

Archiving Social Science Data: A Collaborative Process

Tom Zelenock

Kaye Marz

Inter-university Consortium for Political and Social Research

When a dataset arrives at ICPSR, its "journey" — which began at the point of data collection and which will end on a researcher's desktop — is nearly over. Yet many important tasks remain. During the ICPSR "leg" of the journey, the dataset will be prepared for public release and archived in perpetuity, thus benefiting the entire social science research community, now and into the future.

In this article, we describe some of the steps involved in the archiving process as ICPSR makes data more useful to the research community. Individuals interested in depositing data with ICPSR will find some helpful guidelines on preparing data for archiving, which, when followed, will ultimately help to expedite data release.

ICPSR's Role

Archiving involves much more than simply making copies of data. ICPSR staff examine each dataset from the viewpoint of a potential user and then apply both social science knowledge and technical expertise to add value to and enhance the usefulness of the dataset for secondary analysis.

Before ICPSR makes a dataset available for distribution, a number of steps are taken to ensure that it will meet users' needs. ICPSR staff perform several critical tasks:

- Evaluate the data collection for completeness, suitability for public release, and readiness for use
- Back up and store both the original data and the processed data at an external site
- Determine whether any problems of respondent confidentiality exist, checking for problems arising either from direct or indirect identification
- Assess whether the documentation is complete and accurate
- Test the technical characteristics of the documentation against the data and make sure data and documentation match
- Prepare finding aids, including searchable study descriptions and bibliographic citations, to assist in locating the collection within ICPSR's archive
- Consult with the principal investigator (PI)/data producer, when necessary, to remedy any problems uncovered during the review of the data
- Convert paper documentation into electronic documentation
- Create SAS and SPSS data definition statements to facilitate analysis
- Standardize missing data codes
- Reformat data to achieve more efficient transmission and storage
- Standardize coding schemes across elements of a study where they are different
- Check observed frequencies against reported frequencies
- Check for consistency of survey responses and skip patterns
- Check data against documentation for completeness, wild codes, and missing codes
- Correct data in consultation with the PI/data producer when errors are found

Sometimes these processes take longer than anticipated — particularly in cases where there are problems with a dataset — and misunderstandings may arise. There will always be some users who expect instant turnaround, and ICPSR does strive to release data in as timely a manner as possible.

But in reality, these important archival operations do take time if they are per-

formed carefully, with attention to detail. The volume of data that make their way through ICPSR must also be taken into account: in fiscal year 1996 alone, ICPSR acquired and released 171 new studies, consisting of over 4,000 discrete data files.

Guidelines for Data Deposits

To aid researchers and data producers in submitting data for archiving, ICPSR has published a booklet entitled *Guide to Social Science Data Preparation and Archiving*. The *Guide* includes a Data Deposit Form, which must be completed and signed before ICPSR can release a data collection. Filling out the form in a thorough and accurate manner can greatly help staff to ascertain the scope and contents of a collection.

Available in print and in electronic form on the ICPSR Website (<http://www.icpsr.umich.edu/ICPSR/Archive/Deposit>), these documents provide useful information on how to prepare data for deposit in a public archive and how to meet ICPSR's specific deposit requirements. ICPSR recommends that individuals involved in data collection read these materials at the beginning of their projects for advice on good archival practices.

Transmittal Media

ICPSR receives data on several types of media: tapes, diskettes, CD-ROMs, FTP, even via e-mail. *ICPSR encourages data depositors to send data electronically via FTP* (to initiate this process, send email to deposit@icpsr.umich.edu) but other media are also acceptable.

File Formats

Similarly, data are currently sent to ICPSR in a variety of formats: raw ASCII data, SAS transport files and datasets, SPSS portable (export) and system files, dBase, Paradox, etc. Because of the time involved in extracting raw data

from some of these types of files, *ICPSR recommends that data be sent in ASCII form or as SAS transport or SPSS portable files*. Receiving these file formats speeds throughput time considerably. Note that, if at all possible, data files should be accompanied by complete variable labels, value labels, and missing data specifications.

Documentation

Documentation supplied to ICPSR is examined thoroughly by the staff for completeness and consistency. The most basic pieces of documentation that staff look for initially are items relating to the data collection as a whole: number of cases, number of variables, number of data files, record length, data structure (flat files or hierarchical), and how multiple files are linked. Without this information, it is not possible for ICPSR to know if it has received the entire dataset from the data depositor. Recently, a study was sent to ICPSR with documentation indicating that the data were structured as a hierarchical file with seven record types. After reviewing the data, only three record types were found. A discussion with the PI resulted in a resupply of both the data and the documentation.

