

Acquisition Criteria at the Murray Research Center: A Center for the Study of Lives

Jacquelyn B. James, PhD
Associate Director

Murray Research Center: A Center for the Study of Lives
Radcliffe Institute for Advanced Study
Harvard University

A paper prepared for presentation at the CODATA 18th International Conference, Montreal, Canada, 2002.

Abstract

The Murray Research Center is a repository for social and behavioral sciences data on the in-depth study of lives over time, and issues of special concern to American women. The center acquires data sets that are amenable to secondary analysis, replication, or longitudinal follow-up. In determining whether or not to acquire a new data set for the archive, several criteria are used. The criteria can be roughly grouped into five general categories: content of the study, methodology, previous analysis and publication record, historical value, and cost of acquiring and processing the data. Each of these will be described with an indication of the relative importance of each criterion, where possible.

Founded in 1976, the Murray Research Center: A Center for the Study of Lives is a national repository for social and behavioral science data on human development and social change, with special emphasis on the lives of American women. Data housed at the Murray Center are made available to qualified scholars and researchers for secondary analysis, replication, and sometimes follow-up studies. Presently, our archive includes approximately 280 data sets with a wide range of topics, samples, and designs. Many of these studies include in-depth interviews or at the very least, some open-ended survey questions. We make it a priority to acquire data for our collection that have not been exhaustively analyzed, which contain qualitative or interview data, or which are longitudinal in design. The purpose of this paper is to document the ways that the center selects and archives these kinds of research materials.

The Murray Research Center: A Center for the Study of Lives

The issue of underuse (or waste) of data has been a concern of funders and research administrators for some time and has recently received renewed attention (McArdle, 2000; Schaie, 2000; James & Zarrett, 2000; Ferraro & McNally, 2000). As Kozlowski (1993) [cited in Colby, James & Hart, 1998] has observed, "The further one gets from a project, the greater is the chance of loss. Those details which once seemed too obvious to note become fragmented or lost." For this reason and because grants tend not to fully cover thorough and complete analyses of complex data sets, projects are often left with a great deal of valuable data unanalyzed. A central purpose of the Murray Center is to turn data that might otherwise be wasted into a rich and accessible resource for new research.

In order to be an effective resource for new research and contribute to minimizing the waste of data, it is important not only to preserve and document data, but also to let the research community know about the availability of the data and to provide some training in how to use it. The latter is especially important, because methods for secondary analysis, especially secondary analysis of qualitative data, are unfamiliar to many researchers and not taught in many graduate study programs (Colby, James and Hart, 1998). Thus, the Murray Center serves as both a repository and a research center that provides opportunities for training in the use of existing data.

At least in the United States, the Murray Center's archive is unique in several ways. First, the Murray Center is the only archive that preserves the original subject records as well as coded, machine-readable data. As far as we know, the Murray Research Center is the only repository in the U.S. that is designed to offer a wide range of data sets with original, qualitative records, many of which are longitudinal.

The Center is also unique in that it allows for the possibility that samples from many of the studies it holds are available for further follow-up by a new investigator. The potential to conduct a follow-up study is very valuable in that it allows a new investigator to design the outcome measures used. Moreover, by allowing follow-up studies to be conducted, a study that was not longitudinal can become so. In general, the availability of multiple longitudinal data sets within the center's holdings makes it possible to add a new cohort to a single cohort longitudinal study or to integrate two data sets into a single multi-cohort study (Colby and Phelps, 1990).

Approximately 85 of the center's holdings are longitudinal studies. New studies and new

waves of existing studies are also added to the archive each year. In addition, the center has, over the years, received funding for special acquisition initiatives.

One such initiative, funded by the U.S. National Institute of Mental Health and the John D. and Catherine T. MacArthur Foundation, was to develop a major collection of longitudinal studies of mental health including, for example, Baumrind's *Family Socialization and Developmental Competence Project*, Brunswick's *Harlem Longitudinal Study*, Glueck and Glueck's *Crime Causation Study*, the Institute of Human Development's *Intergenerational Studies*, Terman's *Life Cycle Study of Children of High Ability*, and Vaillant's *Study of Adult Development* (The Grant Study).

