Federal Statistical Policy and Standards (OFSPS) took action to establish a mechanism for using the staff resources contributed by various federal agencies. The result was the establishment of a federal task force, which in turn, would coordinate federal efforts to develop acceptable standards for cataloging MRDF. Under the task force's direction, a small interagency working group developed standards for statistical data files as they apply to creating bibliographic citations and abstracts. These procedures are presently being applied by several agencies in an effort to produce more informative and reliable directories of federal MRDF.

In October 1979, the Bureau of the Census issued a new inventory of their holdings, entitled *Directory of Data Files*.13 With this *Directory*, the bureau has incorporated the task force's standards for citation and abstracts. The citation may be characterized as a "minicatalog" entry which includes the International Standard Bibliographic Description
Integration of Catalog Records

(ISBD) punctuation. This effort was a tremendous breakthrough on the federal level, and it made the link between descriptive practices of the federal sector and the existing bibliographic standards of the library and information science community much closer.

In a related effort, the OFSPS established an Interagency Committee on Data Access and Use. This committee, in turn, initiated a multi-agency project to adapt these same standards and produce a comprehensive directory of federal statistical data files. The Directory of Federal Statistical Data Files, issued jointly by the Machine-Readable Products Division of the National Technical Information Service (NTIS) and OFSPS, contains more detailed bibliographic information than past efforts. (For a further description of the directory, see Duncan's article in this issue of Library Trends.)

OFSPS has now moved to the Office of Management and Budget (OMB), and it is expected that OMB will issue a Statistical Policy Directive on Standards for Abstracts of Public Use Statistical MRDF. Such a directive would help to institutionalize the Directory as a regular periodic publication and to establish uniform standards among federal statistical agencies. Standing behind the directive will be Technical Paper No. 3: Procedures for Preparation of Abstracts of Public Use Statistical Machine-Readable Data Files.14

Cataloging Manual

The cataloging manual that had been used in the IASSIST-sponsored cataloging test had to be revised. After the ALA subcommittee issued its final report, the Joint Steering Committee of AACR2 made further changes and recommendations. With the appearance of chapter nine in the second edition of AACR, a new manual was planned, and its scope was extended to include basic procedures for proper bibliographic control and additional levels of recordkeeping associated with library management of data files. Its objectives were broken down into five broad areas: (1) to provide guidelines for establishing bibliographic conventions for MRDF (especially for those data producers or distributors in need of guidance or structure in this area); (2) to suggest integrated levels of recordkeeping for MRDF; (3) to bring into sharper focus the AACR2 rules as they relate to cataloging of computerized files; (4) to provide notes, examples and interpretations of MRDF cataloging which would otherwise not be available; and (5) to provide working tools for those cataloging MRDF for the first time.

Assistance was sought to support the work on this new manuscript entitled Cataloging Machine-Readable Data Files: An Interpretive
Manual, and in August 1979, I received funding from the Council on Library Resources (CLR). The grant from CLR was funded under the auspices of its Bibliographic Service Development Program (BSDP), which has focused on the development of a set of strategies aimed at establishing bibliographic control of materials in libraries and sharing the bibliographic data that is produced. What followed was a period of investigation and research into the informational needs of both catalogers and users of MRDF. It was determined that the interpretive aspects of the manual should fall on the side of the many intricacies of computerized information in general and on the unique characteristics of certain classes of MRDF in particular. Experts were consulted in the areas of computer hardware and software, computer cartography, language/text processing, simulation models, federal statistics and survey data. Site visits were completed to the Library of Congress and to academic and research libraries, including Princeton University, Columbia University, and University of Michigan. In addition, an exchange of information among those data centers and libraries engaged in early cataloging efforts was developed. As a result, the manual contains explanatory information plus cataloging examples on many different types of MRDF, including survey data, federal statistical files, cartographic programs, econometric models, computerized dictionaries, Greek text files, and economic time series. Associated terminology is defined and a glossary of MRDF-related terms is provided.

The biggest difference between the cataloging of books and the cataloging of MRDF is that the cataloger normally does not have an "object in hand" which he is able to describe; and even if he did, it would not do him much good. External descriptive labels on magnetic tapes are not permanent, nor do they carry the customary prominence or authority associated with external labels for other media (e.g., sound recordings). According to AACR2, the chief source of information for an MRDF is the internal user-header label (an option available on standard labeled magnetic tape reels). Lacking this label, the chief source of information for an MRDF is the accompanying documentation generated by the creator, producer, etc., of the file. Documentation is a generic term covering a wide range of descriptive items, such as a data dictionary, tape layout, codebook, and user's guide. Both an internal user-header label and documentation external to an MRDF are discussed in the manual, and selected types of documentation are provided as examples.

With any medium, the quality of cataloging depends on the so-called authority or prominence of the source from which bibliographic
Integration of Catalog Records

information can be obtained. In the past, very little attention has been
given to the importance of providing complete bibliographic informa-
tion to an MRDF or its external documentation. Many of these external
descriptive sources relating to the content and organization of a particu-
lar file have little or no file-specific bibliographic information which
could provide some authority for the cataloger. Without a standardized
title page, the number of useful descriptive elements varies from file to
file. In addition, certain numeric data files have no titles at all, while
others may have two or three loosely associated titles. Before satisfactory
cataloging efforts for MRDF can take place, some external controls
must be exercised over the information describing this new medium,
and guidelines establishing proper bibliographic conventions must be
outlined. To address this problem, chapter three of the manual provides
guidelines on how to create a descriptive title, title page, bibliographic
citation, and data abstract.

Additional levels of recordkeeping are required to maintain MRDF
in any library collection. Some of the MRDF recordkeeping practices
currently in operation by libraries and information centers are reviewed,
and suggested integrated levels of recordkeeping for MRDF are
outlined.

Because the new cataloging rules for MRDF have not been tested on
a large scale, it was necessary to match the rules with specific examples.
Extensive applications of the rules were tested on a variety of files and
programs and the results were documented and explained throughout
the text. Specialists at the Library of Congress were consulted on rule
interpretations, and LC policy interpretations as they relate to MRDF
have been noted.

The first draft of the manual was reviewed in late December 1980
and early 1981. It is expected that the final version of the manual will be

MARC Format for Machine-Readable Data Files

On June 1, 1979, the Library of Congress Network Development
Office, in cooperation with the LC Automated Systems Office,
announced that it would begin work on compiling a MARC format for
machine-readable data files. The project was under the direction of
Lenore Maruyama, and to assist her in this effort, an advisory committee
of individuals who were actively involved with standards for biblioi-
ographic control of MRDF was established. The mandate of the commit-
tee was to provide input and advice on the elements to be included in the
format, review the drafts and comments from other organizations or key
individuals on the recommended format, and make recommendations on how the completed format should be updated and maintained.

The final draft of MRDF/MARC format has been designed to incorporate multiple levels of information. The data elements included in the format reflect a broad interpretation of informational needs beyond the traditional catalog record. This concept is outlined by Maruyama in the introduction to the *Machine-Readable Data Files: A MARC Format*:

The MRDF format has been designed to accommodate the data elements specified in the second edition of the *Anglo-American Cataloguing Rules* (AACR2), but the data elements included in the format have not been limited to those described in AACR2. Also, the explicit identification (or content designation) of these elements has been designed to accommodate a variety of products, e.g., a data inventory...[a sales catalog,] a union catalog, in addition to a catalog record [in the form of a printed card].

**ICPSR's Automated Cataloging System for MRDF**

Also in 1979, the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan received a grant from the National Endowment for the Humanities (NEH) to create a multipurpose automated cataloging system for its current holdings. The entire project is intended to be a pilot project which will provide a full-scale test of the cataloging rules for MRDF as prescribed in chapter nine of AACR2, identifying any needed modifications and revisions both in the cataloging code and in the newly formulated MRDF/MARC format. The ICPSR system will implement many of the data elements provided in the format and be operational in an interactive mode via the Michigan Terminal System (MTS) version of the Stanford Public Information Retrieval System (SPIRES). The automated cataloging system will act as a resource data base for information on thousands of available data files relevant to the social sciences. Included among the products that will be available from the system are detailed data abstracts. These data abstracts will be compiled in the consortium's annual *Guide to Resources and Services*.

**Cataloging-in-Source**

Another important information service to be derived from the ICPSR system is "cataloging-in-source" (also known as "cataloging-during-production"). Although modeled after the Cataloging in Publication (CIP) scheme, this MRDF effort operates outside the jurisdictional directives of the Library of Congress. The importance of
Integration of Catalog Records

the scheme is that it allows major data producers to provide cataloging information at the early stages of a file's development. The cataloging takes place either at the site of the MRDF production by an in-house librarian or by a professional cataloger at another location. The final results are printed on the verso of the title page of the file's documentation.

The first implementation of "cataloging-in-source" was carried out in 1978 by the National Opinion Research Center (NORC) under the direction of Patrick Bova, NORC's data librarian. The NORC effort included not only the proper cataloging for the file itself, but the cataloging for the file's printed documentation as well. Also included on the verso was the proper bibliographic citation for the file, to which a user may refer when citing the MRDF in the literature.

Since 1978, two other major data producers have implemented the "cataloging-in-source" scheme—the Bureau of the Census and ICPSR. An example of ICPSR's cataloging copy on the verso of the title page of the file's documentation is given in figure 1. The MRDF "catalog-in-source" is important because such information is more likely to be accurate, and it assures that the original issue of an MRDF is cataloged.

Informational Needs of MRDF Users

The design of any information system should take into consideration the needs of its potential users. Improved access to information on the existence and availability of MRDF has been at the forefront of all efforts to bring bibliographic control to MRDF. The value of cataloging is ultimately proved not by how well each MRDF is uniquely defined, but by how efficiently the user is directed to the resource he needs. What follows is an examination of the informational needs of social science users and the resulting data elements that have been included in the MRDF/MARC format.

At the Conference on Cataloging and Information Services for Machine-Readable Data Files held at Airlie House, Warrenton, Virginia, in March 1978, a special session was devoted to MRDF user needs with respect to the creation of a national information system for computerized files and programs. The session was a combination of creating "wish lists" and reacting to already existing "catalogs" for MRDF. While the user input group agreed that a more descriptive emphasis than is usually inherent in a traditional catalog entry was essential to users, there was substantial difference of opinion as to how extensive the
Janda, Kenneth


1 data file (158 logical records) + codebook (268 p.)

Summary: Data on the characteristics of 158 political parties operating in 53 nations from 1950 through 1962.
1. Political parties. I. Title.

Printed codebook only

Janda, Kenneth


268 p., 23 cm.

This codebook is to be used in conjunction with the machine-readable data file by the same title.
1. Political parties. I. Title.

BIBLIOGRAPHIC CITATION

All manuscripts using this data file and/or codebook should contain the following citation:

Janda, Kenneth. Comparative Political Parties Data, 1950-1962 [machine-readable data file]. Principal investigator, Kenneth Janda; [generated as part of] the International Comparative Political Parties Project, Northwestern University. - Ann Arbor, Mich. : Inter-university Consortium for Political and Social Research. 1 data file (158 logical records) and codebook (268 p.).


Copyright © 1980, The University of Michigan, all rights reserved.

Printed in the United States of America.

Copyright restrictions do not apply to member institutions of the ICPSR. All or part of this codebook may be reproduced for use at member institutions with appropriate citation to the principal investigator and the ICPSR.

Fig. 1. Example of Cataloging-During-Production for MRDF
Integration of Catalog Records

descriptive information in a national information system should be.\textsuperscript{16} One contributor to this conference explained:

Users feel strongly that there must be an emphasis on the data summary or descriptive phase of the presentation of information on data files. The most important section of the cataloging card is the abstract. The abstract should be as detailed as possible and give as much information on the machine-readable data file as is consistent with the limits of the catalog entry. As users we would be content with many fewer details than are suggested by the AACR II cataloging rules and would prefer more extended comment in the files themselves. A simple and related point is the need to have some identification of the genesis of the file and its history. In this way it would be possible to link slightly modified files and to recognize when a data set is similar to one that is already in hand. Additionally, some key-word structure would yield a great deal of information on the data files themselves. Finally, we believe it is useful and important to link data files and software, in those situations in which a particular software has been created to manage and/or operate with a particular file.\textsuperscript{17}

Another contributor expressed user’s needs in this way:

User requirements center upon data element retrieval through definitive data file description and data base documentation. Comprehensively, this means users require:

- Knowledge of the existence of data.
- Knowledge of the source of data.
- Knowledge of the applicability of data to solving specific problems or analytic needs.

These expressions of user’s needs and many others were taken into consideration when formulating the MRDF/MARC format. Examination of Machine-Readable Data Files: A MARC Format will indicate data elements beyond those required for describing a monograph or a serial and include those needed to depict the special characteristics of this medium and the particular needs of MRDF users.

Conceptually, data elements for MRDF can be broken down into at least six levels: (1) those needed to identify MRDF (e.g., bibliographic elements); (2) those needed to describe the contents of MRDF (e.g., descriptive summary, data abstract, or in-depth subject analysis per item or variable); (3) those needed to classify MRDF (e.g., appropriate classification codes, indexing or subject descriptors necessary to group like data files together); (4) those needed to access MRDF (e.g., physical characteristics such as recording density and computer/software compatibility); (5) those needed to analyze or use MRDF (e.g., citation of documentation and related reports, how/when the data were collected, unit of analysis, sampling procedures); and (6) those needed to archive
or maintain MRDF (e.g., in-house records pertaining to the processing, storage and use of the data file).

Selected data elements for these six levels as represented in the MRDF/MARC format are given below: 19

Level 1: Bibliographic Identity
—corporate or personal author (e.g., principal investigator, program director, etc.)
—title, subtitle, and other title information (i.e., statement of responsibility)
—general material designation (i.e., machine-readable data file)
—edition, plus appropriate statements of responsibility relating to edition
—production statement, including place, organization and date of production
—distributor statement (if appropriate), including place, organization and date of distribution
—size of file (including number of files, number of logical records, and statement indicating the presence of accompanying documentation
—series statement (title and numbering within series, if appropriate)
—notes
—unique identification numbers

Level 2: Data Abstract
—unique identification number
—type of file (numeric, text, computer programs, etc.)
—bibliographic citation of MRDF
—methodology (universe, sampling, unit of analysis, etc.)
—geographic coverage
—time period (chronological coverage of MRDF)
—date(s) of data collection (if unique from other dates)
—summary (subject matter description)
—derived source of data (if derived from printed sources or other MRDF)
—file size (number of observations, cases, variables, and any special file characteristics)
—bibliographic citation of accompanying documentation
—primary publications based on the use of the MRDF
—terms of availability
—contact person

Level 3: Classification
—Library of Congress Classification Number
—Dewey Decimal Classification Code
Integration of Catalog Records

- LC Geographic Classification Code
- Subject category codes (applied locally)
- Descriptors or index terms (applied locally)
- Geographic headings (applied locally)

Level 4: Access of MRDF (technical information)
- mode of access
- type of data carrier or storage medium
- recording density, blocking factors, etc.
- computer compatibility
- software compatibility
- peripheral requirements
- special formats or system files

Level 5: Analysis or Use of MRDF
- file structure/sort sequence
- condition of data
- restrictions on use
- intended audience or level of expertise
- applications of the file or program
- linkage with other files or programs
- unit of analysis
- sampling procedures
- citation and location of documentation

Level 6: Archiving or Maintaining MRDF
- archival study number
- personal or organizational donor of MRDF
- date received
- date processed and entered into collection
- retention status (if temporary)
- access code (publicly available, restricted, etc.)
- cost for file duplication/dissemination
- frequency of updates or additions
- holdings note (for serials or serial-like MRDF)
- processing history (changes, revisions, modifications, etc.)
- documentation number or shelf location

These data elements are not meant to be all inclusive, rather they are provided here to demonstrate the feasibility of an integrated approach to MRDF descriptive information.

Local applications of the MRDF/MARC format include the generation of several distinct products from one record, including a bibliographic citation, catalog entry, and data abstract. Examples of these
products along with associated content designations, are given in the appendixes to this paper. (At the time of this writing, no bibliographic utility had incorporated the MRDF/MARC format into its system, but two local applications of the format have been realized—at the Social Science Data Library, University of North Carolina and at the Inter-University Consortium for Political and Social Research, University of Michigan.)

Where Do We Go From Here?

There is no doubt that machine-readable data will play an even greater role in research and development programs of the future. More and more data needed for government and private research will appear in computerized form. Researchers and scholars should not have to spend additional time and dollars locating and acquiring appropriate MRDF. This is a function that can best be provided by a bibliographic utility. Such a utility already has the expertise in online network access and the data-base management programs for MARC-formatted files to offer the following products and services.

Shared cataloging. The machine-readable version of the 1970 censuses has undoubtedly been cataloged and described hundreds of times at as many libraries and data centers. The process of shared cataloging reduces this work to a one-time effort. Participants in a bibliographic utility system could benefit from the work performed by others.

Authority control. Access to authorized forms of author, uniform titles, author/title series and title subject headings used in bibliographic records would be provided by the utility. The primary purpose of an authority file is to accomplish the collocation function of the catalog, that is, to enable the catalog to relate and display together works by the same author, on the same subject, and in various editions regardless of the media.

Acquisition system. A bibliographic data base maintained by a utility can serve many purposes, including providing sufficient information for ordering available MRDF. Centralized access to such information would greatly reduce the time and effort required to locate and purchase MRDF needed by researchers. Some utilities even provide recordkeeping services related to the order process, including accounting functions.

Private file creation. Utilities may offer each participant the capability to create his or her own file of copied (derived) and original records. Such a file cannot be altered in any way by other participants.
Integration of Catalog Records

This would allow participants to have interactive access to files representing their own unique holdings and associated local recordkeeping.

Union list. With several libraries and data centers contributing cataloging information on their uniquely held data files and programs, a union list of MRDF could be established. The union list would operate as a centralized inventory of data resources (who has what and where) across the country. Participants compiling the union list would be registering new acquisitions on an ongoing basis, thus providing a constantly updated and comprehensive list of unique MRDF, including a list of libraries and agencies which maintain these files.

Products. Derived products from an MRDF bibliographic data base maintained by a utility might include catalog records, book catalogs (with multiple indexes), data abstracts, distributor lists, orders in process, new acquisitions lists, union lists, special subject bibliographies, and local inventories. Most products can be provided in varying formats, including printed form, microform and machine-readable.

With the cataloging code (AACR2) available for MRDF, with the appearance of a working manual to help catalogers interpret these rules, and with the data elements and content designators defined in the MRDF/MARC format, the “blueprint” is at last in place for the next step—the integration of MRDF records into any of the existing bibliographic utilities. The benefits of utilizing existing bibliographic utilities to provide information on available MRDF is evident. However, other problems must be addressed before such a step can be implemented.

Problems Related to This Effort

Compiling an Expanded Record for MRDF. The question has been raised whether catalogers of MRDF can be persuaded to provide information beyond the briefest bibliographic record—especially since the number of characters required to compile an expanded bibliographic record for MRDF has been approximated at 2500 characters. There are several reasons offered here as to why catalogers should be persuaded to create such a record. First, catalogers of printed materials must deal with a large volume of works, and there is usually a backlog of works to be cataloged. By comparison, the volume of MRDF to be cataloged will be low. With a low volume of input, the cataloger should theoretically have more time to compile an expanded record. Without such a record, the identified needs of MRDF users will not be met. Second, catalogers of printed materials normally are not required to look beyond the title.
page for cataloging information. However, catalogers of MRDF will of necessity be required to examine documentation beyond the title page to extract information not only to identify a particular MRDF but also to provide information on its nature and use. It is predicted that catalogers will find it necessary to provide more information in the note area of the bibliographic entry than they would for other media. The result will be that much of the information that goes into compiling an MRDF catalog entry may also be used to prepare a data abstract. In practical terms, there appears to be no reason why these same data elements should have to be compiled twice. Third, with automated bibliographic systems, we are no longer bound to the three-by-five card mentality, nor to the concept of meeting only one informational need. By thinking in terms of multiple applications of one system, the shared benefits go up and the cost of duplication goes down. The intended design of the MRDF bibliographic record as described here is to serve as an “organic record” from which several products can be derived without duplication of effort.

_Lack of expertise._ Before there can be widespread cataloging of MRDF, participating catalogers must be given the opportunity to become more familiar with this technical medium. Workshops and training sessions must be developed and offered as needed.

_Lack of professionalization._ At the present time there is no professional group within the library profession to speak to the needs of MRDF catalogers nor to be a vocal group for changes related to the cataloging code or the MRDF/MARC format. IASSIST (through its Classification Action Group) is the only visible group currently addressing these needs, but it has a limited voice in the library world. An organization similar to the International Association of Music Libraries should be organized for data librarians. This librarian group could also represent the needs of the user, publicize problems and promote sharing.

_Inspired participation._ Several research libraries are already cataloging MRDF (including Yale, Princeton, UCLA, and the University of British Columbia), but other libraries and centers maintaining MRDF must be persuaded to participate in the effort to compile bibliographic records for their unique holdings. Such participation is crucial to the goal of a union list for MRDF. Also crucial to the effort for social science researchers is the participation of the federal data collecting agencies and the support of the Office of Federal Statistical Policy and Standards (OFSPS).
Integration of Catalog Records

Conclusion

The cost of computers and communications technology is declining steadily. With smaller computers and better software becoming more readily available, and with scholars familiarizing themselves with these tools and their applications, a new dimension to the information explosion is now apparent; and with it comes an increasing demand for access to more and better-documented data files. Before a data file can have value, however, it must first be communicated to the potential user. Communicating the availability of usable data is an inseparable part of research and an integral part of librarianship. In the near future, libraries will have no choice but to become more involved with computerized files and programs. The nature of this involvement might well depend on demonstrated need, creative planning and available resources; and while it is not yet feasible to expect libraries to provide a full range of services related to MRDF, they are prepared to provide better access to information on the availability of data files. Taeuber sums it up this way: “While it is difficult to single out one function which is more important than any other, if libraries participated in the data revolution in no other way, preparation of a union list of data resources would be a major contribution to research. This could be a first step in increased library participation while training for the technical functions proceeds.”

ACKNOWLEDGMENT

The author would like to thank Ann S. Gray, research assistant in the Social Science Data Library at the University of North Carolina, for her work in preparing the examples contained in this paper.
Appendix A

MARC-Formatted Bibliographic Record for an MRDF

Appendix information taken from the Directory of Data Files prepared by the U.S. Bureau of the Census, Washington, D.C.

did ***** C-40  
stn  14-01-FOURTH-USCEN-70  
tit#ac Census of population and housing, 1970 housing summary statistic file # conducted by the Bureau of the Census  
edn  DUALabs ed.  
pro  Arlington, Va. * Data Use and Access Laboratories (DUALabs) #1972  
col#ae 3 data files (ca. 210000, 220000, 90000 logical records) #1 ccdebcck  
noq  This is a series of summary statistic files each containing detailed housing characteristics by geographic area based on the 1970 census sample questionnaires. Each file contains records which correspond to an individual geographic area.  
noq/2 Also known as: Fourth Count Ecensus Summary, 1970 Census of Population and Housing  
tof  Numeric (summary statistics)  
nox  This file, known as the "fourth count" has three individual files, each containing identical subject matter. Some of the variables included in the tables are total number of housing units, year structure built, tenure, race of head, gross rent or value of unit, persons per room, heating, air conditioning equipment, and presence of the following: clothes washer, clothes dryer, dishwasher, food freezer, television set, and battery operated radio. Separate tables with similar information are provided for the Spanish population.  
tim  Data contained in the files pertain to the date of the census, April 1, 1970, except for selected items which relate to historical periods.  
for  Data is 'compressed'. Use DUALabs' MOD-series program to process. Use DUALabs' DDLIST program to produce a listing of DUALabs' Data Descriptor List  
not  Title from: Directory of Data Files prepared by the Bureau of the Census  
chn  14-13  
acc  I  
uni  The universe consists of all housing units. The data are based on 5-, 15-, and 20-percent samples.
Integration of Catalog Records

Appendix A—Continued

sof Three files; File A contains approximately 210,000 logical records representing about 35,000 tracts for the U.S. File B contains approximately 228,000 logical records representing 37,500 MCL's for the U.S. File C contains approximately 90,000 logical records representing about 13,000 summary areas for the U.S.

qec The three housing files in fourth count have different levels of geography. File A presents housing summary statistics for all census tracts. File B presents housing summary statistics by minor civil divisions (or census county divisions). File C presents housing summary statistics for States, counties, places of 2,500 or more, SMSA's and component parts of SMSA's, urban/rural non-farm, and rural farm components.

ref#bef 1970 Census of Population and Housing Fourth Count Housing Summary Tape (Sample) Arlington, Va.: Data Use and Access Laboratories (DUALabs), 1972. This is DUALabs' version of the Bureau of the Census' documentation and technical guide for this file.


sul#ayx Housing—United States—Statistics
suh Census—Population and Housing Data
aec United States. Bureau of the Census
aec/2 Data Use and Access Laboratories (DUALabs), Arlington, Va.
tie Fourth count housing summary, 1970 census of population and housing
doe 81/04/09#ag
ced d-636 - d-638
Appendix B
Catalog Entry Derived from Bibliographic Record for MRDF

3 data files (ca. 210000, 228000, 90000 logical records) + 1 codebook.

This is a series of summary statistic files each containing detailed housing characteristics by geographic area based on the 1970 census sample questionnaires. Each file contains records which correspond to an individual geographic area.

Title from: Directory of Data Files prepared by the Bureau of the Census.
Also known as: Fourth Count Housing Summary, 1970 Census of Population and Housing.
Data is 'compressed'. Use DUALabs' MOD-series program to process. Use DUALabs' ELIST program to produce a listing of DUALabs' Data Descriptor List.

This is DUALabs' version of the Bureau of the Census' documentation and technical guide for this file.

Geographic coverage: The three housing files in fourth count have different levels of geography. File A presents housing summary statistics for all census tracts. File B presents housing summary statistics by minor civil divisions (or census county divisions). File C presents housing summary statistics for States, counties, places of 2,500 or more, SMSAs and component parts of SMSAs, urban/rural non-farm, and rural farm components.

Data contained in the files pertain to the date of the census, April 1, 1970, except for selected items which relate to historical periods.

Summary: This file, known as the "fourth count" has three individual files, each containing identical subject matter. Some of the variables included in the tables are total number of housing units, year structure built, tenure, race of head, gross rent or value of unit, persons per room, heating, air conditioning equipment, and presence of the following: clothes washer, clothes dryer, dishwasher, food freezer, television set, and battery operated radio. Separate tables with similar information are provided for the Spanish population.


C-40
Integration of Catalog Records

Appendix C

Data Abstract Derived from Bibliographic Record for MRDF

MRDF ABSTRACT

ID Number: DIE-C-40

Type of File: Numeric (Summary statistics)

Citation: Census of population and housing, 1970 housing summary statistic file 4 [machine-readable data file] conducted by the Bureau of the Census. DIALabs ed. Arlington, Va: Data Use and Access Laboratories (DIALabs), 1972.

Universe: The universe consists of all housing units. The data are based on 5-, 15-, and 20-percent samples.

Geographic Coverage: The three housing files in fourth count have different levels of geography. File A presents housing summary statistics for all census tracts. File E presents housing summary statistics by minor civil divisions (or census county divisions). File C presents housing summary statistics for States, counties, places of 2,500 or more, SMSA's and component parts of SMSA's, urban/rural non-farm, and rural farm components.

Time Period: Data contained in the files pertain to the date of the census, April 1, 1970, except for selected items which relate to historical periods.

Summary: This file, known as the "fourth count" has three individual files, each containing identical subject matter. Some of the variables included in the tables are total number of housing units, year structure built, tenure, race of head, gross rent or value of unit, persons per room, heating, air conditioning equipment, and presence of the following: clothes washer, clothes dryer, dishwasher, food freezer, television set, and battery operated radio. Separate tables with similar information are provided for the Spanish population.
Appendix C—Continued

**File Size:** Three files; file A contains approximately 210,000 logical records representing about 35,000 tracts for the U.S. File B contains approximately 228,000 logical records representing 37,500 MCD's for the U.S. File C contains approximately 90,000 logical records representing about 13,000 summary areas for the U.S..

**Special Formats:** Data is 'compressed'. Use DUALats' MOD-series program to process. Use DUALats' DDLIST program to produce a listing of DUALats' Data Descriptor List.

**Primary Reference:** 1970 Census of Population and Housing Fourth Count. Housing Summary Tape (Sample).

**Contact Person:** Judith Pcole, Research Consultant, North Carolina State Data Center, Institute for Research in Social Science, Manning Hall 26A, University of North Carolina, Chapel Hill, N.C. 27514 (919) 966-3546.
Integration of Catalog Records

References

Accessing Social Statistics

JOSEPH W. DUNCAN

Introduction

The Federal Statistical System collects, compiles and publishes a vast amount of statistical information on social characteristics and conditions of the population in the United States. These social data cover such areas as population; health and nutrition; housing and environment; transportation; public safety; education and training; work; social security and welfare; income and productivity; social participation; and culture, leisure and use of time. These data are frequently underutilized because there is a lack of adequate information about these data sources and access to existing information is sometimes difficult.

