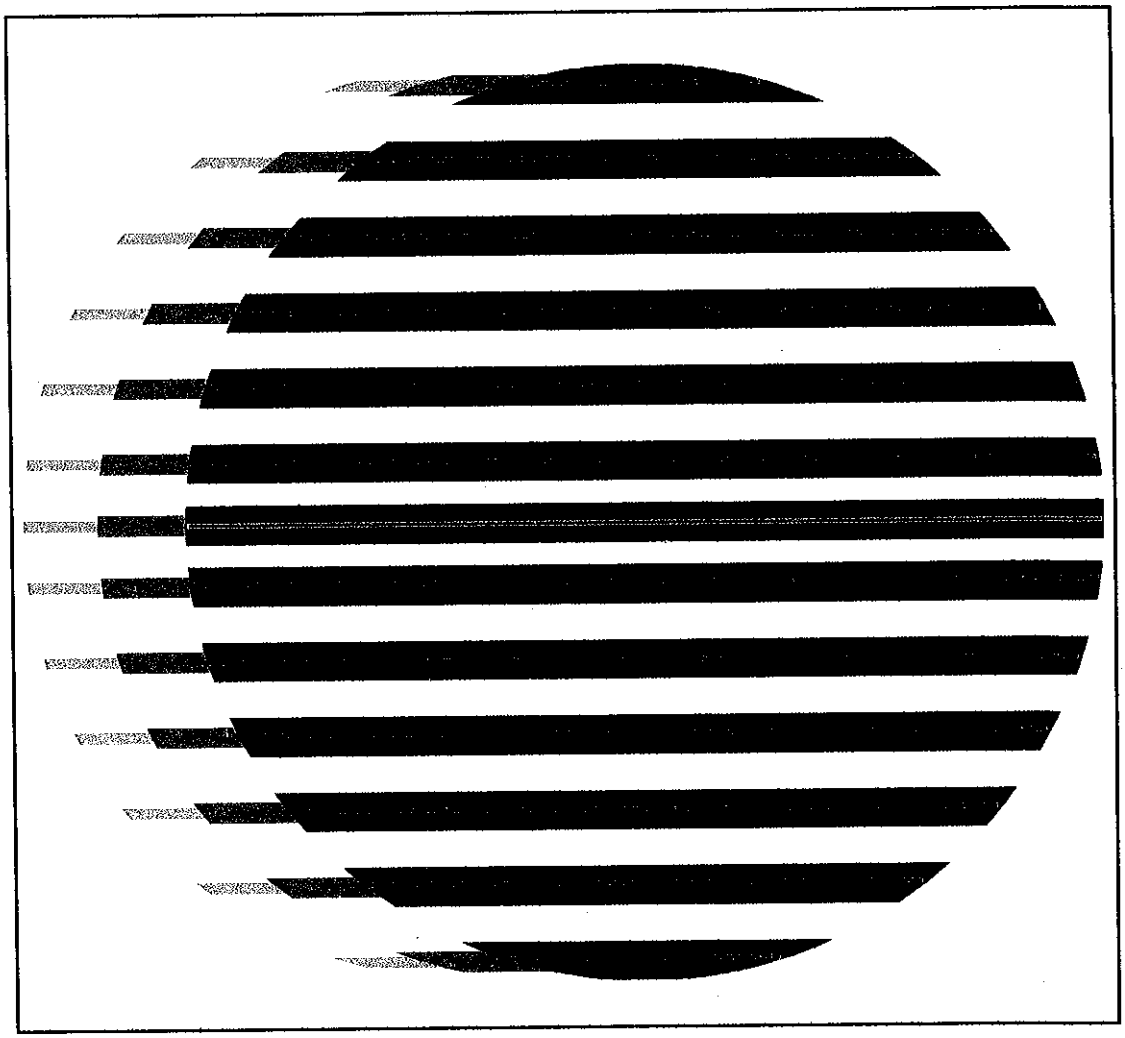
IASSIST

QUARTERLY

VOLUME 24 Fall 2000 NUMBER 3



IASSIST QUARTERLY

The **IASSIST QUARTERLY** represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The **QUARTERLY** reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of **IASSIST**.

Information for Authors:

The **QUARTERLY** is published four times per year. Authors are encouraged to submit papers as word processing files. Hard copy submissions may be required in some instances. Word processing files may be sent via email to jtstratford@ucdavis.edu. Manuscripts should be sent to Editor: Juri Stratford, Government Information and Maps Department, Shields Library, University of California, 100 North West Quad, Davis, California 95616-5292. Phone: (530) 752-1624.

