Data Sources for Library and Information Science Research

Presentation for Library Research Seminar I November 1-2, 1996 Florida State University Tallahassee, Florida USA

Robin Rice
Special Librarian, Data and Program Library Service
University of Wisconsin-Madison
rice@dpls.dacc.wisc.edu

Introduction

It is definitely a pleasure and honor to be participating in this seminar with all of you. I am here, in part, because of conversations I have had with library school faculty members about the lack of use of my library's resources by library school researchers, particularly doctoral students on our campus. I work as a Special Librarian at the Data and Program Library Service at UW-Madison, which serves the Social Science research community on campus through reference and provision of machine-readable statistical data files and accompanying documentation.

I believe that Library and Information Science (LIS) is an inter-disciplinary field, and that to a great extent, Library Science can be characterized as a Social Science. This assumption shapes the rest of this paper, and what I have to offer in the way of research stones unturned. First I will discuss the notion of secondary analysis, then I will discuss some of the major datasets available.

Secondary Analysis

Secondary analysis means using a body of data collected by someone else, for a given purpose, and coming up with new findings based on different approaches to examining the data. For example, the Census of Population and Housing is conducted by the U. S. Bureau of the Census every ten years by Congressional order because of a Constitutional mandate to count the population for the purpose of defining legislative districts. Obviously, this huge body of data is used by many individuals and groups for myriad other purposes. Clearly, it is beyond the ability of each of these other groups and researchers to gather the equivalent

of U.S. census data on their own. At best, a symbiotic relationship occurs in which an expensive data gathering project is given added value by stake-holders who use the data to generate new and more knowledge than the original intended purpose. Secondary analysis is a unique tradition in the Social Sciences compared with the Physical Sciences where primary experimentation is the main arena for research, and the Humanities where less emphasis is placed on empirical proof of theorization.

Using secondary analysis for quantitative research provides the researcher with the following five advantages: (1) The researcher is able to focus on research outcomes rather than collection activity. (2) Economy of scale for large, expensive sample sizes is achieved (e.g. even if the dataset must be purchased by the researcher, the cost is much less than that of conducting one's own study). (3) Because the data is available for public use, results are reproducible, and therefore can be tested for reliability. (4) There are less sampling and non-sampling errors due to the robustness of large datasets as well as the expertise of the data collector and/or checks and balances of the funding mechanism. (Large datasets often include an 'over-sample' of minority populations to achieve statistical significance, and include appropriate "weights" to avoid distorted measurements.) (5) There is a pre-existing network of researchers, both inside and outside one's field familiar with the dataset, to expand one's research 'conversation'.

Disadvantages also exist: Researchers often have to reformulate their original research question when they find out what data is and is not available. They have no control over survey questionnaire wording and other design issues. They lack an intimate understanding of the collection process, and therefore need to study the documentation thoroughly to avoid making errors in analysis. And, because principal investigators sometimes wait until the last stages of a project to take the time to document the data, codebooks can be incomplete or ambiguous, or, excruciatingly voluminous. Also, analysts may have to deal with "dirty" data, (i.e. data files which contain processing errors such as missing records or misplaced variables). The lag-time between when the data is collected and when it becomes publicly available is often more than a year or two. Limits may be placed on data items due to confidentiality concerns, such as geographic location or ability to match individuals across datasets. Finally, funding may be suddenly cut or reduced so that research dependent on future series may have to be dropped.

The last two disadvantages--restrictions due to confidentiality concerns and loss of funding have become more prominent with actions taken by the current U.S. Congress. In 1995 Congress imposed new rules on confidentiality in the Department of Education. One outcome is that the important panel study High School and Beyond has been released in its latest wave only in table, or aggregate form, whereas previously it was released as microdata—a file or files containing each individual's responses to a questionnaire. The inability to access the raw data by individual observations is a serious blow to researchers studying the social outcomes of the HS&B cohort. Fortunately there is a procedure for gaining access to the raw data for bona fide research, but it is cumbersome, slow, and difficult for graduate students to obtain the special permission.

