

DATA TAPES



SOCIAL SCIENCE DATA ARCHIVE / Survey Research Laboratory / University of Illinois

No. 3

January, 1972

Role of the Social Science Data Archive

A new service facility is developing at the University of Illinois to provide certain types of machine-readable data relevant to the data needs that exist throughout the state. The Social Science Data Archive (SSDA) of the University's Survey Research Laboratory has been completely reorganized and will serve as a major source of behavioral and social science data for research, teaching and policy decision-making throughout the state. SSDA will be a primary service organization providing data for the academic community of the University, other universities and colleges throughout the state, governmental agencies (both state and municipal), and various organizations in the private sector.

The present holdings of the archive are being upgraded and made part of a user-oriented information system for the behavioral and social sciences. The archive has collected and made easily available a wealth of data in various fields such as sociology, political science, economics, and history. Data from studies at major survey research organizations and the Survey Research Laboratory have been archived and made available for secondary data

analysis. Other than SRL, the archive holds data from the following:

1. Roper Public Opinion Center
2. Survey Research Center, University of Michigan
3. Inter-University Consortium for Political Research
4. Center for International Studies, M.I.T.
5. International Data Library and Reference Service, University of California at Berkeley
6. National Opinion Research Center, University of Chicago
7. Political Science Research Library, Yale University
8. Public Opinion Survey Unit, University of Missouri
9. U.S. Census Bureau
10. Bureau of Economic and Business Research, University of Illinois

Codebooks and ancillary documentation are available for all of the holdings. Special codebooks containing the marginal frequency distributions for each variable in each study have been constructed for all the SRL studies held in the archive. These data are extremely useful for student and faculty research, and they provide a very effective and inexpensive teaching aid by making live, well-documented data available for class and individual projects.

A Census Data Processing Center

A major part of the archive's re-organization has been to develop the capability of processing all the 1970 Census of Housing and Population data. At a minimum, the full set of machine-readable summary tape census data for Illinois will be acquired and made available to any user. The SSDA staff will be capable of servicing all requests of this data base. Data for additional states will also be acquired as user demand warrants and as resources allow. At this time, it seems quite possible that data for the entire Midwest might be acquired. Long-range plans of the archive do not preclude the acquisition of census data for the entire United States.

Data Availability

The full set of census data for each state is divided into six partitions known as "counts." The primary difference among these counts is the geographic areas and the number of census items that each covers. The First Count summary tapes of 1970 census data are in hand now, and requests for data are being serviced as they are received. It is expected that the Second and Third Counts will be available in February. The Fourth Count will probably be available in April or May. We have received no indication from the Census Bureau concerning the availability of the Fifth or Sixth Counts. (See Appendix A for current availability.)

The first three counts contain complete data from 100 percent of the households in various geographic subdivisions of Illinois. These data cover 5 population and 15 housing items. The Fourth through Sixth Counts contain data from a 15 percent sample and a 5 percent sample of the households in various geographic subdivisions. There are certain common

items in these two samples that can be used to generate a 20 percent sample.

Though Counts 4, 5, and 6 are weighted sample data, many more census items are covered in each geographic area. Thus, the 20 percent sample covers 15 population items and 20 housing items; the 15 percent sample covers 23 population items and 25 housing items; and the 5 percent sample covers 30 population items and 35 housing items. (See Appendix B.)

The six counts contain information as follows:

- First Count: block groups and enumeration districts (average population=820)
- Second Count: census tracts (average population=4,000)
- Third Count: city blocks (average population=90)
- Fourth Count: census tracts
- Fifth Count: ZIP-code areas (3-digit areas where average population size is 260,000 and 5-digit areas where average population size is 10,000)
- Sixth Count: cities with a population of 100,000 or larger

Summary data are also available at the following levels: township, county, and Standard Metropolitan Statistical Areas (SMSAs).

In addition to the six counts of data on census summary tapes, SSDA will be able to provide data from the 1960 1/1,000 public use sample tape, the 1970 1/100 public use sample tape, and the 1960 1/100 public use sample tape, which is in a format compatible with the 1970 version and will therefore be useful for comparative purposes. The public use samples will contain a sample of records for individual households, with the characteristics of household members provided

in considerable detail. In addition, demographic characteristics of the neighborhood of each sample household will be included. All identifying information about an individual, household, or firm is, of course, removed. The public use sample files can be drawn from either the 15 percent sample records or the 5 percent sample records.

