The Historian and Social Science Data Archives in the United States

ALLAN G. BOGUE

DURING THE PAST thirty years social scientists have learned that quantitative data, when converted to machine-readable form in the course of even relatively modest research projects, may also be useful to scholars interested in other research problems as well as to those who wish to use it for replicative research. When a source is so rich that it provides data for many scholars working on different kinds of research in a variety of disciplines, the utility of placing it in standard machine-readable format ready for easy distribution becomes readily apparent. The onerous tasks of conversion would then need be undertaken only once; subsequent investigators can begin their research at the stage of analysis, and an almost incalculable amount of time and labor can be saved within the research community. Such is the rationale that underlies the development of machine-readable data archives. Does it also apply to historical data as well as to data generated by survey analysis and the contemporary activity of governments, business firms and other societal institutions? Certainly there are important historical sources that are quantitative in nature, or that can be converted to quantitative form. Historians active in research are comparable in number to researchers in the more populous social science disciplines and some of them have created or helped to create machine-readable data files. The analogy can be pushed too far, however, and in this article I will discuss the degree to which historians have been involved in the data archiving movement and the challenges it presents to their discipline.

Allan G. Bogue is Frederick Jackson Turner Professor of History, University of Wisconsin, Madison.

This essay also appears, in a slightly different form, in *American Behavioral Scientist* 19:419-42, April 1976, and is reprinted by permission of its publishers, Sage Publications, Inc., © 1976.

Today machine-readable data files of interest to researchers in the social sciences and history are found in a variety of locations. Individuals who have completed research on a project, or advanced it to the point where they are willing to make their data available to other researchers, hold important files in their personal possession. On the other hand one finds large, nationally oriented data archives that maintain a considerable library of data files available for distribution under institutionalized arrangements and which are continuing to add new data files to their holdings as the result of the activities of affiliated or cooperating research groups or individuals. In the latter category the Inter-University Consortium for Political and Social Research (ICPR) and the Roper Public Opinion Research Center are the outstanding illustrations, the former preeminently academic in orientation and the latter operating in the private sector as well. A number of smaller archival agencies stand between the two extremes in the United States, their activities circumscribed by regional, state, institutional or subject matter boundaries.[1]

The smaller agencies emphasize various activities depending on their unique purposes and institutional needs. Thus, investigators in various survey research centers based in universities have been primarily concerned with conducting public opinion surveys required by the researchers of the parent institution, and data archiving has been little more than a storage and servicing operation incident to the needs of local investigators. Other agencies have developed more elaborate functions, and the staff of the Data Program Library Service (DPLS) at the University of Wisconsin try to maintain a library of data files generated by local scholars in the social sciences, as well as to collect special service programs of possible interest to local investigators and researchers elsewhere. The personnel of DPLS maintain reference lists of data in other archives and serve as intermediaries in obtaining research data from such agencies, retaining a backup file of such material when received. Staff members also prepare data packets for classroom use at the request of instructors and assist in the research of students and sometimes faculty by providing advice on coding and program selection, as well as elementary instruction in the use of the program packages available at the university computer center. The Social Science Data Archive of the Laboratory for Political Research at the University of Iowa performs similar functions in serving the research and instructional needs of social scientists, both at the University of Iowa and at the twelve liberal arts colleges which make up the Iowa Regional Computer Network. Since it was estab-

lished in 1969, this archive has acquired more than 300 machine-readable studies in a variety of disciplines, including history. The laboratory also publishes *SS Data*, a newsletter designed to promote fuller use of data archives in general.[2] Forty archives in the United States and abroad now contribute information concerning their activities to this newsletter.

As yet, the historian has been in most respects a very junior partner in most data archiving developments. Because only a relatively small proportion of historians has been engaged in automated data analysis, economists, political scientists and other social scientists have generally shaped the processes of generating and storing machine-readable data. Moreover, these social scientists for the most part have been interested in contemporary society's institutions and problems; their data needs have therefore overlapped considerably and their style of research and use of social science theory have been similar, regardless of discipline. On the other hand, even when historians have produced computer-aided quantitative research, it has often been essentially humanistic in intent and form, rather than designed to test specific social science theory. An alliance between the social scientist and the quantitative historian in the common cause of data archiving is thus not as easy as preliminary consideration might suggest.

Historians have not taken complete advantage of the opportunities that the computer age has offered them. For example, they have not exploited the machine-readable data files issuing from the social scientists' survey research that time is rapidly converting into historical source material. The rigorously designed public opinion surveys of the past thirty to forty years—as contrasted with the heterogeneous collection of public opinion polls conducted by newspapers and magazines extending back to the nineteenth century—now provide a longitudinal dimension of some magnitude. Although splicing data taken from a variety of survey agencies and polls is a task requiring much skill, survey data offers a rich but still ignored field of research to historians interested in quantitative approaches to the history of the United States since 1936. Despite historians' neglect of these resources, some historians have been strongly influenced by the theoretical and methodological trends in the social sciences during the last generation and by the development of computer-aided research during that period.