Documentation of any sort should be sent to ICPSR in electronic format — either in ASCII, WordPerfect or Microsoft Word. Receiving codebooks, reports, data collection instruments, and other documentation in only hardcopy format greatly increases the amount of time to process a study. Hardcopy documentation in most cases will be scanned — a significant duplication of effort since a computer-readable file almost always exists.

Complete copies of reports, if produced, can be helpful to ICPSR staff. Reports are reviewed to confirm case counts, frequencies, and types of data collected and to assist in the preparation of accurate study descriptions. We also refer data users to published arti-

cles. In one case, after reading a journal article sent with the data, staff realized that an entire file of data was missing from a recent supply of data.

Whenever possible, ICPSR tries to release *data collection instruments* with a dataset. To users of the data, instruments provide detail about question wording, content, and skip patterns that cannot be completely reflected in the variable labels and codes. To ICPSR staff, survey instruments are also valuable in resolving questions about how variables are coded.

The Archival Process: A Brief Chronology

Data Acquisition

When ICPSR acquires a dataset, staff check first to make sure that the data shipment is complete and that the data depositor has supplied readable data. Every staff member has stories of data tapes that were blank, diskettes that were corrupted, or files that were FTP'd incorrectly. Technology has made it easier to transfer data, but it is still necessary to inspect files thoroughly upon their arrival.

Data Processing

Confidential information

Staff review all studies sent to ICPSR for the presence of confidential information. *ICPSR strongly recommends that data producers remove all respondent identifiers prior to deposit of their data collections*. Responsible archiving practice demands that ICPSR not release any information that could pose a threat to the anonymity of an individual. ICPSR takes this responsibility extremely seriously.

Sometimes ICPSR receives data that contain direct identifiers such as names, phone numbers, Social Security numbers, or addresses. In these

cases, direct identifiers are always blanked or deleted to safeguard privacy before releasing the data to the public.

Dealing with indirect identifiers is more problematic. Analyses of subgroups or crosstabulations of ascriptive characteristics may result in small groups of individuals who may be identifiable when other geographic and/or demographic information is brought to bear. Or staff may determine that using other data sources in combination with a given dataset may result in identification of individuals.

Recoding data to eliminate indirect identifiers may include converting dates to time intervals; this allows for time lapse analyses without providing exact dates that might permit identification of individuals. Dates of birth are often converted to less specific ages. Variables such as income may be converted to categories.

Recoding data to preserve the confidentiality of individuals is a necessary, but time-consuming, task, and unfortunately in some cases it reduces the amount of information available for analysis. ICPSR staff work closely with PIs on confidentiality issues since they are most familiar with the data. Again, the best solution is for the PI to determine prior to data deposit whether the confidentiality of individuals is being preserved and, if not, to take necessary action.

Study descriptions

Another step in the processing phase is the preparation of accurate and comprehensive study descriptions and bibliographic citations, which will appear in the codebook for each study as well as on the ICPSR Website and in various publications. Using the documentation provided, staff write summaries that include descriptions of the study design, sampling procedures, goals of the research, and technical characteristics.

Electronic document conversion

Most hardcopy technical documentation received by ICPSR is now routinely being converted to electronic form, which adds another step to the processing phase. ICPSR's Electronic Document Conversion Team has achieved many efficiencies by adopting Adobe Acrobat's Portable Document Format (PDF) as the default for most scanned documents. Conversion to PDF is much faster than traditional OCR scanning and also has the advantage of preserving a document's original formatting.

SAS and SPSS data definition statements

ICPSR staff often create SAS and SPSS setups to facilitate use of the data by researchers. Depending on the materials received with a data collection, this can be a trivial process or a lengthy, involved one. SAS and SPSS setups provide column locations and formats for each variable in a data file, variable and value labels, and missing value declarations. The setups can be modified by the user to select only variables of interest or for use with other statistical packages.

Preparing an "ICPSR dataset"

When resources permit, ICPSR processes a collection more intensively. An "ICPSR Dataset," (also called an ICPSR edition, formerly Class I or Class II data) typically contains column-delimited ASCII data files, codebooks and data collection instruments in electronic format, and SAS and SPSS data definition statements. All variables are documented with descriptive labels, codes are labeled, system missing values are recoded and labeled, variables are formatted to a minimum width, and values are reviewed for appropriateness. In terms of documentation, this means that the research itself is well summarized, specifically the

methods and sampling, and that any other information necessary to describe the data are included with the codebook.

Variable labels and value labels.

Whether data are sent as SAS datasets, SPSS portable files, or raw data files, often the variable labels and code descriptors are missing or incomplete. To the original researcher, the variable names and coding scheme may be obvious, but to a new user, more descriptive labeling is necessary. A fundamental responsibility of ICPSR processing staff is to make certain that all variables are thoroughly identified and labeled. When variables are not thoroughly described, staff consult the documentation, questionnaires, and the PI to determine how to identify them.