This article is intended to assist users of federal social data in accessing those data. A general overview of the federal statistical system and the central coordinating office is presented, along with a discussion of its data access policy. Finally, some selected source documents which aid in accessing social statistics produced by the federal statistical agencies are listed.

Overview of the Federal Statistical System

The statistical system of the U.S. government is decentralized, with responsibility and authority for statistical activities divided by subject matter among the agencies. These organizational arrangements for
producing federal statistics include several statistical collection and analytical agencies, many statistical units, and the statistical activities and outputs of major program agencies. Currently, there are over ninety federal agencies authorized to collect, tabulate and disseminate statistical data.

The origin of federal data collection activity, and in particular social statistics, can be traced back to the Constitution of the United States, which required an enumeration of the population within three years after the first meeting of the Congress and every ten years thereafter. Consequently, the first census of population was conducted in 1790. Subsequent legislation requiring the collection of statistics resulted in the formation of new statistical units in the federal government. This division of responsibility for statistical activities necessitates a central agency with responsibility and authority for providing general policy guidance on the development of an integrated statistical system to meet the needs of the federal government policy-makers and other users of federal statistics. The most recent in a series of efforts to coordinate federal statistical activities is the establishment of the Statistical Policy Branch in the Office of Information and Regulatory Affairs (OIRA) in the Office of Management and Budget (OMB).

The statistical policy function was transferred to OMB effective April 1, 1981, under the requirements of the Paperwork Reduction Act of 1980 (P.L. 96-511). Section 3(a) of that act requires the president and the director of OMB to delegate to the administrator of OIRA all their functions, authority and responsibility for statistical policy and coordination under Section 103 of the Budget and Accounting Procedures Act of 1950. These include the responsibilities: "to develop programs and to issue regulations and orders for the improved gathering, compiling, analyzing, publishing, and disseminating of statistical information for any purpose by the various agencies in the executive branch of the Government. Such regulations and orders shall be adhered to by such agencies." In addition to exercising this general statistical policy authority, Section 3504(d) of the Paperwork Reduction Act specifies that the statistical policy and coordination functions are to include:

1. developing long range plans for the improved performance of Federal statistical activities and programs;
2. coordinating, through the review of budget proposals and as otherwise provided in this section, the functions of the Federal Government with respect to gathering, interpreting, and disseminating statistics and statistical information;
3. developing and implementing Government-wide policies, principles, standards, and guidelines concerning statistical col-
Accessing Social Statistics

lection procedures and methods, statistical data classifications, and statistical information presentation and dissemination; and

(4) evaluating statistical program performance and agency compliance with Government-wide policies, principles, standards, and guidelines.

Thus, the government-wide statistical policy function extends to all social and economic statistics and includes activities such as: (1) the planning and coordination of statistical programs, agencies and issues across all departments and all subject areas; (2) the review of statistical forms and reporting plans; (3) the issuance of statistical standards and guidelines for ensuring the quality, comparability, timeliness, and accuracy of federal data; and (4) international coordination between U.S. government agencies and international organizations on statistical matters.

Data Access Policy

Recently the central coordinating unit has assumed an active role in facilitating access to federal data. As part of its effort to develop a coordinated approach to the general problem of data access, an Interagency Committee on Data Access and Use was established in April 1980. "Data access" is a broad term used by agencies to refer to publication policies and practices; development of general informational and reference materials; handling of general, public inquiries; servicing technically sophisticated data users; providing machine-readable data files (MRDF); developing user training programs; fielding user surveys; and so forth. The data access policy seeks to assist users in finding statistical information in a timely fashion for whatever the intended purpose. Users frequently have had short time horizons for locating the data, and they are unaware of potentially useful data. Their problems are complicated further by the lack of either a common format for data or a translator for the technical specifications on the data. Further, there is a shortage of resources for data analysis, and frequently special-purpose data are not always suitable for general applications. The Interagency Committee on Data Access and Use was established to address these common complaints and problems associated with using federal data.

Recommended Good Practices

The committee has developed the following selected recommended good practices for data access, which agencies are encouraged to observe
and which provide general standards against which agencies can measure their performance.

1. Federal statistical agencies should designate a specific organizational unit...[and personnel, with responsibility for improving data access in that agency].

2. All federal statistical agencies and agencies with major statistical programs should publish brochures and flyers which describe the statistics they collect and publish, and how to access the statistics; these descriptive materials should be issued and updated in a timely manner.

3. All federal statistical agencies and agencies with major statistical programs should prepare and publish reference materials which provide the user with clearly written technical guidance for using the agencies' data. Reference materials should also be regularly updated.

4. Agencies which serve sophisticated user communities should establish newsletters if the size and importance of their user community warrants it.

5. Agencies should devote special attention to the problem of adequate documentation for data files which are not available for public use but to which access is provided on some restricted basis.

6. Agencies should ensure that their programs for dissemination of information about data sources and contacts are systematically thought out and coordinated.

7. Statistical agencies are strongly encouraged to develop catalogs and directories which provide access information on major topics irrespective of which agencies hold data on the topics.

8. All federal statistical agencies should have a public inquiry service which answers queries from the general public. Telephone numbers for public inquiry should be published in agency brochures.

9. Statistical programs in agencies where statistics are embedded in a larger organization [which] does not have an inquiry service of its own should examine the larger [agency] inquiry service to ensure that queries of a statistical nature are well handled.

10. Agencies should study the utility to management of routinely logging public inquiries or at least spot-checking them periodically.

11. Agencies should sensitize all staff to proper attitudes in the handling of public inquiries.

12. Agencies should seek to promote the full range of their data products or services which have general utility....
Accessing Social Statistics

13. Statistical agencies and statistical programs must devote special attention to the problem of adequate abstracting and technical documentation for public use machine-readable data files. In particular, each agency should establish within-agency standards for technical documentation, and ensure that the standards are enforced.5

The committee suggests that federal statistical agencies implement these recommended good practices as a positive step toward making statistical data more available to the public in machine-readable as well as other forms, and toward providing greater services to users of social data as well as other data.

Directory of Federal Statistical Data Files

The latest development in aids to accessing data is the issuance in March 1981 of A Directory of Federal Statistical Data Files. The Directory was a joint effort of two agencies within the Department of Commerce, the National Technical Information Service and the Office of Federal Statistical Policy and Standards. The latter agency was the predecessor of the Statistical Policy Branch of OMB. The statistical input to the Directory was overseen by the Interagency Committee on Data Access and Use. It was developed to assist users in accessing machine-readable data, and comprises abstracts describing statistical and related files produced by the federal government.

Since the early 1970s, federal agencies have prepared catalogs and directories on their data files; however, the scope, content, quality, and format of the individual directories varied greatly. The Directory integrates these individual agency sources of information through a centralized source, and represents the first effort to implement a continuing program to describe agency data files. Hence, a more fully coordinated effort involving all federal agencies producing statistical files is now being undertaken by the Statistical Policy staff. Standards are being applied to the descriptive information for the files and the preparation of the abstracts contained in the Directory. Each abstract contains the following eleven information items: bibliographic citation, file reference, general description, geographic coverage, time coverage, technical characteristics, reference materials, related reports, related files, contacts, availability. A sample abstract entry can be found in the appendix to this article.

The Directory is organized into three sections: (1) scope and use of the Directory, (2) file abstracts (organized by agency), and (3) appendixes. Users of the Directory wishing to review all files available from a particular agency should first examine the table of contents to deter-
mine if files from that agency are contained in the Directory. The second step, for agencies with large file holdings, is to refer to its Appendix I, which is a listing of all file titles included in the Directory.

An alternative way to use the Directory is through the subject matter indexes. Two subject matter indexes are included in the appendixes. Appendix II contains an index arranged in alphabetical order by subject matter key-word or key-word phrase. The subject matter keywords contain descriptors for the types of data items contained in the file and related identifier. Appendix III contains basically the same set of keywords as contained in Appendix II, but is arranged in alphabetical order by agency by file. Appendix III permits the user to determine which key-word phrases were associated with which files.

The ultimate objective of the Directory is to include all federal statistical data files which are available to the public or may be processed by federal agencies at a user's request, providing data generated from the file would meet agency releasability standards. The Directory focuses on major statistical data files. Major files include all federal statistical files developed expressly for public dissemination, developed at significant cost, of significant value to the public in general, or of considerable importance due to the level of user demand and the use of the file. The preparation and updating of the Directory is a joint effort of federal statistical and information-producing and -disseminating agencies, with continuing coordination provided by the Statistical Policy Division of OMB.

**Effect of Technology on Data Use**

Two recent efforts underline the importance of newly emerging technology. These are the National Indicators System (NIS) and the Decision Information Display System (DIDS). Under the auspices of the White House Office of Planning and Evaluation, and coordinated by the Statistical Policy Branch, these developments in computer technology seek to make statistical information more useful.

The National Indicators System is a program of briefings designed for systematically informing the president, vice-president and senior White House staff on social, demographic and economic trends in the United States in a policy-relevant format. The objective is to assist the president in making responsible decisions by providing periodic briefings which give an objective background description of the conditions in America. The intention is to develop a communication system which draws on the enormous statistical resources of the federal government to
Accessing Social Statistics

describe national conditions with data that relate to the policies currently pending before the president, cabinet, cabinet councils, or that are anticipated to go before them. The system is designed to be an honest broker of information linked to policy planning, but it is not part of the regular policy advocacy process. The responsibility for developing the briefing materials is assumed by the participating federal agencies under the general coordination of the Statistical Policy Branch.

The Decision Information Display System was conceived as a technique for displaying statistical information on a geographic basis so that White House, executive branch and congressional staff could see the impact of issues across states, counties and other political jurisdictions in the United States. It is designed for rapid graphic presentation of social and economic variables about various geographic regions. DIDS is a cooperative, interagency-funded program for the application of information technology in the statistical community. DIDS currently has approximately 3500 data sets immediately accessible to meet demands for data in a short time frame. The data bases are accessible to the user in a menu-driven approach, so that by entering a simple numeric code from a list of data bases, information on a particular topic is available. Since the information display is oriented toward geographically defined information, DIDS forces a common format on the data bases.

Source Documents

As part of its coordinating effort, the Office of Management and Budget's Statistical Policy Branch prepares publications to inform users of developments affecting federal statistics and to aid them in accessing the data. A general overview of federal social statistics can be found in the following publications issued by the Statistical Policy Branch and its predecessor organizations:

Statistical Services of the United States Government, 1975. This document includes brief descriptions of principal economic and social statistics programs, as well as a list of the principal statistical publications prepared by the federal statistical agencies.

A Framework for Planning U.S. Federal Statistics for the 1980's. This publication presents a comprehensive review of all major statistical programs and data series, as well as the publications that are available as a result of these data collection activities.

Statistical Reporter. This monthly publication contains information on newly issued publications and machine-readable data file.
In addition to the above reports, *Social Indicators III*, prepared by the U.S. Bureau of Census, presents a comprehensive variety of statistical information on the current social situation in the United States, along with selected bibliographies. While these publications contain information on accessing social data from the federal statistical system in general, the following is a selected bibliography by agency for users interested in particular subject-matter areas.

**Bureau of the Census**

*Directory of Data Files.* This publication contains structured descriptions of public use data files and computer software available from the Census Bureau. Available on a subscription basis, the publication is updated regularly.

*Data Access Descriptions.* Issued irregularly, 1967 to present. Contains user-oriented, topical reports on accessing and using census data and products.


*Directory of Federal Statistics for Local Areas: Urban Update,* 1977-78, 1981. Similar to the directory immediately above, except that geographic focus is on cities, standard metropolitan statistical areas, labor market areas, etc.


*Reference Manual on Population and Housing Statistics from the Census Bureau,* March 1977. A comprehensive discussion of census and survey reports and computerized products, with attention to limitations as well as strengths of the various data and to skills required to find specific data.

*Factfinder for the Nation,* Series CFF, issued irregularly. A series of topical brochures describing the range of census materials available on a given subject. In particular, Factfinder No. 5, "Reference Sources," provides a bibliography of Census Bureau data access publications.

*Census '80 Introduction to Products and Services,* November 1979. A general introduction to the 1980 Census of Population and Housing data.

*Housing Data Resources: Indicators and Sources of Data for Analyzing Housing and Neighboring Conditions,* 1980.
Accessing Social Statistics

General Accounting Office
*Federal Information Sources and Systems 1980.* This publication describes federal sources and systems maintained by the executive agencies which contain fiscal, budgetary, and program-related information. The report is prepared as a part of the *Congressional Sourcebook Series* and is an update to a volume prepared in 1977. The report contains descriptions of many statistical files prepared by government agencies.

Bureau of Labor Statistics
*BLS Data Bank Files and Statistical Routines.* This publication generally describes public use data files and computer software produced and distributed by the agency. Prepared in 1978, the report is now out of print. Updates are now being prepared and a revised report may be released by BLS.

National Archives and Records Service
*Catalog of Machine-Readable Records of the National Archives of the United States.* This catalog is a second edition publication which describes machine-readable data files accessioned by the National Archives. A new catalog is now being prepared and is expected to be available in 1981. The publication will contain statistical and other types of files.

National Center for Education Statistics
*Computer Tapes Available from NCES.* This report contains brief summary descriptions of files which are available for public use through the National Center for Education Statistics. The report is updated frequently.

Directory of Federal Agency Education Data Tapes. This publication contains structured descriptions of data files produced by federal agencies which contain education-related data. The report was prepared in 1979.

National Center for Health Statistics
*Catalog of Public Use Data Tapes from the National Center for Health Statistics.* This report generally describes NCHS statistical programs from which data are produced in machine-readable form, and summarizes the content and structure of public use files prepared. The report is updated annually.

*Facts at Your Fingertips: A Guide to Sources of Statistical Information on Major Health Topics,* 4th ed., fall 1979. This guide lists sources of...
statistical information on over 100 major health topics, and references NCHS data on particular topics and other public and private sources of information.

_NCHS Publications on the Health of Minorities_, May 1980. An annotated bibliography of selected NCHS reports which present data on health topics of special interest to women.

_NCHS Publication on the Health of the Elderly_, November 1980. An annotated bibliography of selected NCHS reports which present data on health topics concerning the elderly population.

_National Technical Information Service Directory of Computerized Data Files and Computer Software_. This report contains descriptions of data files and computer software distributed by the agency. Originally prepared in 1974, the publication was updated in 1976; it is now out of date.


_Social Security Administration Research Publications and Microdata Files: Spring 1980_. This report contains a brief summary description for individual microdata files produced by the SSA Office of Research and Statistics. Content, technical features, and availability are described. The catalog is updated frequently.

Lois A. Alexander and Thomas B. Jabine. "Access to Social Security Microdata Files for Research and Statistical Purposes." _Social Security Bulletin_, vol. 41, No. 8, August 1978. This article focuses on the characteristics of SSA microdata files and on the development of a disclosure policy aimed at serving the public interest while protecting the privacy of individuals and the confidentiality of research and statistical information.


"Some Statistical Research Resources Available at the Social Security Administration," April 1979. This brochure describes the nature and
Accessing Social Statistics

availability of data from SSA's Lifetime Earnings Records and Continuous Work History Sample.

Bureau of Justice Statistics
Inter-University Consortium for Political and Social Research. Guide to Resources and Services: 1979-80. Ann Arbor, Mich.: Institute for Social Research, 1980. This Guide provides information on all of the resources and services offered by ICPSR of which the Criminal Justice Archive and Information Network (CJAIN) is a part. These include data archiving, data access functions related to the archives, training in a variety of areas related to ICPSR, and computing assistance.6
Appendix

Sample Abstract From
A Directory of Federal Statistical Data Files


**FILE REFERENCE:** 101.1010/100180

**GENERAL DESCRIPTION:**
The _Census of Population and Housing, 1970 Summary Statistic File 1_ is a series of summary statistic files, each containing detailed characteristics of the U.S. population and housing by geographic area based upon the 1970 census completed count (100%) questionnaires. The files are also known as the 1970 Census First Count Summary tapes. Two types of files—A and B—were prepared separately for each state and include the same subject matter data but contain summary statistics for different levels and types of geography. Each file contains records which correspond to an individual geographic area and contain population data by age, race, sex, and family and household characteristics and housing data descriptive of units and their occupancy.

**GEOGRAPHIC COVERAGE:**
The geographic coverage differs for each of the file types. File A contains summary statistics for enumeration districts and block groups. File B contains summary statistics for counties, minor civil divisions (or census county divisions), places and congressional districts.

**TIME COVERAGE:**
Data contained in the files pertain to the date of the census—April 1, 1970.

**TECHNICAL CHARACTERISTICS:**
Files are packaged and distributed on a state separate basis. Due to the volume and complexity of the structure of these files, please consult the reference materials cited below for further information on size and technical characteristics.

**REFERENCE MATERIALS:**

**RELATED REPORTS:**
Accessing Social Statistics

RELATED FILES:
a. Other summary statistic files.
b. Public use sample files.

CONTACTS:
Chief, Customer Services Branch
Data User Services Division
U.S. Bureau of the Census
Washington, D.C. 20233
(301)449-1600

AVAILABILITY:
The files are available from the Bureau of the Census for $80 per output reel. File order numbers are CSUM70001a(File 1 A) and CSUM70001b(File 1 B).
References


5. Ibid., pp. 328-32+.

6. Ibid., pp. 341-44.
Models of Data Library Development and Information Systems Services: An Overview

MARGARET O'NEILL ADAMS

The three papers following this article describe variations for the provision of data library services, as they are instituted at the universities of British Columbia, Florida and Wisconsin. The fourth paper describes an online numeric information system and discusses how the reference department in the central library at the University of Kentucky has begun to incorporate services from this resource into its routine. One clear message of these papers is that there is no single administrative structure for services related to social science data files nor for numeric information systems that is ideally suited for all institutional settings. The diversity of academia will be reflected in the variety of facilities that provide data services, with the institutional framework for these services determined by local conditions.

The papers are linked by their underlying assumption that provision of service for machine-readable data files (MRDF) or from online numeric information systems is basically a library activity. Such an assumption has not been as obvious as it may seem, however, nor has it been shared universally throughout the library and information professions. Therefore, my own paper lays considerable emphasis on the links between provision of computer-based data resource services and the general evolution of library reference services. Ray Jones and Laine Ruus each show informatively how the central university libraries of their institutions, the universities of Florida and British Columbia, respectively, have assumed responsibility for these services, albeit within differing frameworks.

Margaret O'Neill Adams is Manager, Kentucky Economic Information System, Center for Applied Economic Research, University of Kentucky, Lexington.
In the case of the University of Wisconsin, the Data and Program Library Service (DPLS) described did not evolve from within the university library structure, nor has it since been incorporated by it. Nevertheless, its staff has maintained close ties with campus traditional librarians, the network of campus special libraries, and the program of the Library School at Wisconsin. As Alice Robbin points out, perhaps the most important library-related aspect of DPLS has been its mandate to provide its services to the widest range of university users, regardless of their status or departmental affiliation. It thus has always operated in the spirit of traditional university library services. DPLS can serve as a viable model for those institutions planning to provide data library services outside the institutional framework of the traditional university library.

It is rather unlikely, however, that this model will be adopted by many institutions during the 1980s. A facility like DPLS was a natural product of expansion such as occurred for the social sciences during the 1960s. In addition, the creation of a data library independent of an academic department, or of a service organization with a long-established tradition of university support, requires a large institutional setting, as well as a strong commitment to the sponsorship of interdisciplinary activities. Aided by considerable vision on the part of the social scientists who secured the necessary funding, all of these conditions came together at a propitious time, and DPLS was founded.

In a period of economic retrenchment such as is being experienced by most educational institutions today, new facilities rarely are established, regardless of their merits. One can argue that creation of interdisciplinary or "umbrella-like" facilities is more warranted during periods of economic stringency than when restraint is not so necessary. Yet the reality is that they do not receive the support they need because providing it means withdrawing support from some other well-established activity. Hence, the model that most colleges and universities will probably follow for at least the next several years, presuming that they have an interest in providing services related to computer-based data resources, is the type of facility described by Laine Ruus or the services of the reference department outlined by Ray Jones, or some variation of these.

Data library services that are integrated into general university library services, or supported jointly by university libraries and computing centers, have the distinct advantage, as Laine Ruus points out, of "deriving...primary funding from...the most stable and secure budgets in the academic environment." Budget-cutting for libraries and com-
Models

puting centers occurs when overall resources decline; it is just about inconceivable that any college or university, no matter its size, would ever eliminate library or computing center services altogether. Incorporation of computer-based resources into library and/or computing center services will undoubtedly require reallocation of some university resources; it need not necessarily imply significant new investment.

Budget considerations are not, however, the only criteria for determining the most appropriate framework for providing a particular type of service. It is equally relevant that provision of data services by university libraries, whether alone or in conjunction with the university computing center, validates them as services that are basic to the general teaching and research activities of the university community. As expressed by Ray Jones, "no significant format or method of accessing information can be excluded in the teaching and service program of a university library." Situated within the university library, data services do not become dependent upon "the grace and wisdom of the deans," to use Alice Robbin's words, nor are such services viewed, again in her words, "as expendable luxuries."

Another common theme in these papers is that there are several problems which all data libraries experience, however they are instituted. Perhaps the most important is the absence of any national or international union catalog for MRDF or for computer-based data resources, such as online numeric data-base systems. As a result there is no integration of information about MRDF into central library card or online catalogs. A related issue is the lack of any real bibliographic control for MRDF, although the standard bibliographic citation format recently adopted in the Anglo-American Cataloging Rules (2d ed.) is a major breakthrough toward the solution of this problem. It would seem that solutions to cataloging problems and related services can be found when the experience of catalogers and the structures of the existing traditional library network are utilized. This development is more likely if the traditional library system has a vested interest in the dissemination of information about MRDF and numeric data-base systems, which obviously occurs when the library offers services for computer-based data resources.

Several developments during the past decade are at least peripherally relevant to the issues discussed here. For example, while the nascent data library "movement" has only gradually evolved, and has some of the same basic problems at the beginning of the 1980s that it had a decade ago, there simultaneously has been a revolution in the library's provision of bibliographic information for published material. This
has occurred, of course, through the online searching of bibliographic data-base systems and the utilization of cataloging networks. An entire profession of “information scientists” has emerged, and there has been a proliferation in the literature related to “information science.” In some places this has complemented the work of traditional librarians; in others, it has supplanted it. Many traditionally trained librarians have become “online searchers”; fewer are now “data librarians.”

In fact, in the information science and online literature, terms such as “data-base systems” are used in contexts which suggest that the normative application of computer technology for libraries is in the areas of bibliography and catalogs. With a few notable exceptions, contemporary librarians and information scientists have seemed generally disinclined to explore the world of MRDF, although this is gradually changing. Computer technology has thus been used during the past decade or so primarily to make efficient the provision of traditional library services; provision of nontraditional (i.e., data library) services has not spread in the same way.

This is all perfectly understandable. The advent of online bibliographic data-base services and online cataloging in the traditional library in large part explains the slower growth of data library services within the traditional structure. Whether consciously or unconsciously, libraries clearly looked to the computer to solve some of the problems they had related to provision of traditional information service, before turning to this technology to provide nontraditional services. The absence of computer-based data resource services in many university libraries is thus not necessarily a sign of their resistance or inertia concerning these things, but rather a case of traditional services thus far having had priority.

The fact that there was no national organization to coordinate solutions to common MRDF problems following the termination of federal funding to the Council of Social Science Data Archives at the end of the 1960s undoubtedly also accounts for the absence of a union catalog for MRDF, and for the problems related to this. The formation during the late 1970s of the International Association for Social Science Information Service and Technology (IASSIST) again provides a forum for data producers, processors and users to coordinate their efforts at solving their common problems. However, future developments in the United States regarding the coordination of services for computer-based information resources will be affected by the policies of the federal government regarding information dissemination.

If the federal government views the provision of public information, such as statistical and survey data collected and processed at public
Models

expense, as one of its basic responsibilities, then one can also foresee renewed involvement by the federal government in supporting the search for solutions to the problems alluded to here. If, however, as it seems to be doing, the federal government abdicates its responsibilities for the provision of public information and turns this task over to the private sector, data librarians, information scientists and the community of public data users will find themselves at the mercy of the market. With such a scenario, the future for the availability of public data and for the solution of the problems faced by the providers of computer-based information services is bleak, for it is unlikely that the private sector will ever be interested in coordinating or providing the totality of information which is publicly produced. It is in the interest of all those professionally involved in public information servicing to guarantee that the federal government retains its traditional commitment to the provision of information as one of its most basic services. Otherwise, private companies will control the availability of data generated at public expense according to their own assessment of marketability.
The Data Library in the University of Florida Libraries

RAY JONES

A UNIVERSITY LIBRARY has three major goals: (1) to collect the most significant information resources; (2) to organize these resources; and (3) to service them for faculty, students and a number of "outside" publics. Machine-readable data files (MRDF) are now a growing segment of available information resources. The private sector, the public sector, individuals engaged in research, and many institutions are producing files for historical use, research and decision-making purposes. The academic library can play a major role in archiving and servicing these files. For the culture of print, an international infrastructure of producers, distributors and collectors has long been developed. National systems of inventory and description such as the Library of Congress and OCLC make the professional responsibilities of librarians much easier. For the hundreds of producers and very few collectors of MRDF, there is very little of this infrastructure available. Major federal, state and regional governments are producing time-series data and single surveys of importance in decision-making and research. The private sector is very advanced in these areas. Without a developing network of the creators, collectors and archivists of these data, chaos is inevitable.

In the area of data services, the lack of a network or infrastructure has created a number of critical issues. The overriding question concerns what groups shall take the responsibility of archiving and servicing MRDF. This writer believes that the large university library may be one of the appropriate institutions. He retains this opinion even when

Ray Jones is Data and Social Science Librarian, Department of Reference and Bibliography, University of Florida at Gainesville.
few major academic libraries are now participating in the field. Selecting, acquiring, cataloging and servicing MRDF parallel many of the librarian's professional responsibilities with printed and microform materials. The challenge of joining the traditional skills of librarianship and the new skills in computer technology is a compelling one. It is also of prime importance to both library administration and library staff. Ideally, librarians should know of the existence of all types of information regardless of format and method of access. This is the basic assumption of the University of Florida Libraries administration.

The operation and services of the Data Library in the University of Florida Libraries depend upon a functioning team from several major units within the library system. Actually there is no Data Library as such. Responsibilities for basic functions are decentralized; no single librarian or systems professional has overall authority. The central team consists of two reference librarians in the Department of Reference and Bibliography and a systems programmer from the Systems Group within the library. The major operations, services and day-to-day decision-making are handled by these three persons. A tape library or archive for machine-readable data has been established for several years under the basic control of the Systems Group. The latter have the responsibility for archiving the collection of computer tapes, data sets and software programs. The Social Sciences Reference Librarian and his departmental colleague are responsible for the general coordination of the public services.

This decentralized management environment is a product of the history of data services at the University of Florida Libraries. In 1971, the Department of Reference and the Systems Group formed a team to serve as a summary tape processing center for the 1970 Census of Population and Housing. \(^1\) At that time three reference librarians served as the interface for everyone requiring census data in machine-readable format. They held the interviews and did the coding in the appropriate software. The Systems Group was responsible for the quality control and submissions. Both Data Use and Access Laboratories (DUALabs) software and the major census counts were purchased and utilized.

Approximately two years ago, another layer of organization and services was “integrated” within the library system when the latter assumed responsibility for the Inter-University Consortium for Political and Social Research (ICPSR) data sets. The social sciences reference librarian is now responsible for services. The social sciences reference librarian became the ICPSR representative for the university and supervisor of a graduate student serving as part-time data manager. The consortium membership is paid by the universities of the State Univer-
The University of Florida

sity System (SUS). Together, this network in Florida is called the Florida Consortium for Political Research.

The library serves as the ICPSR archive within the state. The computing systems of the various members are dissimilar and require different technical formats for the tapes. Requests for Consortium data sets from each university ICPSR representative are sent to the reference department to provide a copy compatible with the school's particular computer system. The librarian involved is responsible for filling the request for the MRDF, either from the archive or through contact with the ICPSR in Ann Arbor. The part-time manager for the Florida Consortium, a doctoral candidate in political science, handles the recopying and mailing of the tapes. Codebooks are maintained in reference and can be sent through interlibrary loan when necessary. The librarian represents the University of Florida at yearly meetings of the Florida consortium. The Systems Group provides the blank tapes and handles all archiving. Difficult technical problems are referred to representatives of the Computing Center at the University of Florida. All reference services are the responsibility of the librarian.