In retrospect, it is now clear that the late 1950s and early 1960s were crucial years in the development of quantitative analysis in

economic, political and social history in the United States. Developments in economic history were more spectacular in various respects than those taking place in political and social history; even in these latter areas, however, some fundamental work was being done, various seminar directors were stressing the yield to be expected from the application of social science theory and quantitative methods in history, and the groundwork was also being laid for a truly impressive collaborative effort in the development of machine-readable data archives. As a number of researchers came to appreciate the importance of nineteenth-century voting returns for the development of a "new" political history, they also became aware of the wasted motion and resources involved in repeated trips to the basic sources, then in "a disreputable state of scatteration."[3] Researchers individually laboriously abstracted data, prepared code sheets, and had the data keypunched before data analysis could begin. Why not therefore develop some sort of collaborative venture that would retrieve the data necessary to prepare a master file of machine-readable popular voting returns for all the states for as much of our national history as possible, and which would, once completed, be freely available to all interested researchers?

Stated so simply, such a query had only one sensible answer; in the late 1950s, however, the audacity of the proposition implicit in it was startling. Historical research had been typically the work of lonely prospectors. Collaborative effort was not without precedent among historians, however; they had long utilized the resources of their national association to print collections of manuscript sources and to produce bibliographic compendia and guides to manuscript holdings. Moreover, the understanding by political scientists that one body of quantitative data—e.g., the results of a presidential panel survey analysis—might serve as the basis for a considerable number of studies done by different scholars was easily extended to the popular source that the election returns promised to be. Such thinking crystallized in the conversations of Charles Sellers, Lee Benson, and William Riker (a political scientist) in 1958 when all were fellows at the Center for Advanced Study in the Behavioral Sciences at Stanford, and in various discussions which these men held at professional meetings with historians such as Samuel P. Hays.

General descriptions of the development of the Historical Data Archive of the Inter-University Consortium for Political Research have appeared elsewhere, and the specific details appear in the annual reports of ICPR.[4] As the result of sympathetic reaction on the

part of W. Pendelton Herring, president of the Social Science Research Council, and members of its board of directors (particularly V. O. Key, Jr. and Roy F. Nichols), W. Dean Burnham obtained funds to explore the feasibility of retrieving and preparing machine-readable files of American electoral returns. Warren Miller, director of ICPR, and Angus Campbell of the University of Michigan's Survey Research Center believed that the testing of theories generated in electoral survey analysis against time series, although presenting various analytical and conceptual problems, also held the promise of significant substantive and theoretical advances in the study of American politics. They were therefore sympathetic to the idea that a historical data archive might well be an appropriate activity of ICPR. Cooperating with Miller, Benson performed a remarkable feat of organizational legerdemain by obtaining the approval of the Council of the American Historical Association (AHA) for the organization of an Ad Hoc Committee for the Collection of the Basic Quantitative Data of American Political History (AHAQDC) under the association's aegis.[5] This committee in turn organized a network of state subcommittees to collect county-level electoral data from the published government documents and archival records of every state since 1824 to be forwarded to ICPR for processing into machine-readable form. Both the National Science Foundation (NSF) and the Ford Foundation assisted in financing the project. A historian interested in political analysis, Howard Allen, joined the ICPR staff in 1964 to supervise the development of the Historical Data Archive and worked to solve the special problems involved in processing historical data and in developing an effective archival system for them.

In cooperation with ICPR, the AHA committee sponsored a conference at the Fels Institute in Philadelphia in 1964 in order to discuss the problems involved in the construction from federal census data of collateral series of demographic, economic and sociocultural data at the same county level of aggregation as the electoral series. As these developments moved forward, committee members learned that the roll calls of the U.S. Congress from 1789 through the 1930s were available in a form suitable for machine processing as the result of the work of personnel in a WPA project directed by Clifford Lord during the later years of the Great Depression. As a result of these developments, researchers were able to obtain from ICPR data from three major historical data series by the late 1960s. By then, officers in the foundations instrumental in providing funding for retrieval and processing of the historical data series had concluded that a more

conservative approach to the support of data processing was in order. As a spokesman of the National Science Foundation put it, the time had come to see whether the very considerable investment made in the Historical Data Archive would be justified by the degree to which scholars used it. Since the late 1960s major funding agencies have, in general, subsidized data retrieval and archiving only as an aspect of substantive research projects.

