The Role of Secondary Data Analysis in Teaching the Social Sciences

JEFF SOBAL

TEACHING THE SOCIAL sciences includes several goals. As in any other discipline, information is transmitted to students, instilling social facts into their pool of knowledge. Theories are taught to provide generalizations about social facts. Finally, methods of conducting social research are central components of social science instruction, where the logic of inquiry and the procedures for analysis of data are passed on to students. The goals of information and theory are served adequately, but not entirely, through books, journals and other written sources. Inquiry and methodology are processes, and are best learned by actually conducting research. For this purpose, books are not sufficient, and teachers must turn to other sources. Secondary data analysis, the reanalysis of machine-readable data, is one of the best supplements to traditional teaching methods, especially for teaching research methodology and statistics.

The general state of instruction with secondary data today may be seen in light of the growth of the "secondary data movement." Since the 1960s, there has been a growing emphasis upon the use of secondary data in research, with important developments in social indicators analysis, the rise of survey archives, and the overall development of quantitative social research all playing a part. As noted by Nesvold,[1] much of the development of the use of secondary data in teaching has been a "trickle down effect" from research. As more people became involved in collecting and analyzing quantitative data, their work spilled over into their teaching.

Jeff Sobal is Assistant Professor and Assistant Director of Research, Department of Family Medicine, University of Maryland School of Medicine, Baltimore.

Before proceeding to a discussion of the benefits and problems of teaching with secondary data, we should be clear about exactly what constitutes secondary data. Any data which have been collected for another purpose and later reanalyzed may be seen as secondary data. These would include field notes from ethnographers, coding sheets from content anlaysis, results from experiments, friendship choices from sociometry, or questionnaires from surveys and censuses, as well as data collected for administrative or other purposes. While teaching may well occur using any of these types of secondary data, this discussion will only consider quantitative, machine-readable data of the kind available from censuses, surveys or administrative records. These are the most common data, and often the most useful type available.

Such quantitative data are often difficult to acquire, especially for students, because they are expensive and time consuming to collect if high-quality information is desired. Yet quantitative training is becoming almost a necessity for all of the social sciences. Since the introduction of the computer into the social sciences in the 1950s, quantitative analysis of data has become an almost essential skill in the disciplines of psychology, sociology, anthropology, political science, economics, geography, history, and education, as well as in related fields such as public health and marketing. In both graduate and undergraduate training, and often in the more sophisticated secondary schools, students in the social sciences are required to learn statistics, the logic of data anlysis, and have some exposure to the use of the computer. Yet simply going through the motions of calculating statistics is not enough, because statistics and analysis are meaningful to social scientists as tools for understanding data rather than ends in themselves. Thus data become an important part of methodological training, and the reanalysis of previously collected data is one of the best ways of teaching research methodology.

Exposure to the use of the computer is becoming almost ubiquitous in the social sciences, especially during graduate training, although undergraduate study and even secondary school teaching are not far behind. Secondary data offers instructional opportunities for computing without the problems of data collection. Some form of data is usually necessary in training students in computing, and the availability of secondary data sets enhances this aspect of computing instruction. Once students understand data analysis with the computer, then they need data to utilize their skills. The availability of a number of easily accessible secondary data sources makes it likely that students' computing capabilities will be utilized, and even extended, rather than relegated to memories of skills that they once possessed.

Some disciplines are more quantitatively oriented than others and therefore need a wider variety and higher quality of secondary data sets. This does not imply that less quantitatively oriented disciplines are to be ignored when secondary data are acquired, because these disciplines are often those in which the use of secondary data can provide the furthest-reaching innovations in teaching methods, and are those whose practitioners are the least likely to have sufficient awareness of the availability of secondary data to request their acquisition. The availability of secondary data relevant to a discipline and the publicizing of their availability among teachers can help to add important quantitative perspectives in classes where there previously had been no exposure to the perspective. In addition to noting the range in quantitative analysis among disciplines, there is also a great deal of variation in their traditional reliance on secondary data and future receptivity to teaching with secondary data. People in economics and geography have traditionally worked with quantitative data collected by others and accept the use of secondary data in the classroom with relative ease compared to fields such as history, which may rely on available data but infrequently use quantitative materials, or psychology, which commonly stresses quantitative analysis but infrequently utilizes other people's data. The reasons for these variations among disciplines range from the type of subject matter, to the favored methodology, to norms about how teaching and research should be done in that discipline. These variations are mentioned both so that data archivists and librarians can be sensitive to problems that will be encountered, and so that unexplored areas of teaching can be considered.