Essentially, these samples offer the researcher an opportunity to create special tabulations--that is, to cross-tabulate variables of unique interest for large areas within the state, such as counties or SMSAs. The 1/1,000 public use sample for 1960 is currently available and the 1/100 public use samples for 1960 and 1970 should be available during the first quarter of this year.

Data Processing Capabilities

SSDA is implementing a package of computer programs to perform various types of data file manipulation and management and to provide "easy to read" and "easy to work with" output for its users. The package consists of six computer programs: four are presently operational and well documented, and the remaining two are under development. These programs are described in Appendix C.

The SSDA programming staff also has the ability to write and implement special-purpose computer programs for special tabulation requests that cannot be serviced using any of these six programs.

SSDA will have available a copy of the Address Coding Guide (ACG) and the Dual Independent Map Encoding (DIME) file. The ACG tape covers the urbanized portion of a Standard Metropolitan Statistical Area. The file describes a block side by street name, the address range along the block

side, and the associated geographic codes such as census tract, block, and municipality. The ACG can be used to match locally collected data to small areas such as blocks or block sides. The ACG will enable planning agencies, municipal governments, health agencies, school districts, and transportation agencies to code data to local geographic units.

The DIME file is an improvement over the ACG. DIME contains the same information as the ACG and it identifies the intersection (node) at each end of the street segment. The node's x-y coordinates provide the ability to produce computer and plotter displays of statistical maps. By linking geographic file codes to demographic records, one can map those areas having a statistical relationship to each other. DIME also permits mapping of special areas.

SSDA will also have available a copy of the ADMATCH package, which operates on either the ACG or DIME for purposes of geocoding addresses. In other words, this package is designed to assign geographic codes to machine-readable data records containing street addresses. Geographic codes such as census tracts, school districts, or traffic zones can be assigned to the data records in files of construction permits, school records, police arrests, etc. ADMATCH links the data records to the ACG or DIME records by matching the address on the data record to the address range in the ACG or DIME. When a match is found, geographic codes from the ACG or DIME are assigned to the data records.

SSDA also has the capability of producing computer maps using the SYMAP program. The coordinate information produced in address-matching permits each address to be plotted on a computerized map. Address-matching makes it possible to generate maps for those areas with a statistical relationship.

Use of Census Data

Census data are extremely useful for teaching, research, market analysis, and general policy-making. Demographic data are valuable for keeping current the content of courses taught in the behavioral and social sciences. Both students and faculty members will be able to use these data for classroom presentations, term projects, and out-of-class research interests. Administrators from many different areas--education, business, medicine nonprofit organizations, government, etc.--will find that census data can provide an invaluable aid to decision-making, policy formation, and developmental planning.

The Development of a Statewide Information System

The reorganization of the archive has taken place with the intention of developing a comprehensive information system for the state of Illinois. This information system will integrate several components to provide a wide range of social, political, demographic, and economic data. The system will be composed of several computer-based, easily and economically accessible data files that can be machine manipulated to provide optimal service to system users.

A systemic approach to information storage and retrieval allows a great deal of flexibility, while preserving optimal capability for servicing user data requests. It allows SSDA to bring several data files "under one roof" and, in effect, create a "data supermarket." This saves the potential user much time and effort in locating the appropriate data to meet his needs, and then allows him to pick and choose information from several data files. The SSDA staff will work in individualized consultation with

each user to inform him of data availability and to tailor the archive's services to meet his requests.

Staff

Daniel James Amick, Chief

Dr. Amick has a joint appointment with the Department of Sociology at Chicago Circle and has had experience in the design and implementation of computerized information systems at the University of Pittsburgh.

Judith Peyton, Systems Analyst

Mrs. Peyton received a B. A. in mathematics and computer science. She worked for two years as a systems analyst for a private firm and worked for a year as a computer programmer at the Computer Center of the University of Illinois at Chicago Circle.

Noel Uri, Research Assistant

Mr. Uri is presently pursuing a Ph.D. degree in economics at the University of Illinois at Urbana-Champaign. Mr. Uri received a B.A. in mathematics and an M.A. in economics from San Diego State College in California. He has worked as a computer programmer for the Systems Development Corporation and the United States Army.

Workshops and Training Programs

SSDA offers workshops at the University of Illinois for interested students, faculty, and administrators to explain the structure and content of the census data files, their availability and accessibility, and their uses. In addition, special training programs for the same purpose are offered to users from governmental and nonprofit agencies.