The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press *Manual of Style* or Kate L. Turabian's *Manual for Writers*. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". *Journal of the American Society for Information Science 30(2):77-82, March 1979*. Announcements of conferences, training sessions, or the like, are welcomed and should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event.

Title: Newsletter - International Association for Social Science Information Service and Technology

ISSN - United States: 0739-1137

CONTENTS

Volume 24 **Number 3** **Fall 2000**

FEATURES

The Social Science Electronic Data Library: Serving the Needs of Data Librarians and Users

by Michael Carley & Josefina J. Card *

The last decade has witnessed enormous strides in two areas: first, the development of numerous social science data sets of high quality; and, second, the development of the computing hardware and software capability and infrastructure needed to locate and analyze these data sets for minimal cost and to communicate data findings in interesting and easy-to-understand fashion. Hand in hand with these advances in data development have come technological advances which allow social science research and teaching laboratories, with the hardware and software needed to analyze the best data in a given field, to be set up with ease by an academic department or even by an individual professor. Additionally, sophisticated data analysis software packages formerly available only for mainframe computers have become available for microcomputers at a much reduced cost. Taken together, these developments make it possible for academic departments, research institutes, and government offices of all sizes and levels of financial resources to access and analyze exemplary data sets for research, teaching, and program- and policy-development purposes.

Data archives, in both the private and public sectors, allow easy and open access to many hundreds of the best health and social science data sets covering a broad range of topics, study populations, and making use of a variety of research designs. The data available from many of these archives are clean and the documentation user-friendly. For these and other reasons, researchers and instructors who are considering the use of secondary data welcome the functions served by a well designed data archive. This data is used to:

- conduct secondary analyses of outstanding data sets to serve the needs of policy, practice, or basic research;
- perform meta analyses based on access to multiple original raw data sets;
- prepare research proposals on various issues;
- write publications comparing and contrasting results from related data sets;
- prepare masters theses and doctoral dissertations;
- produce classroom materials for teaching substantive, methodological, and statistical concepts from real-world data.

In this paper, we will discuss three areas of major concern for data librarians and data users wishing to obtain and use data for secondary analysis: data quality, format, and dissemination. We review and contrast the issues and concerns of data librarians and data users, outlining areas of similarity and difference. We will explore how one large data collection, the Social Science Electronic Data Library (SSEDL), compiled over the last 17 years by Sociometrics Corporation, has addressed each of these issues and the conflicts and problems that arose during that process. Finally, we peer into the future, assessing how data providers can bridge knowledge gaps via recent technological advances.

The Social Science Electronic Data Library (SSEDL)

The Sociometrics Social Science Electronic Data Library is a premium health and social science resource that consists of seven topically focused data archives. With over 300 data sets from 200 different studies comprising seven topically-focused collections, it is a unique source of high quality health and social science data and documentation for researchers, educators, students, and policy analysts. The Electronic Data Library was made available in 1999 on a set of CD-ROMs and includes an online membership with free access to datasets for downloading by members.

The Collections:

The **Data Archive on Adolescent Pregnancy and Pregnancy Prevention** (DAAPPP) was established by the US Office of Population Affairs (OPA) in 1982 as the repository for the best social science data on the incidence, prevalence, antecedents and consequences of teenage pregnancy and family planning. In 1994, the scope of DAAPPP was expanded to include studies that focus more broadly on adolescent sexual health issues, thereby including studies examining behavioral factors related to sexually transmitted diseases (STDs) in addition to pregnancy. DAAPPP currently holds data from over 150 premiere studies (many of them longitudinal) on sexuality, health, and adolescence.

State-of-the-art research data on the American family are

available through the **American Family Data Archive** (AFDA). AFDA, funded by the National Institute for Child Health and Human Development, contains data and documentation from 20 nationally recognized studies on important issues relating to American family life, demographics, and family patterns. Among the topics covered are educational, economic, health, social, and psychological indicators, child welfare, family violence, marriage, divorce, child care and child custody.

