The Henry A. Murray Research Center: Alternative Data Sources - Unique, Yet Less Visible Archives and Programs

by Erin Phelps [1]
The Henry A. Murray Research Center
of Radcliffe College

BACKGROUND INFORMATION ON THE MURRAY RESEARCH CENTER

The Henry A. Murray Research Center of Radcliffe College was established in 1976 as a national repository for data from the fields of psychology, psychiatry, sociology, anthropology, economics, political science, and education. Unlike most other data banks, the Murray Center archives original subject records as well as coded, machine-readable data. These original records often include transcripts of in-depth interviews, behavioral observations, responses to projective tests, or other information that can be used profitably for secondary analysis. This makes possible a restructuring of the subject records and mitigates the degree to which one is locked in to the theoretical assumptions under which the data were collected.

In spite of the clear advantages of making raw data available, most data banks offer only coded computer data. Some major longitudinal studies, for example the New York Longitudinal Study (Thomas and Chess, 1977), are archived in a manner that allows access to the records by outside investigators, but in general each of these studies is housed separately. To our knowledge, the Murray Research Center is the only repository that is designed to offer a wide range of longitudinal data sets, many of which include raw data.

Data sets may be reanalyzed in order to explore questions other than those addressed by the original investigators. Reanalysis may involve recoding of the raw data. Data from studies that employ comparable designs and instruments may be combined to provide a larger, more varied sample than would be possible otherwise. In addition to reanalysis, these studies can be used as baseline data for replication studies in order to assess the effects of social change. In many cases, samples can be recontacted for further longitudinal follow-up. The availability of samples which can be followed up encourages the collection of longitudinal data that would otherwise be too costly or difficult to obtain. Furthermore, the data collected early in this century can be used to address questions of interest to social historians. Reviews of existing data can facilitate exploratory research by allowing a researcher to refine research questions, to assess the best means for addressing those questions, and to develop new research instruments or

coding schemes. In addition to using the Murray Center's data sets in the ways mentioned above, faculty and students use these data in courses to illustrate how the data are collected and analyzed.

The resources of the Murray Center are open to undergraduate and graduate students, faculty, and other researchers from around the world. Staff members conduct several introductory workshops each semester and are available weekdays to provide individual consultation on using the Center's resources. Users are not charged for access to the data. They are charged only for special services they require (such as duplicating) and for the cost of computer time. Our Guide to the Data Resources of the Henry A. Murray Research Center is available for $12 and provides information about the more than 190 data sets that are completely processed and those that are still in the process of acquisition. In addition, an Index to the Guide is available for $5, which includes a detailed listing of the methods of data collection, content areas, and an index of the data sets according to these characteristics.

An important advantage of the Murray Center is that the archive is located within an active research center. The Murray Research Center offers staff assistance to data users, seminars and conferences on methods for making productive use of existing data, and a visiting scholar's program which hosts several researchers each year in residence at the Center. This kind of setting fosters intellectual collaboration and helps to ensure that the Center's data resources are used, and used well, for important new research.

DATA FORMATS
Original Records
We currently provide information in two forms: 1) copies of original paper records and 2) coded computer readable data. In addition, we are beginning the process of microfilming paper records, and where possible, converting mainframe computer files to desktop computer floppy disk format. We are also investigating the possibility of archiving videotapes of social interaction.

In the case of studies that include original subject records, processing the data includes several steps. The first is arranging for duplicating and shipping of the

records held by the original investigator, checking the quality of the photocopies and obtaining replacements when necessary, and removing names and other identifiers. Then the material is organized by subject and measure and stored in acid free folders and boxes. A complete inventory of the material is created in order to document the occurrence of missing data, and a detailed documentation is written, describing the data collection procedures and other information about the data set that a user might need. Finally, a "user file" is created that contains the legal agreement between the contributor and Radcliffe College, the inventory and documentation, copies of measures, publications, and information on machine-readable data.

With funding from the National Institutes of Mental Health and the MacArthur Foundation, we have begun to create microfiche copies of original record data. The availability of microfiche copies will make the data more easily transferrable outside our local area. For reasons of confidentiality, however, use of raw data off-site is not allowed without the written consent of the data contributor and will be allowed only when the data are completely anonymous with respect to the subjects' identities. This rule is designed to protect the privacy of the study participants. In some cases, however, when the subjects' identities could not possibly be revealed by the records, data contributors do give permission for off-site use. In such cases, the Murray Center's clerical staff must photocopy all of the material to be sent, since using a copy service might result in loss of the records. This is obviously a very expensive and cumbersome procedure. With microfilm records available, the process of screening and controlling off-site use would be conducted with equal care, but in cases where no risk to confidentiality is involved, the ease of transfer would be greatly enhanced.