With the completion of the processing of the major historical data series and a change in attitude at several major funding agencies, the first major phase in the development of the ICPR Historical Data Archive ended. The process should be considered a rather striking achievement. Although the state subcommittees of AHAQDC occasionally supplied data for processing that was inadequately described or otherwise deficient, the widespread network generated a sense of involvement and performed an advertising and educational function that might have been lacking if dependence had been placed solely on a central organization. At one point, AHA committee and ICPR personnel had hoped that the electoral data could also be aggregated at the level of minor civil subdivisions, and ICPR staff experimented with Wisconsin voting returns at the precinct level. It became clear, however, that the funds required for aggregation of voting data at this level were not available. Some scholars have argued that minor civil-subdivision data allow electoral analysis of higher quality than do county returns, but this line of criticism not only overlooks some of the problems involved in the use of minor civil-subdivision data, but also exaggerates the deficiencies of county level data and underestimates the range of analysis possible in its use. Furthermore, it ignores the possibility of using minor subdivision sample data in conjunction with analysis of the larger units, and disregards the cost constraints that have been involved. There are few—if any—instances in which researchers have convincingly refuted general conclusions carefully derived from county level analysis on the basis of analysis of smaller electoral units.

The second phase of development of the Historical Data Archive of ICPR was less spectacular than the initial stage, but perhaps has been even more impressive.[6] Requests from investigators for data from the basic historical files have increased in number rather steadily, reaching a figure of 21,081,895 card images in 1973-74, approximately one-half the number of card images distributed from the Survey Research Archive of ICPR. About 50 percent of the applicants are professional historians, with the remaining number of applicants

from the social science disciplines in general. Meanwhile, the archive has continued to expand. Retrieval and archiving activity funded in conjunction with research proposals supported by NSF and the National Endowment for the Humanities (NEH) in French social and political history has produced important series of data from the *Statistique générale de la France.* The latter foundation also provided funds to the Wisconsin State Historical Society and ICPR that allowed staff members of these institutions to extend the American county level electoral series from 1824 back to 1789 insofar as these data have been found to exist. ICPR personnel have also obtained data files bearing on the political and social development of other nations that are comparable to the basic American series in the Historical Data Archive.

ICPR personnel now routinely invite historians and social scientists holding data files of interest to the historical and social science communities to present copies to the Historical Data Archive. Although in the late 1960s the Consortium Council approved the creation of an advisory subcommittee headed by the historian member of the council to provide policy recommendations and identify specific data sets of interest for the Historical Archive, relations with the AHA committee remained close. Participants at conferences held in 1967 at Ann Arbor by the AHA committee surveyed the quantitative data resources of other nations in papers ultimately published under the editorship of Lorwin and Price,[7] and it was in these conferences that the *Statistique générale* project originated. The staff of the ICPR Historical Data Archive used the committee's survey of available data files in the hands of researchers during 1973 in solicitation efforts. No other data archive currently approaches the ICPR Historical Data Archive in the number, size and utility of its machine-readable file holdings in the field of history.

One of the great resources of the historian has always been the body of records accumulated in the departments and other government agencies in the national capital and, since 1934, the special archival concern of the National Archives. By the late 1950s, as government agencies in Washington turned increasingly to the use of machine processing of data, it became clear that the world of the archivist, well-ordered conceptually if not in fact, was also becoming a great deal more complicated. In an earlier era, federal archivists had decided that the Hollerith cards used for recording individual census returns and other quantitative data were not to be included in the categories of records retained by the national government. Never-

theless, the very magnitude of the movement to convert data into machine-readable form for analytical purposes convinced various farsighted individuals at the National Archives, notably Meyer Fishbein of the Records Appraisal Division,[8] that much of the basic data concerning governmental functions preserved during earlier periods in some form of written, typed or printed records would be lost if archivists ignored the thousands of tapes and other machine-readable data files then existing in the governmental community.

Despite such interest, it was not clear during the early 1960s that the National Archives would serve as the agency through which basic machine-readable records of the federal agencies were to be preserved. Social scientists active in the American Economics Association and the Social Science Research Council promoted the idea of a federal data center which would be responsible for coordinating the management and preservation of machine-readable data files in the Washington agencies and bureaus. This point of view was advanced most notably in the report of the so-called Ruggles Committee of SSRC.[9] These advocates of a federal data center were primarily concerned with facilitating the access of investigators in the research community to important bodies of contemporary economic data being developed or held by government agencies; discussion of the role of the National Archives in the continued preservation of government-generated data was masterfully unclear in the Ruggles report. In the past, its authors noted, the National Archives had preserved basic records (such as corporation income tax returns since 1909), but had also discarded data derived from intermediate stages of processing. Machine-readable data analysis, however, had produced a situation in which intermediate records were sometimes more valuable than the final data. Although it was noted that National Archives personnel were becoming aware of the problem, "the problem is so vast that it may require completely new procedures and new policies in the future."[10]

Unfortunately, the idea of a federal data center was much more reasonable to the research community than to the members of Congress and the growing numbers of Americans who saw computers and data banks as a threat to individual privacy. That society should try to understand itself, and that procedures could be developed to enable scholars to contribute to that end without invading the privacy of the individual in an embarrassing or harmful way, seemed to be maxims that carried more conviction in the conference rooms of the Social Science Research Council and the symposia at social science confer-

ences than in congressional committee hearings or the minds of both conservative and liberal congressmen. If not dead, the idea of a federal data center was certainly in cold storage by the late 1960s.