Another threat to Social Science research comes from HR1271, the "Family Privacy Protection Act," passed in the House of Representatives last year as part of the Contract

With America, but not yet passed in the Senate (Library of Congress, 1996). This act targets research on youth by requiring written parental permission before minors are surveyed by any program funded in whole or in part by federal funds. The previous norm for youth research was the reverse--parents could choose to exclude their children's participation by writing a note. This bill would practically assure the demise of the long-term study, Monitoring the Future which has surveyed adolescents' at-risk behavior and attitudes since 1976. The law would grossly skew the sample intended to equally represent all American youth by inordinately increasing the pool of nonrespondents (i.e. students whose parents fail to grant written permission to participate).

While privacy is a legitimate concern for survey respondents, research norms have long protected individual identification in datasets by stripping out name, address, social security number etc. and replacing them with a unique identifier variable. In a Senate Government Relations Committee hearing in August 1996, Senator Glenn defended the contributions of Monitoring the Future and outlined less drastic precautions that could be taken to protect privacy rights, such as distributing the questionnaire content to parents in advance (Library of Congress, 1996).

The Act explicitly names the following as off-limit questions for minors: "Parental political affiliations or beliefs. . . mental or psychological problems. . . sexual behavior or attitudes. . . illegal, antisocial, or self-incriminating behavior. . . appraisals of other individuals with whom the minor has a familial relationship. . . [and] religious affiliations or beliefs" (Library of Congress, 1996). It is as if Congress believes that suppressing knowledge about teenagers' use of drugs, sexual activity, and exposure to domestic violence etc. will make the associated social problems go away.

Funding of social science research has become politicized too, as Congress has withheld funding for various data collection agencies during this year's Budget showdown. Targeted agencies included the Bureau of Labor Statistics, Department of Education, the Social Science Directorate of the National Science Foundation (which Representative Walker claimed should not be funded because Social Science is not really a science), and it is even being proposed that the Department of Commerce, which includes the Bureau of the Census and the Bureau of Economic Analysis be "dismantled" and its duties farmed out to various agencies. As of September, the Government Accounting Office is examining claims of "duplication" of collection and dissemination of federal data. Questions have been trimmed from the long-form survey of the Year 2000 Census, and the notion of piggybacking research questions on the Census has been called into question, threatening that symbiotic relationship between government data producers and researchers mentioned earlier.

I. Sources of Institution-Level Data

As a social science, it seems to me that LIS has two potential universes to study: people and institutions. (There is also the universe of documents or text, which is outside the scope of this paper.) Data can come from government or private producers. First I will discuss various sources for studying institutional data. Possible areas of interest for study are: comparisons among libraries of the same type; among libraries of different types (i.e.

public, school, academic, special); comparisons among parent institutions (corporations, school districts, universities, state governments, and municipalities to which libraries belong); and international comparisons.

U. S. Government Sources

The following four datasets are collected by the National Center for Education Statistics (NCES)--under the Department of Education, in their Library Statistics Program. It is worth mentioning that the National Data Resource Center (NDRC) has been established by the NCES to provide free, custom data analysis on NCES datasets for the public, with a turnaround time of 4-6 days (NCES, 1996).

The Public Libraries Survey is collected and disseminated annually through the Federal-State Cooperative System for public library data (FSCS). All known public libraries-- about 9,000, are included in the survey, identified by State Data Coordinators at state library agencies. Data requested from each library includes "Information about staffing; operating income and expenditures; type of governance; type of administrative structure; size of collection; and service measures such as reference transactions, public service hours, interlibrary loans, circulation, and library visits," (Davis 1995, 467). These data are available at the individual library level, and aggregated at the state and national levels. The most recent year released is for 1993.

This geographically-specific dataset could be used to track attributes of public libraries in relation to other societal factors of towns and cities, and the demographics of the people who live there. Examples of interesting questions might be: how do service hours affect the number of library visits, and what breakdown of hours tends to serve the most people? How does a local unemployment rate affect library visits? How do different types of library governance affect types of expenditures? What is the difference in number of reference transactions or inter-library loan for libraries with more support staff and less professional staff? Do libraries' collections reflect the ethnic composition of the local residents? Do towns that have libraries with a higher operating budget per capita, or higher circulation rates per capita, have other commonalities, such as a highly- educated population, or a strong school district, or low unemployment, or other, perhaps surprising factors, such as a large immigrant population, or a certain form of governance structure, and so on.

