

## 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.<sup>1</sup> 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*;<sup>2</sup> their *Elections and the Political Order*;<sup>3</sup> or the more recent *American National Election Studies Data Sourcebook 1952-1978*.<sup>4</sup> 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.<sup>5</sup> 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,<sup>6</sup> 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,<sup>7</sup> 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.<sup>8</sup> 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 contemporary 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.<sup>9</sup> Throughout the world academicians and policy-makers are performing secondary analysis of existing data files to address questions 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 importantly, 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 questions dealing with employment status, another may be concerned primarily 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 computation 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 International Association for Social Science Information Service and Technology (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 questions 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 every 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

1. For descriptions of summary tape, see U.S. Bureau of the Census. *Directory of Data Files, Bureau of Census, 1979*. Washington, D.C.: U.S. Bureau of the Census, 1979.
2. Campbell, Angus, et al. *The American Voter, An Abridgement*. New York: John Wiley & Sons, 1964.
3. ————. *Elections and the Political Order*. New York: John Wiley & Sons, 1966.
4. Miller, Warren E., et al. *American National Election Studies Data Sourcebook, 1952-1978*. Cambridge, Mass.: Harvard University Press, 1980.
5. For a more detailed discussion concerning selected literature and data resources relevant to the study of voting behavior, see Goudy, Frank W. "American Political Behavior and the Election Process: A Bibliographic Essay." *RQ* 19(Spring 1980):237-44.
6. Inter-University Consortium for Political and Social Research. *Guide to Resources and Services 1980-1981*. Ann Arbor: University of Michigan, 1981.
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.
8. Hyman, Herbert H. *Secondary Analysis of Sample Surveys*. New York: John

JUDITH ROWE

Wiley & Sons, 1972; and Smith, Kent W., and Rowe, Judith S. "Using Secondary Analysis for Quasi-Experimental Research." *Social Science Information* 18(1979):451-72.

9. Miller, et al., *American National Election Studies*; Converse, Philip E., et al. *American Social Attitudes Data Sourcebok 1947-1978*. Cambridge, Mass.: Harvard University Press, 1980; and Nie, Norman, et al. *The Changing American Voter*. Cambridge, Mass.: Harvard University Press, 1976.