

Development of a Departmental Data Archive for Teaching and Research

Thomas M. Brinthaupt and John T. Pennington
Middle Tennessee State University

The analysis of archival data offers many opportunities for students, teachers, and researchers. However, the use of major data archives can be complicated and formidable. We describe the development and use of a departmental data archive that our faculty and students can use. This archive holds demographic as well as self-report behavioral and attitudinal data collected from mass-testing sessions conducted each semester with introductory psychology students. We discuss possible uses and the strengths and weaknesses of such archives for psychology teachers and researchers.

Nearly 30 years ago, Bryant and Wortman (1978) encouraged the development of departmental data archives in psychology. However, since that paper was published, we have found little discussion about these archives and little evidence of their development. This article describes the creation of such an archive and illustrates some of its uses for teachers and researchers.

There are numerous large national and international social science data archives available to psychologists (e.g., Altman et al., 2001; Ryssevik & Musgrave, 2001; Schoenfeldt, 1970). Despite the depth and breadth of existing archives, there are several barriers to their use (Zaitzow & Fields, 1996). As Altman et al. pointed out, creating, sharing, and disseminating archival data is often difficult and expensive for the archivists (see also Sieber, 1991). Also, the size and complexity of some data archives make them difficult to access and therefore less appealing for teachers and students. Unless there is a data set that closely matches one's needs or interests, most students (and probably many faculty) are unable or unwilling to complete the steps necessary to use one of these major archives. In addition, there may be usage costs involved, special programs or technology required, or limitations on who can use the data.

An alternative to the use of national or international data archives is to create and use a local data archive. One advantage such an archive has over larger archives is that the content of the data set is determined by the research-relevant measures and questions provided by local researchers and instructors. There will be a better match with local faculty research interests, and data reflecting the research interests of one's faculty are likely to be of greater interest to students. A second advantage is that control over the archive's operation, content, and usage is local. Local control allows greater flexibility regarding the inclusion of new measures and the development of the archive. Third, a local archive might be a good first experience for students interested in using larger archives. For example, students might learn about some of the advantages of using archival data and, as a result, develop an interest in larger archives. Finally, a local archive's content can be less narrow and restrictive than a larger archive and less dependent on specific funding agency goals.

Our department has used a psychology research pool for more than 15 years. Psychology faculty and students conduct research requiring human participants on a regular basis. The Research Pool contains students from introductory psychology classes who receive descriptions of available research studies. They may choose to participate in several studies in a semester for course credit, or they may submit written critiques of empirical articles.

Like many other colleges and universities, we found that researchers who use the pool often need to recruit a specific sample of students (e.g., married students, those with low self-esteem, smokers) or require baseline information most easily measured before participants arrive for a study. For these and other reasons, our department established a mass-testing procedure and began collecting data from Introductory Psychology students in Spring 1999. At the beginning of each Fall and Spring semester, students may obtain one of their required research credits by participating in one of these 30-min data collection sessions. In these sessions, participants provide demographic information about themselves and complete a variety of personality and attitudinal scales and measures. Whereas we repeat most of the demographic items each semester, many of the other measures change periodically to suit researchers' needs.

At this writing, we have 11 semesters of mass-testing data collected from nearly 5,000 respondents. To organize and facilitate the use of these data, we created a user-friendly mass-testing data archive. This database permits cross-sectional research (e.g., examining differences in alcohol and tobacco use over time), analysis of large personality data subsets, and student use of a "real-world" multivariable data set to learn how to conduct and interpret statistical analyses. Each semester we need approximately 3 weeks to obtain Institutional Review Board (IRB) approval for the data collection, set up and run mass-testing sessions, clean the data, and update the archive. Assistance from graduate students and participating faculty members helps lessen this workload.

Nature of the Archive

Our goal in creating the archive was to allow our faculty and students quick and easy access to the data we have collected in our mass-testing sessions. As with other research situations, researchers who want to use the archive database first must gain IRB approval for their intended project. After researchers receive IRB approval, they submit an archival research request form and receive the requested data files with the condition that they delete the files once they complete their project.

There is an interesting variety of data that researchers can use. Demographic information collected every semester includes respondents' sex, year in school, date of birth, age, race or ethnicity, marital status, and smoking status. Personality, attitudinal, and behavioral measures currently in the archive (many of which appear across multiple semesters) include alcohol use, attachment style, fear of death, homophobia, psychological reactance, racism, religiosity, satisfaction with life, self-esteem, self-talk, and spider phobia. Because students

sometimes repeat Introductory Psychology, they might appear more than once in our archive. Mass-testing participants report whether they have taken Introductory Psychology at our university previously so that researchers may exclude them when conducting cross-sectional analyses or combining data from multiple semesters.

The archive data files are available in SPSS and Excel® formats, with standardized variable labels. We included Excel files because of that program's widespread availability and use among our faculty. We removed all personally identifiable information, created copies of the original survey forms completed by participants, and developed descriptions of each individual scale (or attitude survey) and its scoring instructions. We also cleaned the data by checking for missing or erroneously scanned data points. Because of copyright restrictions, we do not include the scales in the online surveys or descriptions. Researchers can obtain these measures from us. The data files reside on the first author's computer and are not directly accessible by other researchers.