Charges for Services

SSDA will charge for its services based upon retrieving its own costs for processing requests and covering its operating expenses. Prospective users are urged to contact SSDA to discuss their data requirements. Such discussions can help clarify in the user's own mind what types of data will best fit his needs. After these discussions, SSDA can prepare estimates on the cost of furnishing the required data.

Inquiries

Inquiries regarding the services provided by SSDA should be directed to one of the following persons:

Daniel James Amick
Survey Research Laboratory
University of Illinois at Chicago
Circle
P. O. Box 6905
Chicago, Illinois 60680
Phone: (312) 996-5304

or

Noel Uri
Survey Research Laboratory
University of Illinois at Urbana-
Champaign
Room 11, David Kinley Hall
Urbana, Illinois 61801
Phone: (217) 333-6572

Those in the Midwest may inquire at one of the academic census processing centers listed in Appendix D.

APPENDIX A

12/1/71

AVAILABILITY AND NUMBER OF 1970 CENSUS SUMMARY TAPES

	1st Count		2nd Count		3rd Count	4th Count						6th Count	
	File A	File B	File A	File B		Population			Housing			Pop.	Hsg.
						File A	File B	File C	File A	File B	File C		
UNITED STATES													
ALABAMA	2	1	2		3								
ALASKA	1	1	1		1								
ARIZONA	1	1	1		3								
ARKANSAS	2	1	1		2								
CALIFORNIA	8	1											
COLORADO	1	1	2		4								
CONNECTICUT	1	1	2	2	4								
DELAWARE	1	1	1	1	1								
DIST. OF COLUMBIA	1	1	1	1	1								
FLORIDA	3	1	3		11								
GEORGIA	2	1		4	3								
HAWAII	1	1	1	1	1								
IDAHO	1	1	1		1								
ILLINOIS	5	2	5		11								
INDIANA	3	1	2		5								
IOWA	2	2	1	5									
KANSAS	2	1	1		4								
KENTUCKY	2	1	1	3	2								
LOUISIANA	2	1	2	2									
MAINE	1	1	1	2	1								
MARYLAND	2	1	2	2	4								
MASSACHUSETTS	2	1			8								
MICHIGAN	4	1	4		9								
MINNESOTA	2	2	2	8									
MISSISSIPPI	1	1	1		1								
MISSOURI	3	2	2	5	4								
MONTANA	1	1	1		1								
NEBRASKA	1	1	1	5	2								
NEVADA	1	1	1		1								
NEW HAMPSHIRE	1	1	1		1								
NEW JERSEY	3	1	4		10								
NEW MEXICO	1	1	1		1								
NEW YORK	6	1			21								
NORTH CAROLINA	2	1	2	4	3								
NORTH DAKOTA	1	1	1		1								
OHIO	5	2	5		11								
OKLAHOMA	2	1	2		3								
OREGON	1	1	1	2	4								
PENNSYLVANIA	5	2			13								
RHODE ISLAND	1	1	1	1	2								
SOUTH CAROLINA	1	1	1		3								
SOUTH DAKOTA	1	1	1		1								
TENNESSEE	2	1	1	3	5								
TEXAS	5	1	5		15								
UTAH	1	1	1		2								
VERMONT	1	1	0										
VIRGINIA	2	1	2		5								
WASHINGTON	2	1	2		6								
WEST VIRGINIA	1	1	1	2	1								
WISCONSIN	2	1	2		6								
WYOMING	1	1	0		1								
PUERTO RICO	2	1											

APPENDIX B

Subject Items Included in the 1970 Census

(Percents indicate size of sample for which the question is asked; 100 percent is complete coverage)

Population items

100 percent	Relationship to head of household Color or race Age Sex
20 percent	Marital status State or country of birth Years of school completed Number of children ever born Employment status Hours worked last week Weeks worked in 1969 Last year in which worked Occupation, industry, and class of worker Activity 5 years ago Income in 1969
15 percent	Country of birth of parents Mother tongue Year moved into this house Place of residence 5 years ago School or college enrollment (public or private) Veteran status Place of work Means of transportation to work
5 percent	Mexican or Spanish origin or descent Citizenship Year of immigration When-married Vocational training completed Presence and duration of disability Occupation-industry 5 years ago

Housing items

100 percent	Number of units at this address Telephone Private entrance to living quarters Complete kitchen facilities Rooms Water supply Flush toilet Bathtub or shower Basement Tenure Commercial establishment on property Value Contract rent Vacancy status Months vacant
20 percent	Components of gross rent Heating equipment Year structure built Number of units in structure and whether a trailer
15 percent	Farm residence Source of water Sewage disposal Bathrooms Air conditioning Automobiles
5 percent	Stories, elevator in structure Fuel--heating, cooking, water heating Bedrooms Clothes washing machine Clothes dryer Dishwasher Home food freezer Television Radio Second home