With additional funding from NIMH along with a grant from the National Science Foundation, Murray Center staff established another major archive, this one designed to enhance the racial and ethnic diversity of the archive.

A few examples of these data sets include: Brunswick's *Harlem Longitudinal Study*; Eccles' *Prince George's County Study of Adolescent Development in Multiple Contexts*; and Suarez-Orozco and Suarez-Orozco's *Immigration, Family Life and Achievement Motivation Among Latino Adolescents*. This one is ongoing.

Most recently, we have a grant from the W.T. Grant Foundation to build an archive of studies of adolescents and youth.

A few examples of this archive are: Eccles, J. *Childhood and Beyond*; Fox, G., *Intrafamilial Sexual Socialization Outcomes and Patterns*; and Gilligan, C., *Girls' Development Studies*. Some studies of course qualify for placement in several archive divisions.

Ultimately, all special collection initiatives are folded into the general archive.

The Selection Process

In determining whether or not to acquire a new data set for the archive, several criteria are used by an Acquisitions Committee appointed by the center's director. The criteria can be roughly grouped into five general categories: content of the study, methodology (instrumentation and design), previous analysis and publication record, historical value, and cost of acquiring and processing the data. Each of these is described below with an indication of the relative importance of each criterion, where possible.

Content. The information in a data set must be relevant to the study of American women and/or basic issues in human development. Topic areas include: women and work, education, the family, psychological development and psychological processes, stress and coping, mental health; physical health, reproduction, and sexuality, political participation; and social policy. In addition to these ongoing topics of importance to the archive, smaller initiatives develop around a special topic as a focus for new acquisitions. The issues are chosen to reflect issues of current and future research importance. For example, in the recent past, studies of divorce and remarriage were identified as of particular interest. When a target area is chosen, staff members make a special effort to locate relevant studies by reading journals, attending conferences, and

contacting researchers who have worked in the targeted area. Finally, as I mentioned, during the last decade or so, we have secured grants for large topic-oriented acquisition initiatives, such as the Mental Health Archive, the Diversity Archive and the Adolescence and Youth Archive.

Instrumentation. As I have said, the Murray Center is unusual in that it acquires and makes available original subject records as well as coded computer data. The availability of extensive open-ended material offers the possibility of recasting the material in new terms, recoding for new variables, and addressing questions that are very different from those of the original investigators. Since the acquisition of qualitative material is a unique and valuable feature of the Murray Center's archive, emphasis is placed on acquiring studies that include such material. Studies that consist solely of coded computer data are de-emphasized since such data sets are readily available through other archives. Videotapes, which also allow for great flexibility in recoding, are also desirable acquisitions, assuming that funding for processing is available and that issues of informed consent and confidentiality can be resolved satisfactorily.

Because they offer greater opportunities for secondary analysis, studies that include a wide range of measures are more valuable than those with fewer measures. Measures with well-documented reliability and validity are preferred.

Design. Because of their special value for secondary analysis, data generated using certain kinds of research designs are preferred. Longitudinal studies have the highest priority. The possibility of further longitudinal follow-up by new investigators increases the value of a study for acquisition if a follow-up appears to be feasible and appropriate. High attrition rates seriously affect the quality of longitudinal research and are evaluated carefully when longitudinal studies are considered. Likewise, the extent of missing data is a factor in evaluating a study for acquisition. In addition to longitudinal designs, replications, surveys, and case studies may provide useful opportunities for secondary analysis. If the topic is one for which replication would be valuable, particularly for examining social change, a data set may be acquired with the objective of providing the base-line for a replication study. Similarly, studies with one or two high-quality replications already completed are valuable for archiving.

Finally, large surveys with national samples have a high priority if they are of particular content relevance, of very high quality, and not available elsewhere. Smaller-sample surveys are sometimes acceptable as well, particularly if they include responses to open-ended as well as closed-ended questionnaire or interview items. Cross-sectional studies that sample systematically across a wide age range are of moderate priority. Intervention and experimental studies will rarely be acquired because they usually are not well suited to secondary analysis. Experimental studies in particular are often designed to test very specific hypotheses, leaving manipulated samples and little room for reanalysis.