Obviously, more data besides the Public Libraries Survey would be needed to get at these types of questions. Local Census demographic breakdowns would be useful, along with other institutional data such as the Annual Survey of Governments, combined with the quinquennial Census of Governments, by the Bureau of the Census. These datasets provide detailed employment and financial data for local municipalities as well as county and state governments, and even school districts. Here, comparisons are possible with other budgetary outlays and among geographic locales. Other local-level datasets such as the Uniform Crime Reports collected by the Federal Bureau of Investigation, which reveals crime volume and types for local areas, may also shed light on how the public prioritizes local funding for libraries in relation to other public services such as police force, schools, and social agencies.

The Academic Libraries Survey (ALS) conducted by NCES examined academic libraries on a three-year cycle from 1966 to 1988. Since then, ALS has been integrated with the IPEDS (Integrated Postsecondary Education Data System) and is on a 2-year cycle. (IPEDS began in 1986 as the successor to HEGIS--Higher Education General Information Survey). Three thousand, five hundred academic libraries are surveyed in this census of accredited higher education institutions as well as non-accredited institutions with a program of four or more years. Variables focus on "number and salaries of full-time equivalent staff, by position; circulation and interlibrary loan transactions; book and media collections; public service hours and number served; and operating expenditures by purpose," [NCES 1995, 459].

As the future of higher education receives more and more scrutiny by social scientists and university administrations the role of university and college libraries becomes a crucial issue for members of the LIS field. Academic librarians today are grappling with many significant decisions. For example, 'To what extent should academic libraries reduce costs of rising serial prices by staking out an area of an easily-shared, distributed collection across research libraries, versus maintaining a diluted, but more rounded and accessible local collection?' A similar dilemma is 'What proportion of resources should be used to enhance users' access to electronic information versus building the print collection?' Early indicators of successful trends may be able to be discerned through comparisons among academic libraries of similar size or status. The IPEDS as a whole can be used to relate universities and colleges' administrative structure, student completion rates, or strengths of different departments, to characteristics of the institutions' libraries.

School libraries/media centers were surveyed by NCES in 1985-86 in the School Library Statistics Survey and in the 1990-91 Schools and Staffing Survey (SASS). Questions asked of both public and private elementary and secondary school libraries and media centers included "number of students served and number of professional staff and aides; at the district level, number of full-time equivalent librarians/media specialists, vacant positions, positions abolished, and approved positions; and amount of librarian input in establishing curriculum," (NCES 1995, 468). In the 1993-94 follow-up survey, the questionnaire was revised and included questions about facilities, technology, career histories and work load of the Media Specialist/Librarian and perceptions of the profession and workplace. (Davis 1995, 122). The units of analysis for the SASS as a whole are school districts, schools, administrators, and teachers.

Non-government Sources

Many surveys of libraries (and library professionals) have been conducted by LIS researchers on their own or with funding from private foundations or government agencies. Unfortunately, it has not been a tradition for LIS researchers to archive the raw data to make available for public use. However, the advent of the World Wide Web means that principal investigators can more easily make their findings accessible to the public with enhancements such as hypertext links, tables, graphs, and images of survey instruments. They could even, if they wish, put up their data files to be downloaded for secondary analysis. One current example of principal investigators doing this is the web version of The 1996 National Survey of Public Libraries and the Internet: Progress and Issues: Final

Report (Bertot, McClure and Zweizig). This timely survey builds on data collected in a previous survey by the same principal investigators in 1994—another valuable aspect for a study focused on rapidly changing technologies.

Library associations sometimes distribute survey data they have collected about their members. The American Library Association (ALA) Library & Research Center (LARC) maintains a set of fact sheets on basic library statistics on the World Wide Web (ALA 1996). Check your state or regional library associations and chapters for further (local) sources. The Association of Research Libraries (ARL), through its Statistics and Measurement Program, collects, analyzes and distributes yearly institutional data about its 119 large academic and non-academic research libraries in North America covering 54 variables. They have collected statistics beginning in 1961, and maintain previous series dating back to 1907. "The current ARL data include categories for library characteristics, collections, service activities, personnel, expenditures, and university data" (ARL 1995).