Despite a limited knowledge of Web page design, we developed a Web site for the archive. The *Archive Overview* page provides information on the archive's history, structure, potential uses, and conditions for use. The *Contents* page provides a semester-by-semester listing of data collected, sample sizes, and hyperlinks to the surveys used. In addition, there are "Caution!" notes within the surveys, highlighting items that may be problematic or that respondents might have misinterpreted. On the *Scales Info* page, we list each of the scales and measures used. For each measure, we give a brief (two to three sentence) description, instructions for scoring, and the full citation (including Web links if available). The fourth page is the *Data Request Form*, which researchers must complete to gain access to the data files. On this form, they provide a brief description of the project, the specific data files they are requesting, proof of IRB approval, and the preferred file format.

The archive has the potential to facilitate classroom education and research in several ways. It is clear that the analysis of archival data is an effective way to teach students about research (e.g., Lutsky, 1986). To date, students in graduate-level quantitative methods classes (e.g., factor analysis, multivariate analysis) have used the archive to test the factor structures and validity of measures. In addition, undergraduates in advanced research methods have conducted correlational studies, and graduate students have employed the archive to conduct psychometric and cross-sectional analyses on various attitude scales. Graduate students have also used archive data for their master's theses.

Strengths and Weaknesses of Local Archives

Among the limitations of most data archives are that the data are not precisely what one needs and the original investigators' perspectives on a topic or problem might differ from one's own (Elder, Pavalko, & Clipp, 1993; Zaitzow & Fields, 1996). By their very nature, local archives can offset some of these disadvantages. For example, in response to researchers' and students' needs and interests, we have added or reworded demographic items in our mass-testing data collection. We also have complete knowledge of the data collection materials and methodology. In addition, students

might be more interested in data that come from their own school and fellow students.

A disadvantage (which may also apply to larger data archives) is that we include no longitudinal data in our archive, meaning that only cross-sectional analyses are possible. There is also some inconsistency of measures used from semester to semester. Typically, because of time and space limitations in data collection, we include only smaller (10- to 25-item) standardized measures. Whereas our archive includes data only from human participants, there is an interest in and need for animal data archives (e.g., Church, 2002; Kurtzman, Church, & Crystal, 2002).

Why would departmental faculty want to create their own data archive? We believe that an archive would be beneficial if a department already collects mass-testing data (or intends to begin doing so), and faculty can tie the data to their class assignments and projects in subsequent semesters. The effort required to establish a mass-testing data collection procedure would probably be worthwhile only if enough researchers needed selection or baseline data for their projects. If departments have mass-testing, we would certainly recommend the development of a local archive. Once an archive is established, perhaps its best use is by faculty who teach research methods and statistics courses. These instructors might include key items each semester with an eye toward establishing data subsets for undergraduate and graduate student assignments and projects. Few drawbacks exist when using the archive for these purposes.

In summary, we believe a local data archive is easy enough to implement that it can be beneficial for students and teachers. It allows control over and access to data that reflect the interests of local research-active instructors. Because the archive represents the varied interests of our students and faculty, our students can conduct many possible projects, with a wide variety of measures. Researchers who regularly collect mass-testing data might consider the benefits of creating a local archive rather than having those data languish.

References

- Altman, M., Andreev, L., Diggory, M., King, G., Sone, A., Verba, S., Kiskis, D., & Krot, M. (2001). A digital library for the dissemination and replication of quantitative social science research: The Virtual Data Center. *Social Science Computer Review*, 19, 458-471.
- Bryant, F. B., & Wortman, P. M. (1978). Secondary analysis: The case for data archives. *American Psychologist*, 33, 381-387.
- Church, R. M. (2002). The effective use of secondary data. *Learning and Motivation*, 33, 32-45.
- Elder, G. H., Jr., Pavalko, E. K., & Clipp, E. C. (1993). *Working with archival data: Studying lives*. Newbury Park, CA: Sage.
- Kurtzman, H. S., Church, R. M., & Crystal, J. D. (2002). Data archiving for animal cognition research: Report of an NIMH workshop. *Animal Learning and Behavior*, 30, 405-412.
- Lutsky, N. (1986). Undergraduate research experience through the analysis of data sets in psychology courses. *Teaching of Psychology*, 13, 119-122.
- Ryssevik, J., & Musgrave, S. (2001). The social science dream machine: Resource discovery, analysis, and delivery on the Web. *Social Science Computer Review*, 19, 163-174.
- Schoenfeldt, L. F. (1970). Data archives as resources for research, instruction, and policy planning. *American Psychologist*, 25, 609-616.
- Sieber, J. E. (1991). Openness in the social sciences: Sharing data. *Ethics & Behavior*, 1, 69-86.

Zaitzow, B. H., & Fields, C. B. (1996). Using archival data sets. In F. T. L. Leong & J. T. Austin (Eds.), *The psychology research handbook: A guide for graduate students and research assistants* (pp. 251–261). Thousand Oaks, CA: Sage.

Notes

1. Development of the archive and writing of this article were greatly facilitated through the support of a Middle Tennessee State University Faculty Research and Creative Activity Committee Grant awarded to Tom Brinthaup and John Pennington.
2. We thank Doug Cothorn, Dennis Kramer, and Will Langston for their comments on an earlier version of the manuscript.
3. Readers can view the archive site at the following address: http://mtsu32.mtsu.edu:11311/archive/psych_data_archive.htm.
4. Send correspondence to Tom Brinthaup, P.O. Box X034, Department of Psychology, Middle Tennessee State University, Murfreesboro, TN 37132; e-mail: tbrintha@mtsu.edu.