System missing data and missing values. Data are often supplied to ICPSR with blanks or system missing values. This is not necessarily a problem but can result in ambiguity. Does system missing mean "No Response," "Not Applicable," or "Out of Universe"? Is it a catch-all for every category of missing data? Occasionally data are sent containing blanks that have a substantive meaning. Where this is the case, staff typically recode the blanks to a value and label them appropriately. System missing data without substantive meaning are usually recoded to a series of "9s" and labeled "Unknown" or "Missing".

Reformatting variables for best fit.

Data files are frequently sent with default formats much larger than necessary. ICPSR reformats data files to the smallest possible size for optimum speed in file transfer over the WWW. While reformatting a few variables is a simple process, reformatting thousands is time-consuming; this is another area in which data producers can be helpful by reformatting data before deposit.

Undocumented codes. Reviewing frequencies is the primary way that staff

determine that all variables have descriptive labels, that all categorical variables have code labels, that missing data have been standardized, and that there are no undocumented codes.

Undocumented codes take a great deal of time to investigate. When undocumented codes are identified, the processor will generally refer to the documentation again, but it is almost always the case that additional contact with the PI is necessary. Then the PI must determine whether a new code was created or whether a data entry error occurred.

Investigating undocumented codes, although time-consuming, is relatively easy for categorical variables. Continuous variables, such as salary amounts or ages, are more difficult to assess for proper coding because of extreme values. Is a value three standard deviations beyond the mean really valid, or is it a coding error? Are data that indicate a murder victim is 99 years old accurate, or is 99 an undocumented code for "Not Applicable?" A list of all the variables and their valid values is extremely helpful. Continuous variables can be listed with their valid ranges, categorical variables with their valid codes and labels. Any reserved values (e.g., 98 — Not Applicable) should be listed as well.

Skip patterns and consistency checks. Checks on skip patterns can only be performed when the variables have separate codes distinguishing between "Not applicable" and general missing data. ICPSR staff review subsequent variables for consistency of the response patterns given the coding of the filter variable. Other consistency checks include identifying unlikely relationships among variables (e.g., children older than parents), or respondents outside of the sampling frame (e.g., juveniles included in a study of retirees).

Data Preservation

ICPSR staff copy all data onto disk for distribution, and two backup copies are made on magnetic tape. Copies are checked to assure that errors were not introduced during the copy process, and staff update databases that track the holdings of the archive. When all copies of data and documentation have been verified, the data are ready for release.

Copies of each dataset (both the original data and the final product, if it is altered in any way for release) are stored offsite. This step is essential: were a fire or other disaster to destroy ICPSR's computers, the copies stored offsite would still remain.

In keeping with its mission to the social sciences, ICPSR preserves data for future generations of researchers and strives to communicate the archiving ethic to the larger social science community.

Data Release

After all processing is completed and final versions backed up, data are made available on the Website for direct downloading and CDNet retrieval. The ICPSR Hotline list on the Website informs users of new data releases, as do the Recent Additions and Recent Updates facilities, which provide access to all data released over the past 90 days. The *ICPSR Bulletin*, available in print and electronic format, offers another list of new data releases.

Summary and Conclusion

By the time ICPSR receives a data collection, it is often considered "finished" by the researcher depositing it. Contacting the PI to resolve questions may seem like a simple, quick exercise, but it often is not. Questions are

difficult to answer because materials may have been sent to storage, research assistants may have moved on, or the PI may have started new projects. For all these reasons, ICPSR recommends that data producers and depositors take time up front to assemble complete and accurate information for transfer to ICPSR for archiving.

ICPSR receives data in many formats, and we attempt to make these data available in a consistent format along with complete computer-readable documentation. PIs, data producers, and data depositors can help to expedite ICPSR's archiving tasks by making use of the *Guide to Social Science Data Preparation and Archiving* and by completing the Data Deposit Form. Providing data in recommended formats and providing electronic documentation save considerable time. Documentation containing variables that are completely described with labels and valid codes or ranges is helpful. Removing undocumented codes from the data prior to sending it to ICPSR results in fewer questions and consultations. Issues of confidentiality are also best addressed early on by the PI.

ICPSR is reviewing its internal procedures to become more efficient in preparing datasets by upgrading computing facilities, revising staff training, and developing new tools for processing. Recommendations of the Archival Development Committee of the ICPSR Council may also provide useful guidelines for archiving and processing, including the specification of minimum technical standards for data being deposited.

Archiving and disseminating data is a cooperative effort, and individuals depositing data with ICPSR can help by adopting the suggestions listed above; working together, we can make the last leg of a dataset's journey much shorter. □

Tom Zelenock and Kaye Marz are Research Associates with the National Archive of Criminal Justice Data at ICPSR.

