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.[1] 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 Reseach in Social Science at the University of North Carolina at Chapel Hill. The network also includes Duke University and North Carolina State University.[2]

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

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

References

1. Glenn, Norval. "The Social Scientific Data Archives: The Problem of Underutilization." *American Sociologist* 8(Feb. 1973):42-45.
2. Dodd, Sue A. "Building a Bibliographic/MARC Data Base for Social Science Data Files in a Network Environment." *IASSIST Newsletter* 2(Spring 1978):34-37.

This Page Intentionally Left Blank