If the census services involve a team within the library, the Consortium activities not only involve the same team but also financial and operational relationships with all state members, the University of Florida administration, its Computer Center and the Department of Political Science. Each participating university within the SUS pays a portion of the ICPSR fees. The University of Florida Libraries pays those for the University of Florida. Operating funds for archiving and systems operations are supplied by the executive vice-president of the university, the Computer Center, and the Department of Political Science. The salary of the data manager is paid for by the vice-president and the Political Science Department. Normal computing costs are met by the Computer Center, but the library pays for special programming required by changes to new systems such as Multiple Virtual Storage.

The census activities receive their funding completely from the library system, as do all the non-ICPSR services. Census data in machine-readable format and any necessary software are purchased through normal acquisitions channels and billed to a tape fund. Other data sets are purchased from the tape fund in the library or jointly utilizing both that fund and departmental library funds. The Systems Group bears the operational costs of the organization of the tape library, its maintenance, preservation, tape and file management, and all computer submissions. Just as fees are charged for online bibliographical searches made by the reference staff, fees are also charged for census output, tape copies and subsets of public data sets when done by refer-
ence and systems staff. Cost algorithms are determined by the library administration.

To coordinate this complex network of services, operations and funding at a policy-making level requires a decision-making body. Approximately two and one-half years ago, the library administration formed a Committee on Machine-Readable Data to coordinate all machine-readable functions and make policy decisions. Budgets, planning, acquisitions of data files, new and existing services—all are handled by the Committee on Machine-Readable Data. Its members are the chairs of reference, documents, acquisitions, the systems manager, his unit supervisor and the social sciences reference librarian. The chairman of the committee is the associate director for public services.

The administration of the library system has mandated that data services be provided to the faculty, students, government and “outside” public. The primary emphasis has been on the 1970 Census of Population and Housing in machine-readable format. The 1980 census will also become a central focus for data services. In addition, the program includes the organizing and servicing by the library of a collection of MRDF from public and private sectors and the purchase of specialized software. Moreover, the University of Florida Libraries now archive all MRDF purchased with library funds by any individual, department or group within the university. With the exception of census tape data services, the activities do not include programming, teaching of software systems, or any interpretation of statistical products. The organizing of codebooks, documentation, tape and file management, reference services, and teaching the substantive content of the files themselves are basic responsibilities of the library team.

The development of the collection of data sets and programs has centered about the Census Access Program and the needs and demands of faculty and research groups on and off the campus. Along with counts I-V for the United States, the microdata sets for the 1970 census for all states, the DUALabs software, and the census bureau's geographic base files for Florida, files especially important to demographic research were added. These include the Current Population Survey Annual Demographic File, Survey of Income and Education, and Florida Vital Statistics. The Bureau of Economic and Business Research at the university serves as the official unit for state demographic forecasting and monitors the Florida economy. Certain critical files in the economic area, such as County Business Patterns, are a necessity. Large public data files from the federal and state level are requested by both the bureau and the research centers at the university. The test tapes for the 1980 census have arrived, as well as CENSPAC, the software developed
by the census bureau. When available in 1981, the census tapes themselves will be acquired. The decision determining geographical coverage will be made by the library administration.

The College of Business has transferred many significant private and public data sets to the library. The MRDF for Compustat, Value Line, Federal Reserve System Report of Income and Report of Condition, CRSP Stock Files, and CRSP Government Bond Files are now paid for by library funds. Students and faculty gain access to both documentation and files through the Reference Department. The College of Business's Center for Econometrics recently used a large sum of their allocation for funds to purchase time-series data from the National Archives. The files include the Statistics of Income and U.S. Federal Outlays. Now that departmental library funds can be spent for MRDF, the many departments of the College of Business will add significantly to the collection over the next decade.

The library collection appears to serve as a focus of the need and demand for MRDF. Four departments asked that the National Longitudinal Surveys of Labor Market Experience be purchased. Currently, one dissertation in sociology and three research projects in economics and sociology are being based on these files. Prior to the transfer of ICPSR data sets to the library, the collection reflected the needs of the participating universities, especially their political science departments. There are now requests from the major social sciences for ICPSR sets in the areas of economics, sociology, statistics, gerontology, and anthropology. Within the past year there has been a 30 percent increase in requests for Consortium data from the University of Florida faculty and faculties in the state membership. With the wide variety of ICPSR offerings relevant to many fields, the collection should become broader and reach a larger group of users. The history department is showing interest in the large historical data sets available through the Consortium. These are being acquired slowly. It is hoped that a system of communication with the faculty can be evolved to select from new ICPSR offerings those most relevant to the research needs of many groups.

One of the largest users of MRDF is the Food and Resource Economics Department in Agriculture. The 1978 Florida Census of Agriculture, state-generated data files for municipal and community outlays, and the Food and Agriculture Organization (FAO) Production and Trade Statistics have been acquired at its request. In addition to the Florida Vital Statistics, three current surveys of Florida's socioeconomic development are being archived. An archive collection of significant state data is a collection goal of the library.
A very recent development is the acquisition from the Human Relations Area Files of its specialized software, HRAFLIB, and three of its data sets. The chair of the Anthropology department is most anxious to introduce the faculty and students into the area of MRDF, and especially into cross-cultural hypothesis testing. He has also requested that the library obtain a subscription to the Roper Polls. A special representative has been appointed from the department to develop interest in these files. This department has organized a data laboratory and will teach a graduate class using machine-readable files in January 1982. Anthropology and every major department from psychiatry to the Center for Wetlands eagerly await the 1980 Census of Population and Housing.

The development of technical services for MRDF has not progressed as rapidly as the collection or its use. The problem arises because a very limited staff acquire, process and catalog the data sets. They each have many other responsibilities in the reference and systems areas. Basically these responsibilities are handled by the social sciences reference librarian, the systems programmer, and the part-time Consortium data manager. At times other staff members are asked to help process codebooks and documentation, and write an abstract for a special file. None of the data files nor the documentation is recorded in the card catalog, or in OCLC.

After the Committee on Machine-Readable Data has made the decision to purchase and has allocated funds, the social sciences reference librarian places the order for the data set, including the technical description of the tapes, through the acquisitions department. One copy of the tape and two sets of the documentation or codebooks are always ordered. When received, these are sent to the tape library. A back-up tape is made, and a copy of the documentation is retained for the tape library. When ready for public use, the second copy of the documentation is returned to the social sciences reference librarian to be placed on a special shelf of codebooks in the reference department. He then does the basic cataloging and writes an abstract. Copies of the abstract are placed in a loose-leaf notebook at ready reference and on the special codebook shelf. Codebooks are also stripped for the electronic security system. These procedures apply to census, public and private sector data sets. The reference copy of codebooks is circulated.

A different set of procedures is used for the ICPSR machine-readable files. As yet, the latter are not thoroughly integrated with the data sets primarily discussed. A backup tape is made by the library systems group and stored with the user tape. Both a backup codebook and a circulating one are placed in the reference department. All ICPSR
The University of Florida

documentation is placed in file drawers in the department next to the special shelf containing all other circulating documentation. Cataloging is at an elementary stage. Two years ago, all ICPSR tapes were held at the Computer Center, and a title card file and typed list of titles were maintained at the Political Science Data Laboratory. All tapes and the card file have been transferred to the reference department. The card file is maintained by the graduate student serving as the Florida Consortium data manager. A card which includes title, ICPSR member, chief investigators, and basic technical information is made for each data set. The latest ICPSR Guide to Resources and Services is used as a basic entry into these files. A basic list of study holdings has been printed using the Apple computer with plans now in progress to develop an online system. Every piece of information necessary for internal and patron use will be included.

Sufficient public services and ease of access to data sets must be provided so that faculty, students and others look to the library whenever MRDF are required. As stated before, these services began with the Census Access Program. The latter involved basic data reference services as well as the teaching of both small and large user groups. In 1972-73, a number of seminars were presented for faculty, students, government workers, and those in business. These included a presentation of census file content and the use of DUALabs software. Requests for both information and computer runs have been handled by face-to-face interviews with campus users. Long-distance telephone is the media used by business and government agencies in requesting data. To meet the requests, the librarians have had to learn the contents and structures of the census files, including the microdata ones. A thorough knowledge of the use of DUALabs software and some basic knowledge of SPSS has also been a requirement. Currently very little instruction is given to any group concerning the use of the census software. Graduate planning classes in urban information and quantification have been given three-hour seminars in which a retrieval problem is assigned. Every class in the social sciences requesting an orientation to library services receives a short introduction into the census, ICPSR and other data sets in the collection.

Laine Ruus, Alice Robbin and Lucinda Conger have described several types and levels of user services in a data library. Those at the University of Florida can be described as basic and conservative. The reference concept upon which the services are predicated makes no distinction as to the format or source of the information. The hope is to join the user with the appropriate information, whether from a printed source located through an index or from a data set. To determine if a
data set is held that will meet the needs of faculty or students is difficult. Giving basic information about the availability of a specifically requested file and its accessibility is relatively easy. Data services at the University of Florida emphasize the subject contents of files. With the census activities as the exception, none of the staff involved offer any programming or statistical services to the user. These problems are referred to the appropriate campus facilities.

As with many reference activities, user demands always expand the services. Having access to one of the few tape archives in the state available to many groups, the library MRDF team is actually required to provide clearinghouse services. Users throughout the state and on campus wish to know if certain statistical data exist in machine-readable form. Many times they are sure that these data must be available from "somewhere" or "someone." The training provided by the International Association for Social Science Information Service and Technology (IASSIST) and the Association of Public Data Users (APDU) sessions, catalogs of existing files, and contacts with federal, state and professional organizations are invaluable. In the current state of development for data services, the "old girl and boy" network is absolutely essential. While the user services described may be meager when compared with those hoped for in the ideal state, they are a beginning and have required considerable training, time, effort, and funds. Basically they are also considered a significant part of the general reference services offered at the University of Florida Libraries.

In March 1981, the University of Florida Libraries, the Computer Center and the U.S. Bureau of the Census jointly presented a one-day basic workshop on the 1980 Census. One hundred thirty people attended from throughout the state. There is the possibility that an intermediate to advanced workshop on the 1980 Census will be presented this spring. The Anthropology Department is also considering a joint workshop with the library on the Human Relations Area Files machine-readable products. Of course classes and seminars will be presented to the students, staff and faculty of the university to introduce both the census and other data sets.

With expanded campus interest, a supportive administration, and a cooperative computer center, there still are no royal roads to integrating the various media and formats in a university library. One of the difficult problems facing the data librarian is to learn of the mounting of new surveys and of the availability of MRDF. We were recently informed that surveys of Florida's socioeconomic status are being made by the agricultural extension group on the campus. Several of the questions relate to energy and its use in Florida. Although the file and
documentation were prepared by faculty at the university, none of the originators of the surveys considered archiving the data sets in the library. Later they were added to the holdings. It is equally difficult to learn of the existence of significant MRDF at the state level, yet they can be a major information source for planning in the socioeconomic and environmental areas.

The catalogs published by the various federal agencies and the ICPSR, as well as those from data archives and libraries, are helping meet the problem of identifying new sources. Dynamic professional groups in APDU, IASSIST and certain federal agencies are constantly working to demystify MRDF and aid in their description, location and use. It is hoped that all data libraries, archives and federal agencies can cooperate in developing a union catalog of holdings.

Development of the bibliographical control and cataloging of MRDF is also a critical issue and must proceed quickly. Data services in our own library are hampered because a limited cataloging staff must handle the complexities of rare books, manuscript collection, microform, serials, and current imprints. Bibliographic control and access to MRDF through the card catalog is not feasible for the next few years at the University of Florida Libraries. The real problems raised by insufficient bibliographic control came to light when the ICPSR files were transferred to the library. Card files for the data sets received had been maintained by a succession of graduate students working part-time in the Political Sciences Data Laboratory. Changes in so-called “editions” of the data sets as well as copying of the files on new tapes had not been meticulously recorded. Given their training, librarians may enjoy grappling with these intricate problems of bibliographical control and inventory. Fortunately, the systems programmer handles the tape inventory with its technical information for all non-ICPSR data files. Above all, a team effort of data librarians and systems personnel is required for adequate bibliographical control and technical control.

A significant problem for data services is funding. While the response of faculty and students to one centralized location for MRDF is positive, their demands on the library are increasing. So are the costs of purchasing, staff and computer services. Data sets are basically purchased from library funds, whether allocated to university departments or not. Competition for available funds and the rising costs of materials and processing may play havoc with a very expensive undertaking. The necessary funds to purchase MRDF, train the staff, support it, and service an increasing need for MRDF could become an administrative and public relations problem for the library. For example, to purchase the 1980 census tapes and train the technical and reference personnel
involved will require many thousands of dollars. The number of professional library positions and the book funds and operating expenses probably will not increase within the next few years. To maintain a tape archive is expensive, requiring the continual use of computer funds to map and "dump" portions of the files at intervals for their preservation. An environment where humidity and temperature are controlled is also essential. Competing units within the library, inflation, and the problem of allocation of library funds may take their toll. The purchase of very large data sets may well depend upon their continual use by not only the social sciences but also faculty in business, medicine and agriculture. To judge the research and teaching potential of expensive MRDF is a very important responsibility of the data librarian in any setting.

The recruitment, training and development of both a reference and systems staff for data services is of paramount importance. The systems programmer involved must be oriented toward the public service goals of the library, and have interest in large and small data sets and complex data management problems. The participating reference staff should see the files and software as a basic information source for the public, along with printed sources. Some knowledge of research methodology, statistical techniques, computer programming, and communication skills are extremely helpful in understanding the needs and levels of expertise of a wide range of users. Again, with the exception of census activities, the experience of the staff at the University of Florida is fairly limited. Since 1976 only two reference librarians have the responsibilities for handling users' needs for all census, ICPSR and other MRDF. Only one systems programmer is basically responsible for the tape library, file management and liaison with the Computer Center. Coordination of data activities and supervision of a part-time Florida Consortium data manager are the responsibilities of the social sciences reference librarian. A student assistant will be requested this year to divide his work time between reference and systems, and will be assigned clerical tasks. The increased activities experienced this year will be absorbed by the regular team.

The critical issues of staff and training will be very apparent this fiscal year with the availability of the 1980 census in machine-readable format. Demand from all groups is expected to be heavier than in the seventies, and both 1970 and 1980 data will probably be requested to make comparisons. Moreover, the files will be more complex, and a new, powerful software developed by the census bureau will have to be learned. A knowledge of both printed sources and MRDF will have to be acquired. Another reference staff member, the urban and regional docu-
ments librarian, has been asked to learn the 1980 census machine-readable files. The two experienced reference librarians and the systems programmer have attended the census bureau's training program on census machine-readable files.

The interrelationship between the reference and systems staff and the Computer Center is a critical one. The three reference librarians involved in the 1980 census activities, the systems programmer, and two Computer Center staff have met regularly to discuss both the substantive content of the files and the problems of using CENSPAC. Special tutoring by the systems programmer is being given to the new member of the reference team. These meetings are also used to plan seminars for faculty, students and outside groups for the 1980 census and other related files and are a cogent reminder that data services depend upon training and cooperation among several groups of persons with diverse skills. For all participants, these sessions have given an overview of the reference and systems problems involved in a complex undertaking.

The next five years will be a crucial period for the development of data services at the University of Florida Libraries. With demand increasing and so few existing data libraries and archives, an academic library offering basic data services may suffer from the "sitting duck" syndrome. Many groups trained to utilize MRDF and those who require the data for research, decision-making and grant proposals will expect a centralized apparatus to have larger collections, staff for interface, and reasonable algorithms for fixing charges. The current structural organization will require more formalization. Meetings will need to continue and should include the heads of reference and the Systems Group, along with those who are handling the requests. It is this writer's opinion, however, that a special department or unit to handle MRDF services is not a good choice at this time. Reference librarians must accept and learn the new technologies and the fact that many techniques are required to retrieve today's information. Systems groups need to be aware that their expertise ultimately should lead to satisfying users' needs. A team approach will continue to be used.

Staffing plans for the next five years will need to consider the provision of a half-time data manager whose sole responsibility would concern the growing demands of the Florida Consortium based upon ICPSR holdings. In turn, ICPSR services need to be welded into the total data program. A central ICPSR archive of data sets at the University of Florida Libraries can enable social scientists at the other state universities to carry out significant research without having large data libraries on their own campuses. More standardized computing systems at the smaller schools and a developing remote computer linkage within
the state system will increase the demand for all data activities, including reference services. The coordinator of these data activities must learn to manage a larger team, having different but complementary skills in a library setting.

Within a year and a half, the regular library catalog holdings from 1975 to present will probably be available to users in an online system. Catalog information and abstracts for data sets can be included as a separate online file available to all. Specialists in the reference staff for business, humanities and science can do the cataloging and abstracting for MRDF in their respective fields. During the next five years at least, it is the reference librarian at the University of Florida who will need to develop skills in the rather complex cataloging systems evolving for MRDF.

During the same time period, the reference staff involved with data services can also integrate their teaching efforts within the overall instruction program of the library. A concerted effort to teach students, faculty and outside groups both printed and machine-readable sources of statistical information is a definite goal of the University of Florida Libraries. For all students, but especially for graduate students, this instruction should begin early in their stay on campus. Bibliographic instruction is receiving great attention now. No significant format or method of accessing information can be excluded in the teaching and services program of a university library.

Large data files are "interdisciplinary" by their very nature. Anthropologists, sociologists, psychologists, medical researchers, political scientists, and economists are using many of the same files. If there is a time series involved, the potential use is even greater. Individual surveys are certainly quite important but costly. Researchers from many disciplines can use data from the very large sets in a number of significant ways. The impact of large public data sets and those involved with time-series information will loom large in the selection and funding of data services. In addition to the need for time-series MRDF, faculty are becoming interested in MRDF from abroad. One of our political scientists is manipulating the new German General Social Survey (Nationaler Sozialer Survey). A young historian from Cambridge has asked the writer to locate any MRDF concerning German fraternities in the 1930s and the characteristics of their members. A continued contact with foreign data archives and libraries would seem to be indicated. The development of linkages to data archives abroad and the dissemination of information about them to faculty are responsibilities of the data librarian.
More extensive linkages to outside publics in government and industry will be expected within the next five years. Planners, statisticians, systems analysts, and data-base managers are involved in all levels of government and the private sector. These are potential users of MRDF in data archives and libraries. One does not need to have advanced graduate training to realize that billions of dollars may be allocated to various groups and functions using 1980 census data. The city clerk of a small town applying for a housing grant may not know about analysis of variance, but he knows he must have socioeconomic census data by enumeration district for his proposal. Legislation and application guidelines require this information. If applications for grant money require a comparison of data for 1970 and 1980, think of what linkages may be established between data libraries and the community at large.

In 1980 there were some plans to establish a Florida state data center with the University of Florida Libraries as one of its affiliates. Unfortunately, the center has not materialized. Such a unit could help coordinate the acquiring of both census and state data files. Moreover, a Florida state data center would serve as a stimulus and link in developing a network in Florida. Meanwhile, the library will probably play the role of regional data center without formal designation as such.

Within a few years, networks and consortium for MRDF should be more highly developed. Requests from the Florida Consortium members within the State University System have increased approximately 30 percent. A statewide computer linkage system could permit members to copy and subset ICPSR files, as well as to "dump" codebooks in machine-readable format. Each participating university could also have access to all public data files. How will having a central archive with remote terminals affect the future relations of members and their demand for data services? What will be the acquisitions role of a central data library and archive in relation to smaller facilities? If other universities and agencies have complete access, how will costs be shared? These will be interesting problems to monitor over the next few years.

MRDF are now a major information source. This fact seems apparent whether the demands for access are met by data libraries housed in a computer center, special unit or an academic library. Data services within a large university library are feasible when library administration is supportive, funds are adequate, and reference and systems personnel can perform as a team. For the user to identify the university library as a source of information in all formats is a step forward. It
RAY JONES

constitutes a significant challenge to the profession of librarianship within the next decade.

References

The University of British Columbia Data Library: An Overview

Laine G.M. Ruus

The University of British Columbia Data Library represents, I believe, a unique organizational model among data libraries (by which I mean to include, as well, data archives and data banks) in the manner in which it is jointly operated by the library and Computing Centre of the university. How it came to be as it is is a result of its historical development; it continues to function as it does due to the success of the original model.

In 1963/64, a Statistical Centre for the Social Sciences was established in the university’s Faculty of Arts, primarily through the efforts of the departments of economics, political science, and anthropology and sociology. The purpose of the center was to provide statistical and programming consultation to faculty and graduate students in the Faculty of Arts, i.e., to act as an intermediary between the social scientists and the Computer Centre. By 1965 the Statistical Centre had entered into membership agreements with the, then, Inter-University Consortium for Political Research (ICPR) and the International Survey Library Association (ISLA), the membership arm of the Roper Public Opinion Research Center, then at Williams College. Through these memberships a small local collection of punched cards, magnetic tapes and codebooks was built up, in the Department of Political Science. This Political Science Data Bank was administered by a senior faculty member, and his students on a part-time basis. It was used only by a few faculty and students in the department, and consequently, small as the collection was, it was underutilized. The directors of the Statistical
Centre were not primarily interested in the management of a data library facility; therefore, the Political Science Data Bank remained in the department for a number of years.

Dr. Jean Laponce, who was, for a time, chairman of the Statistical Centre Steering Committee, had adopted the idea that traditional libraries should assume responsibility for the management of collections of machine-readable data files (MRDF)—first promoted, I believe, by Ithiel de Sola Pool.1 As early as 1965, attempts had been made to have the data bank's acquisitions and memberships funded by the library; this move to transfer the financial burden from the Faculty of Arts to the library was not permitted by the university senate. By the early 1970s, however, Laponce had successfully persuaded the university librarian and the head of the Computing Centre that they should jointly bear responsibility for the management of the local MRDF collection. In 1970, a proposal was drafted for the creation of a Data Library to be jointly operated by the Computing Centre and the library and: “to have the functions of acquiring, organizing, storing and servicing machine-readable data files. It would also assume liaison responsibilities in connection with other data libraries....”2 This proposal was accepted by the senate, and in 1972 the Data Library was created.

Its first collections consisted of the data files originally collected through the ICPR and ISLA memberships, and a few other files collected from individual researchers, including a small collection of local 1961 and 1966 Canadian census data. The new Data Library was staffed by a research assistant from the Department of Political Science, transferred to the Computing Centre's payroll, and administered by a systems librarian on a part-time basis. In its second year of operation, the Data Library was finally staffed by a full-time professional computer programmer and a part-time reference librarian, and growth of staff and services since has been continual. The present organizational structure of the Data Library is as follows.

Mandate. The basic mandate of the Data Library, as defined in the original proposal, is “to develop collections and services in accordance with the academic requirements of the University, in parallel with the policies of the Computing Centre and the Library.”3 Because both “parent” institutions are mandated to serve the entire local academic community, this interdisciplinary focus applies to the Data Library as well.

Financing. Financing is derived from a number of sources. The Computing Centre provides a full-time programmer analyst, a computer dollar budget, physical plant, and all that is subsumed by hardware
and software support, as well as some minor office expenditures. The library provides a full-time librarian, a full-time clerical staff person, a real-dollar acquisitions budget, general collections, reference and technical services support, cataloging services, and such incidental expenses as supplies. In addition, to supplement the acquisitions budget, cooperative funding arrangements for the maintenance of very expensive subscriptions or memberships have been negotiated such that, for example, the Department of Political Science contributes part of the cost of the annual Inter-University Consortium for Political and Social Research (ICPSR) membership, and the Faculty of Commerce contributes one-half of the annual subscription costs of CRSP and COMPUSTAT data bases. Thus, the Data Library is in the fortunate position of deriving its primary funding from two of the most stable and secure budgets in the academic environment.

Staff. As noted earlier, the Data Library has a full-time staff of three. Training of these staff has been primarily carried on in-house to complement the experience and academic backgrounds of the three incumbents. Thus the librarian, who had no previous training in statistics or programming, attended both ICPSR summer school sessions in statistics and data management, as well as various local courses. The programmer, on the other hand, whose background included both statistics and programming, received only in-house training. Duties have been divided between these positions such that the librarian has been primarily responsible for acquisitions, reference, library-related technical services, and administration; the programmer has been responsible for user consultation, programming for public and internal purposes, and computer-related technical services; the clerical staff member is responsible for technical services vis-à-vis the codebook and reference collections, tape management, circulation, etc. In addition, for special programming or data management projects, part-time students are periodically hired whose academic backgrounds are suited to the nature of the project in question.

Collections Policy

As is, I believe, common, the collections policy is rather vague and ad hoc. The original mandate made only one stipulation regarding collections—that the Data Library “develop collections...in accordance with the academic requirements of the University, in parallel with the policies of the Computing Centre and the Library.” A policy has evolved over the years, however, which can be outlined as follows: the
Data Library will collect automatically all significant Canadian data files, such as census data, election studies, and other major social surveys, public opinion poll data, etc. All other MRDF are acquired on request, tempered by considered need, potential for future use, and, of course, budgetary constraints. In addition, the library will function as a data archive in the sense that an attempt is made to acquire any original MRDF produced by local researchers, or offered for deposit by outside researchers (depository MRDF), and every effort is made to ensure that these are maintained for posterity. Because of the breadth of the mandate, it has not been necessary to restrict collections to those applicable to any one or limited number of disciplines. Other restrictions, however, are applied of a technical or contractual nature. To clarify, data files are not acquired or maintained which cannot at the very least be made available to any member of the local academic community, nor are MRDF acquired which contain confidential data, such as names, addresses, Social Insurance Numbers (SINs), etc., such that individual privacy is violated. Further, MRDF are not maintained which lack adequate documentation, or which are so "dirty" as to be useless for secondary analysis.

**Composition of the Collection**

Because of the previously mentioned factors, the collection at the present time cannot be called homogeneous, nor is it one of any great depth. It contains a rather comprehensive collection of Canadian census data, a good collection of Canadian Gallup Poll data, and most major Canadian public opinion surveys. On the other hand, there is a small collection of texts in English, French, German, Latin, Greek, and Amerind languages, but no major collection of texts from any one language, author or time period. Our fairly large collection of Canadian socioeconomic time series data is not yet balanced by an adequate collection of Canadian financial data. A small collection of health-related data is not balanced by an adequate collection of vital statistics data. And then there are many "odd-ball" files, such as digitized maps, a catalog of digraphs, and a large collection of satellite imagery of the northwest coast of North America. This sporadic collection is, in essence, then, nothing more than a direct reflection of our mandate—it reflects the research and teaching needs of the local academic community over the past two decades.
Technical Services

These tend to be of a preservative orientation rather than creative. This is dictated by the mandate, by the focus of the raison d' être of the Data Library as a library rather than as an archive, and by the primary user community which is the local academic community, rather than a wider national or international one.

For the purpose of maintenance, the Data Library collection consists of three parts: a collection of magnetic tapes on which are stored MRDF, documentation (i.e., codebooks) and some special-purpose programs. Each collection requires a separate set of technical procedures. They are as follows:

—MRDF are acquired, stored on tape, cataloged, and documented in the data base catalog. When necessary, MRDF are also "cleaned" and "translated" into system files formatted to the requirements of statistical packages such as SPSS or OSIRIS, as required by users.
—tapes are routinely copied, cleaned and rewritten.
—codebooks and other documentation are copied, or (if machine-readable listed to paper or COMfiche (if also very large). They are bound, assigned call numbers, and otherwise processed for circulation. Only very recently has it been possible, given time and staff constraints, to begin to create machine-readable documentation and cleaned (OSIRIS format) MRDF of depository files.
—software may be written for two reasons: (1) to rationalize and support in-house procedures, and (2) to make data access and retrieval by users as efficient and foolproof as possible. Such software is extensively documented, in standard Computer Centre user documentation.
—auxiliary technical services include maintenance of a reference collection, circulation system, personnel records, etc.

User Services

User services account for the major portion of Data Library staff time. The mandate stipulates very clearly that the Data Library's primary users are the local academic community, i.e., faculty, students and staff. However, increasingly in the past decade, the obligation of the university to serve the community at large has been reflected in the service policies of both the library and Computing Centre, and therefore also of the Data Library. Where possible, direct access to the holdings is provided to users in the government, commercial and private sectors;
LAINE RUUS

and MRDF are copied on request for outside individuals and institutions. Our ability to provide these services is, however, very restricted by our contracts with MRDF suppliers, both individual and corporate.

The ideal (and therefore, almost by definition, seldom encountered) user transaction can be considered to consist of seven stages. For the sake of convenience, I will consider our user services within the framework of these seven stages.

Stage one consists of basic orientation in response to an initial user inquiry. Whether this be conducted in the Data Library, or in other information dissemination areas, such as the main library’s or Computing Centre’s information desks, the user has access to basic orientational materials which include information on how to access, as well as generate instruction in the use of, the Data Library’s data base of data file descriptions (SPIRES catalog). Occasionally, this stage consists of in-class orientation lectures by Data Library staff.

