

R. Cressey. 1974. *Criminology*. J. Pippincott.
 Theory of Differential Association
Theories of Deviance, 3rd
 edition, ed. by R. M. Merton, R. A. Cloward, and C. B. Little. Itasca,

eds. 1985. *Theories of De-*
 L. Peacock.

ard. 1986. *Theoretical Crim-*
 ork: Oxford University Press.

iate Professor of Sociology at
 teaching responsibilities include
 of religion, and research meth-
 d in conducting local community
 ie sociology majors. At present,
 e the influence of social networks
 older people and the effects of
 ious beliefs. Address correspon-
 an, Division of Social Sciences,
 i, LA 50140.

CD-ROM TECHNOLOGY AND THE TEACHING OF SOCIOLOGY*

JOSEFINA J. CARD
Sociometrics Corporation

JAMES L. PETERSON
Sociometrics Corporation

NEW TECHNOLOGY

Sociology is a data-intensive discipline. Analyses often are conducted on data sets that comprise thousands of cases and hundreds of variables. Teaching requires the instructor to impart to students both an appreciation and a knowledge of the means of analyzing large bodies of data, but until recently the cost of computing was a barrier to providing most students with hands-on experience in analyzing such data.

The rapid expansion in the use of desk-top microcomputers over the past decade has begun to change this situation. Microcomputers have become much faster and their storage capacities are much larger than in the past. Meanwhile, the price of a basic system has declined sharply. State-of-the-art microcomputer systems cost no more now than their much less powerful counterparts of a few years ago. Such systems, priced under \$3,000, are now within the reach of many academic departments. The savings in mainframe computing charges in a single grant can easily pay for a system.

Most microcomputers come with at least one floppy diskette drive that can read and write information on removable 3 1/2 or 5 1/4 inch floppy diskettes, each of which can hold from 360 thousand to 1.4 million characters (or bytes) of information. Another common storage device, the hard disk, can hold from 10 million to 330 million bytes.

It is the emergence of inexpensive optical storage media for microcomputers, however, that promises to help microcomputer technology produce radical changes in how sociology is taught and how research is conducted. We describe the new technology, cite illustrative applications, and point out possibilities for the future.

* This paper was produced with funds provided by Contract 282-87-0062 between the U.S. Office of Population Affairs and Sociometrics Corporation. The authors thank Aaron Kellem for information about the POPLINE CD-ROM, as well as R. Timothy Reagan and Max Nelson-Kilger for helpful comments on an earlier draft of this paper.

Recently a new microcomputer storage medium, the CD-ROM (compact disk, read-only memory), has been introduced; it is a product of the same laser-optical technology as the popular audio compact disk. The medium is noteworthy for its high storage capacity, permanence, small size, fast access time, and relatively low cost. A single CD-ROM—only 4.72 inches in diameter and 1.2 millimeters thick—can hold 550 million bytes of information, the equivalent of 1,500 floppy diskettes or 150,000 printed pages! With the help of the retrieval software generally provided with a CD-ROM, any piece of information on the disk can be accessed quickly, generally in less than a second.

Low cost is another advantage. Drives for reading CD-ROMs can be added to almost any microcomputer and sell for \$600 to \$1,000; significant price reductions can be expected in the near future. Individual disks, each holding volumes of information, such as an entire encyclopedia or several large national surveys, currently sell for \$99 to \$995. Because price is related closely to volume, the cost likely will fall to less than \$100 as the market expands. This is not much more than the cost of a high-quality hard-cover text or monograph. Even greater cost-effectiveness is realized when a CD-ROM library, consisting of a single drive and a collection of associated disks, is shared by many users (such as the members of a sociology department) via a local area network (LAN) of interconnected microcomputers.

How does CD-ROM technology work? A spiral sequence of small pits is burned onto the disk's surface by laser technology. Information is represented in binary fashion by the lengths of the pits and of the spaces (or "lands") between them. Up to two billion pits can be put onto a single disk. The pits and lands are interpreted by the CD-ROM reader and the computer as letters and numbers. After the surface is pitted, it is coated with a reflective metal

layer and then a protective lacquer. This process ensures that CD-ROMs will be virtually indestructible, making them a suitable medium for preserving information for posterity.

