

TOWARD CREATING THE PROFESSIONAL DATA LIBRARIAN

Alice Robbin

University of Wisconsin - Madison

Abstract

This article describes the need for training the professional data librarian, archivist, and information scientist in a framework of social science research and applications and library and information sciences. The University of Wisconsin-Madison Intersession 1978 course was designed to meet this need. The course is described.

Since President Lyndon Johnson called for a "War on Poverty" in his State of the Union Message of January 1964, we have seen an exponential growth in the production of statistical information in order to allocate resources at the national, state, and local levels of government; to plan, audit and evaluate the distribution of resources; and to ensure adequate planning for human needs and an equitable delivery of benefits. Almost all federal legislation has included requirements to collect, analyze, and report the findings of data gathering. Recent trends in federal reporting requirements suggest an even greater increase in the rate of production of statistical data during the next decade as the need increases for more information by policy planners and analysts in both the public and private sectors.

Most of the data have been collected as part of the administrative record keeping process of governments, but a large portion of the data gathering has been funded as part of the research, policy, and evaluation activities of the Federal government. In the social sciences, an increasing amount of scientific research activity as well as policy planning and

evaluation involves the collection, dissemination, and secondary analysis of statistical machine readable data files (MRDF). Some of these files are produced in the course of individual research projects; others, by organizations in the course of their operations; and still others, by ongoing data collection efforts funded by a consortium of data users.

While a substantial portion of these data are probably not useful for reanalysis, a great body of data continues to be useful for research related to public policy analysis, planning, and evaluation. While some of the data have been transferred to national archives and data centers whose major functions are to preserve, describe, and disseminate these data, many of these data files which are potentially rich sources of information remain outside the public domain.

The problems of access to information about MRDF, the quality of data, and the need for good documentation describing MRDF have been issues discussed by secondary analysts and data archive staffs for a number of years. One aspect of the MRDF problem, however, has not been sufficiently addressed, and that has been the insufficient and

inadequate concerted national efforts to facilitate access to MRDF through the development of training programs for professional librarians and information scientists.

When IASSIST was established in 1976, it was with the recognition that members of data archives and libraries needed a vehicle to communicate information about organizing, managing, and disseminating machine readable data files. In the IASSIST Constitution, the 'Objectives' include the establishment of training courses for data center personnel (NEWSLETTER, 1(1), 1976). This objective not only represents the recognition that data center personnel need assistance, but that established data services have the potential for providing training programs to assist others in understanding the nature of MRDF and the special problems of organization, management, and dissemination associated with this medium. The Inter-university Consortium for Political and Social Research responded to this objective by holding two Workshops on Data Library Management in the Summers of 1976 and 1977, as adjuncts to their regular summer program (Rowe, 1977). The Workshops were taught by Carolyn Geda, Alice Robbin, and Judith Rowe. Participants included trained data center personnel and professional librarians who wanted to become more informed about MRDF and integrating them in a library collection.

These Workshops were a source of satisfaction to their instructors and a good deal of information was communicated and exchanged. But it became increasingly obvious that a Workshop was not the best structure in which to communicate a conceptual framework for organizing and managing MRDF. There were few

incentives to utilize the computing and data processing facilities, and within the time constraints there was little possibility of dealing with major social research and applications concepts needed to understand statistical MRDF, data base management, and organizational behavior. In addition, if the ideas were to reach an audience who could most directly benefit from this learning experience--the professional librarian--the course had to be taught in a university library school environment, and integrated in the library school and social science departments' curricula.

Why two different departments and indeed ones which rarely communicate with one another? Although social scientists continue to demonstrate negative attitudes toward the library profession, it is the library which has the expertise in our society to organize and disseminate information. Libraries are a natural environment for MRDF, because it can be treated as an additional informational resource, albeit in a different storage medium. Library schools train professionals to handle a variety of informational resources. Courses in library automation, systems design, information storage and retrieval, and on-line bibliographic data bases are becoming integral offerings of library schools throughout the country. Thus, a library school is the natural setting for introducing the concept of numeric or statistical MRDF. But while library schools routinely address problems of textual data in machine readable form, they have not addressed the problems of numeric MRDF and future professional librarians and information scientists are ill-equipped to serve the quantitative bent of today's social scientist.