Type of Secondary Data for Teaching

There are many types of secondary data available for instructional use. A number of simulations and games are currently available, such as the EXPER-SIM series,[2] where investigations are undertaken by students on artificially created data. These provide a valuable experience in simulating research, but may not be taken as seriously by students as is working with actual data. At the same level as these artificial models are a number of excellent teaching packages which utilize secondary data. These include SETUPS,[3] the Minnesota Manuals,[4] ICPSR Instructional Subsets,[5] International Studies Association Learning Materials,[6] and other packages available from CONDUIT[7] and the University of Iowa Political Research Laboratory.[8] All of these packages share common features, including easily accessible programming for student analysis, a relatively small sample size, a selection of a small number of

variables interesting to students, and manuals for student use. Some also include additional instructions for teachers. These are easy to use for both pupils and instructors, and well worth including in any data archive or library which houses secondary data. Brighter students can often teach themselves how to use them, and easy availability will facilitate this.

However, the simplicity of these instructional packages can also be a drawback. They provide only a limited amount of material for students and can lead to boredom with repeated use.[9] Especially when different types of packages of the same degree of difficulty are used in several classes, instructional packages are no longer challenging and become routine. This is not to say that all students can easily master their use, especially their intended use of grounding theoretical conceptualizations in data. Some pupils will be so overwhelmed by the mechanics of assessing a computerized teaching package that they will have difficulty learning much from the experience. Special support by teachers and their assistants is often necessary to achieve learning goals and practical mastery of the technique in these cases.

For students who have mastered teaching packages which use secondary data and for other pupils ready for more challenging analysis, the logical next step in teaching is the use of any available secondary data set. Here all of the opportunities and problems of full-scale data analysis are faced by students.[10] Availability is the key to effective instructional use at this level. If data archives or college libraries have codebooks within easy access of goal-directed researchers as well as browsers, their use will be maximized. For teachers to adopt secondary data anlaysis as a pedagogical technique, they need to be aware of the data available. For students with some capabilities in secondary data analysis or those willing to learn about the technique, exposure to the data sets is vital.

Within the diversity of data available for teaching, several characteristics of the data set are important considerations in selecting data for student use. Obviously the topic of the data must be consonant with that of the class, but this can be specified so narrowly as to restrict creativity and fail to accommodate diverse student interests. General omnibus surveys, such as the National Opinion Research Corporation's General Social Survey,[11] usefully serve a wide range of interests for both students and instructors. They are especially useful where a class with a diversity of interests uses a single data source. Yet for more advanced courses and student papers, these broad offerings do not provide sufficient depth and need to be replaced by specialized studies focusing on one topic. A

primary function of archivists and librarians is to aid instructors and students in finding and accessing these diverse data sets, in addition to the management of the omnibus studies used in the bulk of teaching.

Another important consideration in choosing data for teaching is the unit of analysis.[12] It must be appropriate to the topic taught, such as using the individual for psychology, the nation for political science, or the culture for anthropology. In addition, students must be able to comprehend data collected on that unit of analysis easily. Generally, students find data using the individual as the unit of analysis the easiest to fathom, with data collected on counties, organizations or nations much more difficult to conceptualize. Other attributes of secondary data deviating from simplicity, such as hierarchical file structures or the presence of subfiles, also make some data sets less than ideal for most student use.

Levels of Instructional Use

There are many levels of instructional use for secondary data, and all can be facilitated by data archivists and data librarians. They include secondary school teaching, undergraduate introductory courses, undergraduate advanced courses, graduate training, professional training, and continuing education. Although the bulk of secondary data use in teaching occurs with advanced undergraduates and graduate students, pedagogical pioneers are currently extending secondary data instruction into other areas. Archivists and librarians can encourage the development of secondary data teaching among those who do not currently use it by advertising their services to teachers within their institutions and being sure that entire library staffs are aware of teaching opportunities at all levels of instruction.

In addition to being a useful teaching tool at many levels of instruction, secondary data are useful in many types of student training. They are an excellent tool for introducing students to a discipline, providing a valuable supplement to lecturing and reading. By actually allowing students to test their own interpretations of a subject matter using real data, students become active participants in inquiry rather than passive recipients of information. The utility of doing original data analysis is greater for advanced courses in a topic where students are even more sophisticated and immersed in the subject matter. Seminars offer a rich environment for secondary data use, with numerous opportunities for testing the hypotheses and ideas raised in discussion by accessing secondary data. As mentioned earlier, methodology and statistics courses

frequently use available data to make the empirical techniques "come alive" with meaning by applying them to real-world data. All of these classroom situations can involve students with regular assignments of data analysis or use occasional activities in class.

Secondary data analysis can prove useful in individual student projects as well as assignments for entire classes. A reanalysis and interpretation of available data is a valuable alternative to the traditional library term paper which simply repeats or reworks the findings of other researchers. In this case, archivists and librarians can play a vital role in making students and teachers aware of the option of secondary data analysis.