The **AIDS/STD Data Archive** (AIDS) consists of original research data and instruments from 11 premier studies on AIDS/HIV and other sexually transmitted diseases (STDs). The collection was established with funding from the National Institute of Child Health and Human Development (NICHD. Included data sets address the following topics: the incidence and prevalence of specific sexual behaviors (including abstinence, vaginal and anal intercourse, oral-genital sexual activity, masturbation); contraceptive and STD-preventive behavior; attitudes and beliefs regarding sexual behavior and methods of contraception and STD prophylaxis; AIDS/HIV knowledge, attitudes, behavior, and serostatus; current and past episodes of gonorrhea, syphilis, chlamydia, and other STDs; and high-risk behavior, including alcohol/drug use and prostitution.

The **Maternal Drug Abuse Archive** (MDA) brings together seven state-of-the-art research databases on maternal alcohol and drug abuse. Funded by the National Institute on Drug Abuse, the collection includes data on the following topics: the prevalence of drug use among pregnant women and women of childbearing age; demographic characteristics of pregnant drug users; types and patterns of illicit drug use; social, psychological and economic antecedents of pre- and perinatal drug abuse; the effects of pre- and perinatal substance use on pregnancy complications and neonatal status; and the effects of fetal alcohol and drug exposure on children's physical, neurobehavioral, psychological and social development.

The **Data Archive of Social Research on Aging** (DASRA) was assembled with the support of a grant from the National Institute on Aging. DASRA contains data and documentation from three very large nationally recognized studies. These three studies covered a variety of topics including functional status and impairment, living arrangements, caregiving and social support, health attitudes, retirement income and plans, mortality, health, financial resources and assets, expenditures, cognitive ability, medical conditions, housing, health insurance, and personal characteristics

The **Research Archive on Disability in the United States** (RADIUS) was funded by the National Center for Medical Rehabilitation Research (NCMRR) within the National Institute for Child Health and Human Development (NICHD). The purpose of the project is to facilitate access to the best data sets on the prevalence, incidence, correlates, and consequences of disability in the U.S. The heart of the archive is a collection of 19 studies that address the topic of disability. These data sets permit analyses on topics such as: the incidence and prevalence of specific diseases, disorders, and impairments, including deficits of cognition, emotion, physiology, and anatomical structure; functional limitations across a variety of specific organ systems; disabilities in relation to major life roles and activities, such as work, parenting, education, and recreation; societal limitations including physical, attitudinal, and economical barriers that restrict full participation in society; psychosocial and interpersonal factors such as coping with stress, sexuality, feelings of control and productivity, quality of life, and family relations and support; health care and rehabilitation issues such as medical costs, coverage, service utilization, use of orthotic, prosthetic, assistive devices, effectiveness of rehabilitation; as well as a variety of basic demographic factors on respondents such as age, race, sex, income, occupation, marital status, family size, and living arrangements.

To facilitate access to the best contextual data, Sociometrics has developed a **Contextual Data Archive**. By contextual data we mean data that describe the population, social, and economic characteristics of geographic areas, from census tracts to states, in which people reside or work. The contextual data archive consists of a series of files, each organized around a different geographic unit of analysis (such as census tracts, school districts, counties, states, etc). Each file contains variables drawn from various sources, but having one common geographic unit of analysis. Support for this project was provided by the National Institute of Child Health and Human Development.

Data Quality

Both data librarians and data users have an abiding interest in the availability of high quality digital data. The librarian must put her/his limited resources to the most efficient and effective use possible. The resources we mean here are not only financial assets such as approved budgets, but also material and human capital such as shelf space and the person-hours of those who must purchase, assemble, and maintain various data collections. Because all of these resources are limited and precious, data librarians must make wise choices as to how to prioritize their use in order to achieve the highest quality collection of data for the users at their institutions.

Data users are also concerned about having data of the highest possible quality. The users of digital data wish to make the best possible contribution to the body of knowledge in their field. That contribution is placed in jeopardy if the data used is of questionable quality. Data gathered via poor research design, or via a good design

poorly executed are of little use in advancing knowledge. Researchers using such data risk not only making a tainted contribution, but also of generating criticism from colleagues and associates who recognize the problems or limitations of the data being used.