Computer Data
In addition to original subject records, most of the data sets include tapes of computer-accessible data. Data tapes are checked against codebooks, and the necessary data cleaning is completed either by the contributor or Center staff. Our current method of processing a computer data file depends upon the form in which the data are received. If a SAS file or an SPSS system file was created on a mainframe that is compatible with our mainframe, processing simply begins with the given file as it is. Commonly, however, we write a computer program in SAS or SPSSX to read a raw number file. In either case, data are carefully checked. Two forms of computer output are generated and checked against any original interview or questionnaire data accompanying the tape. The first is a frequency distribution of all variables. A computer codebook and any other documentation is used to check for out-of-range values, correct missing value designations, and appropriate labeling. Where original data are available, a case-by-

case listing of actual values for each variable is generated for a random sample of cases. This listing is compared with actual questionnaire and/or interview data to check for consistency between the two forms of data.

Computer tapes of machine-readable data are routinely copied and mailed to investigators wishing to use them. Users are only charged for the cost of the tape and the transfer. Computer data are already used predominantly off-site and are made available as SAS files, SPSSx system files, portable files on computer tapes, or sent via communications networks such as BITNET or AR-PANET. During the past year, requests for data in floppy disk format have been received, a trend that seems likely to continue and increase as desktop computers and appropriate statistical software become more widely available. Because of this, we are beginning to transfer the machine-readable data from many of the Murray Center's data sets to floppy disk format for use by desktop computers. This should facilitate use of the Center's data, as well as decreasing the cost of both transfer and analysis.

Videotape Data
Through a recent grant from the MacArthur Foundation, we are exploring the feasibility of adding another dimension to our archive — videotaped data. Videotapes of social interactions most often record not only those behaviors that are pertinent to the original investigation, but also record a wide variety of additional behaviors which may be coded by researchers investigating other aspects of human interaction. In addition, new systems for coding interaction data are continually being developed, and all require coding of pilot video data to establish preliminary reliability and validity. Existing video data can often serve as pilot data for testing a new coding system.

Because it is extremely time consuming and costly to collect observational interaction data and the data cannot be analyzed exhaustively by one investigator, these data are currently being shared among researchers. The private transfer of data from one place to another is unregulated so generally there are no clear guidelines for who may view the tapes and what conditions viewers must adhere to. An additional problem with private sharing of these data is that many potential users are not aware of the availability of the data sets that may be appropriate to their needs. Thus, in many ways video data seem ideal for inclusion in a data archive such as the Murray Center, which emphasizes the study of lives in context through reanalysis of original records.

Individual researchers and human subjects review boards are just beginning to recognize the ethical problems involved in sharing data that cannot be completely deidentified. A variety of important ethical issues must

be considered regarding the confidentiality of the subjects in video studies. For example, should parents be allowed to give consent to the archiving of identifiable video data of their children? Should subjects be required to view their videotape data before allowing the data to be placed in an archive? These ethical issues exist for any participant in data sharing, whether a private researcher or an institutionally affiliated data archive, although the issues are somewhat more complicated in the case of an archive.

This year we are spending some time thinking about whether or not we should begin a major program of archiving videotaped social interaction data. The first step is to clarify the ethical issues by ascertaining the current range of policies at major universities regarding sharing of these data, determining the possibility of creating procedures for safeguarding the confidentiality of subjects, and making recommendations regarding ethical policies for both private sharing and archiving of video data. If we conclude that it is possible to resolve the many ethical problems that exist, we will begin identifying data sets for potential acquisition.

OTHER SERVICES
Workshops and Conferences
In addition to archiving data, the Murray Research Center sponsors workshops designed to draw attention to its data resources and provide training in the skills needed to carry out effective secondary analyses. Our plan for the next three years is to offer at least three workshops each year in order to facilitate the use of the Murray Center's archive for theoretically innovative research. We have already begun this expanded outreach to the research community. Within the past year four workshops were scheduled in connection with the Murray Center's recently acquired longitudinal studies. They were: 1) A workshop on methods for life course research which was held in June, 1989 and highlighted the data analysis techniques of event history analysis and causal modeling; 2) A July 1989 workshop on methods for coding open-ended archival material for several important personality constructs; 3) A workshop on the secondary analysis of major longitudinal data sets, to be taught by Glen Elder, George Vaillant, and their associates in October, 1989; 4) A workshop on using case studies for the study of individual lives, held in May, 1990.

Grants
In order to promote actual use of archival data, we have a program of research grants. First, there is the Radcliffe Research Support Program, which offers small grants to post-doctoral researchers who wish to use data housed at the Center. In addition, there are three awards for dissertations in the areas of sex differences or female development, women's life choices and patterns, and

personality or "the study of lives."

The Longitudinal Studies Inventory
A major task of the current NIMH grant has been to expand and update the longitudinal studies inventories originally published by the Social Science Research Council (Inventory of Longitudinal Studies of Middle and Old Age, Midgal, Abeles, & Sherrod, 1981, and Inventory of Longitudinal Research on Childhood and Adolescence, Verdonik & Sherrod, 1984). This has involved entering the existing inventories into a computer database, developing a detailed index, and coding the studies according to index categories. At the same time, we have been seeking out information on studies appropriate for listing but not included in the earlier volumes. In the coming year, this process will be completed, and a new volume including studies of childhood, adolescence, and adulthood will be printed and distributed.