In 1967 the Archivist of the United States set up a committee to study the machine-readable data holdings of the federal agencies. The committee's report resulted in the establishment in 1969 of a Data Archives Staff to inventory federal tape libraries and identify files that were believed to have continuing value. From this beginning developed the current Machine-readable Data Division of the National Archives, headed since July 1974 by Charles Dollar, a historian and author of various publications in the field of quantitative history. In April 1975 the division housed fewer than 1,500 reels of tape,[11] clearly a very modest beginning in view of highly tentative estimates that "agencies of the Federal Government store and process information on the equivalent of 11 million computer tapes," and that automated processing is still increasing in the federal government.[12]

The staff of the Machine-readable Data Division of the National Archives faces immense problems in evaluating tape content, in providing storage facilities, in surmounting technological obsolescence of software and hardware, and in offering services to researchers; this great institution is indeed just embarking on a most fateful venture. Obviously, it is one which in the future will affect the lives and work of many historians who have no wish at present to be involved in the computer revolution. In some cases, e.g., automated correspondence files, the research that follows recovery of the items will be little different in the future than the research done by many scholars in the correspondence files at the National Archives today. In other cases, such as analysis of the personnel records of various agencies, the logic of the storage medium will be translated by some researchers into more elaborate quantitative analyses than would otherwise have been the case. For other quantitative series, the rigor of historical analysis will equal that of analysis performed by government statisticians with contemporary objectives in view.

For the most part, the data that concern the Machine-readable Data Division of the National Archives are analogous to the data being generated by survey research. It bears upon the present and will lie within the domain of "recent" history for a generation to come. It is doubtful that the division will have the manpower or funds in the near future to convert the archives' retrospective holdings of quantitative data now in conventional form into machine-readable form. Staff members of the Machine-readable Data Division are presently

interested in serving as an information clearinghouse for files which were derived from federal records or were funded by federal agencies. Two examples of such files are: (1) research tapes using data from the U.S. censuses of the nineteenth century, and (2) the many machine-readable transportation studies of American cities financed by the Federal Highway Administration during the past forty years.

If one can say that the National Archives has begun to adjust to the computer age, one cannot go quite so far in describing the reactions of most state archives. These agencies have typically had difficulty gaining sufficient legislative support to provide housing for their records, let alone organizing them for efficient use. Like federal government agencies, state agencies have harnessed the computer, but apparently only the staff of the Florida State Archives have yet moved beyond the point of inventorying and appraising machine-readable data files to the task of actually preserving them, although some machine-readable indexing and inventorying projects are now underway elsewhere.

In at least one instance, however, a state archives has begun to convert important data series into machine-readable form. In March 1975, John Daly, director of the Illinois State Archives, announced an effort to promote the greater use of state government records of the type useful to practitioners of the "new" histories. He reported that his agency was "preparing an attempt to place on Hollerith Cards all of the data in regard to land purchase entries in the Federal District Land Office tract books for Illinois, as well as similar entries found in the records of the state [land disposal agencies]."[13]

State historical societies have not shown much concern for machine-readable data. Among them, the Ohio Historical Society has made the most impressive effort, organizing the Ohio Data Archives in 1973. Directed by Eugene Watts of the history department at Ohio State University and assisted by an advisory board drawn from the history and social science departments of various Ohio universities, this agency planned to

> conduct a continual search for quantitative material and . . . administer the technical functions of accessioning, storing, and then diffusing such data on a basic cost basis. . . . The major requirements for data set accessions are that the material must be related to some aspect of Ohio, it must have been collected in a professionally competent manner, and it must have a potential interest for other users.[14]

Materials collected are stored on magnetic tape and disseminated to interested scholars. Watts has reported considerable progress toward the agency's goals but somewhat less enthusiasm and cooperation among historians than he had hoped.[15] Antedating the developments in Ohio by some years (but not planned as a continuing enterprise) were the efforts of staff at the Wisconsin State Historical Society in cooperation with the Wisconsin State Archives to prepare a detailed set of machine-readable county level data depicting Wisconsin's economic, social and human resources throughout the state's history. The code books for these data became available in early 1975 and the data may be obtained through the Data Program Library Service of the University of Wisconsin.

The archival and historical agencies supported by American governments are clearly beginning to react to the advent of automated data systems, but the process has not gone far. There are lessons to be learned by both archivists and the agencies whose data they must preserve, as well as institutional adjustments to be made, including recognition of the fact that surrender of agency tapes is expensive to a degree that surrender of paper records was not assumed to be.