Sample. At the present time, it is not feasible for the Murray Center to maintain a broad international focus. Most of the data sets acquired for the archive will include only North American samples. Occasionally, exceptions are made in cases where the non-U.S. data directly parallel or complement specific holdings on U.S. samples.

The sample must be representative of the group to which the investigator wishes to generalize the findings and must be large enough in relation to the design and variables to make meaningful analyses possible.

For survey data, a high response rate is an important determinant of quality, although some assessment of nonresponse bias may offset difficulties with low response rate. Studies of racial and ethnic minorities are currently a high priority, and studies of Radcliffe students or alumnae are given special consideration.

Previous Analyses and Publication. As an indication of the quality of the research, the distinction of the study and the investigator should be exhibited in a productive publishing record for the study. This may be waived in cases where the lack of publications can be attributed to factors other than low quality of data. In any case, the data must not be so exhaustively analyzed as to preclude further analysis.

Historical Value of the Data. All studies will be scrutinized from a long-range perspective in an attempt to predict whether or not the data will remain valuable in the future. Classic or historically important data sets may occasionally be acquired even if they fall short in terms of sample, measures, or other criteria. Data collected at earlier historical periods may be useful to historians looking at social change or at the history of social science.

Cost of Archiving. The cost of archiving a data set will affect the assessment of its desirability for acquisition. Each data set will be evaluated in terms of how much organization, deidentification, duplication, documentation, vault space, and computer programming are necessary to archive it and make it available to users. In some cases, a somewhat less attractive data set will be acquired if it is easy and inexpensive to do so.

Important uses of qualitative data from the Murray Center's holdings.

The users of Murray Center data include sociologists, psychologists, criminologists, historians, educators, political scientists and economists. Elsewhere (James & Sorensen, 2000), we have documented in some detail several creative approaches to the use of existing qualitative data starting with seldom-used methods that have potential for strong contributions: (1) examples of new prospective studies (new issue to observe) created out of existing prospective studies; (2) examples of the use of multiple data sets for multi-cohort designs; and (3) a few examples of follow-up studies. Of course we also point to examples of classic reanalyses, chosen from numerous such examples from users of the center's data, conducted by representatives of different disciplines, and highlighting both long and short-term projects. In short, we find that our methods for selecting data that are of value for reuse are very successful.

Conclusions

In conclusion we believe that: (1) The Murray Center's Acquisitions Committee has a proven track record for identifying social and behavioral science data that is valuable for reuse; (2) these data can be better stored and cared for in a climate-controlled, protected environment than in offices, attics, and basements; (3) data archives can provide rigorous and sustained protection of human subjects whose data are turned over and made available for re-use; (4) others can find important and creative uses for the data (for generations to come). For these reasons, the selection, processing, archiving and sharing of the data is a very complex and rewarding and enterprise.

References

Colby, A., James, J., & Hart, D. (1998). Competence and character through life. Chicago: University of Chicago Press.

Colby, A., & Phelps, E. (1990). Archiving longitudinal data. In D. Magnusson and L. Bergman (Eds.), Data quality in longitudinal research (249-262). Cambridge: Cambridge University Press.

Ferraro, K.F., & McNally, J. W. (2000). Experiences with archiving multiple longitudinal data sets at NACDA. A paper presented as part of a symposium, *Experiences with Archiving and Using Longitudinal Data in Aging Research*, at the annual meeting of the American Psychological Association, Washington, DC, August 4-8.

James, J. & Zarrett, N. (2000). Using Archival Data to Study Aging. A paper presented as part of a symposium, *Experiences with Archiving and Using Longitudinal Data in Aging Research*, at the annual meeting of the American Psychological Association, Washington, DC, August 4-8.

McArdle, J. (2000). Experiences with archiving and using longitudinal data on the WAIS. A paper presented as part of a symposium, *Experiences with Archiving and Using Longitudinal Data in Aging Research*, at the annual meeting of the American Psychological Association, Washington, DC, August 4-8.