A conference held in 1985 at the High Sierra Casino in Lake Tahoe established physical standards for the layout of information on these disks. These standards, now referred to as the High Sierra format, make it possible for a CD-ROM produced by one publisher to be read on a drive manufactured by another.

Optical storage devices are likely to become the medium of choice in three areas: data storage, archival storage, and multi-media disks.

DATA STORAGE APPLICATIONS

CD-ROM data disks have tremendous potential for facilitating data sharing and secondary data analysis in those social sciences that rely on large data bases for much of their empirical knowledge. Not long ago data from such sources as the census, the National Survey of Family Growth, and the General Social Surveys were stored on mainframe computer tapes that usually cost several hundred dollars each to acquire and were analyzable only on costly mainframe computers. Researchers and departments had to pay several hundred dollars per hour for computer time to perform statistical analyses of these data sets with package programs such as SPSS-X, SAS, BMDP, or OSIRIS.

Many of these statistical analysis packages are now available for microcomputers. With CD-ROMs, raw survey data can be stored and processed inexpensively on a microcomputer. Thus, state-of-the-art sociological data can be subjected to repeated analyses virtually free of charge: after the initial investment in hardware and software—for most sociology departments merely the incremental cost of the CD-ROM drive and disks—data analysis can be conducted for only the cost of paper and electricity.

A National Academy of Science report (Fienberg 1985), as well as several editorials and articles in major publications of the American Sociological Association (Baron 1988; Card 1989; Hauser 1987), have extolled the virtues of data sharing. Data formerly in private hands are now public. Investigators' raw data, which formed the basis for publications and policy, can be reanalyzed for replication or to test competing hypotheses.

Two pioneering data-based CD-ROMs that should be of great interest to sociologists are already available: NATASHA and NLSY. NATASHA,¹ the NATIONAL Archive on Sexuality, Health, and Adolescence, is a CD-ROM released in December 1988 that includes data from 82 of the leading studies on adolescent health, sexuality, pregnancy, and parenthood; it sells for \$495. The 82 data sets include such classic demographic and sociological studies as multiple waves of the U.S. National Fertility Studies, the National Survey of Family Growth, several files of the U.S. Current Population Surveys (with fertility supplements), the 1981 Child Health Supplement to the National Health Interview Survey, and three waves of the National Surveys of Children.

In addition to the original raw data from these and many other studies, the NATASHA CD-ROM includes: 1) SPSS/PC statements that serve as machine-readable documentation for the data files and that facilitate analysis with SPSS/PC;² 2) search and retrieval software to assist the user in isolating variables and data sets relevant to the research problem at hand (each of NATASHA's 39,815 variables can be searched by topic, type, study, and any word in its SPSS/PC label); 3) study summaries that describe each of the 82 data files and their major variables; and 4) a program that allows users to conduct simple unweighted cross-tabulations without the need for statistical programs such as SPSS/PC.

The National Longitudinal Survey of Labor Market Experience of Youth (NLSY)³ is an ongoing study of 12,686 young men and women who were 14 to 21 years old on January 1, 1979. More than 90 percent of these respondents have participated in yearly hour-long personal interviews since 1979. Topics covered include respondents' work experiences, military experiences, schooling, family background, income and assets, health, marriages, fertility, child care, and drug and alcohol use. The CD-ROM contains all the data in binary format,

¹ For information contact Dr. J. J. Card at Sociometrics Corporation, 170 State Street, Suite 260, Los Altos, CA 94022, 415-949-3282.

² Users of other statistical packages such as SAS, BMDP, or SYSTAT can modify NATASHA's SPSS program statements to suit the required syntax of their package.

³ For information contact Dr. Frank Mott at the Center for Human Resources Research, 650 Ackerman Road, Suite A, Columbus, OH 43202, 614-442-7300.

ata-based CD-ROMs that interest to sociologists : NATASHA and NLSY. ATional Archive on Sexu- lolescence, is a CD-ROM er 1988 that includes he leading studies on sexuality, pregnancy, sells for \$495. The 82 ch classic demographic udies as multiple waves al Fertility Studies, the Family Growth, several rent Population Surveys ements), the 1981 Child to the National Health and three waves of the f Children.

e original raw data from r studies, the NATASHA : 1) SPSS/PC statements ine-readable documenta- s and that facilitate anal- 2) search and retrieval he user in isolating vari- relevant to the research ch of NATASHA's 39,815 ched by topic, type, study, SPSS/PC label); 3) study rcribe each of the 82 data r variables; and 4) a pro- ers to conduct simple un- lations without the need ms such as SPSS/PC.

