

UCLA

**Volume IV. 1988-89 - Conference on Comparative Ethnicity: The
Conference Papers, June 1988**

Title

Machine-Readable Data Sources for Comparative Ethnic Research: Problems and Prospects

Permalink

<https://escholarship.org/uc/item/9353344c>

Author

Stephenson, Elizabeth

Publication Date

1988-06-01

Institute for
Social Science Research

ISSR
Working Papers
in the
Social Sciences

1988-89, Vol 4, Number 27

**Machine-Readable Data
Sources for Comparative
Ethnic Research: Problem
and Prospects**

by

Elizabeth Stephenson

University of California
Los Angeles

ABOUT THE AUTHOR...

Elizabeth Stephenson received her Masters of Library Science Degree from UCLA in 1977 and is the Data Archivist at the Institute for Social Science Research, UCLA. Her current research interests focus on the impact of technological change on an information society and the emerging information underclass.

A version of this paper was prepared for the UCLA CONFERENCE ON COMPARATIVE ETHNICITY, June 1988. The Conference was coordinated by Institute for Social Science Research and sponsored by The President's Office, Chancellor's Office, College of Letters and Science, Institute of American Cultures, Center for Afro-American Studies, and Department of Geography and Sociology, UCLA; and by the Division of Social Sciences and Communication, the Los Angeles Project and Department of Geography and Political Science, USC.

**MACHINE-READABLE DATA SOURCES FOR COMPARATIVE ETHNICITY RESEARCH:
PROBLEMS AND PROSPECTS**

ELIZABETH STEPHENSON*
University of California, Los Angeles

INTRODUCTION

In the introduction to this volume, Johnson and Oliver (1988) discuss the importance of exploring questions of ethnicity and ethnic group behavior in a comparative context and highlight a specific set of "ethnic dilemmas" requiring immediate attention and remediation. Answers to the types of comparative ethnic questions they raised would ideally require the collection of primary data via large scale social surveys. Because the design and conduct of a social survey is a time consuming and costly undertaking, especially for the lone researcher, social scientists have traditionally attempted to circumvent the problem by undertaking secondary analysis of previously conducted surveys. Given this longstanding research tradition, Stephenson (1988) has compiled an INDEX OF MACHINE-READABLE DATA FILES FOR USE IN COMPARATIVE ETHNIC RESEARCH. The index contains references to UCLA's holdings of surveys, public opinion polls, and both historical and current enumerative data. It will be significantly useful in comparative ethnic research on such topics as: Ethnic assimilation, segregation, and neighborhood change; Labor markets and entrepreneurship; Political and electoral behavior; Health and well-being; Crime; and, Education.

The purpose of this essay is to encourage future comparative ethnic research by highlighting potential uses and limitations of machine-readable data files, such as those referenced in the INDEX (Stephenson 1988). Toward this end, background details are provided on the organizations and agencies that collect or archive publicly available data and describe in detail selected data files. A second section will focus on data collection policies, sampling deficiencies and inherent limitations for research on comparative aspects of ethnicity and ethnic group behavior. In the concluding section the discussion focuses on, among other salient issues, the social scientist's role in future government decisions regarding the collection of data on ethnic groups in America.

SECONDARY ANALYSIS IN CONTEXT

As a research method, secondary analysis provides social scientists with access to resources at much less cost than would be required to conduct original full-scale surveys. It is possible to replicate studies, to compare results from similar studies, to reanalyze studies using different

*Elizabeth Stephenson is Data Archivist at the Institute for Social Science Research, University of California, Los Angeles. She would like to acknowledge the generous support and advice on the paper's content and design provided by James H. Johnson, Jr., Ann E. Gerken, and Marilyn B. Brewer.

methods, to test models and hypotheses, and, to conduct both cross-sectional and longitudinal analyses. Machine-readable data collected in surveys can also be used for more complex and detailed research projects, and the manipulation of larger amounts of information than printed tables and published reports permit. Further, the use of a machine-readable file makes it possible to reassess an original data collection rather than rely on the summary statistics produced from it (Hakim 1982).

Secondary analysis is a potentially useful approach to comparative ethnic research. But, because surveys that contrast different groups have largely not been conducted, identifying and locating appropriate data files can be a difficult task, if not an impossible challenge. As a consequence, addressing ethnic questions in a comparative context may require the manipulation of data from multiple studies, the acquisition of which may be prohibitively expensive, especially for the single researcher.

Researchers can resolve this dilemma by focusing their attention on locating archived collections of data. Major research-oriented universities in the United States have established Data Archives to provide services that include the identification and acquisition of machine-readable files from both government and private sources, and to provide technical support in the use of the files. University based Data Archives can be located within survey research units, libraries, and some are situated as part of the campus computing facilities. Data Archives offer the researcher an opportunity to evaluate the utility and appropriateness of particular files to research needs. By browsing through codebooks and consulting with Data Archivists, researchers can save valuable time and money in identifying primary data resources. The INDEX described earlier in this paper is one example of the type of resources provided by Archivists (Gerken 1988; Heim 1987).