The program maintains an impressive World Wide Web site that offers 1993 through 1995 institutional data files for downloading and displays a few key time-series and regional comparative data as graph and map images. The site also allows interactive, customizable data extraction of univariate, bivariate or multivariate statistics into graphs, spreadsheet, or raw ASCII data files, and maintains excellent online documentation to describe the data. The documentation even notes limitations of the data as collected: "Except for a few services, the library variables still concern the inputs of on-site collections, staff, and expenditures. The data are useful for describing the traditional characteristics of research libraries, but not for assessing emerging uses of technology for access to information" (ARL 1995). ARL apparently is attempting to correct this gap through new 'special projects': (1) the Interlibrary Loan/Document Delivery Performance Measures Cost Study funded by the Andrew W. Mellon Foundation, (2) a new short survey called Innovative Services in ARL Libraries, (3) a grant from the Council on Library Resources "to study ways to collect information on the character and nature of electronic resources and services," (ARL 1996) and (4) recent supplements to the standard questionnaire on items such as non-print resources and government documents.

The Special Library Association, while not offering public-use datasets per se, articulates a research agenda for the study of special libraries. The agenda highlights five areas for focus: (1) futures, (2) current user issues, (3) measures of productivity and value, (4) client/user satisfaction measures, and (5) staffing (SLA 1996).

II. Sources of Individual-Level Data

While institutional data can yield important knowledge about societies and libraries themselves, in many ways it is the study of people that is the fascinating stuff of social science. The field of LIS is interested in people as library service publics, as information-seekers, knowledge-interpreters and problem-solvers, as library users and non-users, as readers, as consumers of media, as citizens, and so on. Do we know enough about the people we serve to serve them well? Quantitative person-level data comes as censuses (full-count) and surveys (samples). Surveys can be timely (such as poll data measuring public reaction to a current event), or in time-series (questions repeated across years), cross-

sectional (random slices of the population), or longitudinal (following a panel of people through time).

U.S. Government Sources

Census data are the first obvious source of data about persons. Most countries conduct some sort of regular census. The full count of the U. S. decennial census is unparalleled as a source of demographic and housing data for this country and is aggregated by the Census Bureau at several geographic levels: nation, regions, divisions, states plus District of Columbia, counties, county sub-divisions, places/remainders, tracts, block groups, and blocks--the smallest unit of analysis available. Indian reservations, Congressional districts, and zip code areas may also be isolated. Census data are regularly applied by librarians in conducting needs assessments of their local populations. But, as mentioned above, they could be used comparatively along with institutional data about libraries and local governments to draw broad conclusions about the responsiveness of local libraries to changing demographics of communities.

Interestingly, the Year 2000 Census will, for the first time, incorporate statistical techniques to estimate more accurate counts of the population than the Bureau has been able to achieve with door-to-door census takers and mail surveys. Ironically perhaps, this decision is welcomed among social scientists because it is expected to more accurately represent true numbers of minority and homeless populations.

The latest census—1990, is the easiest to access in machine-readable form, because it is distributed on CD-ROM as well as magnetic reel-tape, and can therefore be used on a personal computer. All Census CD-ROM products, including the Summary Tape Files (STF), which provide comparative summaries for the 100% population counts, are written in ".dbf" format, a generic version of D-base. They can be accessed with easy-to- use report generating software called GO, and extraction software called EXTRACT. Moreover, census summary data can be accessed via the Internet at the LOOKUP sites, which mount the CD-ROMs online, and in various formats at the growing Census Bureau World Wide Web site (U. S. Department of Commerce 1996). The map-able Census TIGER files are yet another new format from the 1990 Census that can provide a visual, geographic underpinning to the display of many types of data.

PUMS

The Census Bureau not only counts heads, it also surveys samples of the population for information about social and economic characteristics. "The long form" of the census is sent to 5% of the population and released as the PUMS (Public Use Microdata Series), which is available in different sample. Researchers may choose the smallest file producing statistically significant results for their subject of research, which is convenient because the complete 5% sample is several gigabytes in size

CPS

The Current Population Survey (CPS) is a large monthly survey conducted by the Census Bureau on behalf of the Bureau of Labor Statistics. Its primary purpose is to calculate the

current unemployment rate. By asking respondents about their labor activity in the week prior to the survey, and rotating sample groups out every two years, it provides a wealth of current and historical labor market data for economists. LIS researchers could even study librarians in the labor force by using standardized occupation and class of worker codes.