Longitudinal Survey of ience of Youth (NLSY)³ is f 12,686 young men and to 21 years old on January 0 percent of these respon- ated in yearly hour-long since 1979. Topics covered work experiences, military ing, family background, ealth, marriages, fertility, and alcohol use. The CD- the data in binary format,

ntact Dr. J. J. Card at Sociometrics trect, Suite 260, Los Altos, CA

tistical packages such as SAS, in modify NATASHA's SPSS) suit the required syntax of

ntact Dr. Frank Mott at the Center Research, 650 Ackerman Road, 13202, 614-442-7300.

software to search the documentation files and create sample- or variable-extract files, and numerous indices and documentation files that the software accesses (Center for Human Resource Research 1989).

A good current source of information about data bases available on CD-ROMs is *CD-ROMs In Print* (Desmarais 1990). The CD-ROMs listed with data of interest to sociologists include: the 1985 American Housing Survey; the 1988 County and City Data Book; Labor/Stats, a collection of nine data bases on employment, earnings, and prices; and Population Statistics, population and housing data from the long form of the 1980 census.

ARCHIVAL STORAGE APPLICATIONS

Portability and archivability are two advantages to preserving text in paper format. Because books and magazines can be read without a machine, changes in hardware and software do not affect future generations' ability to retrieve information from them. A serious limitation, however, is the storage space required for large collections of printed materials. Another limitation is the relative difficulty of search. Many documents are not indexed; although most books include a table of contents and a key word index, these index search sets tend to be small and incomplete.

Electronic publishing, the first major application of CD-ROM technology, continues to dominate the field. Most CD-ROMs contain collections of texts of general interest. One of the most widely used, the Microsoft Bookshelf, includes all of the following reference documents: *The American Heritage Dictionary*, *Roget's II: Electronic Thesaurus*, *The World Almanac and Book of Facts*, *Bartlett's Familiar Quotations*, *The Chicago Manual of Style*, the *Houghton Mifflin Spelling Verifier and Corrector*, a variety of standard business forms and letters, the *U.S. Zip Code Directory*, *Houghton Mifflin's Usage Alert*, and the names of many business information sources (Microsoft Corporation 1987). The CD-ROM has two significant advantages over the printed page as an archival storage medium: 1) it consumes far less space; and, 2) being machine-readable, it can be subjected to sophisticated search and retrieval techniques that allow information to be pinpointed by exact location, not merely by page or chapter. Moreover, most text-based CD-ROMs are accompanied by software that allows the user to bring to the screen or to a

printer all mentions of a single word or user-specified combination of words.

A text-based CD-ROM of special interest to sociologists is POPLINE, a collection of 165,000 citations and page-length abstracts of books, articles, and unpublished manuscripts in the population field. POPLINE includes domestic and international publications dealing with issues such as population policy, family planning programs, population projections, legal aspects of population issues, and other topics of interest to demographers and population planners. The POPLINE CD-ROM is accompanied by software that allows searches by key word in a thesaurus, by any word in the abstracts, by author, and by publication type. The original CD-ROM and two updates have been issued; thrice-yearly updates are planned. A related CD-ROM, MEDLINE (focused on the medical literature), also is available.⁴ Other examples of text-based CD-ROMs listed in *CD-ROMs In Print* include *Sociofile and Social Sciences Index* (CD-ROM versions of the *Sociological Abstracts* and *The Social Sciences Index*, respectively) and ERIC, the bibliographic data base on education sponsored by the U.S. Department of Education.

MULTIMEDIA DISKS

It is also possible to develop CD-ROMs that contain an integrated combination of text, numbers, pictures, and sounds and that draw on all of these properties to present information about a topic. Although this application has not yet been brought to bear on the teaching of sociology, an integrated CD-ROM that contains two kinds of information—text and data—relating to the field of criminal justice research is currently under development by Abt Books.⁵

CURRENT AND FUTURE DEVELOPMENTS

In the future we are likely to see the burgeoning of CD-ROMs packed with data from a substantively related collection of studies (such as NATASHA) and/or from large-scale recurring

⁴ For information about the POPLINE or MEDLINE CD-ROMs, contact Silver Platter, 800-343-0064.