On the other hand, most social science disciplines offer at least one course and at a growing number of institutions, major course offerings consist increasingly of a number of areas related to quantitative social research. Social science departments train students in methodology, statistics, survey research, modeling and simulation, data handling, and the like, subjects which provide the basis for understanding the construction and analysis of MRDF. In fact, personnel of most of the data services in North America and Western Europe have been (and continue to be) people who trained in one of the social sciences. They have not been professional librarians and have been slow to recognize that the tasks they perform or the problems they encounter in organizing, managing, and retrieving MRDF (and information about MRDF) are tasks which have been traditionally performed by reference librarians who work with other media (see, Carmichael, 1978; and Robbin, 1978).

A course which integrates the theoretical foundations of library and information science and social science research is therefore one which potentially speaks to the formation of a professional information scientist and manager of numeric or statistical MRDF. With this in mind, in September 1977, I recommended that the Data and Computation Center, of which the Data and Program Library Service is a part, design a course which would respond to a perceived need to train professionals (both in the social sciences and library and information sciences) to deal with the explosion of information in machine readable form. It would be an interdisciplinary, graduate level course, one semester in length, which would draw upon the expertise of social scientists,

data library specialists, and computer specialists at the University of Wisconsin-Madison. Because it would be offered during Inter-session 1978, it could be viewed as a potential course for in-service training for library and archive professionals who wished to become familiar with this informational resource.

The course was cross-listed by the School of Library Science and Department of Economics, and for a variety of bureaucratic reasons was entitled, "Micro Data Collection Methods in Economics." Funding for Inter-session 1978 was made possible with the generous support of the UW-Madison, which encourages its faculty and staff to use Inter-session as an opportunity to develop new courses in response to perceived scientific and social changes within and outside the University. The course was designed to meet the needs or interests of social scientists, users, and generators of numeric machine readable data, and professionals engaged in information services, whose present or future responsibilities might include managing large numeric data files or providing data services or information about numeric MRDF to users. The objective of the course was to provide the student with the underlying principles of access to and management of MRDF in a library and archive setting. Students were introduced to social science research and applications, data collection techniques, computing and data processing, statistical analysis and file handling, policy issues and problems regarding data libraries, and bibliographic documentation and control of numeric MRDF. Problem-solving was an integral part of the course, and included exercises in statistical analysis and building a biblio-

graphic data base of numeric MRDF. In addition, there was a heavy dose of daily readings on which the lectures and discussions were based and a final paper (in lieu of an examination).

Lectures included (1) introduction to social science research methods and applications, introduction to statistical processing and file handling, orientation to data library and data processing and computing facilities; (2) introduction to systems analysis: data library as an information management system, complex data bases, data base management systems (DBMS), networking; (3) selected policy issues concerning MRDF and the data library and archive; (4) bibliographic documentation and control of MRDF; (5) planning a data library and information service for MRDF; (6) special issues of concern for MRDF and data librarians: copyright, strategies for file preservation and handling. Practical exercises included (1) statistical analysis of a MRDF specially prepared for the course, using the SPSS package; (2) use of a data base management system; (3) use of networks; (4) data library procedures; and (5) bibliographic documentation and control (building a bibliographic data base of information for MRDF). The Independent Project (final paper) assignment was a choice of (1) designing a research problem, carrying out limited statistical analysis on the data file prepared for the course, and reviewing the findings in a short paper; (2) writing a short analysis of the problems of creating a bibliographic data base for MRDF; and (3) designing an independent project with the approval of the instructors (most chose to do this and selected a wide range of topics on confidentiality and privacy, content analysis, book

reviews, and archive and library problems and their relationship to MRDF).

Primary responsibility for the course was in the hands of Martin David, Department of Economics and Director of the Data and Computation Center; Alice Robbin, Head of the Data and Program Library Service; and Al Schubert, Head of the Program Consulting Service. Members of the Madison Academic Computing Center (MACC) contributed their expertise during the course, as did experts in data base management from the Department of Landscape Architecture and the Center for Demography and Ecology. Lectures on social science research and data collection were given by an economist, sociologists, and a survey methodologist.