The research staff at Sociometrics recognized these needs when compiling the seven data archives in the Social Science Electronic Data Library. It was determined that each archive would be a 'best of the lot' collection, accepting only the best data available in each of the seven topic areas. To accomplish this, we formed National Advisory Panels of research scientists who were experts in both the substantive content of the particular archive and the research methods commonly applied in that field. The panel, usually consisting of six members, was asked to evaluate candidate data sets on the following five criteria:

- **Technical quality**: among the factors to be considered are high response rates, low attrition rates, use of reliable and valid measures, and sound sampling and design elements.
- **Substantive importance to the field**: factors include the potential to address contemporary issues, to break new ground, and to replicate or confirm important findings.
- **Program or policy relevance**: the ability of the data set to answer applied questions on how to improve public policy or shape intervention programs;
- **Potential for secondary analysis**, including: *Scope of sample* - The broader or more diverse the scope of the sample, the greater the potential of the data for generalization. *Size of sample* - Sample size is always an important consideration. This is even more true for data intended for secondary analysis: sample sizes adequate to support the originally intended analysis may be too small to support other analyses, especially if the new analyses focus on data cells that have a very low proportion of cases. *Breadth of variables and constructs covered* - The potential for secondary analysis is directly related to the breadth of variables measured in the data set. The more numerous and diverse the set of variables, the more possibilities there are for new or expanded analyses.
- **Disciplinary balance**: An archive should attempt to be representative of the entire field of research. Variations in state of the art exist between different sub-areas within any discipline. Thus a somewhat flexible standard (as measured by the other criteria above) should be used to ensure that all major areas of the discipline are represented in the archive as a whole.

In order to perform these evaluations, we provided each panel member with briefing materials consisting of a 2-4 page description of the data source, which covered: the purpose of the study; methods (including sampling design, periodicity, unit of analysis, response rates, and attrition); content (description of variables covered, number of variables, and topics covered); limitations; sponsorship; and a bibliography. In addition, we provided copies of original peer-reviewed publications for each data set, which allowed the panel members to review issues we may have not addressed in our briefing documents. Panel members were encouraged to suggest additional data sets for consideration, and have often done so. Panel members did not vote yes or no on each data set, but rather rated each with a 'priority score' from 1-10. Only those data sets receiving an average score of 7 or above were accepted, and higher priority for archiving was given to those receiving higher scores.

In addition to pre-screening the data sets, archivists for the SSEDL data sets perform several other tasks designed to ensure data quality. We review the data thoroughly, checking to make sure that all variable and value labels are included and are sufficiently descriptive. We check the data for internal consistency and completeness, scanning in particular for variables with an excessive number of missing or out-of-range values. We also perform random checks verifying that the skip logic in the original instrument was followed and that the variables are consistent in relation to one another (no variables describing a female as a father or brother or a male as a mother or sister, etc). Finally, we produce a user's guide to the machine-readable files and documentation which notes any remaining limitations or inconsistencies.

Those archiving digital data face many challenges, not the least of which are the limitations on their own, as well as the users' time and resources. Consequently, different data archivists take a variety of approaches to address these challenges. Our 'best of the lot' approach emphasizes quality over quantity. This means that our data collections, while not as large as some of those from other sources, are of higher overall quality and are better documented than is the industry average. This approach limits the size of our collections, but contributes to their popularity among researchers for their high quality and ease of use.

Formats

The needs of data users and librarians diverge somewhat when it comes to format preferences for digital data collections. Users look for data in the most easily accessible form, while librarians must be concerned with the big picture, and look for collections that serve the needs of as many users as possible, both in the present and the future. The format(s) in which data are provided also have an impact on the role the librarian will take in the data distribution process, which could vary from that of a facilitator to an active gatekeeper. The challenge for data

providers is to address both the *preservation* needs of the librarian and the *ease of use* needs of data users.

Distribution Media. Librarians must conserve their many precious resources, including both financial resources and shelf space. However, they also desire that data collections be in a form that is not easily lost, damaged, or misused by careless users. Therefore, a data collection should be durable, and in a format that is not likely to change rapidly with evolving technology. CD-ROM technology meets these requirements. CDs are more durable than diskettes and do not have the associated (at least perceived) transient qualities of internet sites. Data made available on a CD-ROM are not likely to be easily lost as library staff can, should they choose to, maintain tight control over them or copy their contents to a central repository for safekeeping. Diskettes are more likely to become corrupted and internet sites often are revised and require more constant updating on the part of both the data provider and the librarian to keep all links accurate and up to date.

Data users, on the other hand, prefer that the data are made available in the simplest, easiest to access format possible. However data are made available, it must be transferred to the computer where the user will actually be working. With desktop computer speed and hard drive space increasing exponentially, users often prefer that data be available for copying to their own system, rather than residing at a central repository. Internet downloads may be preferred to CD-ROMs as the data can be copied to the user's own computer, then manipulated and transformed as necessary.