To access additional information about the data collections below, please consult the ICPSR Website at <http://www.icpsr.umich.edu>.

- ABC News Abuse and Excuse Poll, February 1994 — ABC News (ICPSR 6619)
- ABC News Bosnia Attack Poll, February 1994 — ABC News (ICPSR 6620)
- ABC News "Nightline" Clinton Somalia Poll, October 1993 — ABC News (ICPSR 6290)
- ABC News "Nightline" Congressional Budget Poll, July 1993 — ABC News (ICPSR 6282)
- ABC News "Nightline" Haiti Poll, October 1993 — ABC News (ICPSR 6291)
- ABC News "Nightline" Somalia Poll, October 1993 — ABC News (ICPSR 6289)
- ABC News Perot Poll, November 1993 — ABC News (ICPSR 6295)
- ABC News Poll, September 1993 — ABC News (ICPSR 6285)
- ABC News Viewpoint Poll, April 1994 — ABC News (ICPSR 6624)
- ABC News Whitewater Poll, March 1994 — ABC News (ICPSR 6621)
- ABC News Whitewater Revisited Poll, April 1994 — ABC News (ICPSR 6623)
- ABC News/NHK G7 Summit Poll, June 1993 — ABC News/NHK (ICPSR 6281)
- ABC News/Washington Post Poll, May 1993 — ABC News/Washington Post (ICPSR 6280)
- ABC News/Washington Post Poll, August 1993 — ABC News/Washington Post (ICPSR 6283)
- ABC News/Washington Post Poll, November 1993 — ABC News/Washington Post (ICPSR 6292)
- ABC News/Washington Post Poll, January 1994 — ABC News/Washington Post (ICPSR 6617)
- ABC News/Washington Post Poll, February 1994 — ABC News/Washington Post (ICPSR 6618)
- ABC News/Washington Post Poll, March 1994 — ABC News/Washington Post (ICPSR 6622)
- ABC News/Washington Post Poll, May 1994 — ABC News/Washington Post (ICPSR 6625)
- African Coup Events Data, 1986–1990 — T.Y. Wang (ICPSR 6869)
- Aging of Veterans of the Union Army (Version M-2): Military, Pension, and Medical Records for Veterans of the Ohio and Pennsylvania Regiments, 1820–1940 — Robert W. Fogel et al. (ICPSR 6837)
- American National Election Study, 1996: Pre- and Post-Election Survey — Steven J. Rosenstone, Warren E. Miller, Donald R. Kinder, and the National Election Studies (ICPSR 6896)
- Census of Population and Housing, 1970 [California]: Summary Statistic File 4A—Population and Housing [Fourth Count] — United States Department of Commerce. Bureau of the Census (ICPSR 6712)
- Comparative Analysis of Small and Large Group Health Care Utilization and Costs, 1988–1990: [Western Pennsylvania] — Wanda W. Young and Neil Hollander (ICPSR 6859)
- Current Population Survey, February 1996: Displaced Workers — United States Department of Commerce. Bureau of the Census (ICPSR 6879)
- Dynamics of Population Aging in Economic Commission for Europe (ECE) Countries, Census Microdata Samples: Czech Republic, 1991 — United Nations Economic Commission for Europe. Population Activities Unit (ICPSR 6857)
- Eurobarometer 44.0: Cancer, Education Issues, and the Single European Currency, October–November 1995 — Karlheinz Reif and Eric Marlier (ICPSR 6721)
- Families of Missing Children: Psychological Consequences and Promising Interventions in the United States, 1989–1991 — Chris Hatcher (ICPSR 6140)
- General Social Surveys, 1972–1996: [Cumulative File] — James A. Davis and Tom W. Smith (ICPSR 6870)
- National Corrections Reporting Program, 1993: [United States] — United States Department of Justice. Bureau of Justice Statistics (ICPSR 6823)

- National Household Education Survey, 1993** — United States Department of Education. National Center for Education Statistics (ICPSR 6877)
- National Household Survey on Drug Abuse, 1992** — United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Office of Applied Studies (ICPSR 6887)
- National Household Survey on Drug Abuse, 1993** — United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration (ICPSR 6852)
- National Medical Expenditure Survey, 1987: Data From the Household Survey, Health Insurance Plans Survey, Survey of American Indians and Alaska Natives, and Institutional Population Component [Research File 40R]** — United States Department of Health and Human Services. Agency for Health Care Policy and Research (ICPSR 6868)
- National Survey of the Japanese Elderly, 1987** — Jersey Liang and Daisaku Maeda (ICPSR 6842)
- Parenthood in Early Twentieth-Century America Project (PETCAP), 1900–1944** — Ralph LaRossa (ICPSR 6876)
- Survey of Income and Program Participation (SIPP) 1993 Panel** — United States Department of Commerce. Bureau of the Census (ICPSR 6886)
- Washington Post Virginia Governor Poll, September 1993** — *Washington Post* (ICPSR 6286)
- Washington Post Virginia Governor Poll, October 1993** — *Washington Post* (ICPSR 6288)
- Washington Post Virginia Voters Poll, January 1994** — *Washington Post* (ICPSR 6616)