In effect, the scholars who initiated the data retrieval and archiving projects that provided the foundation for the Historical Data Archive of ICPR were predicting that considerable numbers of researchers would use the historical machine-readable data series once they became available. None of them were planning research projects that required analysis of more than a small portion of the data to be assembled; nevertheless, they believed that the electoral, legislative and ancillary data files were so important to the analysis of significant political and social processes that they would be widely used once available. As it turned out, Miller, Benson and their colleagues correctly assessed the importance of the electoral and congressional roll call data.

The staff of the Machine-readable Data Division of the National Archives and those of state archives and historical societies must also forecast future needs when making decisions concerning both the preservation of current governmental records in machine-readable form and the conversion of conventional records to that medium, if the latter policy becomes feasible. Although the personnel of the National Archives have made commendable efforts to seek information about automated data processing in the research community, it might be appropriate for the archivist to create a continuing advisory group of scholars qualified to counsel the staff of the Machine-read-

able Data Division, in order to evaluate the adequacy of conventional archival criteria for preservation for application in the automated data area, and to provide a continuing flow of information on scholarly needs. One or more conferences on these matters would perhaps be an appropriate beginning. State archival staffs should follow the same route.

Realistic consideration of the current situation suggests that data retrieval and conversion projects comparable in scope to the original ICPR-AHAQDC venture cannot be financed at present. Perhaps, however, it is time for NSF personnel to review their thinking of the late 1960s. If the grants in support of the ICPR archive are generating research returns in amounts comparable to or greater than equivalent NSF funds spent elsewhere, some reevaluation and reprogramming may be in order at both NSF and NEH. Setting this possibility aside, there would be considerable disagreement among researchers about which data sources should be given first priority in further archiving activities. A national households sample from the manuscript federal censuses from 1850 to 1950, or a sample drawn from the corporation income tax returns preserved by the National Archives, however, would certainly prove highly attractive to investigators. Should the archivists of the Church of Jesus Christ of Latter-day Saints find the resources to automate all or a major part of the magnificent collection of local records collected by that body during the last generation, behavioral historians would find themselves in a delightfully different ball game.[16] Barring unexpected developments, however, usage and cost factors will probably dictate for the time being that much of the historical machine-readable data generated in the near future will be an outgrowth of the research of individuals or teams who are employed by educational or research institutions and derive additional support from private or public research funding agencies.

This latter type of data set or file is as old as the computer age. The history of the Parker-Gallman sample of southern farms and plantations drawn from the 1850 and 1860 federal agricultural and population censuses is illustrative. William Parker began the actual work of drawing this sample from microfilms of the manuscript census returns in 1960 while at the University of North Carolina, and a considerable number of scholars have used this information—notably Parker and Robert Gallman, their students, and Robert Fogel and Stanley Engerman while preparing *Time on the Cross*.[17] The sample has never been placed in a central depository for preservation and

circulation; other investigators have obtained copies by requesting them from the research team that could most conveniently provide them at the time. Gallman is currently taking steps to place a master file on deposit in an appropriate archive. More recently, Roger Ransom and Richard Sutch have developed a similar sample of southern farms and plantations from the 1880 manuscript census and, building upon the initial work of Fogel and Engerman, Fred Bateman and James Foust have prepared a sample of northern farms from the 1860 manuscript census. Obviously, these data bear upon only a limited era in the history of American agriculture, and that incompletely.

The publication of Stephan Thernstrom's study of social mobility in Newburyport[18] gave a great fillip to the systematic quantitative analysis of urban populations. The development of machine-readable data sets drawn from the U.S. censuses of the nineteenth century and related materials has been underway since the mid-1960s. A number of such data files promise to be as interesting to social and political historians as the Parker-Gallman file has been to economic historians. These include: (1) data sets involving the populations of Philadelphia, Hamilton (Ont.), and Kingston, Buffalo and Poughkeepsie (N.Y.), developed by Theodore Hershberg, Michael Katz, Stuart Blumin, Laurence Glasco, and Clyde Griffin; (2) a Pittsburgh file created by Glasco with the encouragement of Samuel P. Hays; and (3) one covering Cincinnati populations developed by Zane Miller and Guido Dobbert. Facing common problems of occupational classification, the group first mentioned have taken pains to make their data sets compatible in order to foster comparative analysis. Other projects that will provide useful data include those of Richard Jensen and colleague, who are developing migration and population files in cooperation with the Institute for the History of the Family at the Newberry Library.