To best accommodate these divergent needs, we found it necessary to make our data sets available in both CD-ROM format and via our internet web site. SSEDL Volume I is distributed via 17 CD-ROMs, along with accompanying support material. Additionally, each purchasing institution is given free web access to all of the data sets, as well as to new data that have not yet been added to the CD collection. By allowing the users to download the data sets or use the CD-ROMs, we were able to provide both the user and the data librarian with flexibility in both data format and data access.

Analytic Software. A user who wishes to use data on a particular topic would be best served by data that can be retrieved quickly and effortlessly with a variety of software. Given the wide variety of statistical software available to users in different fields, this can be a challenge. Users should have the capacity to use the software of their choice, and the ability to access the data quickly with that software. At the same time, data providers must understand that software currently popular may change or become outdated, making files created from these programs unusable or at the least cumbersome and inefficient to use.

Most data collections have taken one of two approaches to this problem. First, the data provider may distribute *raw data* with a *codebook*. The raw data is typically stored in an ascii file which is simply useless text (numbers) without the codebook. The codebook provides the user with the location of specific variables and cases within the raw data file. The advantage to this approach is that it addresses the issue of durability well. Users can access the data by writing a program using the statistical software package of their choice, inserting the variable and case locations given in the codebook to access the data. Changes in software applications do not affect data distributed in this method, as users write their own program with the language in which they have expertise. However, writing the programs to read the raw data can be a time consuming process, causing users to waste much of their resources on mundane tasks. In addition, such writing is prone to error; one misplaced character can cause much of the data to be written incorrectly.

Other data providers address these issues by distributing the data in a pre-packaged format using one of the most popular statistical software packages (typically SPSS or SAS). By distributing these formatted files (usually either complete system files or portable files), the user can access the data directly simply by opening the files in the appropriate software. When portable files are used, the data can be used with different versions of the same software package (either earlier versus later versions or versions for different operating systems) or in some limited cases, in other popular statistical packages. The advantage to this method is clear: quick, easy access to data. The disadvantage is that this approach cannot possibly be flexible enough to address all user needs. Some users wish to access the data with a software package that is not among the most popular. Also, data distributed in this method can become unusable when software packages radically change their formats. Data made available in the most recent format could be inaccessible within just a few years.

To address the limitations in each of the above approaches, data sets in SSEDL are distributed with raw data files and machine-readable set-up statements for use with both SPSS and SAS statistical software. These set-up statements provide for the best of both worlds: ease of use, combined with flexibility. Users use these syntax files to create the system or portable files in whichever software they are using. For those users who are use software other than SPSS or SAS, the set up statements serve essentially the same purpose as the codebook described above. Because the set-up files are machine-readable, they can often be converted for use with other software with a minimal investment of time. Should the syntax requirements of SPSS or SAS change radically, these files would serve also accomplish this goal.

In addition to the above files, SSEDL data sets also are distributed with a machine-readable SPSS data dictionary file and an SPSS frequency and statistics file. These aid the user in making sure that they have created their system or portable files correctly. Users can compare the statistics in the frequency file to their own, thereby preventing mistakes in data analysis. Both the frequency and dictionary files are also useful in reviewing the contents of the data set. Each data set is also accompanied by a printed User's Guide (provided in machine-readable form, in addition to printed form, for the more recent archives) comprised of a standard set of sections and subsections. The provision of standard machine-readable and printed documentation assists users in familiarizing themselves with the Sociometrics data sets. Once a user has worked with one Sociometrics-packaged data set, it is easy for him or her to work with any of the others. The original instrument and codebook are offered as optional, supplementary documentation for each data set, when available. For the more recent archives, the original instrument is distributed in machine-readable form along with the data, as a set of graphics files (page images).

Search and Retrieval Software. As data sets get larger, both in the number and scope of variables covered and in the number of cases, users are faced with an increasingly overwhelming task of reviewing which parts of a study are necessary and appropriate to their needs. Often, users will begin work with a data set containing over 5,000 variables (and often several thousand cases) only to find that their interests only require 30-40 of those variables. It is important for users to be able to quickly sort though the variable list and find those of interest. Given current (though perhaps temporary) limitations in speed and disk space, users also need to be able to reduce the large data set into one with only those variables needed for analysis.