⁵ Available from Abt Books, Inc., 146 Mt. Auburn St., Cambridge, MA 02138, 617-661-1300. Ask for Stephen Rhind-Tutt, marketing manager.

gitudinal investigations (such as the NLSY). The following innovations, already included in NATASHA and in the NLSY CD-ROMs, are likely to be included and augmented in forthcoming data-based CD-ROMs: standardized machine-readable documentation, program statements for facilitating analysis with leading social science analysis packages, software to permit retrieval of both study-level and item-level information, and software to facilitate the creation of sample and variable extracts. Although some customization of the software for search and retrieval or for data-base extracts often will be required for new CD-ROMs, the cost will be minimized through the adaptation of existing software (as has been done already with the KAware2 retrieval system).⁶

Funds for large-scale basic research, especially for expensive primary data collection, have been growing scarcer. In recent years there has been a trend toward consolidation of data collection resources into fewer but larger multipurpose surveys funded by a consortium of government and private sponsors. These trends highlight the importance of data sharing and secondary data analysis for sociology.

The research and teaching potentials of this technological breakthrough include replication of original work to confirm significant findings, secondary analysis of rich data sets (often for purposes other than those for which the data were collected), and meta-analysis (the comparison of related results from separate investigations). Moreover, data-based CD-ROMs can be used to generate comparison data for the evaluation of service programs, as well as to provide community-, county-, and state-level variables for the study of contextual effects.

CD-ROMs with collections of data sets also can be used to help teach students about research methods, statistics, variable construction, data analysis, computers, and sociological constructs. Easy access to such a collection of data sets allows students to deal directly with methodological issues that they will face as researchers, such as when and how to use case weights, how to assess completeness and consistency of data, how to handle missing information, and how to construct measures from

raw variables. Moreover, analysis of real data such as those provided with NATASHA and the NLSY yields interesting, factual information about matters of interest to young adults, such as dating, family life, school life, careers, and the sexual, contraceptive, and pregnancy-related attitudes and behaviors of their peers. Thus, solving research problems can be made engaging to the student, and interpretation of results should be enlightening. Students' elaboration of results into research briefs could be an additional, useful pedagogical exercise.

A single data set, such as the General Social Survey, also might serve some of these pedagogical objectives. Yet only a storage device with the capacity of a CD-ROM can store dozens of data sets, enabling students to 1) select from among the data sets the one study best suited to their research question, 2) conduct parallel analyses on multiple data sets to strengthen or challenge the generalizability of results, or 3) conduct meta-analyses using results from multiple studies.

MULTIMEDIA STATISTICAL TEXTBOOKS

Statistics textbooks currently on the market generally include problems that are solved with the help of paper, pencil, and calculator. A few texts provide hypothetical data to be entered into a microcomputer for analysis. These data generally are "perfect" (no missing values and no out-of-range values). A few workbooks, such as CHIP (Mazur 1988), which is based on the General Social Survey, have gone so far as to incorporate data on diskette with a workbook to permit such analyses. The CHIP data, however, are limited by the storage medium—the floppy diskette—to a subset of the data so that the analyses possible for the student are limited in turn. In contrast, a new kind of statistics textbook, aimed primarily at advanced undergraduates or graduate students, could be developed around problem sets to be solved with complete, exemplary data sets. The textbook with its problem sets could be linked to some or all of the studies in a data-based CD-ROM, such as NATASHA, or to one or more of the yearly waves in a longitudinal study such as the NLSY. Each problem could be indexed in

⁶ Available from Knowledge Access International, 2685 Marine Way, Suite 1305, Mountain View, CA 94043, 415-969-0606, Matilda Butler, President.

ver, analysis of real data ed with NATASHA and resting, factual informa-interest to young adults, life, school life, careers, ceptive, and pregnancy-behaviors of their peers. h problems can be made nt, and interpretation of ghtening. Students' ela-research briefs could be edagogical exercise.

t, such as the General ight serve some of these s. Yet only a storage de-of a CD-ROM can store enabling students to 1) : data sets the one study search question, 2) con-on multiple data sets to e the generalizability of meta-analyses using re-dies.