Other
In addition to the resources already described, we are developing a set of machine-readable files that should be useful for teachers of statistics and methods courses. We also maintain a Measures File containing copies of primarily psychological instruments, and includes as well information about the measure's development, validity, reliability, and scoring. In addition, we maintain a Feminist Critique File and bibliography. This is a collection of critiques of social science methods and theory, which is updated annually.

USERS
Data from the Murray Center archive are actively used for research in psychology, sociology, psychiatry, education, political science, history, economics, and criminology. Reanalyses of the data are carried out in order to explore questions that are very different from those addressed by the original investigator's analyses. Often these reanalyses are cross disciplinary, involving researchers from one discipline using data collected by investigators from another discipline. A recently published book by historian Elaine Tyler May illustrates this kind of cross-disciplinary reanalysis. The book, Homeward Bound: American Families in the Cold War Era, (May, 1988) draws very centrally on Lowell Kelly's 50-year longitudinal study of personality development within marriage, which is archived at the Murray Research Center. In this work, Professor May traces the connections between politics and family life in the 1950s and argues that the cold war affected virtually all aspects of life, from consumerism to sexuality. As Professor May notes, "The participants in the Kelly Longitudinal Study were among the cohort of Americans who began their families during the early 1940s, establishing the patterns and setting the trends that were to take hold of the nation for the next two decades. They entered

marriage when World War II thrust the nation into another major crisis, wreaking further havoc on families. They raised children as the cold war took shape, with its cloud in international tension and impending doom" (p. 12). This study of family life in the context of an important era in America history parallels the ground-breaking work on the psychological impact of the Great Depression carried out by Glen Elder with the Berkeley and Oakland longitudinal studies (Elder, 1979). Clearly, this kind of work requires archival data, including the recoding of original subject records.

Other research conducted at the Murray Center involves more quantitative reanalysis of machine-readable data. For example, political scientist Eileen McDonagh is using congressional voting records to investigate the political mechanisms through which policy innovation can occur at the national level (McDonagh, in press). The project introduces a new measure of constituency issue position — district level referenda votes — to study policy congruence between grass-root electorates and House representatives. The investigation of policy congruence processes in the context of partisan electoral patterns and demographic characteristics provides for multivariate analysis of the direct, indirect, and interactive effect of these major sources of policy formation.

Another example is provided by Janet Giele's reanalysis of several surveys of college educated women (Giele & Gilfus, in press). Among other things, Giele found that college-educated black women led college-educated white women in the shift that began to occur in the late 1950s and early 1960s toward more complex and hetero-geneous life patterns.

Perhaps most ambitious are studies that involve longitudinal follow-up of an existing sample. The Murray Research Center is the only archive that makes this kind of study possible. A recent example is provided by David McClelland and his colleagues' follow-up of the Sears, Maccoby, and Levin "Patterns of Childbearing" sample (Sears, Maccoby, & Levin, 1957). First studied at age 5, and recontacted on multiple occasions in the interim, the subjects were age 41 at the time of the McClelland follow-up. Among the issues addressed by the follow-up are the family origins of empathic concern (Koestner, et al., in press), changes in motivational patterns in adulthood, the relation between agency motivation and health (McClelland, 1989), and the relation of early patterns of parental discipline and warmth to successful adaptation in adulthood (Franz, et al., 1989).

CONCLUSION
In conclusion, I'd like to agree that the Murray Research Center is unique, but I hope soon will not be very much less visible! I'd like to thank Sue Dodd for including us in this session toward that end.

REFERENCES
Elder, G. (1979). Historical change in life patterns and personality. In P. Baltes & O. Brim (eds.)., Life-span development and behavior (Vol. 2, pp. 117-159). New York: Academic Press.

Giele, J., & Gilfus, M. (in press). Race and college differences in life patterns of educated women. In J. Antler & S. Bilken (eds.), Women and educational change. Albany: State University of New York Press.

Koestner, R., Franz, C., & Weinberger, J. (in press). Family origins of empathic concern: 26-year longitudinal study. Journal of Personality and Social Psychology.

May, E. (1988). Homeward bound: American families in the cold war era. New York: Basic Books.

McClelland D. (1989). Motivation factors in health and disease. American Psychologist, 44, 675-683.

Migdal, S., Abeles, R., & Sherrod, L. (1984). An inventory of longitudinal studies of middle and old age. New York: Social Science Research Council.

Sears, R., Maccoby, E., & Levin, H. (1957). Patterns of childbearing. Stanford University Press.

Thomas, A., & Chess, S. (1977). Temperament and development. New York: Brunner/Mazel.

Verdonik, F., & Sherrod, L. (1984). An inventory of longitudinal research on childhood and adolescence. New York: Social Science Research Council.