APPENDIX C

DESCRIPTION OF COMPUTER PROGRAMS

MOD.-1 (operational)

This computer program will permit the user to extract whole tables, appropriately labeled, for any areas or under any data condition from any 1970 census First Count summary tape file. The tables selected may be summarized across many areas. The objective of the program is to enable the user to list any specific tables from the First Count when they meet geographic criteria he sets (e.g., all EDs in a specific tract) and/or contain an enumerated population equal to, greater than, or less than a quantity he specifies (e.g., 20 or more divorced females). Also, the user may sum together tables with the same table number for whatever areas are identified in his retrieval. He may combine his criteria--for example, all EDs in Tract 101 with 20 or more divorced females and 10 or more Negro males and less than 5 persons under 100, or all EDs in Tract 105 and all blockgroups in Tract 106 with 5 or more widowed males and 20 or more widowed females.

MOD.-2 (operational)

The basic features of the MOD.-2 computer program allow the user to:

Describe simple or complex retrieval statements, analyzing any of the thousands of data elements in the 1970 census data base files.

Select specific "tables" (unique data sets) for display.

Summarize specific "tables" for all records selected for processing by the retrieval statements.

Take "interim" totals on specific "tables," initializing the counters each time the value of a selected field of data changes.

MOD.-3 (operational)

The basic features of the MOD.-3 computer program allow the user to:

Generate fully labeled columnar reports according to user specifications recorded on a simple worksheet.

Create new data elements through the interpretation of user "computation statements."

Select records for processing based on any data characteristic, including those generated by the "computation statements."

Generate one level of summary data; i.e., summarizing blockgroups and EDs to the census tract level.

MOD.-4 (being implemented)

The MOD.-4 computer program provides a readily accessible means for generating structuring auxiliary data files from the various 1970 census summary data bases.

MOD.-4 will enable the user to:

Generate output files with up to ten data elements per area record.

Create new data attributes for output through the development of user "computation statements."

Select records for processing based on any data characteristic, including those generated by the "computation statements."

Develop "aggregate" data files summarizing data for individual small areas; for example, summarizing blockgroups and EDs to the census tract level.

Generate files on punched cards, magnetic tape, or any direct access storage device.

MOD.-5 (operational)

The MOD.-5 computer program can save the user valuable time and money by significantly reducing, or totally eliminating, the need to develop unique application programs for creating detailed tabular reports from the 1970 census summary data files.

MOD.-5 will enable the user to:

Generate fully labeled tabular reports very similar to the standard census printed reports with three report format options.

Create new data elements for display in reports through the development of user "computation statements."

Select records for processing based on any data characteristic, including those generated by the "computation statements."

GEOPICK (being implemented)

The GEOPICK computer program generates a subfile extract of selected geographic areas from any of the 1970 census summary files.

GEOPICK will allow the user to:

Generate subfile extracts in the same record format as the census input file.

Use retrieval statements to identify specific geographic areas for extraction.

Correlate census areas with one's own "user areas" by inserting into each extracted record the Set Number contained in each retrieval statement.

Subsequently process the extracted file to link record segments and sequence the file by "user area." Aggregated summaries can then be generated for one's "user areas" with other computer programs.

APPENDIX D

ACADEMIC CENSUS PROCESSING CENTERS
IN THE MIDWEST

Institute of Urban and Regional Research
University of Iowa
102 Church Street
Iowa City, Iowa 52240
Dr. Kenneth J. Dueker, Associate Director

Institute for Social and Environmental Studies
University of Kansas
607 Blake Hall
Lawrence, Kansas 66044
Dr. Robert T. Aangeenbrug, Director

Inter-University Consortium for Political Research
P. O. Box 1248
Ann Arbor, Michigan 48106
Dr. Jerome M. Clubb, Director - Historical Archives

Minnesota Analysis and Planning System
Agricultural Extension Service - University of Minnesota
226 Temporary So. of Coffey
St. Paul, Minnesota 55101
Dr. John S. Hoyt, Jr., Program Leader for Special Project,
Development and Coordination

Demographic Data Service
University of Missouri - Columbia
104 Sociology Building
Columbia, Missouri 65101
Rex R. Campbell, Associate Professor

Institute for Urban Information Systems
University of Cincinnati
Cincinnati, Ohio 45221
Mr. Fred J. Lundberg, Director