Housing Data in Archives

Two strategies of making data available can be used, each with its strengths and weaknesses in encouraging instructional use of secondary data. One is the archival strategy of housing data, where all secondary data and codebooks are housed in a separate room or building from other library materials. This provides a convenient center for those using secondary data and often serves as "headquarters" for a network of users and a clearinghouse for information. Yet it also can become isolated from the general student population and overlooked by potential users who were not socialized into the clique or specificaly assigned to go there for course work. The centralization does allow a secondary data specialist to be available with the codebooks to explain secondary data analysis to students and researchers. This type of organization is most useful for assigning larger classes to carry out secondary data analysis, where they work together in one site, often developing a great esprit de corps and sharing ideas in the process.

One of the major claims of the sciences as disciplines based upon the model of active inquiry is their involvement in laboratories, both in research and teaching. To share this aspect of science, the social sciences need to develop the equivalent of laboratories.[13] This has been done extensively by psychologists, and has supported their perceived legitimacy as a science. The use of secondary data, the development of data archives, social science research labs, and even special rooms in libraries to house secondary data can extend this "laboratory atmosphere" to other social sciences.

Archives have grown significantly in the United States in recent years, but as Nesvold[14] notes, this growth has been largely vertical and not horizontal. While this has achieved an economy of scale for collec-

tions of secondary data, it has served the needs of researchers more than teachers. Researchers are concentrated in larger universities likely to have data archives, and those who are not situated in these sites can individually visit large archives for information and data. This is not the case with teaching, especially on the undergraduate level, where instructors are frequently dispersed from major universities and do not have the resources to establish their own archives or bring classes to larger archives. Librarians are in the position to bridge this gap for social science instructors, expanding the data archive movement horizontally to serve the needs of teachers. Their ability to house, access, understand, and disseminate knowledge about secondary data to teachers will have important consequences in the future of social science instruction.

In existing data archives, an expanded teaching role is also desirable. The awareness of the values of secondary data analysis produced by disseminating information about archives in the classroom will help to ensure continued interest and support of data archives. Within the staff of an archive, the designation of one person as a primary teaching coordinator and liaison will ensure the active involvement of at least part of the archive's resources in the instructional services and support.

Housing Codebooks in Library Stacks

The other arrangement for housing secondary data is decentralized, where codebooks are dispersed within the general collection of a larger unit, such as a research center or building. This encourages browsing by those not specifically seeking secondary data, opening up the technique to a larger audience of potential users by making it part of the standard information system of a library. The drawbacks of this include a lack of immediate guidance for potential users and a separation of those examining and working with the data. For independent scholarly projects, though, this situation is ideal.

Another important role of libraries in promoting the student use of secondary data is the housing and provision of codebooks for general use. In the data archive, documentation is clearly understood as specific to that setting, and there are generally people around to interpret the codebook to novices. The meaningfulness of a codebook to the uninitiated is not to be taken lightly. A thick document with column numbers and *N*'s listed next to responses to questionnaire items may appear meaningless to the casual reader who happens upon it in the stacks. The provision of an inserted page explaining "how to read a codebook,"

plus a note explaining who should be contacted for additional information (including access to the data) can turn a seemingly confusing mass of numbers into a valuable information tool. Despite this, other efforts need to be made to convey the importance of returning to the original data and disaggregating them by important subcategories rather than simply accepting the tabulations presented in the printed information. The complexities of actually accomplishing this are perhaps best left to the instructors, but informed librarians can generate interest.

Perhaps the ultimate solution for maximizing the use of secondary data in teaching and learning is to combine the use of archives and libraries. A separate archive could be established, perhaps connected with the computer center. In addition, codebooks of at least the most popular, if not all, data sets available in the archive could be placed in the library, with instructions inside directing the reader to the archive for more information. To facilitate the finding of codebooks in card catalogs, cards for codebooks could be of a different color from other cards. This combination could provide the benefits of both worlds for instruction with secondary data.

Contributions of Archivists and Librarians to Teaching

As the gatekeepers of secondary data, archivists and librarians not only act as passive facilitators for teachers and students who already have clear interests and goals in secondary data analysis, but also operate as disseminators of knowledge about this type of resource. There is a vast potential for using secondary data in both teaching and research, and a corollary need for training students and those who serve students, including teachers, archivists and librarians. Secondary data analysis, access and stewardship is not taught in most graduate schools, although this omission is slowly being rectified. Social scientists knowledgeable in this area can work with librarians to help them appreciate the value of secondary data, and to describe social scientists' needs and methods of meeting them by including secondary data in library collections and establishing data archives as part of the library system.