Stage two consists of the identification of MRDF appropriate to the user. Because there is no union catalog of MRDF, this involves searching for MRDF in the local collection as well as other sources. To facilitate searching of the local collection, brief bibliographic records of each MRDF are maintained in the University Library’s card and COM-fiche catalogs. In addition, extensive data file descriptions, including variable summaries, are maintained as an online SPIRES data base, which allows indexed and string searching of the contents of the data file descriptions. In the event that a required MRDF is not part of the local collection, the user receives extensive assistance in searching our reference collection to identify appropriate MRDF and sources thereof.

Stage three, then, is the acquisition by the Data Library of required MRDF not in the collection. Acquisition from outside sources is generally a lengthy, time-consuming, and occasionally frustrating process.

Stage four is the provision of documentation. The Data Library attempts to ensure that it has two copies of all codebooks, and at least one copy of all supplementary documentation, pertaining to MRDF in its collection. The SPIRES catalog record includes citations of all documentation, call numbers of codebooks, and other information necessary to access machine-readable codebooks or other system-generated documentation, as well as citations of published works based on that data file. Novice users receive, of course, consultation on “how does a codebook mean?”

Stage five consists of access to the MRDF. All information necessary to mount Data Library-owned tapes is contained in the Data Library User's Guide; the information necessary to access individual data
files—through a cataloged tape mount procedure which is transparent to the user—is contained in the SPIRES catalog record, including such file information as size, format, etc. Files can thus be accessed at all times that the computer is in attended mode. In certain cases of especially complex MRDF, particularly those heavily used by novice users, additional service is provided in the form of special-purpose software written to simplify and rationalize the retrieval of data and its storage in disc files in formats appropriate as input to other statistical programs. Documentation of these programs is also included in the Data Library User's Guide.

Stage six consists of the user's analysis. Service at this point is restricted to consultation, especially to novice users. The Data Library does not perform any analyses for users.

Stage seven consists of the user's publication of the results of the analyses. At this point, service consists primarily of consultation on such matters as citation formats for MRDF for inclusion in bibliographies and footnotes, etc.

There are, of course, other types of users who require other types of services, such as those wishing to deposit MRDF, those conducting surveys or otherwise creating MRDF, and even (once) a television game show producer wishing to upgrade the intellectual content of his show. Services to these users include consultation on questionnaire design, encoding, data file formats, creation of machine-readable documentation, etc., and agreements to archive MRDF for posterity.

Current Issues

Issues facing the Data Library today are numerous, certainly such as to preclude the possibility of sitting back and with complacency maintaining the status quo. Fortunately, few issues are administrative.

Staff recruitment and training—the Data Library staff had, over the past eight years, remained almost totally stable; unfortunately, in the past year two-thirds of the staff have been replaced. Because of previous lack of turnover, no staff training routines have been developed, nor has a full staff manual been written. Both of these lacunae must be filled; this can only be done in-house.

Documentation—whereas, at the present time, data files can be accessed by the local user virtually at all times, this is not true of most codebooks. Documentation which is only available in printed form is only accessible during the limited hours that the Data Library is open—our experience with placing copies of heavily used documentation in
other library locations with longer opening hours has not been entirely satisfactory for a variety of reasons. Many machine-readable codebooks, which are technically as accessible as any quantitative data file, are so large that it becomes very expensive merely to copy the file. In any case, the average user does not require a listing of the total codebook; normally he requires documentation only for a selected subset of variables. Thus, copying the codebook to COMfiche is not an entirely satisfactory solution. In order to maximize user access, we hope to develop software capable of searching machine-readable codebook records and retrieving those records satisfying user-specified conditions. A user could thus search a codebook and retrieve those variable descriptors of interest to him. Once all the codebooks are converted to this format, access will be much more efficient. There are, of course, problems associated with this, not the least of which is the choice of a standard format for codebooks capable of encompassing all types, including those that describe microdata, macrodata, textual data, representational data, models, etc.

Identification of MRDF—not all data archives and libraries publish catalogs or inventories of their holdings. Nor is there a comprehensive bibliography of those that are available. Thus, the identification of MRDF in other collections is a formidable task. This is a problem that we alone can do nothing more about than to disseminate as widely as possible our own catalog (in which admittedly we have been delinquent—our SPIRES catalog will be published in COMfiche format upon completion of retrospective conversion of all MRDF descriptions), and to promote in every way possible the compilation of a comprehensive bibliography of “finding aids” and, eventually, a union catalog of MRDF. This will not solve the corollary problem of identifying MRDF not in existing data library collections, especially those held by private individuals. The only presently viable solution to the problem of identification is the promotion of universal use of standard bibliographic citation formats in all publications.\(^\text{5}\) Steps now have been initiated in Canada, through the Social Science Federation of Canada, to persuade publishers and editors to adopt as part of their editorial policy appropriate formats for MRDF citation. We can only hope for results and push for the adoption of these standards in Canada and elsewhere.

Access—this is often problematic whenever one is not dealing with a regular disseminator of MRDF, whether it be an individual or corporate body. Individuals are often very protective of their data until such time as they have published—sometimes ten or more years later. After
publication, the individual principal investigator often has not bothered to ensure the long-term maintenance of the raw MRDF, and/or not maintained adequate documentation, and thus the data are "lost"; corporate bodies are also often equally unconcerned about the long-term maintenance of raw MRDF. Government, in the Canadian experience, can impose extreme constraints to protect the privacy of the individual. Such policy recently, in combination with other factors, almost resulted in the total restriction of access to MRDF to be generated by the forthcoming Canadian census. Had this situation not been averted, it would have had disastrous consequences for Canadian demographic and social research. Possibly one solution to this type of problem might be the adoption and publication of a uniform code of ethics for data archives/libraries, including such issues as user access, dissemination and redissemination of MRDF and accompanying documentation, and the resolution of the thorny problem of copyrights vis-à-vis MRDF; some preliminary work in this area has already been done. Certainly, government departments should be encouraged to standardize their MRDF dissemination policies, some of which are permissive, and some totally restrictive.

**New services**—large online quantitative data bases are proliferating on the international computer networks. Some data base producers offer to lease to individual users or institutions periodic "batch" editions of these data bases. As opposed to costly online subscriptions, this is a viable alternate mode of access for academic data libraries whose users often do not require "tomorrow's figures today." However, even the reduced academic rates charged for these data bases are high, and this method of updating data bases is inefficient. Some means of rationalizing access must be developed in order to eliminate unnecessary duplication, effort, expense, and time delays, but yet facilitate access to the data in a format allowing statistical analysis of the data without rekeying. We are currently considering the system developed at the University of Western Ontario for online batch access to the CANSIM data base, which utilizes the facilities of both local and national networks.

A somewhat different issue is the proliferation at various institutions of data archives serving specialized non-social science users, i.e., serving specifically one of the humanities disciplines or the hard sciences. The establishment of several types of data service facility within one parent institution will lead to further duplication of staff, facilities and services, the avoidance of which has been the major argument for the creation of data archives and libraries in the first place. Certainly there is a need for improved communication among these various types of data service facilities.
What of the Future?

The past five years have seen, locally, the wider promotion of the Data Library's services, development of needed software, production of the Data Library User's Guide, and the development of the SPIRES database of MRDF descriptions; on the interinstitutional scene, this period has seen, among other things, increased awareness of a data archive or library "profession" and the creation of the International Association for Social Science Information Service and Technology (IASSIST), the establishment of regular courses in data library management, and the development of cataloging and citation formats. I hope that the next five years will see at least some of the developments outlined herein, including the widespread use of MRDF citation formats, the publication of a bibliography of the literature on MRDF management, a directory of existing data archives and libraries, an efficient liaison with online vendors of quantitative data bases, and the incorporation of these services into the normal sphere of the data library. Also, I hope to see increased communication between our data archives for the social sciences, for the humanities and for the sciences, and the coordination of these types of services at all levels.

References

3. Ibid.
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. 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 demand4 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. 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 over 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, 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, 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 bibliographic 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. 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 bibliograph-
ic 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 inventorizing 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 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. 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 than 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-
ALICE ROBBIN

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-
DPLS: Organizing Special Libraries

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 departmentssh 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. 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.

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. 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 telecommunications 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
DPLS: Organizing Special Libraries

Manual and Electronic Communications Links of the System

other institutions
EDUNET institutions

Memorial Library  DPLS  MACC (Computing Center)
Appendix B

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

<table>
<thead>
<tr>
<th>I. ENUMERATIVE STATISTICS</th>
<th>II. COMMUNICATIONS AND MASS MEDIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA Nation</td>
<td>BA Communications and Mass Media</td>
</tr>
<tr>
<td>AB State or Province</td>
<td></td>
</tr>
<tr>
<td>AC County</td>
<td></td>
</tr>
<tr>
<td>AD Legislative District</td>
<td></td>
</tr>
<tr>
<td>AE Standard Metropolitan Statistical Area</td>
<td></td>
</tr>
<tr>
<td>AF City or Other Local Unit</td>
<td></td>
</tr>
<tr>
<td>AG Household</td>
<td></td>
</tr>
<tr>
<td>AH Individual</td>
<td></td>
</tr>
<tr>
<td>AJ Vital Statistics</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. ECONOMIC SYSTEMS, STRUCTURES AND BEHAVIOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA Economic Attitudes and Behavior</td>
</tr>
<tr>
<td>CB Economic Processes and Indicators</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV. ENVIRONMENT AND NATURAL RESOURCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>DA Environment and Natural Resources</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>V. INSTRUCTIONAL DATA SETS AND COMPUTER PROGRAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>EA Instructional Data Sets</td>
</tr>
<tr>
<td>EB Computer Programs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VI. INTERNATIONAL SYSTEMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA Dyadic and Small Group Interaction</td>
</tr>
<tr>
<td>FB Organizations</td>
</tr>
<tr>
<td>FC Structural Characteristics of the International system</td>
</tr>
<tr>
<td>FD Alliances and Military Affairs</td>
</tr>
<tr>
<td>FE Conflict, Violence and Wars</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VII. LEGAL SYSTEMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA Legal Systems</td>
</tr>
</tbody>
</table>

| VIII. POLITICAL AND GOVERNMENTAL SYSTEMS, STRUCTURES AND BEHAVIOR |
DPLS: Organizing Special Libraries

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

KA Primaries, Conventions and Candidate Evaluations
KB Election Studies
KC Election Returns
KD Hersed Electoral and Ecological Data
Public Opinion on Political Matters and Political Participation
LA Public Opinion
Political Parties
LB Political Parties

Legislative and Deliberative Bodies

MA Roll Call Voting Records
MB Decision Making in Deliberative Bodies
HC 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
XI

ALICE ROBBIN

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
SH 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

DPCS 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

WINTER 1982
427
Suggested Bibliographic Citation


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

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


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.
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.

Appendix D
Overview of an Automated Information Management System for the Data Library
References


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.)


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


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.

15. David, "Access to Data"; and Rokkan, "Data Services in Western Europe."
Online Numeric Data-Base Systems: A Resource for the Traditional Library

MARGARET O'NEILL ADAMS

"REFERENCE SERVICE IS the ultimate library service since its object is to insure the meeting of the library patron with the materials or knowledge which the library is established to afford him...."¹ With reference service as the ultimate library service, I propose that the ideal reference department in the latter decades of the twentieth century should coordinate a full range of computer-based services: online library networking, online bibliographic and textual searching and retrieval services, data library services, as described elsewhere in this issue, and utilization of online numeric data-base systems. This model for reference reflects a continuum in the evolution of information services during the past century. Its foundation is a conviction that libraries are better suited, by tradition and structure, to exploit selectively the full gamut of the emerging computer-based information resources than are individuals.

The following discussion analyzes one component of this reference service: online numeric data-base systems for time-series data. For purposes of definition, a numeric data base is described as a "computer-readable collection of data which are predominantly numeric in nature," with numeric data-base systems comprising "a functioning combination of one or more numeric databases...[with] a search system which can retrieve the numeric data...and further process them."² As used here, online numeric data-base systems refer to systems where the content of the data base is composed of discrete items of historic statistical information. Potentially the discussion could easily apply to data bases of statistical projections or forecasted data. A further limitation is

¹ Margaret O'Neill Adams is Manager, Kentucky Economic Information System, Center for Applied Economic Research, University of Kentucky, Lexington.
that only online numeric data-base systems that pertain to the social sciences are considered, although many of the generalizations apply equally to natural science and technical online numeric data-base systems.

The context for a discussion of the library use of online numeric data-base systems is established by briefly tracing the tradition of reference service and by reviewing the ways in which computer technology has enhanced the provision of many library services in recent years. A discussion of some of the latent issues related to libraries and computer technology is also included. Finally, experiences with the Kentucky Economic Information System, an online numeric data-base system for time-series statistics related to the economy of Kentucky, are described to give specificity to the general themes.

According to Rothstein, the conception of an organized reference service descends directly from the ideas of a late nineteenth-century librarian. Samuel Swett Green of the Worcester (Mass.) Free Public Library proposed an explicit program for reference as a responsibility of the library to the 1876 conference of librarians. His ideas are reported to have received general approval at this conference, but the practice of providing reference service seems to have evolved first as a function of special libraries, and only later at university and research libraries. Most of the larger public libraries are said, however, to have adopted reference services by the end of the nineteenth century.3

Perhaps the most intriguing aspect of Green's suggestion, considered now more than a century later, was his perception that provision of a reference or information service, could stimulate library use and therefore garner support. As a public servant, he seems to have been at least partially motivated by the need to demonstrate the public utility of the library as a resource for general information in order to increase the number of patrons and to justify continuing public investment. In other words, his concern seems to have been to formalize reference service as a complement to archival and lending services. Evidently he considered both types of library activities as essential and believed that unless libraries provided an information service, continuing public support for their other tasks was unlikely.

Retrieval and manipulation of statistical information from online numeric data-base systems is but a current version of some aspects of the traditional information service just described. Although use of such systems extends Green's notions beyond anything he could have possibly imagined, it is easy to picture him nodding his approval. After all, the accepted practice of reference librarians has been "supplying...fac-
tual information, mostly out of books such as encyclopedias, directories, dictionaries or the like. Libraries that have the means and the willingness to avail themselves of this new type of resource for numeric information are simply taking advantage of an innovative vehicle to provide efficiently what has become one of the most fundamental of library activities.

It should require no great adjustment for libraries to do this since they are already using computers in a great many ways. Machine production of catalog cards, online in-house catalogs, and management systems for materials circulation and inventory control are but a few examples. One significant library development directly related to technological innovation during the past decade has been the growth of cooperative online networks such as OCLC and RLIN (Research Libraries Information Network). Library networking provides patrons at participating libraries with an online catalog to the holdings of all other network libraries, thus expanding the possibilities for interlibrary lending and sharing of resources.

An equally revolutionary innovation has been the availability of online bibliographic data bases as basic tools in the provision of bibliographic reference services. Online bibliographic data bases achieved this status in most major academic libraries, as well as in many public and other libraries, only in the latter part of the 1970s, although these systems were preceded by computerized batch literature searching in the late 1960s and early 1970s. They emerged as byproducts of computerization in the production of indexes, abstracts, journals, and other secondary source publications. These developments occurred first in scientific and technical fields, but quickly spread to the more general arts, education, humanities, and social science fields, as publishers realized how marketable bibliographic data bases were. Some have suggested that the next logical extension of online library services is full-text retrieval; at least one commercial firm currently markets two such services.

All of the above-mentioned innovations to meld computer technology with library services provide better access to or reference about the information medium that has traditionally been the purview of the library: the printed word. They are creative and logical responses to the challenge of keeping pace with traditional library responsibilities in the face of what is now commonly referred to as the "information explosion." What the innovators have done, however, is simply taken structures and formulas that have been developed for printed information and stored them in a computer-readable and retrievable format. This feat is by no means trivial. Nevertheless, these innovations are limited by
the extent to which they depend upon print media. There is a vast volume of raw and processed information, computer-produced or in computer-readable form, that will never be suitable for publication. Reference service related to these materials is vital if patrons are to have the opportunity of exploiting all the available information resources.

It should be noted here that a commitment to information service for computer-readable nonprint data has been at the heart of the emergence of social science data libraries and archives since the latter part of the 1960s, and at least partially responsible for user-friendly features in some online numeric data-base systems for time-series data developed more recently. When reference departments of traditional libraries adopt information services for computerized nonprint data, they give their clientele "the opportunity of locating information at the one place [they have been]...trained to look for it, the library." Unfortunately, such services are far from commonplace in today's research or public library.

Indeed, as a consequence of the many technological developments of the past twenty years or so, some have predicted the demise of the traditional library altogether. In their crystal balls they have conjured up an amalgam of the library and the computing center. Others envision a future where information seekers sit at their home or office computer terminals, inexpensively calling up and manipulating data at any time of the day or night from some master online data-base system where "everything you always wanted to know [is] on-line." The paperless society becomes the millenium.

Using computer-based resources, especially online numeric data-base systems, need not be viewed as displacing the traditional functions of the library. The memory and manipulative capabilities of the computer cannot provide an appropriate substitute for the library's role in transmitting knowledge through print media, its "archive of learning." It would be anti-intellectual to suggest that computer terminals or video screens alone are appropriate media for the transfer of the "product of the human intellect...[upon which traditionally] the continuation of civilization depends." Contemplative consecutive reading will probably always be fundamental to this transfer.

The area where computer technology provides mechanisms ideally suited for information service comes, however, for reference service related to statistical or other numeric data. It was after all, "transformable information" which Licklider foresaw as comprising the "libraries of the future." Used for this purpose, the computer furnishes a means whereby librarians can provide the information service for numeric data
Online Numeric Data-Base Systems

required by contemporary researchers, planners and analysts, while in no way detracting from the services provided to the library’s traditional constituency of readers.

Nonetheless, the issue becomes muddied further by suggestions from still others that continuing technological developments may soon enable end users to “bypass the library as the traditional information storage center.” If realized, this view would mean enhanced profits for the information industry. As a proposition, however, it ignores the economies inherent in the centralization of information resources, as well as the ideals of the “democratization of information.” Certainly there will be some who are able to afford the luxury of bypassing the library or other centralized resource; there always have been. It is shortsighted to suggest, however, that the society’s requirement for all of its public information, even limited to that in computer-readable format, can be met by individuals accessing master online data bases in isolation.

As a practical matter, this is also an unlikely development, at least in the United States, for the foreseeable future. There are projections for over 100 million computer terminals in the United States by 1995. Only a small portion of these are likely to be dedicated to the types of information services under discussion. Perhaps the greater limiting factor, however, is that truly master online data bases, numeric or otherwise, do not exist. Networks linking compatible statistical data bases, such as might be developed uniquely, but in coordinated fashion, by each of the fifty states, also do not exist.

The one ray of hope seems to be among academic repositories for social science data files. The prototype of a multi-purpose information system developed by the Institute for Research in the Social Sciences at the University of North Carolina includes access to item level responses from selected survey data, and this system is accessible in an online mode through a network. In addition, the Inter-University Consortium for Political and Social Research is working on a similar information system that will provide comparable capabilities relative to the 15,000 machine-readable data files (MRDF) of its repository; this system will also be accessible online through a network.

The federal government is the primary producer of computerized statistical data in the United States, yet there are no plans, as far as the author can determine, to centralize all federal data production efforts, nor to make the products of the decentralized efforts compatible with each other. There are likewise no plans to develop publicly accessible, online, interagency federal data-base systems, even for macrodata. In
contrast, Statistics Canada, through its CANSIM facilities, provides centralized online access to statistical data produced by Canadian federal agencies. Even CANSIM, however, does not approach full integration of all numeric computerized social and economic data produced in Canada.

Most contemporary time-series numeric data-base systems, with social and economic data that are national (United States) in scope, have been developed by the private sector using publicly generated as well as proprietary information. Most are by-products of large-scale econometric modeling efforts. Generally, the range of data they provide and the extent of data availability are determined by their profitability. There is considerable overlap in the content of these data bases, and most are considered expensive to use, especially for individuals. Several of the states have supported development of publicly accessible online numeric data-base systems for state and local area economic and social data. These have been designed for state and substate planning and analysis purposes, and at least one is the byproduct of a state-level econometric modeling effort. In general, they are relatively inexpensive to use. Like the numeric data-base systems that are national in scope, they have potentially unlimited utility for government, commercial or academic research and analysis.

Taken together, all of these developments have resulted in the emergence of a rich variety of public and proprietary online numeric data bases, online data-base systems, and related computerized information resources such as social science data files. The expectations are for continuing growth. There can, however, be only considerable skepticism about the eventual emergence of master online data systems for several reasons. Data producers are not necessarily involved in the maintenance of numeric data-base systems. A national network of compatible data bases is absent and there is an almost complete lack of coordination among the range of computerized information resources.

In any case there is also considerable doubt about the desirability of an eventual emergence of master online data bases, should the many pragmatic problems be resolved. The potentially unlimited volume of data and the plurality of data producers would seem inherently at odds with such a development. Additionally, the prevailing political climate seems not to be one which portends support for centralization of national information services, nor perhaps even for subnational information. The more likely development will probably be greater diversity of computerized information resources than exists currently, mandating even more coordination of these disparate resources by centralized
Online Numeric Data-Base Systems

institutions such as libraries. The library must utilize the wealth of information generated because of computer technology, regardless of the mode in which it is stored, if there is to be any chance of meeting the society’s collective information requirements, as well as individual needs.

A true revolution will have been realized when computer technology is utilized by information professionals for more than coping with or substituting for print media, when it is relied upon to provide numeric data. For online technology allows direct access to data, and in the case of numeric information, eliminates the necessity for publication at all. While this sounds similar to suggestions for virtual journals, electronic mail, and so forth, the concept is quite different. At issue is an understanding that the type of information which the computer can accommodate best and without alteration, and for which is has been specifically designed, is numeric or objectively quantifiable.

Online numeric data-base systems are, therefore, among the most logical and natural computer-based resources available for adoption by library reference departments. Use of them can be viewed as one aspect of a larger range of data library services, or more simply as one of the services offered by a reference department. In other words, where data libraries exist or are planned, online numeric data-base services may be considered as one of their services. Where such a structure does not exist, traditional reference librarians can utilize this service. Since online numeric data-base systems provide access to numeric data in a manner that is similar to bibliographic services provided by online bibliographic data-base systems, both types of service could be offered jointly by reference departments. The same individuals can be trained to do both types of searching, and frequently the same computer terminals can be used.

Retrieval and mathematical manipulation of data, together with table-making, graphing, and mapping capabilities, are standard features of many existing online numeric data-base systems. They enable reference librarians to furnish levels of information service that are far more complete and sophisticated than those that predate the era of the computer in the library. Use of this resource type allows computer technology to act as the catalyst of change, enhancing and extending traditional information service.

All of the considerations discussed thus far have allowed general conclusions about the utility of online numeric data-base systems for traditional libraries. A more complete analysis of this topic should, however, also include some specific description of at least one such
system and the ways in which it can be used by staff at a traditional library. Online numeric data-base systems have to be readily available for library use before any of their theoretical merits matter at all.

Experience with the Kentucky Economic Information System (KEIS) provides some specificity for this discussion because it is an example of a publicly available online numeric data-base system that is used, to some extent, as a resource in the reference department of M.I. King Library, University of Kentucky. This is an addition to its use as a general state and local area statistical resource, accessed directly by state government personnel, nonprofit and private sector analysts, academic researchers, and students throughout Kentucky. Reference librarians at other university libraries in the state, as well as the staff of the Kentucky Department of Library and Archives, are actively considering access to this resource as part of their computer-based services. At Western Kentucky University, Bowling Green, where the Academic Computing Office has for some time coordinated services for social science machine-readable data files with the staff of the university library, assistance in accessing the KEIS has also become a standard service.

Development and maintenance of the KEIS is one of the program activities of the Kentucky Council of Economic Advisors (KCEA). It is not a small undertaking. The KEIS is an outgrowth of the construction of the Kentucky Quarterly Econometric Model. Many other states now support an econometric model, and so it is likely that these states may also support centralized, statewide statistical data-base systems in the future, since their general utility is evident enough. Online numeric information systems are also being developed for a variety of other purposes by states and research organizations. Private economic forecasting firms are building state-level econometric models and selling access to regional online statistical data-base systems.

The KEIS was originally designed as a data system pertaining to state-level economic information, since the modeling effort which it supports is at the state level. Much of the policy and program analysis done by state government agencies requires substate-level data, however, so a county-level data base has also been made an integral part of the system. The public KEIS contains four data banks, distinguished by the frequency of the data series. For example, the annual data bank, ABANK, contains time-series, state-level economic and social data that are annual estimates; comparable quarterly or monthly state-level estimates, when available, are stored in the QBANK and MBANK, respectively. There are approximately 4000 discrete state-level data series. The county-level data bank, CBANK, contains only annual values, with
comparable information stored for each of Kentucky’s 120 counties and 15 Area Development Districts. The content of the county data base is currently limited to approximately fifty data series per county, though given additional funding, there are plans to expand significantly the number of available county series. All of the data banks begin with observations in 1951 and include data up to the most currently released information.

Data for the KEIS data base is gathered from a wide variety of federal, state and private agencies. Occasionally the software or computer programs of the KEIS are used to create new data series, mathematically transformed from data stored in one of the data banks. In general, however, the KCEA staff does not generate data series. Rather, the KEIS should be viewed as a secondary data resource which effectively joins together statistical information produced by numerous primary source agencies. Some of these data series represent published information, but more are unpublished. Even those parts of the data base that represent published information generally include series for a more extensive time period than any single issue of a statistical source document. The multiplicity of sources from which the data are obtained, as well as the sheer volume of data, make this a very rich resource.

A statistical data base, such as that of the KEIS, can be considered in concept as analogous to a statistical abstract, or a collection of statistical abstracts and other statistical documents. Online numeric data-base systems can even be used to support the production of statistical abstracts, as is the case with the Kansas Statistical Abstract.25 In Kentucky, however, no state agency or university publishes a statistical abstract.26 Many state agencies publish statistical data related to their particular mission in an annual report, or in more frequent publications. In many cases the data released through these publications are, in the interest of currency of information, preliminary estimates. Disaggregated data frequently are not published and revised data are often not readily available. The same commentary can be applied to release of statistics by federal agencies, both as pertains to national economic indicators, and to state and local area statistics estimated by them.

Part of the process of maintenance of the KEIS involves direct communication with all appropriate source agencies or data producers, and receipt from them of all revisions of any data series which are included in the data base. The format of these data varies from computer-readable to handwritten. This is perhaps the most archaic part of maintenance of a numeric data base and is by no means limited to the Kentucky situation. It will continue to be the procedure followed by
maintainers of numeric data bases, however, until data producers themselves input their data directly into the data bases of online numeric data-base systems. The technical capabilities for doing this certainly exist, but the evolution of cooperative online ventures between data producers and data-base maintainers has thus far been illusory, as discussed earlier.27

Effectively, then, the KEIS furnishes Kentucky researchers, analysts, librarians, and so forth an online statistical resource which has the potential to serve them in lieu of a statistical abstract. It stores centrally more information than has ever been included in a single publication. This is not to suggest that all possible statistical time-series information about Kentucky are incorporated into the KEIS data base. That is clearly not the case, but it is not outside the realm of possibility. Limitations to the volume of data maintained as a part of any numeric data-base system are more institutional (i.e., staffing and support) than technical. A concerted effort has been made to ensure that all those state-level data series which are commonly needed by a wide range of persons are included in the KEIS data base and are maintained at the most current state of revision.

The nature of the updating or revision capabilities of the KEIS software is such that once an item of information is stored in the public KEIS data base, it is immediately available for retrieval or processing by anyone having an account on either of the two computing networks on which the KEIS is maintained.28 No time elapses between the input of data and their accessibility. Such a feature is extremely important to anyone interested in information at the latest stage of revision. It is particularly relevant with respect to economic data because this type of information is revised frequently by source agencies.

Clearly, however, the most distinct advantage to the use of online numeric data-base systems is not just up-to-date retrieval of data, though this feature is desirable. It is rather the ability to manipulate data mathematically, to produce tables, graphs, or maps from raw or transformed data, or even to build models. Most of these capabilities are as useful to the reference librarian, excepting perhaps the model building, as they are to others who directly access online numeric data-base systems themselves. The arithmetic capabilities allow the librarian who is approached with questions such as "What is the percentage change in per capita personal income over the past ten years?" or, "How can I determine how much of this change is due to growth, and how much to inflation?" to manipulate the basic data from which this information is derived, and assist the searcher to obtain the answers needed. The
librarian's ability to do this is contingent upon his or her having some practical knowledge about fairly simple mathematical manipulation of numeric information, but this knowledge is common enough. Learning to exploit the capabilities of online numeric data-base systems may be a different type of experience for most librarians, but it is intrinsically no more difficult than learning many of the more traditional aspects of reference service. It is undoubtedly more simple than mastering the searching conventions used for most online bibliographic services.

There are a variety of programs in the KEIS which produce standardized tables, each having some unique arrangement of the desired data. Production of these tables is a straightforward activity, requiring only a few easily learned lines of input to the computer. One practical application in utilizing some of these features is for the reference staff periodically to produce printed tables (or hard-copy output) containing the most frequently sought information and to keep these tables at the reference desk to supplement the standard statistical reference materials, many of which are dated by the time they are distributed. To illustrate the usefulness to a reference staff of these various tables, one is included as appendix A of this paper.