To address this need, Sociometrics staff developed powerful search & retrieval software which now accompanies each data archive. This software allows a user to search an entire topically-focused collection, a customized group of data sets created explicitly for a given user, or a single data set; to identify variables of interest across this designated search space and to save located variables as a search set. Users can conduct: (1) full-text keyword searches, including variable names, words in variable labels (question descriptors), and words in value labels (response descriptors); (2) searches by assigned topic and type codes; and (3) searches by study name or assigned data set number. Standard Boolean operators (i.e., "and," "or," "not") can be used to combine search sets. Alongside this software, we provide data extract software which allows users of CD-ROM versions of archived data sets to create customized SPSS or SAS program files containing only those variables of interest to them. This capability permits analyses of subsets of large data sets to be conducted quickly (with rapid turn-around) on most microcomputers. It also saves users significant program development time writing and re-writing SPSS and SAS program statements to define variables used in a given analysis.

Technical Support

The Role of the Data Librarian. The role of the data librarian in this process varies a great deal among institutions. In some cases, the librarian serves as an expert gatekeeper to the data, allowing access to users as s/he deems appropriate and answering a wide range of questions users may have. Others may serve a minimal role, simply providing access to the data and support materials and little else. Librarians also vary in their level of statistical knowledge, as well as their expertise in the various topics that may be covered by the data in their collections.

Our approach to this issue was again to provide the greatest amount of flexibility in the collection as possible. While some data librarians do take on something close to a gatekeeper role, we chose to allow for those who had neither the time nor the expertise to do so. Librarians need to be fully informed, not on all of the topics included in their data collections, but on the process by which they can aid users in finding data of interest to them. Given the wide variety of topics covered in SSEDL, one could never expect librarians to provide users with all of the help they may require. To this end, the Social Science Electronic Data Library includes a variety of tools to facilitate this process. A user's manual and contents manual detail the data sets included in the collection, and a quick start guide offers advice on how to implement the software and the knowledge needed to use the data sets. Most importantly, a 'Guide to the Social Science Electronic Data Library' CD is provided with each collection. This CD takes the librarian through the process of using the data library using a brief step by step tutorial.

SSEDL's Research Support Group. In addition to the support given to the data librarian, users have direct access to help from Sociometrics' archiving and scientific staff through our Research Support Group (RSG). The Research Support Group consists of Ph.D. and Masters level social scientists who provide free technical assistance for users who have questions about accessing or using our data sets. In addition, the RSG occasionally performs consultant work such as the creation of customized data set extracts; user-defined statistical tables and analyses; data archiving, management and analysis services; customized CD-ROMs; and training workshops. These services greatly aid users with limited expertise or resources with which to conduct their own analyses.

Dissemination

The manner and methods by which digital data are disseminated by data providers and eventually by data librarians are crucial to the usability of such data. Users

must be made aware of the availability of data that meets their needs. However, with today's rapidly expanding technologies, the problem of 'information overload' is a crucial one. Even within our own collections, users can become overwhelmed with the sheer amount of information available to them. If these issues are not handled properly, they can inhibit the users' ability to locate and make use of the most appropriate data. Data providers must do what they can to ensure that users have the capability to quickly and easily locate the data sets and even the particular variables their topic of interest requires.

While the search and retrieval software made available for each individual data archive helps users find variables within a study they have already chosen, it does not help when users have not yet selected the study that meets their needs. To address this issue, we created a search mechanism for use on our internet site which is cross archive. This allows for users to search by keyword(s), or designated variable topic or type, for variables of interest in all of the SSEDL data sets. Users can search for words in variable labels (question descriptors) or in value labels (response descriptors). In addition, we include a brief abstract of each study, and users can search for keywords within those abstracts. We chose to make this software available to the general public as well as users on our internet site, in order to allow researchers at non-purchasing institutions the opportunity to find data sets of interest and order them individually.

Data librarians must also make an effort to make help users become aware of available data collections. In order to facilitate this process, we provided not only the above mentioned guide to SSEDL on CD-ROM, we provided informational flyers and brochures to help the librarian make potential users aware of the availability of our data collections. In addition, the Research Support Group provides both librarians and users ongoing advice as to how to find data sets of interest in our collections.

Looking to the Future: New Technologies, New Audiences

The value of any data collection is in part predicated upon its ability to address issues of the day. Therefore, any collection will inherently be of greater value the newer the data are that are contained within it. The preservation of historic data is clearly important, but any collection that aspires to be useful must also be kept up to date with the addition of more recent data. We will continue enlarging the content and capabilities of our data set collections. We will be adding to our current data archives as funds permit, and expanding our efforts by adding new topic areas to the collection. We expect to begin the establishment of a data archive on child well-being shortly. A feasibility study on the formation os a complementary and alternative medicine data archive has just been successfully completed.

