

The Directory will be useful to all types of census data users: data archives, libraries, planning organizations, other Federal agencies, State and local governments, and the private sector. By development of the Directory, the Data User Services Division is providing a new information service to the entire user community.

#### STATE AND REGIONAL DATA ARCHIVES (SRDA)

Murray A. Straus  
University of New Hampshire  
Durham, NH 03824 (603) 862-1888

A vast amount of data is available on a state by state basis, or by region. Political scientists, demographers, economists, and geographers have used this valuable resource. Other social scientists have rarely used this data. Part of the reason is that the data is scattered over many sources. Social scientists are often not aware of its potential. Even those who realize the potential do not know the full range of variables. They also do not have convenient access.

In March 1979 we therefore began to compile a truly comprehensive archive of data on American states and regions. This archive (SRDA) is intended to be the equivalent, for states of the United States, of the Human Relations Area Files and other archives of cross-national data. Well over 2,000 variables have been identified in the preliminary search. Other variables are being added continuously. The data files will be available in machine readable form (SPSS file, card image tape, or cards) or as a printed listing.

Two overlapping archives will make up the SRDA. The State Archive consists of data on each of the 50 states and the District of Columbia. There are now about 1,000 variables actually entered in the State Archive in the form of an SPSS system file. The Regional Archive consists of data on the nine divisions which the U.S. Census uses as its main regional classification. All of the variables in the State Archive will also be in the Regional Archive. However, the Regional Archive will contain additional variables that are not in the State Archive. These are variables for which state by state data could not be obtained. The initial Regional Archive file will be created sometime in 1980.

#### WHY USE STATE LEVEL DATA?

There are many problems connected with the use of state level data for social science research. Among the most obvious is the ambiguity inherent in a statistic which combines, for example, New York City and the Adirondack mountain region. These problems, and also what is to be gained from using state level data, are analyzed in my forthcoming book, State and Regional Analysis in the

Social Sciences. Among the reasons for using state level data despite these problems are:

1. States are the theoretically appropriate unit for issues which involve many aspects of government, politics, taxes, schools, economics, administration, and legal issues, as for example in Hicks, Friedland and Johnson.\*

However, the uses of SRDA data will include many issues for which the states are not the natural unit. In such cases the decision to use state level, as in many research decisions, involves a trade-off: one accepts the problematic aspects of state data in order to be able to do research which would otherwise not be possible or practical. For example:

2. Historical analysis is possible because some state level data goes back to colonial times and the number of variables available has grown exponentially each generation.
3. Causal inferences can sometimes be more clearly established because the same data is available for two or more time periods. This permits the use of time series and cross-lagged correlation analysis.
4. Variables can be linked, even though they are located in different surveys and refer to different respondents. This is possible by first converting the individual level data to state level data: for example, the percent in each state who agree that "Homosexuality is always wrong" from one study, with the percent who oppose the Equal Rights Amendment from another study.
5. Contextual analysis is possible by combining variables from the SRDA with individual level data from specific surveys. For example, Kersti Yllo has constructed a Sexual Inequality Index for each state and used this to find out if the correlates of a male-dominant marriage are the same or different in states where women are generally disadvantaged versus those with greater equality between the sexes.

#### TYPES OF DATA

1. Published lists of state by state data. Many examples are to be found in standard almanacs and the Statistical Abstract of the United States. The largest source of published data is the U.S. Census.
2. Aggregated survey data. These surveys were designed for analysis on an individual by individual basis. For the SRDA, the results are tabulated by state. An example is the percent in each state agreeing with some attitude question.
3. Compilations from documents. Many documents give the state as one item of information. This can be the basis for compiling a new variable. For example, Who's Who in America gives the state of birth. It is possible to obtain a measure of the extent to which each state has contributed eminent persons by tallying the number of eminent Americans born in each state and dividing that by the population of the state.
4. Indexes from other variables. By combining existing variables it is possible to produce entirely new variables, or to produce an index which does a better job of measuring than any one of the variables which are combined to form the index. An example would be a measure of the

---

\*Hicks, Alexander, Roger Friedland and Edwin Johnson 1978 "Class power and state policy: The case of large business corporations, labor unions and governmental redistribution in the American states." American Sociological Review 43 (3):302-315

socioeconomic status of each state's population. This could be made up by combining the median income, education, and occupational prestige values from each state.