In addition, the indexes to the KEIS data banks themselves function as reference documents, assisting the librarian, as well as other users, in determining whether the information sought is maintained in a KEIS public data bank, and therefore whether accessing the KEIS is worthwhile. Familiarity with the structure of the index comes from frequent use, as with all reference materials. A page from the annual bank (ABANK) index is reproduced as appendix B as an example. The reader will note that the index gives the data series name and its number, essential information for most KEIS procedures. A textual description of the data series, which follows as closely as possible the description used by the source agency, is the next piece of information about each data series. If the data are released more frequently than annually, the cross-reference columns, headed QRT (quarterly) and MON (monthly) on the index, will indicate the series number for the comparable data series in one or both of the other state-level data banks. The units in which the data are stored, and the source agency from which the data are obtained, are both given as acronyms in the two columns that follow. These acronyms are defined in the preface to each index. The final piece of information gleaned from the index is the period of time for which each data series is available. This piece of information is automatically revised each time the data series is extended, so the librarian using an index as a reference document will wish to produce a current version of
the index every several months or so. Likewise, when new data series are added to the data base, documentation for them is automatically and immediately incorporated into the data-bank index.

Reference use of a data-bank index will not help patrons understand whether the data stored are suited to their analytical purposes, nor will it tell them about the methodology employed in the production of any particular data series. Other materials, such as the *Kentucky Economic Information System User Manual* or methodological descriptions published by some data suppliers, can furnish some of this information. It seems reasonable to expect also that the user of these data bears ultimate responsibility for determining whether the information he or she seeks in fact provides the indicators sought. No reference librarian should ever be liable for making a patron's methodological judgments for him or her, nor would it even be proper to do so. This is true whether the information the reference librarian assists in locating is stored in a traditional or nontraditional mode.

Online numeric data-base systems that are maintained on computers having interactive, or conversational, mode of access are usually programmed so that they seem more apparently "user-friendly" than systems that are not interactive. On an interactive system, the user can be prompted to answer a few basic questions on the way to obtaining or manipulating the data sought. This way of utilizing a computer system can replicate English-language conversation, and thus poses few barriers to use by nonprogrammers.

One of the computing networks on which the KEIS is maintained, the state government centralized computing system, does offer interactive access. Because of this, an interface program, called the KEIS Procedure, has been written to allow interactive access to the data and statistical software of the KEIS on the state government computer. The KEIS Procedure can also be used to prepare automatically selected KEIS program input for batch submission for those programs which are inappropriate for interactive processing, such as those which produce a large volume of graphically or tabularly presented data, and for production of data-bank indexes. The simplicity, ease of use, and flexibility of the KEIS on the state government network is measured by the extent of its use by a variety of persons in almost every agency of state government. Many of these people have very little, if any, other reason to be accessing the computer.29

University personnel in Kentucky, including librarians, do not have access to the state government computing network. The computing facilities of the public universities in the state are linked to the
Kentucky Educational Computing Network (KECNET). The KECNET computer on which the KEIS has been publicly available since 1976, the University of Kentucky's IBM 370, is currently operational in batch mode only, so users of the KEIS on this network do not have the opportunity to use the KEIS interactively.

Use of the KEIS on KECNET is nevertheless a relatively straightforward operation since the system was designed from the outset for the user who is a nonprogrammer. It allows him either to submit requests (jobs) on punch cards, inputting them into the card reader physically located at the university computing center, or to access the KEIS online from a dial-up or hardwired terminal. To use the KEIS online requires that the user learn CJS (conversational job system), which is the job entry system used for the University of Kentucky's IBM 370. CJS is a command language which closely follows the logic of English; it is not difficult to master.

It should be evident that both of these ways of using the KEIS on KECNET require more learning than most professionals, who do not otherwise need it, would be willing or able to spend the time to accomplish. This was the case for the University of Kentucky reference staff who, although intellectually committed to the widest range of information services, were also already busy enough providing traditional reference service and learning procedures to utilize the computer for a variety of other purposes.

In an attempt to allow all potential users, reference librarians as well as the general public, the most "friendly" access to KEIS on KECNET, some special procedures were written which, from the user's perspective, mimic an interactive mode. The user types in a single line of instructions to the computer, indicating the bank name, the inclusive dates for the requested data, and the data series number(s) having the information sought. The series numbers are obtained from a data-bank index. This single command line automatically activates the procedure, and the user receives the data returned to his or her terminal, more or less immediately thereafter. Lengthy output can be routed to an offline printer. These special procedures are limited to production of a data-base index and retrieval or querying of a data base, including some manipulation facility. A sample of the simplest case use of these procedures is included here as appendix C.

The availability of these special procedures has in fact extended the use of the KEIS to reference librarians, as well as to others who had little or not previous experience with computer-based resources. Formal KEIS training sessions have been held for the staff of the reference and
government publications departments. Use of KEIS by these library personnel in the months since their introduction to it has been limited, as was expected. It can be assumed, however, that as they become more familiar with the KEIS they will proceed from the procedures for simple retrieval of data to the standardized table-producing, graphing, or mapping capabilities, and perhaps to utilizing the statistical capabilities of the system.

Library utilization of online numeric data-base systems, no matter how "user-friendly" they are, will be a slowly developing phenomenon. As long as traditional statistical reference materials that seem to answer the needs of the reference staff continue to be produced, librarians will not necessarily turn to using online numeric systems to assist in the provision of information service. After all, this is not a familiar activity, and it is natural to turn to the unfamiliar only if the familiar does not provide what is sought. When the reference staff perceive, through considerable experience, that accessing online numeric data-base systems provides a more efficient and satisfactory information service for numeric data, they will begin to use this capability more and more frequently.

There can be little doubt that the availability of resources such as the KEIS will increase, though whether this will occur primarily through the public or private sectors remains to be determined. As this happens, reference librarians in general, as well as library administrators, will have to decide whether or not to take advantage of this new resource type. For those who decide positively, accessing online numeric data-base systems will become second nature, just as utilization of online networks for catalog information, or of online bibliographic data-base systems for bibliographic searches are today. And, Samuel Swett Green will rest in peace.

ACKNOWLEDGMENT

The author is indebted for comments and ideas to Charles G. Renfro, former Director, Center for Applied Economic Research, University of Kentucky, and primary developer of the Kentucky Economic Information System; Roy A. Sigafus, Senior Systems Analyst/Programmer, Center for Applied Economic Research, University of Kentucky; and Judith M. Wiza, Head of Data Services, Reference Department, M.I. King Library, University of Kentucky. The author alone is responsible for the content of this article.
### Kentucky Personal Income: Summary

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>9925.58</td>
<td>11797.35</td>
<td>12157.21</td>
<td>14177.13</td>
<td>15851.40</td>
</tr>
<tr>
<td>8</td>
<td>1094.06</td>
<td>1236.68</td>
<td>1433.88</td>
<td>1657.13</td>
<td>1882.67</td>
</tr>
<tr>
<td>9</td>
<td>1533.27</td>
<td>1687.08</td>
<td>1842.09</td>
<td>1928.40</td>
<td>2230.85</td>
</tr>
<tr>
<td>10</td>
<td>460.44</td>
<td>588.02</td>
<td>537.14</td>
<td>517.44</td>
<td>664.69</td>
</tr>
<tr>
<td>11</td>
<td>1067.83</td>
<td>1129.06</td>
<td>1284.94</td>
<td>1410.96</td>
<td>1565.96</td>
</tr>
<tr>
<td>12</td>
<td>538.20</td>
<td>659.86</td>
<td>672.89</td>
<td>650.71</td>
<td>784.22</td>
</tr>
<tr>
<td>13</td>
<td>11905.32</td>
<td>13463.26</td>
<td>15120.29</td>
<td>17111.94</td>
<td>19180.70</td>
</tr>
<tr>
<td>14</td>
<td>9750.62</td>
<td>11052.53</td>
<td>12928.70</td>
<td>14271.33</td>
<td>15047.36</td>
</tr>
<tr>
<td>15</td>
<td>3698.81</td>
<td>3636.73</td>
<td>4161.22</td>
<td>4886.91</td>
<td>5198.04</td>
</tr>
<tr>
<td>16</td>
<td>1826.42</td>
<td>2177.18</td>
<td>2548.99</td>
<td>2807.07</td>
<td>3230.90</td>
</tr>
<tr>
<td>29</td>
<td>1272.39</td>
<td>1463.55</td>
<td>1612.24</td>
<td>1773.05</td>
<td>1967.14</td>
</tr>
<tr>
<td>40</td>
<td>569.72</td>
<td>922.17</td>
<td>1141.66</td>
<td>1280.54</td>
<td>1476.36</td>
</tr>
<tr>
<td>46</td>
<td>723.70</td>
<td>867.71</td>
<td>978.04</td>
<td>1217.28</td>
<td>1338.18</td>
</tr>
<tr>
<td>51</td>
<td>728.20</td>
<td>780.17</td>
<td>829.65</td>
<td>956.45</td>
<td>1073.35</td>
</tr>
<tr>
<td>52</td>
<td>1221.31</td>
<td>1393.26</td>
<td>1504.78</td>
<td>1714.54</td>
<td>1924.02</td>
</tr>
<tr>
<td>61</td>
<td>446.31</td>
<td>530.66</td>
<td>616.07</td>
<td>702.59</td>
<td>799.03</td>
</tr>
<tr>
<td>71</td>
<td>900.68</td>
<td>1086.06</td>
<td>1187.36</td>
<td>1337.92</td>
<td>1530.92</td>
</tr>
<tr>
<td>88</td>
<td>1631.39</td>
<td>1860.92</td>
<td>2068.37</td>
<td>2399.69</td>
<td>2659.36</td>
</tr>
<tr>
<td>101</td>
<td>30.61</td>
<td>35.85</td>
<td>41.34</td>
<td>49.31</td>
<td>57.10</td>
</tr>
<tr>
<td>106</td>
<td>2174.09</td>
<td>2390.73</td>
<td>2591.59</td>
<td>2840.61</td>
<td>3133.35</td>
</tr>
<tr>
<td>5</td>
<td>12462.92</td>
<td>13103.12</td>
<td>15793.18</td>
<td>17762.56</td>
<td>19964.92</td>
</tr>
<tr>
<td>112</td>
<td>664.23</td>
<td>722.07</td>
<td>807.97</td>
<td>924.47</td>
<td>1070.51</td>
</tr>
<tr>
<td>115</td>
<td>1971.52</td>
<td>2206.73</td>
<td>2522.96</td>
<td>2839.35</td>
<td>3259.31</td>
</tr>
<tr>
<td>116</td>
<td>2652.17</td>
<td>2890.42</td>
<td>3081.92</td>
<td>3289.07</td>
<td>3776.13</td>
</tr>
<tr>
<td>1</td>
<td>16537.06</td>
<td>16566.75</td>
<td>20619.50</td>
<td>23187.10</td>
<td>26016.30</td>
</tr>
<tr>
<td>2</td>
<td>16387.00</td>
<td>16119.00</td>
<td>17772.00</td>
<td>19814.00</td>
<td>22191.00</td>
</tr>
<tr>
<td>3</td>
<td>4882.00</td>
<td>5404.00</td>
<td>5946.00</td>
<td>6645.00</td>
<td>7377.00</td>
</tr>
<tr>
<td>4</td>
<td>3.39</td>
<td>3.44</td>
<td>3.47</td>
<td>3.49</td>
<td>3.53</td>
</tr>
<tr>
<td>281</td>
<td>4362.34</td>
<td>4699.28</td>
<td>4235.04</td>
<td>4430.00</td>
<td>4545.29</td>
</tr>
</tbody>
</table>

**Units:**
- Per capita income in current dollars
- Real per capita income in 1972 dollars
- Real personal and real disposable income in millions of 1972 dollars
- Population in millions of persons
- All other series in millions of current dollars

**Source:** U.S. Bureau of Economic Analysis
<table>
<thead>
<tr>
<th>SERIES</th>
<th>NUMBER</th>
<th>NAME</th>
<th>DESCRIPTION</th>
<th>ORI</th>
<th>MON</th>
<th>UNITS</th>
<th>SOURCE</th>
<th>AVAILABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>401</td>
<td>LFRATE</td>
<td>Labor Force Participation Rate</td>
<td>Labor Force/Pop GTE 16</td>
<td>%</td>
<td>DH-EIS</td>
<td>1970-1978</td>
<td></td>
<td></td>
</tr>
<tr>
<td>403</td>
<td>URATE</td>
<td>Employment, Total by Place of Residence (Includes Persons on Strike)</td>
<td>%</td>
<td>DMR-1981</td>
<td>1970-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>404</td>
<td>FLN</td>
<td>Unemployment, Total</td>
<td>%</td>
<td>DMR-1981</td>
<td>1961-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>URN</td>
<td>Unemployment Rate</td>
<td>%</td>
<td>DMR-1981</td>
<td>1961-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>407</td>
<td>NPH</td>
<td>Employment, Total by Place of Residence (Includes Persons on Strike)</td>
<td>%</td>
<td>DMR-1981</td>
<td>1961-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>409</td>
<td>NAGM</td>
<td>Employment, Number of HIRED Workers, Kentucky</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1961-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>410</td>
<td>NAGNM</td>
<td>Employment, Number of NON-HIRED Workers, KY</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1961-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>413</td>
<td>NAG</td>
<td>Employment, Persons Involved in Labor-Management Disputes</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1961-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>414</td>
<td>NAGFM</td>
<td>Employment, Number of Family Farm Workers, KY</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1961-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>416</td>
<td>NAFG</td>
<td>Employment, Manufacturing, TOTAL</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1951-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>422</td>
<td>NTB</td>
<td>Employment, Tobacco Manufactures</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1951-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>427</td>
<td>NMDS</td>
<td>Employment, Men's &amp; Boys Suits, Coats</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1951-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>441</td>
<td>NPS</td>
<td>Employment, Primary Metals</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1951-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>442</td>
<td>NPS</td>
<td>Employment, Blast Furnace &amp; Basic Steel Products</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1951-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>443</td>
<td>NFAB</td>
<td>Employment, Fabricated Metal Products</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1951-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>444</td>
<td>NMFAC</td>
<td>Employment, All Machinery</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1951-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>446</td>
<td>NTHA</td>
<td>Employment, Transportation Equipment</td>
<td>%</td>
<td>TPLH-1981</td>
<td>1951-1980</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Online Numeric Data-Base Systems

#### Appendix C

<table>
<thead>
<tr>
<th>Series</th>
<th>Number</th>
<th>Units</th>
<th>Source</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCOME, PER CAPITA PERSONAL, PIKE CO.</td>
<td>YPC*K098</td>
<td>$</td>
<td>BEA</td>
<td>1970-1975 2341.000 2574.000 2773.000 3172.000 4802.000 5259.000 1976-1979 4898.000 5493.000 6094.000 7337.000</td>
</tr>
<tr>
<td>POPULATION, TOTAL, PIKE CO.</td>
<td>POPTK*098</td>
<td>T</td>
<td>USC-P26</td>
<td>1970-1975 61.059 63.500 65.400 65.700 65.800 69.500 1976-1980 72.300 73.500 73.800 74.300 81.123</td>
</tr>
<tr>
<td>COAL PRODUCTION (SHIPMENTS), TOTAL, PIKE CO.</td>
<td>XPBC*K098</td>
<td>T</td>
<td>MYB-EDR</td>
<td>1970-1975 21299.000 NA 19130.00 00 19090.000 21249.000 19178.000 1976-1978 19002.000 18141.000 15853.000</td>
</tr>
</tbody>
</table>

R: $0.13 16:01:33
References


6. Hoover, Ryan E. “Computer-Aided Services in the Academic Library: Experiences in Organizing and Operating an Online Reference Service.” Online Review 5(Oct. 1979):37. (The full-text retrieval services referred to are products of Mead Data Central known as LEXIS and NEXIS. LEXIS provides interactive searching and retrieval of full texts of state and federal court decisions, statutes, regulations, etc. NEXIS provides similar capabilities for the full texts of selected newspapers, magazines and wire services.)


15. The concept of “democratization of information” is discussed by Bunge, Charles A. “Reference Service in the Information Network.” In Proceedings of the Conference on Interlibrary Communications and Information Networks, edited by Joseph Becker, pp. 109-16. Chicago: ALA, 1971. (The context for this discussion is networking, but the applications of these ideas to a discussion of online numeric data-base systems is appropriate.)

Online Numeric Data-Base Systems


18. Much has been written about the federal statistical system. Perhaps the most comprehensive current treatment is "Improving the Federal Statistical System: Issues and Options," Statistical Reporter 81(Feb. 1981):133-221. (See especially the discussion of online access to aggregate data, p. 210.) For some time there has been the expectation that the Bureau of Labor Statistics system, LABSTAT, would be made publicly available. See, for example, Mendelssohn, Rudolph C. "The New On-Line BLS Data Base and Information System." In Readings in Business and Economic Research Management: Execution and Enterprise, edited by Jayne D. Fischer, pp. 79-85. Madison, Wis.: Association for University Business and Economic Research, 1980. According to Mendelssohn, the LABSTAT data base and software were to have been made publicly available by the National Technical Information Service (NTIS) through a commercial computer. However, because it was later determined that NTIS might be competing with available commercial sources if it did this, NTIS is not, at this writing, offering online access to LABSTAT (information obtained from the Procurement Office, U.S. Department of Commerce, March 1981). The LABSTAT data base has recently been added to the DIALOG Information Retrieval Service, and customers who have access to this vendor have access to the LABSTAT data base, though not to any of the statistical or analytical capabilities of the LABSTAT software; see DIALOG Chronolog 81(May 1981):67-68. One additional development has potential promise for centralizing the availability of data produced by the federal statistical system, at least for data users within the federal government. This is the operation of the Interagency Decision Information Display System (DIDS). Although there is an experimental remote site for the DIDS system, there is a question whether DIDS will be extended to be publicly available outside the federal sector; see "Interview: DIDS and the Federal Statistical Community." DIDS Doings 3(18 Dec. 1980):3, 5-6; and "Late News: Reorganization Bill Signed Into Law." DIDS Doings 3(18 Dec. 1980):1, 6. An earlier appraisal of the DIDS system can be found in Edward K. Zimmerman. "The Evolution of the Domestic Information Display System." Review of Public Data Use 8(June 1980):69-81. For another aspect of the larger question of coordination of online numeric data-base systems being developed within federal agencies, see Coleman, Edwin J., and Cartwright, David W. "Toward a REIS [Regional Economic Information System] for the 1980s." Readings in Business and Economic Research Management: Execution and Enterprise, vol. 2. Madison, Wis.: Association for University Business and Economic Research, forthcoming; and Duncan, Joseph W., et al. "Private versus Public Sector Responsibility for the Collection, Distribution, and Analysis of Statistical Data." Review of Public Data Use 8(Dec. 1980):307-25.

19. The author has in mind the online numeric data-base systems developed by such firms as Data Resources, Chase Econometric Associates, and Wharton Econometric Forecasting Associates. Another national computer-readable data base, although not a database system, is CITIBASE, the CITIBANK Economic Database. This is a direct descendant of the National Bureau of Economic Research (NBER) Time Series Data Bank described by Boschan, Charlotte. "The NBER Time Series Data Bank." Annals of Economic and Social Measurement 1(April 1972):193-209.


Online Numeric Data-Base Systems


23. The Kentucky Council of Economic Advisors is a state government agency, staff support for which is provided by the Center for Applied Economic Research, College of Business and Economics, University of Kentucky. The Kentucky Council of Economic Advisors was created by Executive Order in December 1970 and was confirmed by the Kentucky state legislature in Chapter 147B, Kentucky Revised Statutes in the 1972 regular session. One of its primary responsibilities is to collect, compile, and interpret economic data concerning the economy of the Commonwealth.

24. Examples of these are the Indiana Information Retrieval System (INDIRS) and the Kansas Policy Database System.


26. The Kentucky Department of Commerce, now known as the Commerce Cabinet, has published for many years an annual, Deskbook of Kentucky Economic Statistics, which is not a statistical abstract, but in a limited fashion served some of the same purposes.


28. The Kentucky Economic Information System is maintained on two separate computing networks, and is publicly accessible to anyone with an account on either of these networks. The networks are the centralized state government computing network and the Kentucky Educational Computing Network (KECNET).


30. The notion that these special sets of commands, or execs, mimic an interactive mode is the author's. Roy A. Sigafus wrote the special "user-friendly" procedures for use of the KEIS on KECNET, as well as the KEIS Procedure for use of the KEIS on the state government computing network.
Training of Data Services Professionals:
Past, Present, and Future

Laine G.M. Ruus

Data archives have existed in one form or another for some decades, and the question of the education and training of data archive personnel is not a new one. Yet little has been written to address the issue in the data archive management literature. It would therefore seem appropriate to consider the historical background of data archives* and the current educational scene vis-à-vis data archive personnel, consider some of the pros and cons of the current system, and discuss some alternatives for the future.

During the 1930s the commercial organizations began accumulating large numbers of data files; during the 1940s, the academic research institutions found their storage areas becoming filled with punched cards....Service oriented archives of machine-readable data acquired momentum and importance only after computers became available to a substantial portion of the community of social research scientists: this happened in the late 1950s and early 1960s.1

*The distinction between a data archive and a data library lies in the primary function of a data archive which is to preserve machine-readable data files (MRDF) for posterity, whereas the primary function of a data library is to provide services vis-à-vis MRDF to a community of users. In the context of this paper, however, I use the term data archive to refer to both types of MRDF service facility. Throughout, my remarks concern in the first hand the data archive or library located in an academic institution.

Laine G.M. Ruus is Head, Data Library, University of British Columbia, Vancouver.
Several early data archives, such as the Inter-University Consortium for Political Research (ICPR),* the Roper Center (then at Williams College), and the Zentralarchiv für Empirische Sozialforschung, were established with extramural service commitments, large budgets and large staffs, and could therefore recruit staff with a variety of individual academic and training backgrounds suited to the individual positions they were hired to fill: "Inasmuch as they depend on computers, data archives necessarily must become complex organizations, with staffs that include specialists in computer operations, programming, and data processing techniques, as well as administrators and professional research personnel." As early as 1957, in one of the first major publications to deal with the organization and management of a data archive, York Lucci and Stein Rokkan, in their blueprint for a national American "library data center," planned for an academic as director and a variety of staff, including an archivist, an analyst and a part-time professional librarian.3

As access to computers and use of quantitative research techniques became more widespread, the movement to establish smaller local service data archives extended to many universities: "Most of the existing data archives were founded at centres of social science research, that is, universities. This is particularly true of those created by demands from users at those universities who, in their own work, order data from many and diverse sources." The administrative officers of these local service archives were first and foremost faculty members, with teaching and research responsibilities, who administered the data archive in addition to their primary duties. Any additional staff consisted often of either part-time students or contract research assistants, with possibly some departmental clerical staff assistance.

By the mid-1960s, there was considerable discussion of the feasibility of libraries taking over responsibility for the management of local data services facilities,5 with the corollary expectation that these data archives would be, at least partly, staffed by library personnel: "The library's conventionally trained personnel could learn data-base management and development from the [data] archivists, while the latter profited from the librarian's knowledge of such matters as indexing."6 In the same year, Clifton Brock wrote that "apparently no data archive is operated by or in conjunction with a library."7 This situation has now changed somewhat, but libraries have generally been reluctant to accept this responsibility.

*In 1976 the name was changed to Inter-University Consortium for Political and Social Research.
Training

The earliest efforts to centralize and standardize training were not primarily aimed at data archive personnel. The Council of Social Science Data Archives, one of the first data archive associations, formed in 1962 and consisting of senior data archive administrators, had been primarily concerned with training users, rather than training staff. In 1963, a summer program in statistics and social science research methodologies was established by ICPR at the University of Michigan "out of the belief of Consortium members that it was desirable to supplement the methodological training offered graduate students at a majority of institutions and to permit faculty members to extend their methodological training." A similar summer program in social science research methodologies was established at the University of Essex in 1967.

Efforts such as these were primarily aimed at training a coterie of users of data archives, rather than of data archive personnel, although the training is not wholly incidental to the kind of training required by the latter. A 1967 report to the U.S. National Research Council by the Committee on Information in the Behavioral Sciences stated that: "archives administrators need funds that would permit them to hire service-oriented personnel at the BA level and provide them with the necessary substantive and/or information-processing training to develop a cadre of relatively permanent archival personnel." The method of in-house, on-the-job training of personnel was at that time the only viable means of developing data archive staff. ICPR had conducted occasional short sessions in conjunction with annual ICPR meetings to train official representatives of member institutions in some basic data archive management techniques, but the first major effort to reach those who are in fact managing, as opposed to using, collections of MRDF, to my knowledge, was a two-day workshop on the management of a data and program library held in 1969 at the University of Wisconsin.

Four years later, the situation had not changed substantially, as David Nasatir wrote in his study for Unesco in 1973:

Perhaps the most difficult task in the establishment of a social science data archive is that of staffing the organization. For each of the functions...[performed by the data archive], it is necessary to obtain highly qualified personnel—yet relatively few opportunities exist for individuals to be trained or to gain experience in these activities.

He went on to endorse on-the-job training and the apprenticeship system as viable alternatives: "Due to the difficulty of finding personnel capable of carrying out many tasks of an archive, potential archive
personnel...must often be trained by the archive itself. Archive personnel can also be trained for employment by other archives."

One of the reasons for the lack of earlier development of centralized training for data archive personnel has presumably been money. In the 1960s and early 1970s, jobs were plentiful and job and geographic mobility was high. Many small local service data archives, located as they tended to be in academic institutions, hired personnel from the local major labor force, the student body, which had the advantage that one could select for appropriate academic background, but the disadvantage that this, of all sectors of a mobile work force, was one of the most highly mobile. It is quite natural, therefore, that those in charge of funding should be reluctant to spend any substantial funds to provide training from outside sources for personnel likely to be gone in a year or two. Thus, those for whom any centralized training efforts were aimed were those who had access to travel funds and who had, usually, a more long-term commitment to any one institution—the faculty. Personnel of large archives would have less need for centralized training, having access instead to the expertise of their colleagues and, normally, to an internal training program of some sort, whether more or less formalized.

The last five years have seen several developments. In July 1976, under the aegis of ICPSR, the first (of several) two-week Workshop for Data Librarians was given at the University of Michigan in conjunction with the regular ICPSR summer program, then in its fourteenth year. It is significant that the majority of those attending the workshop had already been managing data archives or data libraries for a number of years. In 1978 and 1979, a similar course was offered at the University of Wisconsin, as a full semester-credit, graduate-level intersession course. In its first year it attracted primarily professional archivists, library school students, and one practicing "data librarian." A rather different recent development has been the introduction of the concept of MRDF as an information resource in library school courses, e.g., at the Graduate School of Library and Information Science, University of Illinois, and at the School of Librarianship, University of Washington. The objective of these courses, however, has not been to train personnel in the skills necessary to manage MRDF, but rather to give future librarians sufficient familiarity with the medium that it can be treated as just another source of information.

There are currently many people working in the field of providing data services—certainly several hundred, and certainly of sufficient numbers and specialization to recognize themselves as a "profession."
Training

Just who are these individuals? Judith Rowe and Carolyn Geda have contended that "some are former programmers, others are trained in the social sciences, and a small number are trained as librarians."16

Recently, the Education Committee of the International Association for Social Science Information Service and Technology (IASSIST) surveyed the members of the association to determine the educational backgrounds and priorities of its members. In the fall of 1980, a questionnaire was sent to all members—characteristic of mail-back surveys, the response rate was low (less than one-third). What follows is a preliminary synthesis of some of these responses—a full report will be released at a later time.17 It must be borne in mind that managing MRDF, for the purposes of the survey, was defined very broadly, so as to include not only those managing an actual collection of MRDF, but also those associated with managing or disseminating information about MRDF, whether or not in direct association with a collection of MRDF. And indeed, about 10 percent of the respondents were not directly associated with any collection of MRDF.

Of the MRDF collections or data archives represented, 70 percent were located in academic institutions. Of these, 56 percent were administered by academic faculties, colleges or departments; 28 percent were administered by libraries and/or computing centers; and 16 percent by independent institutions. Of the nonacademic data archives represented, two-thirds were located in private nonprofit or other research institutions, and the remainder in government agencies at various levels of government. In terms of staff size, 16 percent of archives represented had no full-time permanent staff; 56 percent had "small" staffs of one to three full-time permanent staff members. At the other end of the spectrum were very large data archives with staffs of between ten and twenty-five people (19 percent of respondents); a few "medium"-sized archives with a permanent staff of five to ten were also represented. A full 70 percent, however, of these facilities also employed part-time or temporary staff—these were almost all data archives located in academic institutions.