In putting together the SSEDL, Sociometrics' staff have learned a great deal about data collection methods and ways to improve efficiency and reduce costs to researchers. Currently, we are developing a software product that will aid researchers on this aspect of the process. Sociometrics' Automated Dataset Development Software (ADDS) is an integrated software program that, when completed, will develop and document social science research studies. The program will perform the following functions: 1) Instrument generation—generate a fully formatted research instrument in print, ASCII, and other machine-readable formats. 2) Codebook generation—generate the data set documentation in a printed codebook (also in ASCII and other formats), flow chart (skip map), and data file map. 3) Data entry—provide for data entry from completed questionnaires, with simultaneous error checking. 4) Program file generation—produce a raw data file in ASCII format, and build the program statement files needed to transform the raw data file into SPSS and/or SAS system files. The software will automate tasks best done by computer, improve instrumentation and documentation by providing a complete, high-quality structure and format, and reduce the post data-collection effort of documenting a public-use data set.

Additionally, we are also building an item bank of high quality, commonly used questions, scales, and interviewing tools from the SSEDL collection. This bank will be accessible within the ADDS program to permit users to select questions to develop their own research instruments. The item bank will be filled with several thousand questionnaire items drawn from some of the leading studies in research on the American family. Using questions or scales that have been previously tested will not only improve the choice of questions, but will also lead to greater comparability between studies and over time.

In addition to keeping the Social Science Electronic Data Library current and helping researchers imrpove their methods, we hope to reach new audiences through new technologically innovative products. Bridging the 'knowledge gap' is of prime importance to those who wish to make practical contributions through social science research. Rather than limiting our efforts to trained researchers, we must reach out to other professionals, and, when possible, the lay public as well. We are beginning our efforts to reach the 'paraprofessional' audience with two new products related to the SSEDL. The U.S. Social Surveys: A Sampler of Questions and Responses, will contain searchable, edit-ready, and print-ready machine-readable versions of the demographic, behavioral, and health science instrumentsæquestionnaires, medical forms, interview protocolsæused to collect the data in SSEDL. Questionnaire items will be linked to crosstabulations with age, race/ethnicity, and gender, obtained from the linked SSEDL data archives.

Secondly, the Multivariate Interactive Data Analysis System (MIDAS), will allow online analysis of the data in SSEDL. Online data analytic procedures will include weighted and unweighted frequencies, percentiles, and measures of dispersion and central tendency, as well as two-way and *n*-way tables with measures of association, comparison of means (2-group and ANOVA) and correlations, and the calculation of complex variance estimations. Users will be able to define case subsets, recodes, or aggregations for analysis, and then produce output which can be downloaded or printed. Custom dataset downloads will also be available.

The goal of ADDS is to aid expert researchers in handling the 'front end' of the research process. Through the use of the ADDS software, researchers will be able to reduce their costs and improve the accuracy and efficiency of instrument development, data collection, input, management, and analysis. The goal of Social Surveys and MIDAS is to increase the accessibility of the data to those who are not competent in the sophisticated statistical software packages such as SPSS or SAS. These products will help the 'paraprofessional'--people with college degrees who are not necessarily trained in complex data analysis— avail themselves of exemplary social science data. They will provide a basic introduction to social science methodologies as well, and will be linked to the SSEDL data sets for those who wish to progress to the next stage of data analysis (e.g., advanced undergraduate students).

In sum, the historic progression of SSEDL has been to expand the definition and purpose of a data archive. SSEDL staff have worked to enhance archiving methods to make data easily accessible to researchers. Our advances in this field have helped to make the research process more efficient, especially for those conducting secondary analysis. ADDS will improve the process for primary research as well. Non-researchers will be introduced to data analysis through the new products, Social Surveys and MIDAS. Together, these products will allow us to extract the greatest possible value from our research dollars as data will be used in as many ways as are feasible and by a much wider audience.

* Michael Carley and Josefina J. Card, Sociometrics Corporation, Contact Name and Address: Josefina J. Card Sociometrics Corporation, 170 State St. Suite 260, Los Altos CA 94022, (650) 949-3282, ext. 211, FAX (650) 949-3299, jjcard@socio.com