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 1930's the commercial organizations began accumulating large numbers of data files; during the 1940's, 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 1950's and early 1960's.[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 extrainstitutional 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."[2] 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."[4] 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."[8] 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."[9] 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.[10]

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.[11]

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."[12]

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.[13] In 1978 and 1979, a similar course was offered at the University of Wisconsin, as a full sememster-credit, graduate-level intersession course. In its first year it attracted primarily professional archivists, library school students, and one practicing "data librarian."[14]

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.[15] 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."

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 computer 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 percent 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

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.[18] 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

> librarians, archivists have shied away from prescribing whether archives education should take place in history departments or library schools.[19]

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."[20] 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."[21]

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.[22] 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.[23] 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.

References

1. Bisco, Ralph L. "Social Science Data Archives: Progress and Prospects." *Social Science Information* 6(Feb. 1967):39-40.

2. Ibid., p. 40.

3. Lucci, York, and Rokkan, Stein. *A Library Center of Survey Research Data: A Report of an Inquiry and a Proposal.* New York: Columbia University School of Library Service, 1957, pp. 96-98.

4. Nasatir, David. *Data Archives for the Social Sciences: Purposes, Operations and Problems.* Paris: Unesco, 1973, p. 15.

5. Bisco, Ralph L. "The Research Library and Data Archives for Social Research." Speech prepared for the dedication of the Graduate Research Library, University of Florida, 1967 as quoted in Nasatir, David. *Data Archives for the Social Sciences: Purposes, Operations and Problems.* Paris: UNESCO, 1973, p. 11; and Pool, Ithiel de Sola. "Data Archives and Libraries." In *INTREX: Report of a Planning Conference on Information Transfer Experiments,* edited by Carl F.J. Harman and R. Joyce Harman, pp. 175-81. Cambridge, Mass.: MIT Press, 1965.

6. Bisco, "Research Library and Data Archives."

7. Brock, Clifton. "Political Science." In *Bibliography: Current State and Future Trends,* edited by Robert B. Downs and Frances B. Jenkins, p. 305. Urbana: University of Illinois Press, 1967.

8. Inter-University Consortium for Political Research. *Annual Report, 1968-1969,* Ann Arbor: University of Michigan, 1969.

9. National Research Council. Division of Behavioral Sciences. Committee on Information in Behavioral Sciences. *Communication Systems and Resources in the Behavioral Sciences.* Washington, D.C.: National Academy of Sciences, 1967, p. 41.

10. Adams, M.O., et al., eds. *Proceedings of the Workshop on the Management of a Data and Program Library.* Madison: University of Wisconsin, 1969.

11. Nasatir, *Data Archives,* p. 44.

12. Ibid., p. 54.

13. Rowe, Judith S., and Geda, Carolyn L. "Training the Professional Data Librarian." *Drexel Library Quarterly* 13(Jan. 1977):100-08.

14. Robbin, Alice. "Toward Creating the Professional Data Librarian." *IASSIST Newsletter* 2(Fall 1978):95-100.

15. Mignon, Edmund. "Educating Librarians for Data Services." 1980 (unpublished) and _____. "Numeric Data Bases in the Professional Librarianship Curriculum: Implications for Behavioral and Social Sciences Librarians." *Behavioral and Social Sciences Librarian* 1(Spring 1980):181-87.

16. Rowe and Geda, "Training the Professional," pp. 100-02.

17. Ruus, Laine G.M., and MacDonald, John. "Report on a Survey Conducted by the IASSIST Education Committee," in press.

18. Carmichael, Nancy. "Commentary on Librarians and Machine-Readable Data Files." *Special Libraries* 69(Aug. 1978):306-07; Rowe and Geda, "Training the Professional"; and Robbin, "Toward Creating the Professional."

19. Peace, Nancy E., and Chudacoff, Nancy F. "Archivists and Librarians: A Common Mission, a Common Education." *American Archivist* 42(Oct. 1979):458.

20. Taylor, Hugh A., and Welch, Edwin. "Association of Canadian Archivists: Guidelines Towards a Curriculum for Graduate Archival Training Leading to a Master's Degree in Archival Science." Mimeographed. Association of Canadian Archivists, 1976.

21. Nasatir, *Data Archives,* p. 14.

22. Rowe and Geda, "Training the Professional," pp. 101-02.

23. Nasatir, *Data Archives,* pp. 44-48; and Rowe and Geda, "Training the Professional."

This Page Intentionally Left Blank