When asked if this was the respondent's first job involving managing MRDF, over 60 percent responded yes, and fully half the respondents had been working, in total, in jobs managing MRDF for six years or less (some, indeed, at the time of the survey, for less than one year). The attributes considered most important in getting those jobs were, in order (note that percentages do not total 100 due to the possibility of a respondent giving more than one answer):

- academic training in the social sciences (30 percent),
programming (27 percent),
previous experience managing MRDF (for those into their second or
later job) 27 percent,
quantitative research techniques (21 percent), and
training in library science (21 percent).
In terms of usefulness to the job, however, academic degrees in comput-
er science (5 percent) were rated as uniformly indispensable, those in
library science (20 percent) rated in a range from "occasionally useful" to "indispensable," and all others, i.e., in academic disciplines at the
bachelor's, master's and doctorate levels, ranged the full five-point scale
from "utterly useless" to "indispensable," with 60 percent considering
the degree "frequently useful" to "indispensable."

Regarding additional, job-related training, of those who had taken
additional training (80 percent): 50 percent had taken local courses in
programming, 47 percent had taken local courses in statistics, 45 per-
cent had taken local courses in social science research techniques, 20
percent had taken an ICPSR MRDF management course, and 20 percent
had taken local courses in management techniques. On the other hand,
20 percent of the respondents had taken no additional training
whatever—one assumes, therefore, that for these individuals on-the-job
training suffices.

The object of education is, of course, acquisition of skills. What
skills, then, do the practicing "professionals" consider most important?
Rated "very important" (in order of popularity) were:
data management techniques (70 percent),
familiarity with canned programs (60 percent),
data verification techniques (45 percent), and
secondary analysis techniques (40 percent).
Rated "moderately important" (in order of popularity) were:
statistics (65 percent),
survey methodology (55 percent),
reference (48 percent),
cataloging (48 percent),
indexing and abstracting (45 percent),
personnel administration (43 percent), and
original programming (40 percent).

What we have, then, is a rather young profession, in terms of experience,
which has a good university-level general education needing to be
complemented by additional training in certain skills. Few have faculty
status, and I would hazard that, based on personal experience, few have
access to generous amounts of travel funds. Therefore, the majority have
Training

acquired the additional training they need at the local level where courses are given in a variety of disparate but applicable skills. Because most data archives have very small staffs, I doubt that any formalized on-the-job training programs exist in most cases; on the other hand, this same circumstance demands of the individual a very broad range of skills and general knowledge.

These, then, are the current training options:

1. summer school courses in MRDF management, statistics, survey methods, social science research methods, including the ICPSR training program in the theory and technology of social research (which includes a data management component), and the University of Essex summer school in social science data analysis and collection. Any of these requires a time commitment of two to eight weeks, availability of travel and tuition funds. All of these constitute important skills, the most important being the MRDF management component, but none can stand alone.

2. the semester-long course at the University of Wisconsin, which requires the commitment of a full semester of time, in addition to availability of travel and tuition funds.

3. courses, at any local university, in a variety of subjects, dependent on the size and sophistication of the local institution (and access thereto) and the quantitative orientation of its departments. In this manner, many of the ancillary skills needed can be acquired with a minor commitment of time and money—but not (normally) specific training in MRDF management techniques.

Before considering what might be done, the basic issue of primary responsibility should be addressed. Whose is the fundamental responsibility of acquiring or providing this training? Is it the individual’s responsibility to provide himself with the appropriate training for the job before being hired, or is it the responsibility of the institution hiring him? The answer lies in the balance of supply and demand. When demand is greater than supply, institutions will hire underqualified personnel and train them; when supply is greater than demand, it becomes the individual’s responsibility to acquire the training and then compete for the demand. For this to happen, of course, supply and demand must both be using the same forum. Institutions, however, often prefer to hire internal applicants (more than 50 percent of the survey respondents were hired from internal positions), whether or not qualified, arguing that there are no trained external applicants, and yet external applicants are looking for that chimera, the job opening, and not finding it. If institutions continue to hire untrained internal appli-
cants for such positions, there will be little incentive for individuals hoping to enter the field to spend the time and money to acquire the training beforehand.

What, then, are the alternatives for the future? We can continue to maintain the status quo, endorsing a system consisting primarily of on-the-job training supplemented by local courses in ancillary skills and the continuing summer program courses now being given at the universities of Michigan and Wisconsin (and, I understand, the University of Essex as of 1981). The advantage of this system is that it demands no major commitment of time or money on the part of either the individual or the institution employing him. The disadvantages are that it is not possible to do much more than introduce techniques of MRDF management in the short space of two weeks, it is certainly not possible to give thorough training in these techniques plus ancillary skills, nor is this system conducive to the development of interarchival standards and a professional body of thought. It may, however, be an appropriate means of promoting continuing professional education.

We can adopt Nasatir's suggestion of establishing an apprenticeship system, with the cooperation of the existing large data archives, who would presumably be first in line to be host training centers. But, as venerable as may be the traditions of the apprenticeship system, it is neither efficient nor comprehensive enough to satisfy most training requirements, and it will require fairly substantial investments of time and money. In addition, it is doubtful whether it would contribute much to the development of professional standards.

We can promote the establishment of graduate-level programs dealing with MRDF management techniques as well as other ancillary skills within university schools or departments, such as library schools. Attending such a program would require a major investment of time and money on the part of the individual, or hiring institution, but these should be outweighed by the benefits to be derived from the availability of pretrained staff, thus avoiding the lengthy and often wasteful "trial-and-error" methods of on-the-job training. This is a route that is being favored by many traditional archivists, whose discipline has been attempting to solve a similar dilemma vis-à-vis education since the mid-1930s when the first Society of American Archivists committee on training was formed. Part of their dilemma has been location of such a program:

We find that some archivists perceive librarianship as a profession of low status in comparison to the academic world. This perception may cause them to fear identification with librarians. Recognizing that they cannot be purely academics, yet not wanting to be identified as
I suspect that this perception is not unique to traditional archivists. The 1976 edition of the Association of Canadian Archivists curriculum guidelines neatly begs the issue: "It has been usual to place archival training either in a school of library studies or a department of history, but there is no reason why it could not be part of a school of management sciences. Wherever it is situated it will be necessary to offer a type of programme acceptable to the university as well as to the profession." In 1973, David Nasatir expressed a similar sentiment: "I am assuming that library schools, information science departments, or perhaps social science institutes will develop courses in data library management that do not now exist."

The development of library schools in the latter half of the nineteenth century was a response to several parallel influences, including the immense growth in North America of public lending libraries, growth in the size of library collections, the demise of the apprenticeship system, and the rise of schools of "technical education" as a more efficient means of educating a labor force. At that time new librarians were trained by the apprenticeship system, i.e., in-house training in existing libraries; at that time also, established librarians were complaining of being constantly plagued by others asking for procedural information—reminiscent of a similar plaint raised by Rowe and Geda. Hence, the creation of schools of library economy as a more efficient means of turning out the large number of trained librarians that the market demanded.

If one were to attempt to draw parallels between this development and the contemporary data archive scene, one would first need to know the historical and current growth rates of data archives and data libraries—statistics which, to my knowledge, are not available. One could, of course, hazard some guesses based on the growth of memberships in ICPSR, ECPSR (European Consortium for Political and Social Research, the European membership arm of ICPSR), and ISLA (International Survey Library Association, the membership arm of the Roper Center at the University of Connecticut), bearing in mind that these will not include most specialized data archives, nor those whose budgets are too small to allow memberships in these organizations, nor those data archives in other disciplines, such as the humanities or the physical sciences. Or, one could use the current explosion in the creation of data files as a measure of potential development, if indeed these figures were
available. One would also need to make some predictions as to the effect of recent university funding crises and the current more stringent government funding cuts on the growth of data archives. These developments may, on the one hand, spell the demise of data archives with marginal funding bases; on the other hand, they may spur the development of more securely funded data archives, representing as these do attempts to rationalize data management and acquisition and avoid duplication of effort and duplicate spending of scarce funds. Whatever the future effect of these developments, the present trend seems to be to continue the establishment of local service data archives—some more specialized, others very generalized—at local academic institutions, in government departments, and in the private research and corporate sectors.

The demand, therefore, for trained or willing-to-be-trained personnel persists. This would indicate the need for some form of centralized education system. Whatever form it takes, it must speak to several needs:

1. It must provide training for newcomers to the profession, to an acceptable standard of background knowledge and expertise to allow them to adapt efficiently to their positions.
2. It must complement the on-the-job training of newly practicing professionals, raising their level of expertise to a level so as to allow them to work more efficiently.
3. It should, in addition, be capable of providing continuing education, i.e., training and education in new developments in both technology and techniques, to those with many years of experience in this field, which is so very sensitive to the rapid developments in computer technology.

One of the first requirements is to establish basic standards and curricular guidelines as to what should constitute a basic education for the profession. This is not the place to expound on the eventual contents of these guidelines; some of the requirements basic to this standard have been discussed elsewhere. Others can be extrapolated from the survey conducted by the IASSIST Education Committee. Suffice it to say that the profession must establish the standards to which it should be educated, so that those institutions which may take the initiative to develop such educational programs may develop programs which will meet the needs of the profession.
Training

References

2. Ibid., p. 40.
6. Bisco, "Research Library and Data Archives."
11. Nasatir, *Data Archives*, p. 44.
12. Ibid., p. 54.
23. Nasatir, *Data Archives*, pp. 44-48; and Rowe and Geda, "Training the Professional."
Citation Analysis of Data File Use

HOWARD D. WHITE

Introduction

An argument by no means new is that social scientists who work with machine-readable data files (MRDF) should cite them in their writings, with formal references set apart from main text, just as they now do books, papers and reports. Large-scale suppliers of the files urge this so that their de facto role as data publishers will be properly credited—for nonprofit organizations, an important form of reward. Data librarians have urged it on bibliographic grounds: a data file that is properly identified in a citation, and not just vaguely alluded to in the text, is easier to track down. Moreover, since the advent of Social Sciences Citation Index (SSCI) in 1972, a few commentators have urged it on grounds that such citations—at least those in the journal literature—would be picked up by SSCI and would constitute a "use-history" of data files of great potential interest to all who perform or promote secondary analysis. This paper addresses the last concern—the current state of affairs with respect to citation indexing of data files in SSCI. To anticipate a bit, the situation is chaotic, but not without possibilities for improvement.

A hypothetical example will show how SSCI works and why the inclusion of data files among the cited documents in it is of importance to researchers. The file to be used for illustration is one that I recently cited in a paper of my own:

Howard D. White is Associate Professor, School of Library and Information Science, Drexel University, Philadelphia.
HOWARD WHITE


This is a survey of 2486 adult Americans conducted in early 1970 for the Commission on Obscenity and Pornography; its results were discussed a decade ago in the commission's *Report*, and mentioned as recently as 1979 in Gay Talese's *Thy Neighbor's Wife*. The original analysis of the file was carried out by Response Analysis Corporation of Princeton, New Jersey, and reported in a technical monograph published by the commission as a U.S. government document, *National Survey of Public Attitudes Toward and Experience with Erotic Materials*. But the file has also been used in several secondary analyses reported in the journal literature, and if these papers had cited the file more or less as above (they do not), the Citation Index of SSCI would have routinely picked them up with some such entry as this:

```
Temple U, I Surv Res
Opin Surv Curr 70
Wilson WC J Soc Issue 29 19 73
```

and later, under the same heading:

```
Wilson WC J Sex Res 11 46 75
Wilson WC J Soc Issue 31 69 75
Glassman MB Pub Opin Q 42 161 78
```

My paper would recently have been added to the chain:

```
White HD Library Q 51 192 81
```

One translates these highly condensed entries into full bibliographic listings in SSCI's Source Index; for example, the paper in which Glassman uses the "pornography survey" is, in full:

```
```

But the point is that ideally one can trace the use-history of this file by examining the chain in the various issues of SSCI; or, if one has access to its online version, Social Scisearch, the entire history is available in cumulated form simply by inputting the name of the pornography survey file.

Obviously such cumulated histories should interest data suppliers, like the Roper Center or the Inter-University Consortium for Political and Social Research, since they reveal the use of their offerings in
Citation Analysis

published papers (even if an ill-bred citer fails to mention them as distributors by name). One would expect social scientists to be interested even more, since the use-history registered in SSCI reflects the formation of an identifiable community (of sorts) around a data file. For various reasons (curiosity, rivalry, etc.) geographically dispersed users of the same data want to know of each other’s work. Use-histories thus have considerable human interest value, and the value may be intensified by the highly specialized nature of much secondary analysis. Only a relative few will work with a special-topic file like the pornography survey, or with a particular set of questions (such as those on abortion) in an omnibus file like the General Social Survey. One wants to know who and where they are, what technical problems they encountered, and especially what their findings are. Citation indexing can lead directly to answers to these questions in a way that conventional subject indexing does not.

What, then, is the case: are use-histories of data files available at all? The answer, not widely known, is that citations to data files do appear in SSCI and have for some years. This is not to say that large numbers of social scientists are taking to heart the counsel of such writers as Rowe1 and White,2 and citing files in the style recommended by Dodd.3 They are not. But a fair number of researchers have in one way or another acknowledged use of files—particularly codebooks—in their footnotes or endnotes, and this has been sufficient to leave at least partial use-histories in SSCI (and in Social Scisearch). The rub lies in the phrase “in one way or another.” The lack of a consistent citing style, in combination with editorial practices of SSCI’s publisher, the Institute for Scientific Information (ISI), has resulted in a rather spectacular scattering of the citations to any given file, and only the most determined labors of reassembly—i.e., checking SSCI at many different points—will produce a coherent use-history such as the one above.

The Causes of Fragmentation

There are several ways in which this scattering, or fragmentation, comes about. Basic to the problem is that entries in SSCI are keyboarded directly from the texts of papers in journals, with little or no editorial intervention to correct discrepant citing practices. Then the entries are automatically filed by computer, which is not programmed to reconcile two citations to the same work if they include different elements or begin in different ways.

A major unreconciled difference occurs when researcher A cites a data file with the author as the first element, and researcher B omits the
author and cites it with title first. This has the effect of throwing the two references into wholly different sections of SSCI's Citation Index. A file cited with title first goes into the section reserved for anonymous works, which is something of a bibliographic slum (cf. the opinion of Garfield, ISI's president4). If author is put first, the reference of course goes into an author section, but there are two of these—one for personal authors and another for corporate authors. Unfortunately, many MRDF can be cited by either type of author, and this is where another unreconciled difference occurs. If the file was created in a project with a principal investigator (PI), and if the citer puts the PI—a person—first, the reference will be placed with all the other personally authored works (papers, monographs, etc.) that make up the bulk of the SSCI Citation Index. But if the citer omits the PI and puts a producing or distributing corporate body in author position (as I did with the pornography survey), the reference will go into the corporate author section. (Occasionally, too, a corporate author entry is shunted to the anonymous section of SSCI by mistake.) Thus three different researchers who had worked on the same file, perhaps even on the same set of variables, could find their identically intended citations in three separate sections of SSCI, depending on their choice of first element in citing. Their citations would also be placed in three separate parts of the Cited Reference index to online Social Scisearch.

Further fragmentation occurs within each of the sections. Citers who choose the same first element in their citations often differ in the ways they record titles, or personal or corporate authors. One very common type of fragmentation in SSCI occurs with cited personal authors: researcher A cites by surname and first name (or first initial); researcher B cites by surname, first name and middle name (or initial). As transcribed and computer-filed in SSCI, this causes citations to the same work—say, by James N. Morgan—to be entered in two different places, as the arrows show:

```
Morgan J
Morgan JA
Morgan JB
Morgan JN
```

LIBRARY TRENDS
Citation Analysis

Thus, even in the relatively simple case of a principal investigator as first element, one must look in two places to avoid missing all citations. With corporate authors as first element, the fragmentation is much worse. Librarians over the years have devised elaborate rules for dealing with corporate authors in card catalogs, but even they have had trouble in achieving consistency, and have on occasion changed the rules. Pity then the citers and journal editors: they may follow style guides, but the overall result is bibliographic anarchy. Take, for example, this nonexhaustive list of ways to render the U.S. Census Bureau as author of a file:

- Bureau of the Census
- Census Bureau
- Department of Commerce, Bureau of the Census
- Department of Commerce, Census Bureau
- U.S. Bureau of the Census
- U.S. Census Bureau
- U.S. Department of Commerce, Bureau of the Census
- U.S. Department of Commerce, Census Bureau

The same data file—e.g., the County and City Data Book, 1972, produced from Census Bureau tapes—could be cited under any of these, with the consequence that a thorough searcher must check at least eight different positions in SSCI's alphabet of corporate authors for possible entries.

Inconsistent renderings of titles are no less a problem. There are two major reasons why title citations to the same work may turn up in widely different positions in the anonymous section of SSCI. One is that many data files are actually known by several titles—a source of confusion documented by Dodd—and researchers reflect this diversity when they cite. The other is that even researchers who use the same title do not always record it in the same way. For example, one person may write "General Social Survey 1972" and another, "1972 General Social Survey." The latter style, with year first, will almost surely cause citations to be lost to some users of SSCI or Social Scisearch: the computer puts entries starting with digits wholly outside the alphabetical sequence, and the person searching alphabetically for a title may never think to look in the numeric positions following "Z." Many MRDF can be cited with either a word or a string of digits (such as a year) coming first and determining where the entry will be computer-filed. So, again, the potential for scattering is great. It should also be noted that these problems with titles persist when the title is the second element in citing, after personal or corporate author, since both first and second elements are used in computer sorts.
It should now be clear that anyone who wants to examine the use-history of a data file in SSCI has a time-consuming task ahead. One needs to look in many places to achieve both positive success, which is finding entries, and negative success, which is ascertaining that there is nothing in a particular place to be found. The essential problem with SSCI, whether we want use-histories of data files or anything else, is insufficient vocabulary control. This is a classic problem in creating large and growing bibliographic files: its ramifications were recognized long ago by library catalogers, whose response was to create: (1) authority lists that standardized personal and corporate author names; and (2) uniform titles that conveyed the fundamental identity of works, despite the multiplicity of editions, versions, translations, etc., of these works appearing under diverse names. (Thus, the uniform title *Arabian Nights* in card catalogs unites all editions of this work under a single heading, whether they are titled *Arabian Nights, The Thousand and One Nights, or Tales of Scheherazade.*)

Lubetzky and Hayes in 1969 directed attention to the fact that Science Citation Index was failing to unify references to specific intellectual works because it merely transcribed citers' references to editions of these works.6 (They use a paper by the American physicist F. Willard Gibbs as an example.) The problem remains, in both SCI and its newer companion SSCI. Griffith recently noted that a computer search of SSCI tapes failed to show *Das Kapital* as heavily cited.7 The reason is not that this most influential of writings is not heavily cited, but that the citations it receives are scattered among many different editions and many different citing styles. To a computer, it seems that many different works are being cited only a few times each: one by Marx, another by Marx K; one named *Das Kapital*, a second named *Kapital*, a third named *Capital*, a fourth a volume in Marx's *Collected Works*, and so on. Exactly the same thing has happened with MRDF.

The Institute for Scientific Information is aware of the varieties of fragmentation recorded here. ISI's problem is economic: it is prohibitively expensive to make the copy of thousands of citers conform to authority lists of authors' names and uniform titles. If this work is to be done, it will very likely have to be done by outsiders—a point to which I shall return in closing.

An Experiment in Finding Citations to Data Files

Three major data files—or rather sets of data files—were chosen for an experiment in citation retrieval in SSCI: the General Social Survey,
Citation Analysis

conducted annually since 1972 (except 1979) by the National Opinion Research Center (NORC) in Chicago; the Panel Study of Income Dynamics, conducted annually 1968-78 by the University of Michigan's Institute for Social Research; and the same institute's American National Election Studies, conducted biennially in election years since 1948. Copies of these files are held by hundreds of colleges and universities, and are known to be used by social scientists and their students. One would expect at least some of this use to be registered in references in published papers; and in fact some is. A trial manual search of SSCI, 1973-79, produced 110 citations to the General Social Survey, 47 to the Panel Study of Income Dynamics, and 23 to the Election Studies, as of August 1980. The search was "uncritical" in that all of these files are actually multiyear serials with heterogeneous content, and a citation to any part in any year was counted as a hit. On the other hand, many of the questions in these surveys are repeated over the years, and persons citing files issued in different years may be using the same questions. In any case, the point of the trial search was to find as many citations to the three files as possible, without worrying about refinement by year or by subject.

Table 1 sets forth the various author and title headings under which citations to the three files were found. To keep the table from being unmanageably complex, not all variants in entries have been listed. Even so, the dominant impression from the table is one of complexity and high fragmentation, in sharp contrast to the earlier, idealized example in which a single entry named the pornography survey. Anyone compiling a use-history of the data files in table 1 (or of any others) must in fact search along lines suggested there, and earlier in this paper, if near-completeness is to be attained. Note not only the divergent forms of the same heading, but also the several wholly different headings under which one finds entries in the sections of SSCI.

The search for citations to the General Social Survey (GSS) yielded use data that could be compared to those in NORC's 254-item Annotated Bibliography of Papers Using the General Social Surveys of April 1979. The NORC compilers, while acknowledging that their list is far from complete, state that they included "a computer-assisted check of Sociological Abstracts, Dissertation Abstracts, and the Social Science Index..." in doing their search. It is not clear whether the latter is H.W. Wilson's Social Sciences Index, which cannot be searched by computer, or the Social Sciences Citation Index, which can. Interestingly, however, the manual search of SSCI for the present article turned up fully sixty papers citing one or more annual issues of the GSS that are not
<table>
<thead>
<tr>
<th>Data File</th>
<th>Section and Headings*</th>
<th>Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Social Survey (various years)</td>
<td>Personal Author</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>—Davis J</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—Davis JA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Corporate Author</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>—Nat Op Res Ctr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—Nat Opin Res Ctr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—NORC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—Rop Publ Op Res C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anonymous</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>—Codebook Spring 197-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—Codebook 197- General</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—General Social Surv 197-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—National Data Progra 197-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Panel Study of Income Dynamics, 1968-1978</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Personal Author</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—Morgan JN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Corporate Author</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>—1 Soc Res</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—1 Soc Res Surv Re</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—Mi I Soc Res</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—Mi U Surv Res Ctr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—Surv Res Ctr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—U Mi Surv Res Ctr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—U Mich Surv Res Ctr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—U Mich I Soc Res</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—U Min Surv Res Ctr [sic]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anonymous</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>—Panel Study Income D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—Pan Stud Inc Dyn</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—Pan Study Inc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—Pan Surv Inc Dyn [sic]</td>
<td></td>
</tr>
<tr>
<td>American National Election Study (various years)</td>
<td>Personal Author†</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Corporate Author</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>—Ctr Pol Stud</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—U Mich</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—U Mich Ctr Pol St</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—U Mich Interu Con</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—U Mich Pol Beh</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—U Mich Surv Res C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anonymous</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>—CPR 197- Am National [sic]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—CPS 197- Am National</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—CPS Am National Elec 197-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—SRC 197- Am National</td>
<td></td>
</tr>
</tbody>
</table>

*Author and title headings for this search were derived from title pages of codebooks and Inter-University Consortium for Political and Social Research. Guide to Resources and Services 1979-1980. Ann Arbor: University of Michigan, 1981. Except where noted, hits were found under all headings listed.

†No hits were found under Angus Campbell or Philip E. Converse, both of whom have been principal investigators for this survey.
Citation Analysis

recorded in the NORC bibliography. A few of these are discussions of the GSS as a resource in the librarian's or data archivist's sense. But the great majority are substantive research papers from the same period as those contained in the bibliography. Some are from "major social science journals" of the sort the compilers say they searched (e.g., Social Problems, Human Relations, Social Forces, Sociological Quarterly, Political Science Quarterly); others are from "unpredictable" specialty journals (e.g., Law and Contemporary Problems, Archives of Sexual Behavior, Journal of Communication, Review of Religious Research, Personnel Psychology, Curriculum Inquiry, Journal of Homosexuality), indicating quite vividly the cross-specialty diffusion of GSS data.

These remarks are not intended to derogate the NORC bibliography, which is a valuable work, especially its notes on specific GSS variables employed by researchers. Rather, they are intended to show that SSCI can reveal use of data files even to persons, like those at NORC, who are well placed to know about such uses. However, SSCI's full potential can only be brought out by searchers who know its peculiarities and are willing to look in many different places for citations.

Prospects for Use-Histories

Over the next decade, it may be that both researchers and editors of journals in which they publish will settle on a few more or less standard ways of citing MRDF. The goal is not just proper credit for a file's originator, producer and distributor, but its retrievability as an intellectual work. Citers need to learn to see citation as a contribution to document retrieval—no less so when the "document" is an MRDF than when it is another author's monograph or paper. Toward this end, it would greatly help if journal editors published model citations to data files in their instructions to contributors, just as they now do for works of other kinds. Such model citations should also be incorporated as soon as possible in widely used style guides, such as the Modern Language Association's and Turabian's.

There are examples on which to draw. For the GSS and certain other files, a standard citation now appears in the front matter of the codebook. These are influenced by the style developed for cataloging data files in the Anglo-American Cataloging Rules, 2d ed. (AACR2). Under AACR2 (which is superior as a guide to the ANSI standard for bibliographic references to data files), the choice of initial element in a citation may come down to two: principal investigator's name, like that
of any other author; or title when the PI is not known. Corporate author entries seem to be falling from favor. With some such standardization in the author and title fields, use-histories of data files would be easier to compile.

However, skeptical wisdom suggests that citers and editors will continue to go their idiosyncratic ways, which jointly yield the fragmented entries now in SSCI. Skeptical wisdom also suggests that the fragmentation will not be corrected by human intervention at ISI (although computer algorithms for resolving some of the differences in table 1 seem possible). Those wanting use-histories of data files can only expect to earn them, in the foreseeable future, by hard digging—i.e., by multiple look-ups in a manual search of SSCI and by much consultation of the Cited Reference index in Social Scisearch online. The present article shows that hard searching, using all the entry points implied by table 1 plus any others that appear plausible, does produce hits, and that is perhaps its most encouraging finding.

Researchers should benefit from this knowledge, since they gain from being able to extract use-histories of files of interest. But one hopes that not all use-histories will remain private documents in the hands of researchers. There is a need for published bibliographies of works based on data files, like that of NORC for GSS. The compilers presumably would treat items they found in SSCI, however diverse, as "raw copy" to be bibliographically standardized (in forms of titles, authors' names, etc.) and newly arranged in some appropriate order. It would seem natural for such bibliographic projects to be sponsored by major data suppliers (e.g., the Inter-University Consortium, the Roper Center), at least for data files they know are widely held and used. It would also seem natural for the federal government, which issues so many files as government documents, and which now actively promotes secondary analysis, to take an interest in use-histories as a newly deserving form of bibliography, and to support compilation of them adequately with funds.

ACKNOWLEDGMENT

The author wishes to thank Carolyn Latta, his graduate assistant in the School of Library and Information Science, Drexel University, for the arduous search of SSCI reported above.
References


This Page Intentionally Left Blank
The Role of Secondary Data Analysis in Teaching the Social Sciences

JEFF SOBAL

Teaching the social sciences includes several goals. As in any other discipline, information is transmitted to students, instilling social facts into their pool of knowledge. Theories are taught to provide generalizations about social facts. Finally, methods of conducting social research are central components of social science instruction, where the logic of inquiry and the procedures for analysis of data are passed on to students. The goals of information and theory are served adequately, but not entirely, through books, journals and other written sources. Inquiry and methodology are processes, and are best learned by actually conducting research. For this purpose, books are not sufficient, and teachers must turn to other sources. Secondary data analysis, the reanalysis of machine-readable data, is one of the best supplements to traditional teaching methods, especially for teaching research methodology and statistics.

The general state of instruction with secondary data today may be seen in light of the growth of the "secondary data movement." Since the 1960s, there has been a growing emphasis upon the use of secondary data in research, with important developments in social indicators analysis, the rise of survey archives, and the overall development of quantitative social research all playing a part. As noted by Nesvold,¹ much of the development of the use of secondary data in teaching has been a "trickle down effect" from research. As more people became involved in collecting and analyzing quantitative data, their work spilled over into their teaching.

¹ Jeff Sobal is Assistant Professor and Assistant Director of Research, Department of Family Medicine, University of Maryland School of Medicine, Baltimore.
Before proceeding to a discussion of the benefits and problems of teaching with secondary data, we should be clear about exactly what constitutes secondary data. Any data which have been collected for another purpose and later reanalyzed may be seen as secondary data. These would include field notes from ethnographers, coding sheets from content analysis, results from experiments, friendship choices from sociometry, or questionnaires from surveys and censuses, as well as data collected for administrative or other purposes. While teaching may well occur using any of these types of secondary data, this discussion will only consider quantitative, machine-readable data of the kind available from censuses, surveys or administrative records. These are the most common data, and often the most useful type available.