It is impossible to estimate how many current research projects will generate machine-readable data files that should ultimately be made available for use by other interested researchers. Swierenga recently noted that more than 300 historians have reported ongoing computer projects to various clearinghouses since 1965;[19] of course, sociologists, political scientists and other social scientists have meanwhile been developing various data files of a historical nature as well. Although restricted in its circulation by focus and by the medium of publication, a questionnaire circulated by the AHAQDC in 1973 elicited 225 responses, and approximately 90 of the respondents reported that

they held data sets of interest to other researchers which they were willing to archive.[20] In contrast, a recent listing cites only seventy-four data sets "complete and available for use" in machine-readable archives, although some of the ICPR files listed are massive and could be used by large numbers of individual scholars.[21] This is not to imply that all machine-readable data sets should be archived or maintained in readiness for prompt distribution. Some are of such limited interest to other investigators that preservation by a central agency would surely represent a waste of resources. Identification of such files is sometimes difficult, however, and is clearly one of the continuing problems of the era of machine-readable data.

It is easy to maintain that historians who develop machine-readable files in the course of their research should be willing to make them available to other researchers at an appropriate point in the investigation. It is much more difficult for the reader to check the work of the researcher who uses computer analysis than one who uses conventional sources. Theoretically, the researcher should welcome critical examination of his work and be willing to facilitate it. Moreover, many data sets can be used for a variety of types of secondary analysis that the original investigator often has no intention of performing. Considerations of this sort have led most funding agencies of the federal government to specify that data collected in projects for which they provide funds are to be considered government property rather than the property of the individual researcher, and that such data sets should be available on request to other interested parties. No requirements are made as to the form or general condition of the data when they are surrendered, however.

Many data sets have coding idiosyncrasies or other troublesome characteristics that the original compiler tolerated because of familiarity with the material, but which lessen their utility to others. Few data sets have arrived at ICPR that did not need some degree of cleaning or reclassification. Other data sets reflect the idiosyncratic computer facilities of the researcher's institution. In the current state of the arts and ethos, few principal investigators, having completed a research project, are prepared to spend additional weeks or months reformatting data and code books for their maximum usefulness in an archival depository. If a file is allowed to sit on the researcher's shelf for a few years, however, tape or card deterioration or computer processing developments may render it useless. Those interested in the development of data archives have long argued that funding agencies should require funded researchers to place their data in a

depository named in the initial grant application within a specified period of time.[22] Nevertheless, the data archivist's only weapon in extracting data sets from scholars remains moral suasion; too often it has proven inadequate to the task.

American historians are presently much less apt to use relevant and available machine-readable data in their research—or to develop machine-readable data when appropriate—than are social scientists in general. Behaviorism came late to history and, in contrast to the social sciences, relatively few of the great army of Ph.D.s trained during the 1960s were committed to the use of quantitative analysis in their research. The proportional increase of such individuals was rather considerable, however, and by the end of the 1960s many history departments were becoming interested in hiring individuals who could teach courses in quantitative analysis and data processing. Unfortunately, the onset of the academic depression blunted this development to a considerable degree. Established historians have also essayed quantitative research, but they have often found the road rocky. Decisions made during the Nixon administration to restrict or eliminate the postgraduate training programs of the National Science Foundation and other federal agencies have made it difficult for interested groups to serve the middle-aged scholar who wishes to retool.

With this situation in mind, it is not surprising that the ethos of cooperative research and an understanding of the importance of secondary analysis are less pervasive among historians than among social scientists. Although the ICPR staff contacted all individuals expressing willingness in the AHAQDC 1973 survey[23] to make their data files available for the use of other researchers, few of them have yet deposited their materials in a generally accessible archive. Because most historians are still learning the niceties of coding and data processing "on the job," their files may well be more difficult to clean and service than those obtained from social scientists.

Historical data presents investigators with somewhat different problems of research, design, coding and manipulation than those faced in many social science studies. Developing a file for a panel survey in which respondents are interviewed over a relatively short period of time is considerably different, for instance, from working with historical census data in which individuals may or may not be represented in a series of enumerations separated by substantial time lapses, or from developing a file in which data from several different sources must be merged. Historical data bodies may be massive. Is the

drawing of a sample of individual cases an appropriate technique, or is the recording of complete populations of cities, wards, counties or townships a better strategy? There are differences of opinion as to the most appropriate strategies to follow in processing data in such cases, and the divergence extends to highly technical issues. The editor and authors of the *Historical Methods Newsletter* have performed yeoman service in presenting such matters, but the inexperienced historian and some who can lay claim to considerable experience in such matters may be confused by the claims and counterclaims of enthusiasts, or fail to understand that technically elegant solutions may not be appropriate to their circumstances. Part of the solution to these problems rests in the hands of the departments that produce the Ph.D.s of the history profession. Both undergraduate and graduate programs must adapt to computerization so that adequate training in historical data processing and quantitative analysis can be provided. This is, of course, a long-range solution; for the present data archives and interested associations, or agencies such as the new Social Science History Association, must work to raise the level of expertise within the profession. Should history departments and other agencies fail in such efforts, the answer may be that social science departments, now increasingly interested in the historical dimension, will rear their own breed of social science historian, a development already well advanced in the field of economics.

