

## 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.<sup>1</sup> 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

*The University of Florida*

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,<sup>2</sup> Alice Robbin<sup>3</sup> and Lucinda Conger<sup>4</sup> 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

## *The University of Florida*

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-

*The University of Florida*

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.

## *The University of Florida*

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

1. Jones, Ray, and Wittkopf, Barbara. "Computerized Census Data: Meeting Demands in an Academic Library." *RQ* 19(Spring 1980):246-51.
2. Ruus, Laine G.M. "User Services in a Data Library." *IASSIST Newsletter* 4(1980):29-33.
3. Robbin, Alice. "The Pre-Acquisition Process: A Strategy for Locating and Acquiring Machine-Readable Data." *Drexel Library Quarterly* 13(Jan. 1977):21-42.
4. Conger, Lucinda D. "Data Reference Work with Machine-Readable Data Files in the Social Sciences." *Journal of Academic Librarianship* 2(May 1976):60-65.