Such quantitative data are often difficult to acquire, especially for students, because they are expensive and time consuming to collect if high-quality information is desired. Yet quantitative training is becoming almost a necessity for all of the social sciences. Since the introduction of the computer into the social sciences in the 1950s, quantitative analysis of data has become an almost essential skill in the disciplines of psychology, sociology, anthropology, political science, economics, geography, history, and education, as well as in related fields such as public health and marketing. In both graduate and undergraduate training, and often in the more sophisticated secondary schools, students in the social sciences are required to learn statistics, the logic of data analysis, and have some exposure to the use of the computer. Yet simply going through the motions of calculating statistics is not enough, because statistics and analysis are meaningful to social scientists as tools for understanding data rather than ends in themselves. Thus data become an important part of methodological training, and the reanalysis of previously collected data is one of the best ways of teaching research methodology.

Exposure to the use of the computer is becoming almost ubiquitous in the social sciences, especially during graduate training, although undergraduate study and even secondary school teaching are not far behind. Secondary data offers instructional opportunities for computing without the problems of data collection. Some form of data is usually necessary in training students in computing, and the availability of secondary data sets enhances this aspect of computing instruction. Once students understand data analysis with the computer, then they need data to utilize their skills. The availability of a number of easily accessible secondary data sources makes it likely that students' computing capabilities will be utilized, and even extended, rather than relegated to memories of skills that they once possessed.
Secondary Data Analysis in Teaching

Some disciplines are more quantitatively oriented than others and therefore need a wider variety and higher quality of secondary data sets. This does not imply that less quantitatively oriented disciplines are to be ignored when secondary data are acquired, because these disciplines are often those in which the use of secondary data can provide the furthest-reaching innovations in teaching methods, and are those whose practitioners are the least likely to have sufficient awareness of the availability of secondary data to request their acquisition. The availability of secondary data relevant to a discipline and the publicizing of their availability among teachers can help to add important quantitative perspectives in classes where there previously had been no exposure to the perspective. In addition to noting the range in quantitative analysis among disciplines, there is also a great deal of variation in their traditional reliance on secondary data and future receptivity to teaching with secondary data. People in economics and geography have traditionally worked with quantitative data collected by others and accept the use of secondary data in the classroom with relative ease compared to fields such as history, which may rely on available data but infrequently use quantitative materials, or psychology, which commonly stresses quantitative analysis but infrequently utilizes other people's data. The reasons for these variations among disciplines range from the type of subject matter, to the favored methodology, to norms about how teaching and research should be done in that discipline. These variations are mentioned both so that data archivists and librarians can be sensitive to problems that will be encountered, and so that unexplored areas of teaching can be considered.

Type of Secondary Data for Teaching

There are many types of secondary data available for instructional use. A number of simulations and games are currently available, such as the EXPER-SIM series, where investigations are undertaken by students on artificially created data. These provide a valuable experience in simulating research, but may not be taken as seriously by students as is working with actual data. At the same level as these artificial models are a number of excellent teaching packages which utilize secondary data. These include SETUPS, the Minnesota Manuals, ICPSR Instructional Subsets, International Studies Association Learning Materials, and other packages available from CONDUIT and the University of Iowa Political Research Laboratory. All of these packages share common features, including easily accessible programming for student analysis, a relatively small sample size, a selection of a small number of
variables interesting to students, and manuals for student use. Some also include additional instructions for teachers. These are easy to use for both pupils and instructors, and well worth including in any data archive or library which houses secondary data. Brighter students can often teach themselves how to use them, and easy availability will facilitate this.

However, the simplicity of these instructional packages can also be a drawback. They provide only a limited amount of material for students and can lead to boredom with repeated use. Especially when different types of packages of the same degree of difficulty are used in several classes, instructional packages are no longer challenging and become routine. This is not to say that all students can easily master their use, especially their intended use of grounding theoretical conceptualizations in data. Some pupils will be so overwhelmed by the mechanics of assessing a computerized teaching package that they will have difficulty learning much from the experience. Special support by teachers and their assistants is often necessary to achieve learning goals and practical mastery of the technique in these cases.

For students who have mastered teaching packages which use secondary data and for other pupils ready for more challenging analysis, the logical next step in teaching is the use of any available secondary data set. Here all of the opportunities and problems of full-scale data analysis are faced by students. Availability is the key to effective instructional use at this level. If data archives or college libraries have codebooks within easy access of goal-directed researchers as well as browsers, their use will be maximized. For teachers to adopt secondary data analysis as a pedagogical technique, they need to be aware of the data available. For students with some capabilities in secondary data analysis or those willing to learn about the technique, exposure to the data sets is vital.

Within the diversity of data available for teaching, several characteristics of the data set are important considerations in selecting data for student use. Obviously the topic of the data must be consonant with that of the class, but this can be specified so narrowly as to restrict creativity and fail to accommodate diverse student interests. General omnibus surveys, such as the National Opinion Research Corporation’s General Social Survey, usefully serve a wide range of interests for both students and instructors. They are especially useful where a class with a diversity of interests uses a single data source. Yet for more advanced courses and student papers, these broad offerings do not provide sufficient depth and need to be replaced by specialized studies focusing on one topic. A
primary function of archivists and librarians is to aid instructors and students in finding and accessing these diverse data sets, in addition to the management of the omnibus studies used in the bulk of teaching.

Another important consideration in choosing data for teaching is the unit of analysis. It must be appropriate to the topic taught, such as using the individual for psychology, the nation for political science, or the culture for anthropology. In addition, students must be able to comprehend data collected on that unit of analysis easily. Generally, students find data using the individual as the unit of analysis the easiest to fathom, with data collected on counties, organizations or nations much more difficult to conceptualize. Other attributes of secondary data deviating from simplicity, such as hierarchical file structures or the presence of subfiles, also make some data sets less than ideal for most student use.

Levels of Instructional Use

There are many levels of instructional use for secondary data, and all can be facilitated by data archivists and data librarians. They include secondary school teaching, undergraduate introductory courses, undergraduate advanced courses, graduate training, professional training, and continuing education. Although the bulk of secondary data use in teaching occurs with advanced undergraduates and graduate students, pedagogical pioneers are currently extending secondary data instruction into other areas. Archivists and librarians can encourage the development of secondary data teaching among those who do not currently use it by advertising their services to teachers within their institutions and being sure that entire library staffs are aware of teaching opportunities at all levels of instruction.

In addition to being a useful teaching tool at many levels of instruction, secondary data are useful in many types of student training. They are an excellent tool for introducing students to a discipline, providing a valuable supplement to lecturing and reading. By actually allowing students to test their own interpretations of a subject matter using real data, students become active participants in inquiry rather than passive recipients of information. The utility of doing original data analysis is greater for advanced courses in a topic where students are even more sophisticated and immersed in the subject matter. Seminars offer a rich environment for secondary data use, with numerous opportunities for testing the hypotheses and ideas raised in discussion by accessing secondary data. As mentioned earlier, methodology and statistics courses
frequently use available data to make the empirical techniques "come alive" with meaning by applying them to real-world data. All of these classroom situations can involve students with regular assignments of data analysis or use occasional activities in class.

Secondary data analysis can prove useful in individual student projects as well as assignments for entire classes. A reanalysis and interpretation of available data is a valuable alternative to the traditional library term paper which simply repeats or reworks the findings of other researchers. In this case, archivists and librarians can play a vital role in making students and teachers aware of the option of secondary data analysis.

**Housing Data in Archives**

Two strategies of making data available can be used, each with its strengths and weaknesses in encouraging instructional use of secondary data. One is the archival strategy of housing data, where all secondary data and codebooks are housed in a separate room or building from other library materials. This provides a convenient center for those using secondary data and often serves as "headquarters" for a network of users and a clearinghouse for information. Yet it also can become isolated from the general student population and overlooked by potential users who were not socialized into the clique or specifically assigned to go there for course work. The centralization does allow a secondary data specialist to be available with the codebooks to explain secondary data analysis to students and researchers. This type of organization is most useful for assigning larger classes to carry out secondary data analysis, where they work together in one site, often developing a great esprit de corps and sharing ideas in the process.

One of the major claims of the sciences as disciplines based upon the model of active inquiry is their involvement in laboratories, both in research and teaching. To share this aspect of science, the social sciences need to develop the equivalent of laboratories. This has been done extensively by psychologists, and has supported their perceived legitimacy as a science. The use of secondary data, the development of data archives, social science research labs, and even special rooms in libraries to house secondary data can extend this "laboratory atmosphere" to other social sciences.

Archives have grown significantly in the United States in recent years, but as Nesvold notes, this growth has been largely vertical and not horizontal. While this has achieved an economy of scale for collec-
Secondary Data Analysis in Teaching

tions of secondary data, it has served the needs of researchers more than teachers. Researchers are concentrated in larger universities likely to have data archives, and those who are not situated in these sites can individually visit large archives for information and data. This is not the case with teaching, especially on the undergraduate level, where instructors are frequently dispersed from major universities and do not have the resources to establish their own archives or bring classes to larger archives. Librarians are in the position to bridge this gap for social science instructors, expanding the data archive movement horizontally to serve the needs of teachers. Their ability to house, access, understand, and disseminate knowledge about secondary data to teachers will have important consequences in the future of social science instruction.

In existing data archives, an expanded teaching role is also desirable. The awareness of the values of secondary data analysis produced by disseminating information about archives in the classroom will help to ensure continued interest and support of data archives. Within the staff of an archive, the designation of one person as a primary teaching coordinator and liaison will ensure the active involvement of at least part of the archive's resources in the instructional services and support.

Housing Codebooks in Library Stacks

The other arrangement for housing secondary data is decentralized, where codebooks are dispersed within the general collection of a larger unit, such as a research center or building. This encourages browsing by those not specifically seeking secondary data, opening up the technique to a larger audience of potential users by making it part of the standard information system of a library. The drawbacks of this include a lack of immediate guidance for potential users and a separation of those examining and working with the data. For independent scholarly projects, though, this situation is ideal.

Another important role of libraries in promoting the student use of secondary data is the housing and provision of codebooks for general use. In the data archive, documentation is clearly understood as specific to that setting, and there are generally people around to interpret the codebook to novices. The meaningfulness of a codebook to the uninitiated is not to be taken lightly. A thick document with column numbers and N's listed next to responses to questionnaire items may appear meaningless to the casual reader who happens upon it in the stacks. The provision of an inserted page explaining "how to read a codebook,"
plus a note explaining who should be contacted for additional information (including access to the data) can turn a seemingly confusing mass of numbers into a valuable information tool. Despite this, other efforts need to be made to convey the importance of returning to the original data and disaggregating them by important subcategories rather than simply accepting the tabulations presented in the printed information. The complexities of actually accomplishing this are perhaps best left to the instructors, but informed librarians can generate interest.

Perhaps the ultimate solution for maximizing the use of secondary data in teaching and learning is to combine the use of archives and libraries. A separate archive could be established, perhaps connected with the computer center. In addition, codebooks of at least the most popular, if not all, data sets available in the archive could be placed in the library, with instructions inside directing the reader to the archive for more information. To facilitate the finding of codebooks in card catalogs, cards for codebooks could be of a different color from other cards. This combination could provide the benefits of both worlds for instruction with secondary data.

Contributions of Archivists and Librarians to Teaching

As the gatekeepers of secondary data, archivists and librarians not only act as passive facilitators for teachers and students who already have clear interests and goals in secondary data analysis, but also operate as disseminators of knowledge about this type of resource. There is a vast potential for using secondary data in both teaching and research, and a corollary need for training students and those who serve students, including teachers, archivists and librarians. Secondary data analysis, access and stewardship is not taught in most graduate schools, although this omission is slowly being rectified. Social scientists knowledgeable in this area can work with librarians to help them appreciate the value of secondary data, and to describe social scientists' needs and methods of meeting them by including secondary data in library collections and establishing data archives as part of the library system.

Librarians and archivists who have worked with secondary data can make important contributions to the instructional programs of social science departments. Most social science departments offer a course on research methods, which is often required of all students. Librarians and archivists can seek out the instructors of these courses and make presentations in their classes, describing the use and access of secondary data to the students as well as offering to help individual students in the future. Seminars on data holdings can also be offered to
Secondary Data Analysis in Teaching

faculty. Especially in graduate departments, which are preparing individuals who will teach their discipline, it is useful to familiarize future instructors with the opportunities available in secondary data analysis, including instructing them how to teach with secondary data. Such activities may include descriptions of successful teaching activities by instructors who use secondary data, who can serve as role models in the next generation of instructors.

While this paper is directed at archivists and librarians rather than practicing teachers of social science, a mention of sources for ideas about how to teach with secondary data is in order. Perhaps the best source of information is the colleague who is currently using secondary data in his or her classes. If there is no such person in one's school (including other departments), then professional meetings offer a place to share ideas and "nuts-and-bolts" teaching suggestions, and an increasing number of sessions at professional meetings are being devoted to teaching. Other sources are professional literature on teaching using secondary data, such as Nesvold, Sobal and Treinen, as well as journals such as Teaching Political Science, Teaching Sociology and Teaching Psychology. While librarians do not have to immerse themselves in this literature, it will be useful for them to know which people are teaching with secondary data at their own institution so that they can refer others to those individuals, and to be aware of teaching sources so that they can refer interested instructors to them.

While teachers, archivists and librarians have been discussed separately thus far, in reality they make up a teaching team. They all have independent primary roles, yet all are interdependent in supporting the highest quality of instructional activities. This division of labor permits a separation of tasks, but also requires a great deal of communication among the various groups involved in teaching with secondary data. Occasional meetings which involve teachers and their assistants, librarians, archivists, and computer center personnel cannot only foster additional teaching activities in this area, but also solve problems, and even head off future difficulties.

Conclusion

The future of secondary data instruction is promising. With the growing demand for quantitative analysis skills among social scientists, faculty will increasingly turn to secondary data to teach them these skills. The changing job market is placing an emphasis on policy and applied research skills, for which secondary data are a valuable resource students need to be trained to work with. As research budgets tighten,
investigators will often look to available data to do the research which will spill over into their teaching. Librarians and archivists need to prepare for this rising demand for secondary data for the laboratory and the classroom, and to be sensitive to the needs of this new constituency of secondary data users.

In the end, secondary data are valuable in themselves as well as models for collecting one's own data. While the presence of quality information without the costs of data collection is a valuable tool, a user of secondary data is a captive of the sample collected and a prisoner of the variables available. By teaching students about the value of data analysis, they will learn a great deal in the process and be sufficiently excited about research to use other data and to collect original data. With a commitment to the advancement of instruction with secondary data by social scientists, librarians and data archivists, this goal should become a reality.

References

5. Director for Educational Activities, Inter-University Consortium for Political and Social Research, P.O. Box 1248, Ann Arbor, Mich. 48106.
7. CONDUIT, P.O. Box 388, Iowa City, Iowa 52240.
8. Political Research Laboratory, University of Iowa, Iowa City, Iowa 52242.
14. Nesvold, "Instructional Applications."
15. Ibid.
Data Confidentiality, Social Research and the Government

DAVID F. LINOWES
MICHELE M. HOYMAN

This paper addresses the issue of confidentiality and privacy of information contained in archives and libraries. To do so in a meaningful way requires an analysis of the broad issue of information privacy in general prior to examining its relationship to the functions of the librarian and archivist.

There is no generally accepted definition of privacy. No less a figure than Supreme Court Justice Louis Brandeis spoke, in 1890, of the individual's "right to be let alone" as being the most valued right of civilized man.¹ Twenty years before that, a Michigan judge ruled that privacy was a "constitutional" right. In fact, privacy rights are not specifically spelled out in the Constitution, although the implication is clearly there. The Third Amendment prohibits the lodging of soldiers in private homes without the owner's consent. The Fourth Amendment protects citizens against arbitrary government search. Furthermore, courts have been interpreting a right to privacy from the Fifth Amendment, which protects against self-incrimination; the First Amendment, which guarantees freedom of speech and assembly; and the Ninth Amendment, which reserves to the people all rights not specifically delegated to the states and federal government.

The Privacy Act of 1974, which established the Privacy Commission² and placed certain constraints on federal agencies, for the first time...

¹ David F. Linowes is Professor of Political Economy and Public Policy, School of Social Sciences, and Michele M. Hoyman is Assistant Professor, Institute of Labor and Industrial Relations, University of Illinois at Urbana-Champaign.
gave statutory recognition to a right of privacy, but did not define it. The present, rather urgent concern for this undefined right results from two phenomena in America that are having revolutionary impacts on society: the rapid advances of computer technology and the ever-increasing expectations which the individual has of both government and business. Technological developments have been so drastic that it is difficult for persons of our generation to comprehend them. We now have the technology to store 100 million pieces of information on an inch-square silicon chip. And the cost of this storage on a piece of silicon has been declining at such amazing rates that it is sometimes argued it is cheaper to store data than to destroy them. The power of silicon has proven remarkable in its ability literally to transform society.

The computer was developed in the 1940s. Early models contained large, bulky vacuum tubes, so that a 1950s computer filled an entire room. With the invention of the transistor, computers grew smaller, but the next generation witnessed a quantum leap forward with the introduction of the silicon chip. Now the inch-square “miracle chip” has the calculating capabilities of an entire roomful of computer hardware of the 1950s. A million-dollar computing capacity of three decades ago costs twenty dollars today, and is 100,000 times faster. Work that required one day then is now done in less than one minute. Data contained in a computer data bank are being transmitted across nations and oceans by way of satellite with the speed of light.

Throughout history, societies have had to adjust when great technological breakthroughs have occurred. It happened with the invention of the printing press, the steam engine, the electric light, the automobile. One of the major convulsions in this generation is being triggered by computer technology. Unfortunately, the law and organizational practices have been lagging behind technological developments. The irony of technology may be that it enables society to gain control over everything, except technology.

The other phenomenon, that of individuals demanding an increasing number of services from all institutions, continues to grow at what appears to be a never-ending pace. From the government the public expects social security, unemployment compensation, guaranteed mortgage loans, and all levels of welfare. From business, the public expects credit cards which give instant credit approval any place in the world, and the ability to make plane reservations in a matter of minutes for any kind of trip to anywhere. Libraries themselves are being called upon to render more and more personal and community services. There is a trend toward computerization to provide faster user service within a
library, and there is a trend toward networks to provide services across libraries.

Administrators responsible for furnishing these services must satisfy themselves of an individual applicant's eligibility by demanding and getting more personal, often sensitive, information. Thus, more and more confidential data are being injected into the system of government and business, never to be destroyed.

Today, data have become, in effect, a new element. They are almost never destroyed, and any one item can be retrieved in seconds. During the manual era, data were regularly destroyed, largely because of the cost of storage. Manual retrieval of one record out of a million was costly and time-consuming, often requiring months. Therefore, masses of accumulated personal data had very limited utility.

Threats to Privacy

The continuation of these developments means that certain practices have developed which are threats to privacy. The following are some of those which a federal policy of privacy protection should address.

List Compilers

Among the organizations that thrive in this country are some that monitor the activities of individuals and report thereon to their subscribers for a fee. For example, there is an organization outside of Chicago that professes to identify those persons in this country who are known to be "attacking or ridiculing a major doctrine of the Christian faith or the American way of life." These include authors of books and articles, speakers, and even signers of group advertisements in leading newspapers. In this organization's files are even the names of those individuals who had been involved with the long-defunct House Un-American Activities Committee. If a person's name appears in its file, he or she is characterized as a person with anti-American or anti-Christian attitudes, and investigative companies using its service so report to their clients.

These lists are developed by obtaining names and addresses from public records such as census tract data and automobile registrations, and by renting lists from private industry, such as magazine and book publishers, credit card companies, and charitable organizations. They are then combined into various configurations by computer to develop desired profile groupings. The final profiled lists are rented out for about three and one-half cents a name.
The inclusion of the name and address of a person on one of those profile listings is the basis for an individual being so characterized, whether justified or not. This information in a person's file could be a contributing factor for an adverse decision, be it an important appointment, promotion or granting of credit. The point is that the affected individual does not know that this information is the basis of an adverse decision nor that this information is being kept. A federal privacy policy, following the general principles established by the Privacy Commission, provides for the right of access of an individual to his or her records and the right to "correct" these records if they are inaccurate. A privacy policy would also limit the right of certain holders of information, such as employers, to violate the confidentiality of employees by providing third parties, such as the list compilers, with such information without the employee's knowledge or consent.

Financial Records

Most people regard their finances as a strictly personal and somewhat sensitive matter. They believe, perhaps innocently, that a financial dealing is a confidential matter between them and their banks, creditors or credit card companies. Unfortunately, most people have little concept how seriously their expectation of confidentiality has been compromised.

Checking account and credit card records for the average person constitute, in effect, an economic and social diary, and yet they are increasingly exposed to a wide array of other persons, such as employers, landlords and just curious neighbors. The recently passed Financial Right of Privacy Act of 1978 places limited constraints on some government agencies' access to financial records, but in most areas the government has almost unbridled access to such records. For private sector inquirers, there are few limitations.

Medical Records and Insurance

Everyone is the subject of medical files, usually more than one. Yet, many people are not aware of how available this information is to insurance companies, employers or anyone else who might have an interest in an individual's medical history for virtually any purpose.

Denver District Attorney Dale Tooley found that private medical records can be and have been improperly obtained from most hospitals in the Denver area, not to mention a "remarkable number of clinics and doctors' offices." He tells of people posing as doctors, nurses and even clerics to get medical records which they can sell. Some insurance companies, employers and others are a market for this information.
Confidentiality, Research & the Government

Organizations that use such services may know more about a person's medical condition than the individual, since medical practitioners do not generally allow a patient to see his or her own records. Laws are inadequate to protect against this kind of behavior, which apparently is sponsored throughout the country by some of the largest companies. The need for more protection of the individual in this area is obvious. The Medical Records Act, which is currently in Congress, provides for guarantees of the privacy of an individual's medical records by forbidding their disclosure without the employee's written permission.

Broader Implications of Privacy

In general, it is widely believed that the balance of power in our society is becoming more and more dangerously weighted in favor of large institutions—government and industry alike. A chief reason is that they are the ones with the information.

In the political arena, computerized capabilities have given pressure groups the power to influence candidate selection and key legislative issues in ways not available before. Massive direct-mail campaigns are key weapons in a lobbyist's arsenal. Information regarding the likes and dislikes, political leanings and preferences of specific groups of Americans is so comprehensive that in some cases an election can be determined before the voting begins. Confirmation by fast information retrieval and the importance of the media in reporting this information are together fundamentally altering the nature of the political process in ways which we are just beginning to know.

Furthermore, this is not a problem which is confined to the United States; in fact, the technological problem itself may create problems for the relations between countries. For instance, some nations want to create electronic barriers to halt the flow of information. They consider information within a country a national resource, much like copper or oil. If information does cross their borders electronically, they want to charge a tariff on it.

The lack of controls over information transmission for processing or use in another country leaves developed nations concerned and developing nations alarmed. Economic data, government data, data from home offices of multinational corporations are beamed through the sky in the normal course of business today. Technology in the United States has advanced so far that many developed countries, as well as Third World countries, lag behind. For example, much information is coming into the United States from Canada for processing, classifying and analyzing because it can be handled much more effectively and
economically here. It is just as cheap to beam data across a border or an ocean as it is to beam it next door. Hundreds of millions of dollars of foreign exchange are exported from Canada to the United States to pay for this service. By 1985, it is estimated that it will cost Canada $1.5 billion per year in foreign exchange, and Canada will lose 25,000 jobs.\(^5\) Several nations have established government agencies to administer privacy and trans-border data laws. The impact of this development on librarians and archivists will be most far-reaching.

**Role of Libraries in an Information Society**

The general image of the librarian's role is one of guardian of circulation records as well as researcher of reference questions for library users. As such, the library has access to certain information about users which may be considered confidential. In the circulation and reference capacity, there is a simple direct link between the user whose confidentiality needs to be protected and the professional who is the protector of this confidentiality. However, the increasing computerization of both circulation and reference systems means that access to these records has increased.

Librarians also may have other roles than just in a circulation or reference capacity. The librarian may be an archivist, in which case the professional's role becomes more complicated. The role of the archivist differs substantially from that of the librarian as regards confidentiality. The job of the librarian is to make available all materials to the user, guaranteeing the nature of the user's research question and the particular sources used as confidential. The role of the archivist differs significantly from that of librarian in that he or she exerts control over who can use the collection, and must protect the "implicit trust" of the deposit of the records by assuring that only serious scholars use the collection.\(^6\) For instance, the librarian would not think of querying the user as to why he or she was interested in a certain topic. However, an archivist will not only question potential users, but will make a professional judgment as to which person will be permitted access to the collection. Therefore, the librarian is concerned solely with defending the intellectual freedom of the reader and his/her right to privacy, whereas the archivist plays a gatekeeper role, sometimes blocking the researcher's access if the researcher is not considered a "serious" scholar. Moreover, the librarian will not necessarily release information on who is researching a certain topic, yet the historical archivist will as a matter of courtesy and ethics indicate to a serious researcher the names of other researchers who have used the collection.
Another role of the librarian as computerization increases is as data archivist for large social science data collections. At the moment this may not be a primary concern of the profession, but as an increasing number of centralized information systems are established, the skills that librarians have will be needed—skills such as cataloging, retrieval and reference. Secondly, the more these resources are developed, the more libraries may be called upon to include in their catalog and reference service a list of social science data sources available. Thus, a librarian can function in two capacities: directly as an archivist who catalogs the vast amount of information in a data set, or as a reference librarian who can help the user find the appropriate codebook and study in order to research a topic. The goal of an archivist is to maximize the use of the data, but the increased use will pose an increased risk of violating confidentiality. Therefore, the role of an archivist in protecting privacy becomes critical.

A final role of libraries which should be mentioned is their role as employers. As employers, they face the same issues regarding the privacy of personnel records as private sector employers. The suggestion here is that library personnel practices, like the practices of many private sector employees, may violate the employee's right to privacy. This may be because of the lack of confidentiality of personnel records, or because of other employment practices which violate employee privacy.

Public Visibility of Information Privacy Problems for Libraries

More specific and more visible aspects of information privacy problems involving libraries and archives usually come to the attention of the public through a controversial incident, such as when a librarian refuses to reveal to a law enforcement agency the name of a person who checked out a certain book. Some of the incidents concerning the confidentiality of circulation records became quite controversial, attracting the attention of an entire community.

In 1970, in both Milwaukee and Atlanta, U.S. Treasury agents requested all slips and inquiries for books on explosives. In Milwaukee, the city attorney ruled that such records were “public records,” at which point the librarian complied. In the Atlanta Public Library, the same request was denied in the absence of a subpoena. In another case, the Seattle Public Library in 1974 released its 1970 circulation records to the FBI when the agency presented a subpoena for the records in connection with a forgery case. In 1974 in Los Alamos, Texas, FBI agents requested the librarian to release the circulation records of certain individuals included on a “subversive” list. The
library refused and said that a court order would be required.  

In 1979 when a police officer in Sudbury, Massachusetts, found a bag of marijuana hidden in the *Oxford Book of American Verse* and asked who had checked the book out, the library director, Helen Lowenthal, refused to tell him.  

Lowenthal cited the code of ethics that librarians have regarding the confidentiality of the user-librarian relationship. The library’s board of trustees subsequently adopted the ALA policy passed in 1971 regarding confidentiality.

In many cases the reason for the controversy is not the release of circulation records, but the disclosure of a person’s reference question—information which librarians consider confidential under their code of ethics. In 1979 in Connecticut, police investigating the burning of a cross asked the library for names of persons using materials on the Ku Klux Klan.  

In 1979, state criminal investigators in Iowa asked the Des Moines Public Library to provide names of borrowers of books on occult practices. The officers were investigating cattle mutilations thought to be the result of cult rites.  

In 1980, in Texas, police officers asked a public library to provide the names of all persons who had borrowed chemistry manuals found at the site of an illegal drug lab.

Not infrequently, privacy problems stem from private citizens who want to spy on one another, and have nothing to do with law enforcement. For example, in 1978 a Kansas newspaper editor demanded access to library circulation records as records open to the public. He wanted to know whether city council members who had rejected a new library building used library services.  

A divorced father in Illinois wanted access to a library’s story-hour records to make certain his child was using his name and not that of the mother’s second husband.  

In 1977 a newspaper editor in Washington State demanded access to the records of a community college library in order to prove that tax dollars were wasted on projectors and other equipment available for loan.

### Technological Changes Affect Libraries

The computerization of librarians and archives poses problems for several reasons: (1) there tends to be more information being accumulated and preserved with computers than without; (2) there are more points of access, therefore, more points to be controlled; and (3) more people are able to share the same material that has been placed into a computer data bank than is possible when only one or several hard copies are available. Hence, with computerization there is more need for monitoring of confidentiality safeguards than with manual files, yet science has not yet given us adequate protective technology.
For example, networks which allow more and more individuals to access the system at the same time compound the problem. Computerized reference service networks are especially vulnerable. For instance, a program in California funded by the National Science Foundation seeks to determine if libraries can be used as linking agents between the general public and information in computer data bases. This raises new issues of determining who uses computers to access what information. Confidentiality takes on intensified concerns when an inquiry is of a sensitive nature, such as a request for planned parenthood information or information on a drug rehabilitation program.