Revisions/Updates

- A Sentencing Postscript: Felony Probationers Under Supervision in the Community, 1983** — Mark A. Cuniff and the National Association of Criminal Justice Planners (ICPSR 8647)
- Drug Use Forecasting in 24 Cities in the United States, 1987–1995** — United States Department of Justice. National Institute of Justice (ICPSR 9477)
- Euro-Barometer 39.0: European Community Policies and Family Life, March–April 1993** — Karlheinz Reif and Anna Melich (ICPSR 6195)
- Expanded United States Supreme Court Judicial Database, 1946–1968 Terms** — Harold J. Spaeth (ICPSR 6557)
- Expenditure and Employment Data for the Criminal Justice System [United States]: Extract File, 1983** — United States Department of Justice. Bureau of Justice Statistics (ICPSR 8455)
- Federal Court Cases: Integrated Data Base, 1970–1995** — Federal Judicial Center (ICPSR 8429)
- National Corrections Reporting Program, 1983: [United States]** — United States Department of Justice. Bureau of Justice Statistics (ICPSR 8363)
- National Crime Surveys Extract: Personal Crime Longitudinal Files, 1976–1982** — Christopher A. Innes (ICPSR 8315)
- National Crime Surveys: National Sample, 1973–1983** — United States Department of Justice. Bureau of Justice Statistics (ICPSR 7635)
- National Hospital Ambulatory Medical Care Survey, 1994** — United States Department of Health and Human Services. National Center for Health Statistics (ICPSR 6824)
- National Jail Census, 1983** — United States Department of Justice. Bureau of Justice Statistics (ICPSR 8203)

- National Judicial Reporting Program, 1988: [United States]** — United States Department of Justice. Bureau of Justice Statistics (ICPSR 9449)
- National Judicial Reporting Program, 1992: [United States]** — United States Department of Justice. Bureau of Justice Statistics (ICPSR 6509)
- Offender Based Transaction Statistics (OBTS), 1980: California, Ohio, New York, Pennsylvania, Utah** — United States Department of Justice. Bureau of Justice Statistics (ICPSR 8248)
- Offender Based Transaction Statistics (OBTS), 1981: California, Ohio, Pennsylvania, Utah, Minnesota, New York, Virginia** — United States Department of Justice. Bureau of Justice Statistics (ICPSR 8277)
- Offender Based Transaction Statistics (OBTS), 1982: California, Ohio, Pennsylvania, Utah, Minnesota, New York, Virgin Islands, Virginia** — United States Department of Justice. Bureau of Justice Statistics (ICPSR 8408)
- Offender Based Transaction Statistics (OBTS), 1984: Alaska, California, Delaware, Georgia, Minnesota, Missouri, Nebraska, New York, Ohio, Pennsylvania, and Virginia** — United States Department of Justice. Bureau of Justice Statistics (ICPSR 8675)
- Penal Code Citations: Sentencing in 18 American Felony Courts, 1983** — Mark A. Cuniff (ICPSR 8396)
- Sentencing Outcomes in 28 Felony Courts, 1985 [United States]** — Mark Cuniff (ICPSR 8708)
- Uniform Crime Reporting Program Data: [United States]** — United States Department of Justice. Federal Bureau of Investigation (ICPSR 9028)
- Uniform Crime Reporting Program Data [United States]: County-Level Detailed Arrest and Offense Data, 1994** — United States Department of Justice. Federal Bureau of Investigation (ICPSR 6669)
- Uniform Crime Reports, 1958–1969, and County and City Data Books, 1962, 1967, 1972: Merged Data** — United States Department of Justice. Federal Bureau of Investigation, and United States Department of Commerce. Bureau of the Census (ICPSR 7715)
- United States Supreme Court Judicial Database, 1953–1995 Terms** — Harold J. Spaeth (ICPSR 9422)
- Urban Poverty and Family Life Survey of Chicago, 1987** — William Julius Wilson et al. (ICPSR 6258)

Publication Archive Additions

- National Black Election Study Pilot, 1996** — Katherine Tate (ICPSR 1125)
- Strategic Timing of Position-Taking in Congress: A Study of the North American Free Trade Agreement** — Janet M. Box-Steffensmeier, Laura W. Arnold, and Christopher J.W. Zorn (ICPSR 1126)

Biennial Meeting of Official Representatives

**October 23–26, 1997
Ann Arbor, Michigan**

"ICPSR and the Research Community: Evolution and Revolution"