Librarians and archivists who have worked with secondary data can make important contributions to the instructional programs of social science departments. Most social science departments offer a course on research methods, which is often required of all students. Librarians and archivists can seek out the instructors of these courses and make presentations in their classes, describing the use and access of secondary data to the students as well as offering to help individual students in the future. Seminars on data holdings can also be offered to

faculty. Especially in graduate departments, which are preparing individuals who will teach their discipline, it is useful to familiarize future instructors with the opportunities available in secondary data analysis, including instructing them how to teach with secondary data. Such activities may include descriptions of successful teaching activities by instructors who use secondary data, who can serve as role models in the next generation of instructors.

While this paper is directed at archivists and librarians rather than praticing teachers of social science, a mention of sources for ideas about how to teach with secondary data is in order. Perhaps the best source of information is the colleague who is currently using secondary data in his or her classes. If there is no such person in one's school (including other departments), then professional meetings offer a place to share ideas and "nuts-and-bolts" teaching suggestions, and an increasing number of sessions at professional meetings are being devoted to teaching. Other sources are professional literature on teaching using secondary data, such as Nesvold,[15] Sobal[16] and Treinen,[17] as well as journals such as *Teaching Political Science, Teaching Sociology* and *Teaching Psychology*. While librarians do not have to immerse themselves in this literature, it will be useful for them to know which people are teaching with secondary data at their own institution so that they can refer others to those individuals, and to be aware of teaching sources so that they can refer interested instructors to them.

While teachers, archivists and librarians have been discussed separately thus far, in reality they make up a teaching team. They all have independent primary roles, yet all are interdependent in supporting the highest quality of instructional activities. This division of labor permits a separation of tasks, but also requires a great deal of communication among the various groups involved in teaching with secondary data. Occasional meetings which involve teachers and their assistants, librarians, archivists, and computer center personnel cannot only foster additional teaching activities in this area, but also solve problems, and even head off future difficulties.

Conclusion

The future of secondary data instruction is promising. With the growing demand for quantitative analysis skills among social scientists, faculty will increasingly turn to secondary data to teach them these skills. The changing job market is placing an emphasis on policy and applied research skills, for which secondary data are a valuable resource students need to be trained to work with. As research budgets tighten,

investigators will often look to available data to do the research which will spill over into their teaching. Librarians and archivists need to prepare for this rising demand for secondary data for the laboratory and the classroom, and to be sensitive to the needs of this new constituency of secondary data users.

In the end, secondary data are valuable in themselves as well as models for collecting one's own data. While the presence of quality information without the costs of data collection is a valuable tool, a user of secondary data is a captive of the sample collected and a prisoner of the variables available. By teaching students about the value of data analysis, they will learn a great deal in the process and be sufficiently excited about research to use other data and to collect original data. With a commitment to the advancement of instruction with secondary data by social scientists, librarians and data archivists, this goal should become a reality.

References

1. Nesvold, Betty A. "Instructional Applications of Data Archive Resources." *American Behavioral Scientist* 19(1976):455-66.

2. Main, Dana, and Stout, Robert. *EXPER-SIM: Model Builder (MESS) and Model Driver*. Iowa City, Iowa: CONDUIT, 1978.

3. American Political Science Association Division of Educational Affairs, 1527 New Hampshire Avenue N.W., Washington, D.C. 20036.

4. Flanigan, William H., and Krislov, Samuel. *Little, Brown Manual in Political Analysis*. New York: Little, Brown and Company, 1972 (not being marketed currently).

5. Director for Educational Activities, Inter-University Consortium for Political and Social Research, P.O. Box 1248, Ann Arbor, Mich. 48106.

6. Council for Intercultural Studies and Programs, Learning Resources in International Studies Project, Suite 1026, 60 E. 42nd Street, New York, N.Y. 10017.

7. CONDUIT, P.O. Box 388, Iowa City, Iowa 52240.

8. Political Research Laboratory, University of Iowa, Iowa City, Iowa 52242.

9. Wahlke, John C. "SETUPS: Supplementary Empirical Teaching Units in Political Science." *Teaching Political Science* 7(April 1980):353-68.

10. Hyman, Herbert H. *Secondary Analysis of Sample Surveys*. New York: Wiley, 1972.

11. Davis, James A. *Cumulative Codebook for the 1972-1980 General Social Surveys*. Chicago: National Opinion Research Center, 1980.

12. John Kolp to Sobal, personal communication, 1980.

13. Hans Mauksch to Sobal, personal communication, 1980.

14. Nesvold, "Instructional Applications."

15. Ibid.

16. Sobal, Jeff. "Teaching with Secondary Data." *Teaching Sociology* 8(Jan. 1981): 149-70.

17. Treinen, Heiner. "Notes on an Experience with Secondary Analysis of Survey Data as a Teaching Device." *Social Science Information* 9(April 1972):123-32.