There are already in existence today long-distance, high-speed, interlibrary facsimile links to keep scientists in one laboratory in touch with the literature resources of a distant facility. One particular service enables rapid access to scientific information and exchange of research documents over telephone lines between marine biology centers in Florida and Massachusetts. The digital facsimile transceivers by Rapi-com link not only the 170,000-volume, 2500-medical journal library of the Health Center at the University of Florida’s main campus in Gainesville, Florida, with the C.V. Whitney Laboratory for Experimental Marine Biology and Medicine on the Florida coast five miles away, but also with the Marine Biological Laboratory in Woods Hole, Massachusetts.

One of the more difficult areas in terms of issues of professional ethics and the rights of a user to know is in the area of medical research questions. Often the librarian is trapped between trying to determine whether to provide information to the user consistent with the role of librarian, or whether to refer the question to a medical doctor. Generally, librarians view their role as that of providing information wherever possible, provided they are not called upon to make diagnoses or judgments that are more appropriately the domain of a medical doctor. Some libraries may have their staff prepare answers to medical questions but do not open their medical collections to the public. An interesting recent survey showed that 82 percent of the publicly funded medical school libraries are open to the public. Thirty-two percent offer public services other than access.

The kinds of developments which increasingly pose disturbing potential threats to the confidentiality of sensitive data have prompted a strong professional response from librarians. The American Library Association adopted a “Policy on Confidentiality of Library Records” in 1971 and amended it in 1975. The policy statement sets forth three basic principles for the guidance of its members: (1) the obliteration of all patron records when there is no longer a bona fide need for them;
(2) the use of an identifier other than a social security number; and
(3) the development of safeguards to eliminate unofficial monitoring of
communications channels used in library research.\textsuperscript{22} The problem with
self-regulation through professional training and codes of ethics, how-
ever, is that an increasing number of library personnel are nonprofes-
sional employees. Also, given the special vulnerabilities that
computerization brings, the expectation of voluntary compliance spear-
headed by librarians may be naïve.

Given these kinds of problems, the Privacy Commission recom-
mended many different actions. Although the specific recommenda-
tions of the commission total over 160, they embody only a handful of
guidelines and principles to be applied to all information involving
people. The goals of the Privacy Commission are threefold: (1) to
minimize intrusiveness, (2) to maximize fairness, and (3) to create a
legitimate and enforceable expectation of confidentiality where such
expectation is warranted.

To accomplish these objectives, there are certain principles to
which administrators should adhere. First, they should develop an
appropriateness test for the collection of information; second, they
should provide the protection of confidentiality; and third, they should
guarantee the right to disclosure. Only information that is relevant to
the decision at hand should be collected, and it should only be used for
that purpose. Before an organization transfers these data to a third
person, it should obtain the approval of the person whose record it is.
The individual should be informed which sources will be contacted to
get information, how the data will be used, and to whom the data will be
disclosed. No information should be obtained under false pretenses, or
through the impersonation of others. All individuals should have the
right to see and copy records about himself or herself from any organiza-
tion that keeps a file on the individual, including an employer. If the
individual questions its accuracy, the person should have a right to
correct the record. Where the point is in dispute, the individual’s state-
ment of his/her position should be made part of the permanent file.
Secret files should be outlawed, so that individuals always have knowl-
edge of the existence of records on them.

Government officials who want to gain access to a person’s records
should be required to present proper authorization before being permit-
ted to do so, and the person should be notified when such disclosure is
made. Organizations should only employ service and support firms
whose privacy standards and principles are equivalent to their own.
Confidentiality and MRDF

Alice Robbin, president of the International Association for Social Science Information Service and Technology, has written:

Some of the statistical and research activities of the social scientist have depended on access to and use of information on data subject in individually identifiable form. Similarly, some of the information collected by official data-gathering agencies for their research, statistical, accounting, or administrative purposes has the potential for increasing intrusiveness and harm by parties either associated or unassociated with the original data gathering effort, through compulsory, advertent, or inadvertent disclosure.

The Privacy Protection Study Commission addressed this issue, and observed that activities of the social scientist have depended on voluntary cooperation of the individuals in providing accurate and reliable (confidential) information, with assurances that the information will not be released by third parties in individually identifiable form in a manner whereby inadvertent or unauthorized disclosure of the information would place the data subject at risk.

The accumulation of machine-readable data files (MRDF) on human subjects by government agencies for administrative functions provides much rich data for the social scientist. The Privacy Commission recognized this and recommended to protect an individual

from inadvertent exposure to an administrative action as a consequence of supplying information for a research or statistical purpose...[and] to protect the continued availability (supply) of research and statistical results which are important for the "common welfare,"...there must be a clear functional separation between research and statistical uses and all other uses.... The principle must be established that individually identifiable information collected or compiled for research or statistical purposes may enter into administrative and policy decision making only in aggregate or anonymous form. The reverse flow of individually identifiable information from records maintained by administrators and decision makers to researchers or statisticians can be permitted, but only on the basis of demonstrated need and under stringent safeguards.

Thus official data-gathering agencies must develop a "specific set of standards and guidelines for...practices [which] limit...exposure to risk of the individual who contributes information, either directly or indirectly, to a research or statistical activity," and which, moreover, distinguish among different types of information and types of release. Data librarians who collect, organize and disseminate the contents of MRDF, many of them issued by various government bodies, have dem-
DAVID LINOWES & MICHELE HOYMAN

Demonstrated recognition of the ethical considerations which must come into play when striking a balance between the individual's right to privacy and society's need for knowledge. "Proposed ethical standards" for data archivists and data librarians which treat these concerns have appeared in E. Mochmann and P. Müller's recent volume *Data Protection and Social Science Research: Perspectives from Ten Countries.*

Although the Privacy Commission did not single out the library or archival function for a specific set of recommendations, it did make such recommendations for the related function of research. These follow the same general principles described earlier. These recommendations, in the form of several different pieces of legislation, are now proceeding through Congress.

**Research Activities Recommendations**

In view of the previous discussion, six recommendations for research activities can be made.

First, records and information gathered for research purposes should never be used to influence any decisions or actions directly affecting one of the individuals surveyed, unless that person so authorizes their use. Research organizations should establish a special set of rules to ensure that this will not happen. This means that there should be technical, administrative and physical safeguards against unauthorized or inadvertent disclosure, and the information should be rendered anonymous by being stripped of identifiers as soon after collection as possible.

Second, the organization conducting the research may only disclose individually identifiable records without the consent of the individual identified if certain conditions are met:

1. such disclosure is necessary to accomplish the purpose of the undertaking,
2. the disclosure yields enough social benefit to warrant the increase in the risk to the individual of such exposure,
3. safeguards against unauthorized disclosures are established, and
4. further use or redisclosure without the express consent of the individual identified is prohibited.

Third, no one should be required to divulge information about himself or herself for a research or statistical purpose unless the law requires it. To ensure this, the individual should be informed:

1. that his or her participation is at all times voluntary;
Confidentiality, Research & the Government

2. that the data collection has a specific purpose, and what this purpose is;
3. that there is a possibility that the information may be used in individually identifiable form for additional purposes, research or not; and
4. that if disclosure for purposes other than research is required, the individual will be promptly notified.

Fourth, there should be a review process or a special representative in every research organization responsible for applying the above principles in special cases, specifically, in order to protect people who are not competent to give consent for fear of some loss of benefit or some potential retaliation (for example, prison inmates, employees, welfare recipients, or students). Also, this process should provide safeguards in cases where the research requires that the people being studied are unaware of the existence, purpose or specific nature of the research.

Fifth, if and when these guidelines are followed, an individual should have access to whatever information is used or disclosed in individually identifiable form for any purpose other than a research or statistical one (for example, an inadvertent unauthorized disclosure). Fairness demands that people be able to find out what individually identifiable information about them has been made available. Of course, the research organization should keep an accurate accounting of all such disclosures.

Sixth, if any of the information is disclosed without an assurance that it will not be used in any decision or actions directly affecting the individual concerned, or without a prohibition on further use of disclosure (for example, to a court or an audit agency), the individual should be notified of the disclosure and of his right to access to the record.

The Privacy Commission urged the implementation of these principles; it did not recommend the creation of another regulatory agency to enforce them. Rather, the commission recommended that individuals be given the right of action against persons and organizations who violate these principles. Such legal action would be not only for court costs and actual damages, but also for general damages of between $1000 and $10,000.

Conclusion

The substance of these privacy recommendations is to chip away at the centuries-old property right that organizations have always asserted toward the personal information they maintained in their files about individuals. It is the belief of the Privacy Commission that the time has
DAVID LINOWES & MICHELE HOYMAN

come to create an enforceable claim which the individual can assert to gain access to his or her records. There may be concerns that these privacy protections are too extreme or that they will interfere with the efficient administration of a library or a business. These changes may be viewed as yet another unnecessary drag on the day-to-day operation of commerce.

The goals of efficiency and privacy are not mutually exclusive. Those who think there is a basic conflict between long-term management effectiveness and safeguarding personal privacy rights must either be inexperienced in the art and science of management or ignorant of the consequences of personal privacy abuses. Personal privacy protection is as necessary to the vigor of a successful organization as it is to a nation. One of the significant features distinguishing a totalitarian regime from a democratic one is the deprivation of the individual's right to privacy. Over a century ago, de Tocqueville warned, "If the private rights of an individual are violated...the manners of a nation" are corrupted, putting 'the whole community in jeopardy.' [28]

The findings of the Privacy Commission, as well as recent research at the University of Illinois, produced evidence that the private rights of the individual are currently being violated in the United States. Further, public opinion polls reflect this. A Harris survey released in December 1978 revealed that 64 percent of the people are concerned about threats to their personal privacy, up from 47 percent one year earlier. [29] Thus, until a comprehensive federal privacy policy takes shape, it is up to libraries as organizations to assume the responsibility of examining their practices for abuses of privacy, and voluntarily modifying, if necessary, their policies and procedures.

References

1. Warren, Samuel D., and Brandeis, Louis D. "Right to Privacy." Harvard Law Review 4(15 Dec. 1890):193. (They stated, "And now the right to life has come to mean the right to enjoy life—the right to be alone." [This was in contrast to early interpretations of the right to privacy as the right to life.])
4. Ibid.
Confidentiality, Research & the Government


11. Ibid.


The Council of the American Library Association strongly recommends that the responsible officers of each library in the United States:
1. Formally adopt a policy which specifically recognizes its circulation records and other records identifying the names of library users...to be confidential in nature.
2. Advise all librarians and library employees that such records shall not be made available to any agency of state, federal or local government except pursuant to such process, order or subpoena as may be authorized under the authority of, and pursuant to federal, state or local law relating to civil, criminal or administrative discovery procedures or legislative investigatory power.
3. Resist the issuance or enforcement of any such process, order or subpoena until such time as a proper showing of good cause has been in a court of competent jurisdiction. Point 3,...means that upon receipt of such process, order, or subpoena, the library's officers will consult with their legal counsel to determine if such process, order, or subpoena is in proper form...; or if good cause has not been shown, they will insist that such defects be cured. (Adopted 20 Jan. 1971 by the ALA Council.) Revised 4 July 1975.


15. Ibid.

16. Ibid.

17. Ibid.

18. Ibid.

19. Ibid.


24. Privacy Protection Study Commission, Personal Privacy, chap. 15.

25. Ibid., p. 573.

26. Ibid., p. 574.


Social Science Data Archives and Libraries: A View to the Future

BARTON M. CLARK

Since the mid-1960s there has been serious concern about developing linkages between social science data archives and libraries. These projections have ranged from foreseeing a modest bibliographic and reference role for the library to envisioning total integration of the social science data archive into the library administrative structure.

A major premise upon which the total integration model is built assumes that social science data archives are underutilized. In contrast, libraries are not. As experienced information specialists, librarians would make the resources of the social science data archive more readily accessible, and thus increase the usage of the archive. Although a few libraries (such as the University of Florida) have successfully incorporated major social science data functions into their regular operations, it is doubtful that this model will become commonplace.

Several factors militate against the integrated social science data archive model becoming a reality in the near future. First, the presence of particular materials within the library does not ensure usage of the materials, although it may, in fact, increase access. One only has to examine the limited success of the Human Relations Area Files to attest to the fact that the availability does not necessarily lead to increased usage of materials.

Second, centralized information services located in the library do not normally follow the model provided by social science data archives.

Barton M. Clark is Assistant Director of Public Services for Social Sciences Libraries and Associate Professor of Library Administration, University of Illinois at Urbana-Champaign.
In particular, libraries are not inclined to provide interpreted information to patrons as is often done when machine-readable data files (MRDF) are statistically manipulated. One only has to examine the librarian's approach to medical and legal reference services to emphasize this fact.

Third, few librarians have the requisite skills necessary to work with social science data archives. Most do not have the training in statistics and survey research methods to manipulate MRDF effectively. Although computers are becoming "friendly," effective manipulation of data requires an understanding of the data and its appropriate context. Failure to comprehend this can easily lead to misinterpretation of data through the misuse of statistical techniques. In addition, librarians are not trained in the mechanics of file maintenance, e.g., "cleaning-up" data sets for public use.

Fourth, incorporating social science data archives into the library will not necessarily ensure economic stability of the archives. Libraries are not immune to budget cuts, particularly in these times of economic retrenchment. It is therefore doubtful that many libraries would be willing to accept the additional staffing and materials expenses required to sustain a social science data archive.

Fifth, in most instances, organizational structures separate from libraries already exist to manage social science data archives. In some cases, there is actually more than one unit within an organization which manages social science data archives. At the University of Illinois, the Survey Research Center is the official repository for the 1970 U.S. census tapes, while the Social Science Quantitative Laboratory holds the university's membership in the Inter-University Consortium for Political and Social Research. It is doubtful that already established units such as these would relinquish the management of social science data archives to a library regardless of economy of scale of operations.

Besides these restrictions, one other factor leads one to believe that a scenario which fully integrates the social science data archive into the library will not come to fruition in the near future. The availability of the 1970 U.S. census tapes provided an excellent opportunity for librarians to become actively involved in the social science data archive business. Librarians have had a long experience with U.S. census materials in paper format. A few librarians, of course, did avail themselves of the opportunity to acquire the 1970 U.S. census in machine-readable format. These libraries have often used the census materials as a base to expand their activity with MRDF. Again the chance to become involved with machine-readable files opened with the production of the 1980 U.S. census tapes, but still the number of libraries involved in the
handling of machine-readable files remains small. This lack of involvement leads one to assume that the majority of libraries are satisfied not to integrate fully social science data archives into their systems.

What, then, will be the role of librarians in relationship to social science data archives in the future? For at least the next ten years, it is likely that librarians will perform those functions for users of social science data archives which are now performed for users of more traditionally oriented resources. In particular, librarians will serve as bibliographic brokers, i.e., while librarians will be working toward developing more effective access to MRDF, they will not actually possess the files, which will remain located in social science data archives. The exception will be in the case where there are online numeric data bases readily available through vendors commonly used by the library. These data bases will ordinarily be used to extract factual data which do not require statistical manipulation. It is assumed that these data bases will not in any way replace the need for social science data archives as they now exist.

Of paramount importance will be the necessity to acquire all of the appropriate data access tools for all of the data files held by the institution. These resources would include indexes, inventories and newsletters, all of which provide general descriptions and listings of data sets. Also important are codebooks and questionnaires which provide more detailed information, including revealing which statistical manipulations are possible with the data. These data access tools, of course, would also be located in the social science data archives. Library reference collections, however, should not limit themselves to just acquiring those resource tools which are related to the holdings of the local archive. Indexes, inventories and newsletters which describe MRDF at various worldwide locations should also be acquired to help potential users identify materials which might be useful. To a lesser degree, codebooks of materials not owned by the library should be collected. In addition, libraries also should be responsible for acquiring the paper version of MRDF if it exists.

The adoption of the second edition of the *Anglo-American Cataloging Rules* has supplied new standardized rules for cataloging MRDF. By providing records (including locations) in the card catalog of MRDF entries, access to these materials can be greatly enhanced. The development of online catalogs will improve ease of access even more.

Bibliographic access to MRDF through both reference service and cataloging as described here is easy to accomplish and economically possible, and therefore represents a highly feasible role for librarians to play in the future of library and social science data archive relations.
The role of the librarian could be greatly enhanced, however, if an interactive online bibliographic system could be developed.

The physical structure of a bibliographic data base for MRDF would be much like any other bibliographic data base, e.g., ERIC. Separate searchable paragraphs would exist for author, title, originating institution, year, descriptors, and abstract. Allowance of a maximum of 200 words for an abstract, as is done in the ERIC data base, would provide ample space for a detailed description of the main features of each MRDF in the data base. In addition to the standard searchable paragraphs, a field could be established to include reference materials associated with the MRDF. A prospective user of a file would thus be able to identify all needed ancillary materials relevant to using the file.

The easiest way to develop such a data base would be by using the MARC records. Preliminary steps toward developing such a system at the local level have been successfully tested by the Social Science Data Library of the Institute for Research in Social Science at the University of North Carolina at Chapel Hill. The network also includes Duke University and North Carolina State University.

Use of a standard automated cataloging system would allow for the possibility of a single data base for all the bibliographic records. Considering the large number of producers of public and private data files, a system which could cumulate all of the bibliographic records into a single data base would be highly desirable. It is, however, highly unlikely that such a system will develop. Instead, a series of bibliographic data bases based on the holdings of an institution or of producers is more likely to arise. It is hoped that cooperation will prevail and the number will remain small while encompassing the majority of MRDF. Whichever direction computer-based bibliographic control of MRDF may take, it is essential that the bibliographic data bases be incorporated into the catalogs of one or more of the major vendors, e.g., BRS, Lockheed. Only in this way can increased utilization of the data sets themselves be assured. Since a subject can be searched through a data base such as BRS's CROS, which allows multiple data bases to be searched at the same time, it would be possible to identify germane MRDF at the same time one was retrieving citations to monographs and serials, thus greatly enhancing information capabilities for the social sciences.

Although the bibliographic broker model projects a more conservative involvement of the library with social science data archives than the total integration model, it provides a more realistic view of the
A View to the Future

future. The librarian is thus perceived as a skilled bibliographer who actively contributes to a larger knowledge system.

References

Partial List of Library Trends Issues in Print

<table>
<thead>
<tr>
<th>Title</th>
<th>Editor</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 N 1 Library Boards</td>
<td>J. Archer Eggen</td>
<td>July 1962</td>
</tr>
<tr>
<td>11 3 Law Libraries</td>
<td>Bernita J. Davies</td>
<td>Jan. 1965</td>
</tr>
<tr>
<td>11 4 Financial Administration of Libraries</td>
<td>Ralph P. Parker</td>
<td>Apr. 1963</td>
</tr>
<tr>
<td>V. 12 N 1 Public Library Service to Children</td>
<td>Wimfred C. Ladlev</td>
<td>July 1963</td>
</tr>
<tr>
<td>12 2 Education for Librarianship Abroad in Selected Countries</td>
<td>Harold Lassieur</td>
<td>Apr. 1963</td>
</tr>
<tr>
<td>12 3 Current Trends in Reference Services</td>
<td>Margaret Knox Goggan</td>
<td>Jan. 1964</td>
</tr>
<tr>
<td>12 4 European University Libraries: Current Status and Developments</td>
<td>Robert Vosper</td>
<td>Apr. 1964</td>
</tr>
<tr>
<td>V. 13 N 1 Research Methods in Librarianship</td>
<td>Guy Garrison</td>
<td>July 1964</td>
</tr>
<tr>
<td>13 2 State and Local History in Libraries</td>
<td>Clyde Walton</td>
<td>Oct. 1964</td>
</tr>
<tr>
<td>13 3 Regional Public Library Systems</td>
<td>Hannis S. Smith</td>
<td>Jan. 1965</td>
</tr>
<tr>
<td>13 4 Library Furniture and Furnishings</td>
<td>Frazer G. Poole</td>
<td>Apr. 1965</td>
</tr>
<tr>
<td>V. 14 N 1 Metropolitan Public Library Problems</td>
<td>H.C. Campbell</td>
<td>Jul. 1965</td>
</tr>
<tr>
<td>14 2 Junior College Libraries</td>
<td>Charles L. Trinkner</td>
<td>Oct. 1965</td>
</tr>
<tr>
<td>14 3 Library Service to Industry</td>
<td>Katharine G. Harris</td>
<td>Jan. 1966</td>
</tr>
<tr>
<td>14 4 Current Trends in Branch Libraries</td>
<td>Eugene B. Jackson</td>
<td>Apr. 1966</td>
</tr>
<tr>
<td>V. 15 N 1 Government Publications</td>
<td>Thomas S. Shaw</td>
<td>Jul. 1966</td>
</tr>
<tr>
<td>15 4 Bibliographies: Current State and Future Trends, Part 2</td>
<td>Frances B. Jenkins</td>
<td>Apr. 1967</td>
</tr>
<tr>
<td>V. 16 N 1 Cooperative and Centralized Cataloging</td>
<td>Esther J. Percy</td>
<td>Jul. 1967</td>
</tr>
<tr>
<td>16 2 Library Uses of the New Media of Communication</td>
<td>Robert L. Talmadge</td>
<td>Oct. 1967</td>
</tr>
<tr>
<td>16 3 Abstracting Services</td>
<td>C. Walter Stone</td>
<td>Jan. 1968</td>
</tr>
<tr>
<td>16 4 School Library Services and Administration at the School District Level</td>
<td>Sara K. Sergygl</td>
<td>Apr. 1968</td>
</tr>
<tr>
<td>17 2 Young Adult Service in the Public Library</td>
<td>Audrey Biehl</td>
<td>Oct. 1968</td>
</tr>
<tr>
<td>17 3 Development in National Documentation and Information Services</td>
<td>H.C. Campbell</td>
<td>Jan. 1969</td>
</tr>
<tr>
<td>V. 18 N 1 Trends in College Librarianship</td>
<td>H. V. Dale</td>
<td>Jul. 1969</td>
</tr>
<tr>
<td>18 2 University Library Buildings</td>
<td>David C. Weber</td>
<td>Oct. 1969</td>
</tr>
<tr>
<td>19 3 Book Storage</td>
<td>Mary B. Cassata</td>
<td>Jan. 1971</td>
</tr>
<tr>
<td>19 4 New Dimensions in Educational Technology for Multi-Media Centers</td>
<td>Philip Lewis</td>
<td>Apr. 1971</td>
</tr>
<tr>
<td>V. 20 N 1 Personnel Development and Continuing Education in Libraries</td>
<td>Elizabeth W. Stone</td>
<td>Jul. 1971</td>
</tr>
<tr>
<td>20 2 Library Programs and Services to the Disadvantaged</td>
<td>Helen H. Lyman</td>
<td>Oct. 1971</td>
</tr>
<tr>
<td>20 3 The Influence of American Librarianship Abroad</td>
<td>Cecil K. Byrd</td>
<td>Jan. 1972</td>
</tr>
<tr>
<td>20 4 Current Trends in Urban Main Libraries</td>
<td>Larry Earl Bone</td>
<td>Apr. 1972</td>
</tr>
<tr>
<td>V. 21 N 1 Trends in Archival and Reference Collections of Recorded Sound</td>
<td>Gordon Stevenson</td>
<td>Jul. 1972</td>
</tr>
<tr>
<td>21 3 Library Services to the Aging</td>
<td>Eleanor Phinney</td>
<td>Jan. 1973</td>
</tr>
<tr>
<td>V. 22 N 1 Analyses of Bibliographies</td>
<td>H.R. Simon</td>
<td>Jul. 1973</td>
</tr>
<tr>
<td>22 2 Research in the Fields of Reading and Communication</td>
<td>Alice Loher</td>
<td>Oct. 1973</td>
</tr>
<tr>
<td>22 3 Evaluation of Library Services</td>
<td>Sarah Reed</td>
<td>Jan. 1974</td>
</tr>
<tr>
<td>22 4 Science Materials for Children and Young People</td>
<td>George S. Bonn</td>
<td>Apr. 1974</td>
</tr>
<tr>
<td>Title</td>
<td>Editor</td>
<td>Date</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>V. 23 N. 1 Health Sciences Libraries</td>
<td>Joan Titley Adams</td>
<td>July 1974</td>
</tr>
<tr>
<td>23 2 Library Services in Metropolitan Areas</td>
<td>William S. Badlington</td>
<td>Oct. 1974</td>
</tr>
<tr>
<td>23 4 Resource Allocation in Library Management</td>
<td>Wolfgang M. Freiag</td>
<td>April 1975</td>
</tr>
<tr>
<td>23 H. William Asford</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. 24 N. 1 Federal Aid to Libraries</td>
<td>Genevieve M. Casey</td>
<td>July 1975</td>
</tr>
<tr>
<td>24 2 Library Cooperation</td>
<td>Pearce S. Grove</td>
<td>Oct. 1975</td>
</tr>
<tr>
<td>24 4 Commercial Library Supply Houses</td>
<td>Harold Roeh</td>
<td>April 1976</td>
</tr>
<tr>
<td>V. + 25 N. 1 American Library History 1876-1976</td>
<td>Howard W. Winger</td>
<td>July 1976</td>
</tr>
<tr>
<td>25 4 Trends in the Scholarly Use of Library Resources</td>
<td>Benjamin F. Page</td>
<td>April 1977</td>
</tr>
<tr>
<td>25 D.W. Krummel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. 26 N. 1 Library Services to Correctional Facilities</td>
<td>Jane Pool</td>
<td>Sum. 1977</td>
</tr>
<tr>
<td>26 2 Trends in the Governance of Libraries</td>
<td>F. William Summers</td>
<td>Fall 1977</td>
</tr>
<tr>
<td>26 3 Institution Libraries</td>
<td>Harris C. McClaskey</td>
<td>Win. 1978</td>
</tr>
<tr>
<td>26 4 Publishing in the Third World</td>
<td>Philip G. Altbach</td>
<td></td>
</tr>
<tr>
<td>26 Keith Smith</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. 27 N. 1 Films in Public Libraries</td>
<td>John A. McCrossan</td>
<td>Sum. 1978</td>
</tr>
<tr>
<td>27 2 State Library Development Agencies</td>
<td>Phyllis Bux</td>
<td>Fall 1978</td>
</tr>
<tr>
<td>27 3 Libraries and Society</td>
<td>Margaret F. Stieg</td>
<td>Win. 1979</td>
</tr>
<tr>
<td>27 4 Study and Collecting of Historical Children’s Books</td>
<td>Selma K. Richardson</td>
<td>Spgr. 1979</td>
</tr>
<tr>
<td>V. 28 N. 1 Economics of Academic Libraries</td>
<td>Allen Kent</td>
<td></td>
</tr>
<tr>
<td>28 2 Emerging Patterns of Community Service</td>
<td>Jacob Cohen</td>
<td></td>
</tr>
<tr>
<td>28 3 Library Consultants</td>
<td>K. Leon Montgomery</td>
<td></td>
</tr>
<tr>
<td>28 4 Current Trends in Rural Public Library Service</td>
<td>Margaret Monroe</td>
<td></td>
</tr>
<tr>
<td>28 5 Map Librarianship and Map Collections</td>
<td>Kathleen M. Hein</td>
<td></td>
</tr>
<tr>
<td>28 6 Public Library Right</td>
<td>Ellsworth E. Mason</td>
<td></td>
</tr>
<tr>
<td>28 7 John M. Houlihan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. 29 N. 1 Current Library Use Instruction</td>
<td>A.P. Marshall</td>
<td>Sum. 1980</td>
</tr>
<tr>
<td>29 2 Library Services to Ethnocultural Minorities</td>
<td>Leonard Wertheimer</td>
<td>Fall 1980</td>
</tr>
<tr>
<td>29 3 Map Librarianship and Map Collections</td>
<td>Mary Lyntte Langaard</td>
<td>Win. 1981</td>
</tr>
<tr>
<td>29 4 Public Library Right</td>
<td>Perry D. Morrison</td>
<td>Spgr. 1981</td>
</tr>
<tr>
<td>V. 30 N. 1 Bibliometrics</td>
<td>William Gray Potter</td>
<td>Sum. 1981</td>
</tr>
<tr>
<td>30 2 Conservation of Library Materials</td>
<td>Gerald Lundern</td>
<td>Fall 1981</td>
</tr>
<tr>
<td>30 3 Data Libraries for the Social Sciences</td>
<td>Kathleen M. Hein</td>
<td>Win. 1982</td>
</tr>
</tbody>
</table>

*A complete list of back issues is available from *Library Trends*, Publications Office, 249 Armory Building, 505 E. Armory Street, University of Illinois, Champaign, Ill. 61820.

† Also available in clothbound editions.
Library Trends

Forthcoming numbers are as follows:

Spring 1982, Mental Health Information, Libraries and Services to the Patient. Editor: Phyllis Rubinton, Librarian, Payne Whitney Psychiatric Clinic, New York Hospital-Cornell Medical Center.

Summer 1982, Standards for Library and Information Services. Editor: Terry L. Weech, Associate Professor, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.