Please join us as ICPSR's Official Representatives, Council members, and staff convene to exchange information on timely topics relating to the rapidly evolving research environment. Tentative sessions include:

- Accessing ICPSR's Data Resources: WWW, CDNet, CD-ROM (Preconference workshop)
- Orientation Session for New ORs
- New ICPSR CD-ROM on Minorities Data
- A Conversation With the Executive Director, Staff, and Council
- The Role of the OR in the 21st Century
- Using the Internet for Research
- Documentation Formats and Products
- ICPSR Data in the Classroom
- Innovative Research Methodologies
- Discussion of Proposed Changes in ICPSR's Archiving Priorities

Program committee members include members of the ICPSR Member Relations Committee: **John Garcia**, University of Arizona; **Warren Miller**, Arizona State University; **Samuel Patterson**, Ohio State University; **Wendy Watkins**, Carleton University; and ICPSR staff members **Zack Allen**, **Pamela Schwarzmman**, **Mary Vardigan**, and **Hank Heitowit**.

All sessions for the 1997 meeting, as well as breakfasts and lunches, will be held in the Michigan Union on the University of Michigan campus; Saturday night's banquet dinner will take place at the Michigan League.

Note that registration for the meeting will be handled electronically this year; the online registration form, along with other information about the meeting, will be available on the ICPSR Website at <http://www.icpsr.umich.edu>

Council Member Resigns

Ross Matsueda, University of Iowa, has resigned from Council due to workload and illness. A replacement will not be designated since elections to Council will take place in the fall.

New Members Join ICPSR

Since late 1996, ICPSR has welcomed several institutions into the membership. We are pleased that faculty and staff at these schools now have access to ICPSR's rich data resources and educational opportunities.

New members include:

- National Chung Cheng University
- Rush University
- University of Basle (joining the Swiss National Membership)
- Hong Kong Baptist University
- Budapest University of Economic Sciences
- Bucknell University

The ICPSR staff, Council, and Consortium members extend a sincere welcome to the new member schools and ORs and look forward to working with them in the future.

ICPSR Implements New File Management Scheme

ICPSR's Computing and Network Services has made some recent enhancements to the data management system underlying ICPSR's servicing activities.

Under the new scheme in CDNet, all supporting files associated with a data file (e.g., codebooks and data definition statements) are grouped under the same part number. In terms of ordering data, this means that a data part now encompasses all of the files associated with it; an order for a data part will result in the receipt of all necessary information to make use of that file.

Users may have noticed this change reflected in more recent study descriptions (called "abstracts" on the Website) or in the front matter for codebooks; the descriptions no longer include separate part numbers for supporting files.

1997 Summer Training Program Information Available

ICPSR's Website now offers information describing the 1997 Summer Training Program. An electronic application form is also online for prospective applicants to download and complete.

Access to instructors' syllabi and course descriptions from the *1996 Summer Program Bibliography* is also possible through the Website. Users can select any of the 45 course titles to obtain course-specific information presented by instructors from last year's Program.

Mailing Sent to ORs

The *1995-1996 Annual Report* and Appendix was mailed to ORs in March.

Forthcoming...

As we go to press, we anticipate that the data collections listed below will be available in the coming weeks.

- National Assessment of Gangs in Correctional Facilities, 1992
- National Household Survey on Drug Abuse, 1985
- National Health Interview Survey, 1994: AIDS Knowledge and Attitudes Supplement
- Robert Wood Johnson Foundation Family Health Insurance Survey, 1993

ICPSR Summer Program, 1997 (June 23–August 15)

First Session

(June 23–July 18)

Lectures

Mathematics for Social Scientists I
Mathematics for Social Scientists II
Introduction to Computing
Advanced Topics in Social Research*

Workshops

Quantitative Historical Analysis
Introduction to Statistics and Data Analysis I
Mathematical Models: Game Theory
Introduction to Regression Analysis
Regression Analysis
Multivariate Statistical Methods
Scaling and Dimensional Analysis
Maximum Likelihood Estimation I
GIS/Spatial Data Analysis

One-Week Workshops

Network Analysis (June 16–20)
Criminal Justice Methodology and Analysis
(June 23–27)
Hierarchical Linear Models (July 7–11)
Use of Multiple Datasets for the Study of Aging
(June 2–6)
"LISREL" Models: Introduction (July 21–25)
"LISREL" Models: Intermediate (July 28–August 1)
Categorical Data Analysis: Introduction (June 9–13)
Categorical Data Analysis: Advanced Topics
(June 16–20)
Research on Mental Health Services for Youths
(July 28–August 1)
Modeling Response Data: Surveys, Attitudes, Testing
(August 11–15)

Second Session

(July 21–August 15)