Fourteen registered students and auditors began and completed the course. One half of the students were professional archivists (From the Wisconsin State Archives) and library students and the other half were graduate students (primarily from developing countries) in economics, sociology, political science, business, and history. One of the students is a professional data librarian from the University of California at Los Angeles.

Most of the students had never worked with a computer before, but within a few days were keypunching control cards and submitting statistical runs. Although the course was highly concentrated and very demanding (students attended lectures from 8-11 a.m., Monday through Friday, and worked on assignments with the help of a teaching assistant from noon, often until 10 p.m.), enthusiasm and commitment never flagged. It was an exciting time for instructors and students. A detailed evaluation

instrument was completed by the students. The recommendation was that the course be integrated in the University's curriculum and extended into a full four week program during Summer School or the regular academic year.

The response was so positive that DACC submitted a request for refunding of the Intercession 1979 program. (At UW, all Intercession courses go into competition for funds.) We have recently learned that DACC has successfully competed for funds and the course will once again be offered and cross-listed by the Department of Economics and School of Library Science. Several changes are anticipated on the basis of what the instructors learned last year. First, the title of the course has been changed to reflect the actual course contents. It will be officially titled, "Management of Machine Readable Numeric Data for the Social Sciences." (People were more than a little mystified last year to learn that "Economics 615, Micro Collection Methods in Economics," would deal with the subjects I just described.) The import of the title change and new course number(s) should not be underestimated: While we have no assurance that the Divisional Committee of the College of Letters and Science will approve a new course offering (these changes are very hard to come by), it does suggest the recognition of the need to respond to information and technological changes in our society; universities must continue to broaden their course offerings to meet professionals' needs and to respond to technological change. Indeed, the concept of data services, libraries, and archives has too long remained the purview of special support facilities within the university and the notion that services to preserve and disseminate

statistical MRDF are a function of the general or special library within the university or in both the government and private sectors has not been widely accepted.

Second, the course structure was too demanding, for both the students and instructors. The workload will be reduced. Rather than building a bibliographic data base, time will be spent analyzing the structure and syntax of its contents. Some of the time allocated to the data base will now be devoted to exercises in documentation and control and records management of MRDF (known earlier by its misnomer, "accessioning"). Less time will be devoted to understanding the development of statistical software and more time to understanding problems (through statistical techniques) of statistical data (assaying the quality of data). The number of required readings will be reduced. However, the basic structure of the course remains the same. Once again, a set of instructional materials will be created, but its size (242 p.) will be reduced.

What is evident from the enthusiastic response of the students and instructors is that the course DACC offered last year is much needed. It responds to the recognition that the information explosion must be managed. Access and retrieval of the vast quantities of information in statistical machine readable form are becoming important issues in the information and library sciences' professions and social science disciplines. The course also demonstrates that the complexity and dimension of information in numeric machine readable form are such that no one individual has the expertise to teach future professionals to organize and manage collections of statistical MRDF; rather, the approach to teaching

must be an integrated, interdisciplinary one, where the quantitative social researcher and analyst, data archivist, and computing specialist communicate their particular expertise. In the future it is hoped that Schools of Library and Information Sciences will recognize that numeric machine readable data and the non-librarian professional have a place in their curricula.

NOTE: For further information on the University of Wisconsin-Madison's course, "Management of Numeric Machine Readable Numeric Data for the Social Sciences," to be offered during Intersession 1979 (May 29- June 15, 1978), write the Data and Computation Center, 4452 Social Science Building, 1180

Observatory Drive, University of Wisconsin-Madison, 53706.

References

- Carmichael, Nancy. "Commentary on Librarians and Machine Readable Data Files." Special Libraries 69(8), 1978, 306-07.
- Robbin, Alice. "The Impact of Networking on the Social Science Data Library." IASSIST Newsletter 2(1), 1978, 3-13.
- Rowe, Judith. "Training the Professional Data Librarian." In H. D. White (ed.), Drexel Library Quarterly 13(1), 1977, 100-108.