#### DATA SOURCES

Another way of classifying the data in the SRDA is according to the accessibility of the source. Sections A and B list readily available data sources, each of which contains many variables. The main value of including them in the SRDA is convenience. Researchers using the SRDA data do not have to punch, verify, provide variable levels, etc. Instead they can acquire a proofed, clean, labeled, ready-to-run data set.

The full value of the SRDA, however, will derive from the inclusion of data from sources which are not readily available and which often will not even be known to researchers. These come from dozens of books and research reports, reports of government agencies, special topic reports by the census, and reports of private organizations such as the Institute of Life Insurance, the Audit Bureau of Circulation, the Boy Scouts of America, the National Womens Political Caucus, etc.

Finally, the least accessible data of all are the state level statistics created by aggregating individual level surveys to provide rates and averages for the states and regions. This is a long and expensive process for which the procedures are now being developed.

The initial archive will consist mainly of data given for the 50 states (and the District of Columbia) in the sources below.

#### A. Multi-topic Compendia

- Bacheller, Martin A. (Ed.)  
1979 The Hammond Almanac of a Million Facts, Records, Forecasts. Maplewood, N.J.: Almanac, Inc.
- Book of the States 1978-1979  
1978 Council of State Governments. Lexington, Ky.
- County and City Data Book  
1947- Washington, D.C. Bureau of the Census.
- Demographic, Social and Economic Profile of States: Spring, 1976  
1979 Current Population Reports, Series P-20, No. 334. Washington, D.C.: Bureau of the Census.
- Information Please  
1979 Information Please Almanac, Atlas, & Yearbook, 33rd Edition. New York: Information Please Publ.
- The World Almanac  
1979 The World Almanac & Book of Facts 1979. New York: Newspaper Enterprise Association.
- Roswe, Arthur E. (Ed.)  
1978 1978-79 Help: The Useful Almanac. Washington, D.C.: Consumer News.
- Statistical Abstract of the United States 1878- Washington D.C.: Bureau of the Census, U.S. Department of Commerce.

#### B. Specific Topics (Printed Sources)

- Almanac of American Politics  
1978 New York: E. P. Dutton
- Bureau of Labor Statistics  
1978 Handbook of Labor Statistics. Washington, D.C.: U.S. Department of Labor.

- Gottfredson, Michael R., Michael J. Hindelang, and Nicolette Parisi (Eds.) 1978 Sourcebook of Criminal Justice Statistics, 1977. Washington, D.C.: Criminal Justice Research Center, Law Enforcement Assistance Administration, U.S. Department of Justice.
- Grant, W. Vance and C. George Lind 1979 Digest of Education Statistics 1979. Washington, D.C.: U.S. Department of Health, Education, and Welfare Education Division.
- Rosten, Leo (Ed.) 1975 A Guide to the Religions of America. New York: Simon and Schuster.
- Vital Statistics of the United States 1978 Washington, D.C.: National Center For Health Statistics, HEW.

C. Aggregated Survey Data

- Physical Violence in American Families 1976 Survey conducted by Response Analysis Corp. for Murray A. Straus, principal investigator.

PROGRAM PUBLICATIONS IN PROGRESS  
AT THE UNIVERSITY OF NEW HAMPSHIRE

- Straus, Murray A. Codebook for the State and Regional Data Archives. (Anticipated availability for Part I, Variables 1 to 1,499 is March, 1980; for Part II, Variables 1,500 to 2,999, September, 1980).

State and Regional Data Bulletin

This quarterly publication will replace the Codebook beginning with variable 3,000. The BULLETIN will improve on the Codebook in three ways: (1) Quarterly publication will make materials available more quickly. (2) In addition to documenting the source and nature of the data, it will include a printed listing of the statistics for each state and region. (3) The BULLETIN will include news items about the SRDA and occasional commentary and analyses of data included in that issue. The planned publication date for the first issue is January 1981.