Certain months of the CPS are often devoted to special topics. Of particular interest to LIS researchers may be the supplements on adult education; multiple job holding, flextime, and volunteer work; or the November 1994 supplement on computer ownership and use. The CPS and other labor statistics can be tapped through a series of interactive forms at the BLS website (U. S. Dept. of Labor 1996).

NLS

One famous longitudinal dataset, which is sponsored by the Bureau of Labor Statistics is the series of National Longitudinal Surveys of Labor Market Experience, (NLS) which have followed cohorts (mature and younger men, mature and younger women, and youth) across several years. The National Longitudinal Survey of Youth, known as the NLSY, began in 1979 and has been continued up to the present. An indication of the heavy use this dataset receives is that the bibliography of publications based on the NLS is 447 pages long. Fortunately, the main index is grouped by descriptors, based on groups of variables (Fahy 1995).

A quick search of the NLSY codebook on CD-ROM on the terms "library" and "reading" yielded a few results. In 1979, it was asked of 14 year-olds, "Do you or anyone else living with you have a library card?" to which 8,943 responded yes and 3,690 said no. This question was not asked again in later years. In the same year the approximate number of catalogued volumes in the school library of the respondent was recorded. On a perhaps less encouraging note, a question was asked once in 1981 about leisure reading. "Other than school or work, how much time did you spend yesterday reading books, magazines, or newspapers?" The answers were broken down into dimensions based on number of hours/minutes spent. The number of respondents answering "zero" far outnumbers the other answers combined. In years 1988-92 it was noted if the "Respondent cannot read," (Ohio State University 1993).

A new cohort, referred to as NLSY96 has been started and is scheduled to continue through 1998. "This group of young people, ages 12 through 17, will be interviewed annually to study how they make the transition from full-time schooling to the establishment of their families and careers. Subsamples of this representative national sample of adolescents and young adults will include: a Department of Labor sample of approximately 12,000 young people; a Department of Education oversample of 2,500 disabled students; and two Department of Defense samples, as well as transcript and other school-based information," (Data and Program Library Service 1996a).

NCES Datasets

Before turning to private sources, one more large government data producer that needs to be mentioned here is the National Center for Education Statistics. Besides the datasets produced in the Library Statistics Program, LIS researchers may be interested in some of

the individual-level surveys produced by this agency. Several longitudinal studies produced by NCES are designed to shed light on educational outcomes by following students out of high school and into higher education or the workplace. The first of these, the National Longitudinal Study of the High School Class of 1972, followed a cohort of high school seniors from 1972 through 1986. To compliment this study, NCES began conducting High School and Beyond, which followed sophomores and seniors in the year 1980 to the present, including dropouts. The third major longitudinal study sponsored by the NCES is the National Education Longitudinal Study of 1988. This time, a base-year cohort of eighth-graders was selected. The last interview was in 1992 and another is scheduled for 1998. The Beginning Postsecondary Student Longitudinal Study was begun in 1990 in order to include older students, not just recent high school graduates in the study of transitions to post-high school schooling. Survey years will alternate with the the Baccalaureate and Beyond Longitudinal Study, which will follow college graduates into their education and work experiences (Davis 1995).

NCES has also produced interesting cross-sectional studies. The National Household Education Survey was conducted in 1991, 1993, and 1995. As a household survey, questions span from pre-school to adult education," (Davis 1995, 16). Another recent and valuable survey is the National Adult Literacy Survey so far conducted only in 1992, to a nationally representative sample of about 15,000 individuals aged 16 and older and to 1,000 adults incarcerated in federal and state prisons. The survey is designed to measure literacy "along three dimensions: (1) prose literacy--the ability to understand and use information from connected texts that include editorials, news stories, poems; (2) document literacy--the ability to locate and use information contained in documents, such as job applications or payroll forms, bus schedules, maps, tables, indexes, and (3) quantitative literacy," i. e. basic math skills (Davis 1995, 65]

Non-government Sources

Since non-government produced datasets are not conducted for an administrative purpose they sometimes provide more probing variables about people. As with poll data though, the headlines generated by new studies rarely provide enough information to be used for scientific purposes. Frequencies can show a pattern of an entire group, but individual cases must be examined to understand variations in the responses.