If all historical researchers were now ready and willing to deposit their data files, there would probably not be archiving facilities sufficient to clean, catalog, maintain and circulate them. Essentially, these are library functions, but few university libraries have yet established machine-readable data collections or moved to incorporate the data archives that have developed on campus in response to social science research and teaching needs during the past twenty years.[25] This is easy to understand. The flood of publication during the last generation and the striking increase in the number of student and faculty library users have strained the capacities of university and college libraries, while inflation has eroded the value of the library funds available for the purchase of library materials. Library administrators generally lack the specialized knowledge necessary to supervise machine-readable data archives and positions in such agencies cannot be adequately filled by conventionally trained library or archival personnel. Some librarians fear that control and copyright problems equivalent to those already being encountered in providing photocopy services may develop in this new field of service.

There are indications that the data archives may indeed gravitate to the control of the university library. For instance, the budget of the Princeton University Library includes the campus ICPR membership fee, as is true in a few other cases. Some library personnel are becoming acquainted with the technical problems involved in managing machine-readable data. Developments in the National Archives may serve as a model in this respect, and the availability of 1970 U.S. census data tapes has brought some librarians face to face with the new age.[26] University and college library personnel are becoming acquainted with computer capabilities in other connections as well, notably in the automation of ordering procedures and circulation systems and in the development of computerized regional cataloging, interlibrary loan, serial control and processing information systems such as the Ohio College Library Center.[27] As the machine-readable data file becomes better recognized as a research resource, and literary works increasingly come to have their machine-readable editions, the logic of making data archives a part of library services will become convincingly apparent.

This article suggests that the computer revolution has not yet stimulated massive and imaginative response in either the historical profession or those supporting agencies that have typically served its members. Neither the potential magnitude of change that the computer promises in historical research nor the unique problems faced by its historian users are well understood by either the academic community or the public and private agencies that supervise and sustain it. It is certainly normal for disciplines to experience periods of crisis when both its members and the public question its utility. Some believe that time-oriented studies currently face such a crisis today; shrunken college and university enrollments in history courses, elimination or reduction of history requirements in teacher training, and public disinterest in the writing of most historians are cited as evidence. To some historians, the use of quantitative methods and the computer is part of a broader effort to develop a more theoretically oriented and rigorous variety of history that will assist in understanding human development and contemporary society. This, they believe, is the appropriate answer to history's malaise and they find confirmation of this view in the fact that after a behavioral revolution that was notably anti- or ahistorical in tone, many social scientists are moving to reintroduce a historical dimension in their research. Within a short time, a relatively small number of quantita-

tive historians have substantially altered the appearance of considerable stretches of historical terrain.

Like most innovators, these historians face the distrust of colleagues who see no cause for alarm in the profession, or believe that the remedy lies in doing old things better. More serious is the fact that these "new" historians do not fit into the current research establishment—their research proposals often lack the theoretical component that NSF demands, but at the same time are too quantitative to be received enthusiastically at NEH, where the code word for success is now *humanistic* (or so at least disappointed applicants sometimes believe). Such historians are also frustrated because their research concerns are not always taken seriously in other government agencies, where they should be. The efforts of the U.S. Bureau of the Census to block the opening of the population schedules of 1900 culminated a generation of frustration for historians who had seen the bureau terminate its former practice of certifying qualified historians as bureau clerks for the purpose of research, and successfully recommended destruction of the manuscript agricultural census schedules for 1900 and succeeding years. Research conducted in the open census schedules for the years 1850-80 has amply demonstrated that it is fallacious to assume that the contemporary analysts of any given census will indeed ask and answer all of the questions that seem important to succeeding generations of scholars seeking to understand their society. Historians recognize that the citizen's privacy must be protected, but they also know that this can be done without destroying the census rolls or establishing excessively long periods of cloture. In making suggestions to these ends, spokesmen of the Bureau of the Census, historians believe, are trying to offer the historian as a sacrifice to those who would deny society its right to understand itself.

Several years ago, Angus Campbell published a wry description of the dangers threatening the social scientist who tried to reach a New Jerusalem along the Glory Road opened by the computer: cost problems, data problems, organizational problems, confidentiality problems, etc.[28] Five years later the road is still there—and so are the obstacles. Relatively few historians are on that road yet, and it may be more difficult for them than for social scientists. Given the state of their discipline, however, it is essential that they push on.

References

1. Robbin, A.A. "Social Science Data Access Problems: A Brief Description of the Wisconsin Experience." *In* William G. Tyler, ed. *Data Banks and Archives for Social Science Research on Latin America* (Publication No. 6). Gainesville, Fla., Consortium of Latin American Studies Programs, 1975, pp. 11-25.