Lectures

Nonlinear Systems
Introduction to Computing
Matrix Algebra
Dynamic and Longitudinal Analysis
Advanced Topics in Social Research*

Workshops

Simultaneous Equation Models
Regression Analysis
Time Series Analysis
Mathematical Models: Rational Choice
Introduction to Statistics and Data Analysis II
Categorical Analysis
Advanced Maximum Likelihood Estimation
"LISREL" Models: General Structural Equations
Advanced Analysis of Variance
Quantitative Analysis of Crime and Criminal Justice
Quantitative Analysis on Latin America

*Advanced Topics

Resampling Techniques: Jackknife and Bootstrap
Graphical Presentation and Analysis of Data
Missing Data Analysis
Data Visualization
Bayesian Modeling
Measurement in Social Science

For a copy of the 1997 ICPSR Summer Program brochure and application, contact: ICPSR Summer Program, P.O. Box 1248, Ann Arbor, MI 48106-1248, Phone: (313) 764-8392, E-mail: sumprog@icpsr.umich.edu, or consult the ICPSR Website at <http://www.icpsr.umich.edu>.

ICPSR's Internet Resources

ICPSR Website

ICPSR's site on the World Wide Web offers extensive information about ICPSR's resources and services. Users may search and browse ICPSR's catalog of data holdings to identify data of interest. Detailed descriptions of collections in ICPSR's social science archive are available, and documentation is searchable and downloadable. Subsetting and analysis capabilities are available for selected data collections. Users may also participate in forums related to specific data collections. To access the ICPSR Homepage, use the URL:

<http://www.icpsr.umich.edu>

Through the Homepage, users may link to other ICPSR websites. The site for the National Archive of Criminal Justice Data (NACJD) enables the user to browse, search, and download data free of charge from NACJD's over 500 criminal justice data collections:

<http://www.icpsr.umich.edu/NACJD>

The National Archive of Computerized Data on Aging (NACDA) maintains a website offering information to assist researchers in gerontology and geriatrics. Users may search and browse NACDA's holdings and download a limited selection of studies from this site:

<http://www.icpsr.umich.edu/NACDA>

Other websites are also under development at ICPSR.

Data Retrieval and FTP

ICPSR's holdings are now available for FTP retrieval by ICPSR Official Representatives (ORs) directly through the ICPSR Website as well as through the Consortium Data Network (CDNet), ICPSR's custom ordering facility. For more information on using these services, contact ICPSR User Support.

FastTrack

FastTrack permits **anonymous FTP** retrieval of some of ICPSR's newly-acquired data collections, usually series data. FastTrack data are made available as soon as they are acquired, with minimal checking, before being processed by ICPSR. ORs interested in this service should consult the **Hotline** or subscribe to the ICPSR **series** mail list, which carries announcements of the availability of FastTrack datasets. To retrieve FastTrack data, connect to ICPSR's anonymous FTP site ([ftp.icpsr.umich.edu](ftp://icpsr.umich.edu)) and enter the username "anonymous", using your actual e-mail address as the password. Then change to the FastTrack subdirectory: **cd/pub/FastTrack**.

Discussion Lists

Several mail lists are offered to promote discussion and information exchange.

series. This list provides an easy way to obtain information on the status of studies that are parts of serial data collections, including those studies not yet released by ICPSR.

or-announce. This list, automatically available to all ORs, carries announcements from ICPSR staff regarding new developments at ICPSR.

eurobarometer and **wls**. These lists offer a venue for researchers to discuss experiences using data from the Euro-Barometer series and the Wisconsin Longitudinal Study and to receive updates.

orl. Hosted by the University of Alberta, orl is available to ORs and others interested in exchanging information or sharing ideas about ICPSR, its data, and its services.

To subscribe to a list, send e-mail to majordomo@icpsr.umich.edu using the text: **subscribe series [or eurobarometer or wls]**. (Note: For orl, send mail to majordomo@majordomo.srv.ualberta.ca using the text: **subscribe orl**.) Then exit as you normally would from e-mail; you will receive mail back from majordomo.

Hotline

Newly released data collections and updates to existing collections are announced via the ICPSR Hotline facility, which can be accessed through the Website.

Publication-Related Archive

This service enables users to deposit and retrieve replication datasets (i.e., data and information necessary to permit replication of a corresponding published article, book, or dissertation) via **anonymous FTP**. To submit such datasets, send mail to: pra@icpsr.umich.edu to receive an electronic form that must be completed and returned. To deposit the data, connect to ICPSR's anonymous FTP site ([ftp.icpsr.umich.edu](ftp://icpsr.umich.edu)) and enter the username "anonymous", using your actual e-mail address as the password. Then change to the Publication-Related Archive subdirectory: **cd/pub/PRA**. To retrieve data, connect to the same site for further instructions.