GSS

"The General Social Survey (GSS) is an almost annual, 'omnibus,' personal interview survey of U.S. households conducted by the National Opinion Research Center (NORC) with James A. Davis and Tom W. Smith as principal investigators (PIs). The first survey took place in 1972 and since then more than 35,000 respondents have answered over 2,500 different questions," (Inter-university Consortium for Political and Social Research 1996a). As a cross-sectional study, the GSS attempts to study social 'indicators' by replicating the same survey questions each year, or in rotating years, with as little variation in wording as possible. At the same time, it has absorbed requests from social scientists and funders to add new items, so that the current questionnaire takes 90 minutes to administer. Also, one-time supplements have been added to keep the balance between replication and new

questions.

The GSS has always been a user-friendly dataset, since its purpose is to avail itself to secondary analysis. With the advent of a complete website earlier this year, it is now easier to use. GSS-DIRS (Data and Information Retrieval System) is equipped with a search engine, online codebook, subject and module indices, searchable bibliography, reports, trend tables, discussion forum, and interactive extraction of the cumulative data file (1972-94).

A brief search on the term 'library' reveals a number of questions from several years asking whether the respondent would agree with community members who wanted to remove a book with dangerous ideas from the public library: the questions pertain to books against religion, for Communism, for homosexuality, for militarism (doing away with elections), for racism (touting genetic inferiority of Blacks), and for socialism (government ownership of property). Obviously, these questions are part of the Civil Liberties component of the GSS which also includes other questions about tolerance and freedom of speech. In the searchable bibliography on the GSS-DIRS site, I found that Howard D. White analyzed these and related questions such as whether birth control information should be made available to teenagers in a Library Journal article (White 1986).

The GSS also has an international component--the International Social Survey Programme (ISSP), offering valuable cross-country replication of questions, for the following countries: Britain, Bulgaria, Canada, the Czech Republic, Germany, Hungary, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, the Philippines, Poland, Russia, Slovenia, Spain, and the United States. Recent modules of the ISSP are Family/Sex Roles--1994, Environment--1993, Social Inequality--1992, Religion--1991, and Role of Government--1990 (Davis 1996).

ANES

"The American National Election Studies (ANES) are national surveys carried out by the Survey Research Center (SRC) and by the Center for Political Studies (CPS) of the Institute for Social Research at the University of Michigan. This time-series collection has been fielded continuously since 1948, presenting data on Americans' social backgrounds, enduring political predispositions, social and political values, perceptions and evaluations of groups and candidates, opinions on questions of public policy, and participation in political life," (Inter-university Consortium for Political and Social Research 1996b). Like the GSS, the ANES is also available on the web with complete documentation and interactive data extraction capability. For browsing portions of the ANES results in online format, see the NES Guide to Public Opinion and Electoral Behavior (National Election Studies 1996).

The ANES is to political scientists what the GSS is to sociologists--a mountain of individual-level data with a solid design approved by experts in the field, replicable over time. If the CPS can inform LIS about changes in the labor market affecting library use, and the GSS can inform the field about cultural preferences and attitudes by groups of potential library users, than the ANES can tell us something about voting, public opinion,

and political participation in a mass democratic society. The ANES and other surveys, such as James S. Jackson's National Black Election Panel Study, also provide demographic breakdowns about who votes, reads newspapers or magazines, and watches political campaigns on television or listens to them on the radio.

The Institute for Research in Social Science at the University of North Carolina-Chapel Hill is the repository for Louis Harris poll data as well as many state polls. The IRSS provides a unique Internet resource called the Public Opinion Poll Item Index, which allows question-level keyword searches across surveys and years (IRSS 1996). While the raw data is not freely available, it can be useful as well as interesting to view the frequencies of the answers to the questions. The resource can also be used in designing questionnaires to see how similar questions have been worded by pollsters.