STATISTICAL BOOKS

urrently on the market blemis that are solved pencil, and calculator. ypothetical data to be omputer for analysis. are "perfect"(no miss-of-range values). A few CHIP (Mazur 1988), General Social Survey, o incorporate data on ok to permit such anal-owever, are limited by he floppy diskette—to that the analyses pos-re limited in turn. In of statistics textbook, anced undergraduates could be developed o be solved with com-ets. The textbook with e linked to some or all based CD-ROM, such e or more of the yearly al study such as the could be indexed in

multiple ways, such as by topic, difficulty, or data sets appropriate for solution. Only one copy of the CD-ROM per department need be acquired: multiple stations could access the CD-ROM via a local area network, or data extracts could be copied onto floppy diskettes for a particular assignment or exercise. A prototype for such a workbook has already been developed in the field of family research (Card, Reagan, Nelson-Kilger, and Peterson 1989).

Optical scanners now can transfer information from a printed document to a microcomputer. The scanned material is converted by optical character recognition software into a machine-readable text file. This technology can be applied to create CD-ROMs that contain the entire classic and contemporary literature for a given subfield of study. Instructors could use these CD-ROMs as resource material for a class on the topic. Useful elements of such a disk include 1) the relevant scientific literature, 2) illustrative data sets capable of shedding light on the topic, and 3) programmed teaching and learning aids such as search and retrieval software, model lesson plans at both the undergraduate and the graduate level, and interactive computer-aided instruction programs. Such lessons would refer to publications and data sets on the CD-ROM.

All of the capabilities outlined above already exist individually, either fully developed or in prototype. What makes the CD-ROM such a powerful tool is that all of these separate capabilities can be brought together and integrated carefully in a single storage medium. The potential of CD-ROM technology for teaching and research is beginning to be appreciated only now. The future is wide open and promises great excitement and innovation.

REFERENCES

- Baron, J. N. 1988. "Guest Editorial—Data Sharing as a Public Good." *American Sociological Review* 53:vi-viii.
- Card, J. J. 1989. "Facilitating Data Sharing" *ASA Footnotes* 17:8.
- Card, J. J., R. T. Reagan, M. Nelson-Kilger, and J. L. Peterson. 1989. "Establishment of a Family Data Archive: SBIR Phase I Final Report." Los Altos, CA: Sociometrics Corporation.
- Center for Human Resource Research. 1989. *NLSY: National Longitudinal Surveys Compact Disc User's Guide Data Base Access, Release 1.1*. Columbus: The Ohio State University Center for Human Resource Research.
- Desmarais, N. 1990. *CD-ROMs in Print*. Westport, CT: Meckler Corporation.
- Fienberg, S.E., M. E. Martin, and M. L. Straf, eds. 1985. *Sharing Research Data*. Washington, DC: National Academy Press.
- Hauser, R. 1987. "Guest Editorial—Sharing Data: It's Time for ASA Journals to Follow the Folkways of a Scientific Sociology." *American Sociological Review* 52:vi-viii.
- Mazur, A. 1988. *CHIP: Marriage and the Family 246*. Syracuse, NY: Scottholm Terrace.
- Microsoft Corporation. 1987. *Microsoft Bookshelf User's Guide*. Redmond, WA: Microsoft Press.

Josefina J. Card is President of Sociometrics Corporation, a research and development firm specializing in social science research applications and in the use of microcomputers to further social science research. Sociometrics has pioneered in the establishment and operation of topically-focused data archives in the adolescent pregnancy, family, criminal justice, and aging fields of study. Address correspondence to J.J. Card, Sociometrics Corporation, 170 State Street, Suite 260, Los Altos, CA 94022.

James L. Peterson is Principal Research Scientist at Sociometrics Corporation. His primary interests and experience have been in research focusing on children, youth, and families. He has served as Study Director at the Institute for Survey Research at Temple University, Staff Associate at the Social Science Research Council's Center for the Coordination of Research on Social Indicators, and Assistant Research Director and Assistant Director of Child Trends, Inc.