2. Kolp, John G. Personal communication, June 5, 1975.

3. Campbell, Angus. "Some Questions about the New Jerusalem." *In* Ralph L. Bisco, ed. *Data Bases, Computers, and the Social Sciences.* New York, John Wiley and Sons, 1970, p. 43.

4. Inter-University Consortium for Political Research. *Annual Report.* Ann Arbor, Mich., Inter-University Consortium for Political Research, 1963- . *See also* Bogue, Allan G. "United States: The 'New' Political History," *Journal of Contemporary History* 3:5-27, Jan. 1968; and Clubb, J.M. "Historical Politics: American Elections, 1824-1970," *Social Science Research Council Items* 25:46-50, Dec. 1971.

5. The committee's name changed somewhat through time and its scope was broadened to embrace quantitative data in history generally. The Consortium's name has recently become Inter-University Consortium for Political and Social Research.

6. Silbey, Joel H. "Clio and Computers: Moving into Phase II, 1970-1972," *Computers and the Humanities* 7:67-79, Nov. 1972.

7. Lorwin, Val R., and Price, Jacob M., eds. *The Dimensions of the Past; Materials, Problems and Opportunities for Quantitative Work in History.* New Haven, Conn., Yale University Press, 1972.

8. *See* Davis, L.E. "Preservation of Machine-readable Records" (Project NN-570-1). Washington, D.C., Records Appraisal Division, National Archives and Records Service, 1969 (ditto); and Fishbein, Meyer H. "Appraising Information in Machine Language Form," *American Archivist* 35:35-43, Jan. 1972.

9. Ruggles, Richard, *et al.* "Report of the Committee on the Preservation and Use of Economic Data to the Social Science Research Council." 1965. (mimeographed)

10. *Ibid.*, p. 8.

11. U.S. National Archives and Records Service. *Catalog of Machine-readable Records in the National Archives of the United States.* Washington, D.C., National Archives Trust Fund Board, *et al.*, 1975.

12. Dollar, Charles M. "Computers, the National Archives, and Researchers." p. 3. (unpublished)

13. Daly, John. Personal communications, March 20 and April 21, 1975.

14. Ohio State Historical Society. *The Ohio Data Archives: Sponsored by the Ohio Historical Society.* Columbus, Ohio State Historical Society, 1973.

15. Watts, Eugene J. Personal communication, April 23, 1976.

16. Wimmer, Larry T., and Pope, Clayne L. "The Genealogical Society Library of Salt Lake City: A Source of Data for Economic and Social Historians," *Historical Methods Newsletter* 8:51-58, March 1975.

17. Fogel, Robert W., and Engerman, Stanley L. *Time on the Cross.* Boston, Little, Brown, 1974.

18. Thernstrom, Stephan. *Poverty and Progress; Social Mobility in a Nineteenth Century City.* Cambridge, Mass., Harvard University Press, 1964.

19. Swierenga, Robert P. "Computers and American History: The Impact of the 'New' Generation," *Journal of American History* 60:1051, March 1974.

20. Bogue, Allan G. "The Quantitative Data Questionnaire," *AHA Newsletter* 11:27-30, Sept. 1973.

21. Falk, Marvin, and Kolp, John G. "Quantitative Data," *AHA Newsletter* 12:7-8, Dec. 1974.

22. Adams, M.A., *et al.* "Proceedings of the Workshop on the Management of a Data and Program Library, University of Wisconsin, June 19-20, 1969." Madison, Wis., Data and Program Library Service, University of Wisconsin, 1969, pp. 29-34. (mimeographed)

23. Bogue, "The Quantitative Data Questionnaire," *op. cit.*

24. *Historical Methods Newsletter.* Pittsburgh, Department of History and University Center for International Studies, University of Pittsburgh 1967- .

25. Fussler, Herman H. *Research Libraries and Technology; A Report to the Sloan Foundation.* Chicago, University of Chicago Press, 1973, pp. 51-72; and Weber, David C. "Information for Contemporary Times." *In* Association of Research Libraries. *Coping with Change: The Challenge for Research Libraries* (Minutes of the Eighty-Second Meeting, May 11-12, 1973, New Orleans, Louisiana). Washington, D.C., ARL, 1973, pp. 31-38.

26. Rowe, Judith S., and Ryan, Mary. "Library Service from Numerical Data Bases: The 1970 Census as a Paradigm," *College & Research Libraries* 35:7-15, Jan. 1974.

27. Aman, Mohammed M. "Computer Applications in Academic Library Operations." *In* E.J. Josey, ed. *New Dimensions for Academic Library Service.* Metuchen, N.J., Scarecrow Press, 1975, pp. 86-99; and Kilgour, Frederick G. "The Ohio College Library Center: A User-oriented System." *In* Josey, *op. cit.*, pp. 250-55.

28. Campbell, *op. cit.*