A quick search on the stem-word 'library' does yield some interesting items. A University of Georgia poll asked a state sample "whether or not you know where the public library nearest to your home is located?" (88% said yes). An Indiana University poll asked "In the past year, have you used the services of your public library?" (60% said yes.) A University of Maryland poll asked "When did you last visit a Maryland public library?" (45% said in the last month, 22% in the last 6 months, 8% in the last year, and 24% more than a year). Other questions found asked about quality of service, types of services used, whether their local library had non-book items to lend, and whether they used inter-library loan. Several poll questions affirmed that a strong majority of people supported increased local or state taxes to fund libraries. On the other hand, in a 1994 Rutgers University poll asking, "Please tell me whether you favor or oppose having your town share that service [public libraries] with a neighboring town if it held down your property taxes," 82% said they would favor this proposal (IRSS 1996).

One fascinating private research project for LIS is currently in progress at Carnegie Mellon University. The HomeNet Project is an experiment of Internet use by over 100 Pittsburgh households who were given "computer equipment, subsidized access to the Internet and training in using both their computers and the Internet," (The HomeNet Project 1996). The study, begun in February 1995 and expected to last three years, claims that "through detailed, ongoing questionnaires and electronic data collection, Internet usage and its effects on participants' lives can be studied and analyzed in unprecedented detail," (Kraut et al. 1995).

Conclusions from a preliminary report published on the World Wide Web are: "When ordinary people are given access to the Internet from home, half of them use it regularly after 5 months. Teenagers are central to Internet use at home. . . Though studies show that high-income, educated white males dominate the Internet, the HomeNet study shows that once financial barriers are lowered, lower income and less well educated people are as likely to become enthusiasts. Race and gender, however, remain associated with Internet usage, perhaps because the Internet's mainly white, male users has created a resource environment most attractive to men and whites. . . People gravitate toward their idiosyncratic interests. Since most individuals will be interested in only a few of the thousands of services offered them, they need easy ways to mark their information space to reflect their personal interests. . ." (The HomeNet Project 1996)

Conclusion

The studies presented here merely scratch the surface of public-use social science datasets, and yet they may leave LIS researchers wishing for more. While Internet technology brings more and more information to the scholar's desktop, locating variable- specific information about public-use datasets is never simple. Currently there are many academic compilations of World Wide Web links that can provide knowledgeable intermediary guidance, though of course no one site 'has it all.' One such compilation is the Internet Crossroads in the Social Sciences site maintained by the Data and Program Library Service at the University of Wisconsin-Madison (Data and Program Library Service 1996b).

For scholars with an affiliation with a research University, the Inter-university Consortium for Political and Social Research is the largest data archive in the United States, in which all datasets are freely accessible to institutional members. One simply needs to find the Official Representative for their campus---sometimes a librarian, sometimes a faculty member. The ICPSR has a well-designed searchable web site with broad subject headings and abstracts as well as file-specific information (ICPSR 1996c).

Another way to increase the pool of secondary sources for LIS researchers is to begin to yield influence over data producers on the survey questions and design. The GSS in particular solicits input from scholars across the social sciences. Smaller surveys such as Wisconsin Opinions, a monthly core telephone survey of Wisconsin households begun in 1992 conducted by the Wisconsin Survey Research Laboratory, solicits questions to be added in the monthly polls by organizations, interest groups, or individuals. Other state survey organizations may have similar policies. Certainly LIS researchers might express their needs for both institutional and individual-level statistics to the NCES, which conducts so many important studies for LIS. The arena of networked information in particular may open up more dialogue among LIS researchers and sociologists, anthropologists, political scientists, economists, etc., as each discipline grapples with the societal changes brought on by a networked environment, hopefully in a collaborative and inter-disciplinary way

Reference List

American Library Association. 1996. LARC Fact Sheets. Chicago, Illinois: American Library Association. Internet URL: http://www.ala.org/library/larcfact.html.

Association of Research Libraries. 1995. 1992-95 ARL Statistics. Washington, D.C.: Association of Research Libraries. Internet URL: http://viva.lib.virginia.edu/socsci/arl/1994/95doc.html.

Association of Research Libraries. 1996. CLR Grant to ARL. Washington, D.C.: Association of Research Libraries. Internet URL: http://arl.cni.org/stats/Statistics/arlstat/clrgrant96.html.

Bertot, John Carlo, Charles McClure, and Douglas L. Zweizig. 1996. The 1996 national survey of public libraries and the Internet. Washington, D. C.: National Commission on Libraries and Information Science. Internet URL: http://istweb.syr.edu/Project/Faculty

/McClure- NSPL96/NSPL96_T.html.