Electronic References

Electronic versions of most of ICPSR's printed publications are available through the ICPSR Website. These publications include:

- *Guide to Resources and Services*
- *ICPSR Bulletin*
- *Summer Training Program Bibliography and Announcement*
- *Annual Report*
- *Catalogs and newsletters of ICPSR's topic-related archives, NACJD and NACDA*
- *Guide to Social Science Data Preparation and Archiving*

ADDRESS CORRECTION REQUESTED

Moving? Please send us your new address,
along with your old mailing label.

ICPSR COUNCIL MEMBERS, 1996-1997

Suzanne Bianchi, University of Maryland
bianchi@bss1.umd.edu

Kenneth A. Bollen, University of North Carolina,
Chapel Hill
bollen@unc.edu

John A. Garcia, University of Arizona
jag@u.arizona.edu

Ann S. Gray, Cornell University
asg3@cornell.edu

Charles Hirschman, University of Washington
charles@u.washington.edu

Margaret Levi, University of Washington
mlevi@u.washington.edu

Edward Nelson, California State University, Fresno
ed_nelson@csufresno.edu

Samuel C. Patterson, Past Chair, Ohio State
University
spatte@ohstmva.acs.ohio-state.edu

Carole Shamas, Chair, University of California,
Riverside
lhr14@ucr.ac1.ucr.edu

Elizabeth Stephenson, University of California,
Los Angeles
libbie@ucla.edu

Wendy Watkins, Carleton University
wwatkins@ccs.carleton.ca

Halliman A. Winsborough, University of
Wisconsin, Madison
winsboro@ssc.wisc.edu

To reach all Council members:
council@icpsr.umich.edu

ICPSR ASSOCIATE DIRECTORS

Heinz Eulau
Stanford University

Norval D. Glenn
University of Texas, Austin

M. Kent Jennings
University of California, Santa Barbara
University of Michigan

Warren E. Miller
Arizona State University

ICPSR ADMINISTRATION AND STAFF

Richard C. Rockwell, Executive Director
richard@icpsr.umich.edu (313) 764-2570

Pamela Schwarzmann, Senior Financial Analyst
pschwarz@icpsr.umich.edu (313) 764-2570

Michelle Humphres, Administrative Secretary
michelle@icpsr.umich.edu (313) 764-2570

COMPUTING AND NETWORK SERVICES

John E. Gray, Director
jgray@icpsr.umich.edu (313) 763-3482

EDUCATIONAL RESOURCES

Henry Heitowit, Director
hank@icpsr.umich.edu (313) 764-8392

ARCHIVE STAFF

Erik W. Austin, Director, Archival Development
erik@icpsr.umich.edu (313) 936-1752

Janet Vavra, Technical Director
jan@icpsr.umich.edu (313) 763-5010

Nancy H. Fultz, Assistant Archival Director
nancyf@icpsr.umich.edu (313) 936-1752

Peter Granda, Assistant Archival Director
peterg@icpsr.umich.edu (313) 936-1752

Christopher S. Dunn, Assistant Archival Director
cdunn@icpsr.umich.edu (313) 763-5011

Mary Morris, Research Associate, User Support
morris@icpsr.umich.edu (313) 763-5010

JoAnne McFarland, Manager, Substance Abuse
and Mental Health Project
jmcfar@icpsr.umich.edu (313) 936-1752

Fredric D. Wolinsky, NACDA Resident Scientist
Saint Louis University
wolinsky@wpogate.slu.edu

May 1997 Vol. XVII No. 4

The Inter-university Consortium for Political and Social Research (ICPSR), located at the Institute for Social Research in Ann Arbor, is the world's largest repository of computer-readable social science data. For over 30 years, the Consortium has served the social science community by acquiring, processing, and distributing data collections on a broad range of topics. Researchers at the Consortium's member institutions may obtain any of these data collections at no charge; researchers at nonmember institutions may also use the data, after paying an access fee. To find out more about ICPSR's holdings or about a specific data collection, access the ICPSR Website at the URL: <http://www.icpsr.umich.edu>.

The *ICPSR Bulletin* is published four times during each academic year (in September, December, February, and May) to inform Official Representatives at the member campuses, ICPSR Council Members, and other interested scholars of activities occurring at ICPSR and at other member institutions and to list the data collections most recently released or updated by ICPSR. For subscription information, contact the Editor.

Subscription Price: \$15 per year

ICPSR
Institute for Social Research
426 Thompson St., Ann Arbor, MI 48109
P.O. Box 1248, Ann Arbor, MI 48106-1248

Phone: (313) 764-2570

Fax: (313) 764-8041

E-mail: netmail@icpsr.umich.edu

ICPSR Website: <http://www.icpsr.umich.edu>

Mary B. Vardigan, Editor
Amy B. Garber, Associate Editor



Printed on recycled paper