Data and Program Library Service. 1996a. NLSY96. DPLS News, March 1996. March. Madison, WI: University of Wisconsin-Madison. Internet URL: http://dpls.dacc.wisc.edu /pubs/mar96news.html.

Data and Program Library Service. 1996b. NLSY96. DPLS--Internet Crossroads. Madison, WI: University of Wisconsin-Madison. Internet URL: http://dpls.dacc.wisc.edu /internet.html.

Davis, Celestine and Bill Sonnenberg. [eds] 1995. Programs and plans of the National Center for Education Statistics, 1995 edition. Washington, D.C.: U. S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics.

Davis, James A. and Tom W. Smith. 1996. NORC--General Social Survey. Chicago, Illinois: National Opinion Research Center. Internet URL: http://www.norc.uchicago.edu /gss.htm.

Fahy, Terry W. [ed.] 1995. The National longitudinal surveys of labor market experience: an annotated bibliography of research, 1968-1995 edition. Columbus, Ohio: Center for Human Resource Research.

The HomeNet Project. 1996. Welcome to HomeNet: studying Internet family use. Pittsburgh, Pennsylvania: Carnegie Mellon University. Internet URL: http://homenet.andrew.cmu.edu/progress/index.html.

Institute for Research in Social Science. 1996. Public Opinion Poll Item Index. [Online database] Chapel Hill, N.C.: University of North Carolina. Internet URL: http://www.irss.unc.edu:80/data_archive/pollsearch.html.

Inter-university Consortium for Political and Social Research. 1996a. General Social Surveys-- Introduction. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Internet URL: http://www.icpsr.umich.edu/gss/about/gss/gssintro.htm.

Inter-university Consortium for Political and Social Research. 1996b. The National Election Studies. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Internet URL: http://www.icpsr.umich.edu/NES/nutshell.html.

Inter-university Consortium for Political and Social Research. 1996c. Archival holdings. [online database]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Internet URL: http://www.icpsr.umich.edu/archive1.html.

Kraut, Robert, William Scherlis, Tridas Mukhopadhyay, Jane Manning and Sara Kiesler. 1995. HomeNet Paper. September 1995 Newsletter. Pittsburgh, Pennsylvania: Carnegie Mellon University. Internet URL: http://homenet.andrew.cmu.edu/progress/report2.html.

Library of Congress. 1996. THOMAS: Legislative Information on the Internet. [Online database]. Washington, D. C.: Library of Congress. Internet URL: http://thomas.loc.gov/.

National Center for Education Statistics. 1996. The National Data Resource Center. [pamphlet] Washington, D.C.: U. S. Department of Education. Office of Educational Research and Improvement.

National Center for Education Statistics. 1995. Digest of education statistics, 1995. Washington, D.C.: U. S. Department of Education. Office of Educational Research and Improvement.

National Election Studies. 1996. NES guide to public opinion and electoral behavior. Ann Arbor, Michigan: National Election Studies. Internet URL: http://www.umich.edu /~nes/resourcs/nesguide/gd-index.htm.

Ohio State University. Center for Human Resource Research. 1995. National longitudinal surveys of labor market experience, youth cohort: 1979-1993. [machine-readable data file] Columbus, OH: Ohio State University. Center for Human Resource Research. [distributor]

Special Library Association. 1996. SLA research agenda. Washington, D.C.: Special Libraries Association. Internet URL: http://www.sla.org/research/research_age.html.

U. S. Department of Commerce. Bureau of the Census. 1996. 1990 Census Lookup (1.4). Washington, D.C.: U. S. Department of Commerce. Bureau of the Census. Internet URL: http://venus.census.gov/cdrom/lookup/.

U. S. Department of Labor. Bureau of Labor Statistics. 1996. Labor force statistics from the current population survey home page. Washington, D. C.: U. S. Department of Labor. Bureau of Labor Statistics. Internet URL: http://www.bls.gov/cpshome.htm.

White, Howard D. 1986. Majorities for censorship. Library Journal 111 (July): 31-38.

Robin C. Rice
rice@dpls.dacc.wisc.edu

To subscribe to LIBRES send e-mail message to listproc@info.curtin.edu.au with the text: subscribe libres [your first name] [your last name]

Return to [Libres Home Page](#)