RELEVANT DATA FILES: SOURCES AND AREAS OF COVERAGE

The federal government is the largest single source of machine-readable data in the United States. State and local governments also collect information, primarily through administrative records, but generally rely on the data collected by various agencies of the federal government. Agencies disseminating the most useful information for comparative ethnicity research are: Bureau of the Census, Bureau of Labor Statistics, Bureau of Justice Statistics, Social Security Administration, National Center for Health Statistics (NCHS), National Center for Education Statistics (NCES), and, the Immigration and Naturalization Service (INS). Each of these agencies is mandated to collect information on the U. S. population as needed to carry out the business of government and to provide legislators with information to make informed' decisions. The information collected is publicly available and can be used for a variety of demographic and policy-related research. In the following paragraphs, some of these publicly available data files are described.

Immigrants and Immigration: Population Size, Distribution and Settlement

As Johnson and Oliver point out in the introduction to this volume, immigration is an issue which lies at the root of many of the "ethnic dilemmas" confronting American society today. As large numbers of people

from Central and South America, Asia and the Middle East arrive in the United States annually, the needs of these population groups and their impact on American society must be continually assessed. The Immigration and Naturalization Service (INS) gathers data in three publicly available surveys: IMMIGRANTS ADMITTED TO THE UNITED STATES, NON-IMMIGRANTS ADMITTED TO THE UNITED STATES, and, ALIEN ADDRESS REPORTS. The INS data files include information on port and date of entry into the U.S., age, country of birth, marital status, sex, occupation, nationality, and intended place of residence. Similar details are gathered under the Alien Address Reporting System. Additional immigrant data are collected by the Social Security Administration's Office of Refugee Resettlement. Some researchers also use the supplemental immigration data. Supplemental immigration data are collected in June of each year by the Bureau of the Census, in its CURRENT POPULATION SURVEYS. Relevant questions included in the survey cover age, country of birth, citizenship status, date of entry into the U.S., and number of children born outside the U.S.

Ethnic Population Characteristics, Distribution and Demography

Many researchers rely on survey data, but enumerative data is useful in comparative ethnic research for tabulating the numbers of people in different ethnic groups; for assessing population characteristics, and geographic diversity and differences; and for studying income, housing, and migration patterns of different ethnic groups. The Bureau of the Census gathers data on the entire population every ten years in the decennial CENSUS OF POPULATION AND HOUSING. In accordance with current U.S. laws, the first use of Census data is to apportion representation in Congress. In addition, information about the ethnic composition of specific geographic areas (such as congressional and school districts, cities, counties, neighborhoods) is used to provide education, housing, employment and federal benefits (Lowry 1982, p. 43). There has been a race question on the Census questionnaire since 1790 (U.S. Bureau of the Census 1988, p. 4). The Census questions focus on demographic items including: age, sex, race, Hispanic origin, education, place of birth and citizenship, ancestry, language spoken at home, disability, fertility, veteran status, employment, occupation and work experience. Housing questions are also asked covering the building or structure, energy sources and costs, and use of utilities and fuels.

Labor Force, Economic Status and Mobility

Data on income distribution, labor force participation/non-participation, occupational mobility, and entrepreneurship are central to our understanding of the comparative economic status of ethnic groups in America. The Bureau of the Census SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP), was instituted in 1983 to provide measures of income distribution, wealth, poverty, and the use and effect of federal assistance programs for families in the United States (U.S. Bureau of the Census 1987, I-2). In the first SIPP survey, a sample of 28,000 households were interviewed every four months for a period of 24 months, resulting in a nine wave study. Another useful study is the CURRENT POPULATION SURVEY (CPS), also carried out by the Bureau of the Census. Conducted monthly on a

sample of 60-70,000 households, this survey studies labor force activity for the week prior to the survey. Core topics covered include employment status, occupation and industry of those aged 14 and older. Supplemental questions, asked in May of each year, focus on topics such as displaced workers and occupational training. Some researchers also use the Census Bureau compilations of data found in the COUNTY AND CITY DATA BOOK (CCDB), and the COUNTY BUSINESS PATTERNS (CBP). These two are used to study local area trends and to address broader issues such as the restructuring of labor markets in a multi ethnic environment.

Fertility, Health Status and Well-Being

The health and well-being of selected U.S. ethnic groups can be studied through a variety of surveys conducted by the National Center for Health Statistics (NCHS). The data collected by NCHS covers health and nutrition; health practices and behaviors; use of medical professionals, services and facilities; medical care costs; and, fertility. NCHS also conducts special surveys focusing on certain population groups or specific health issues. In some surveys questionnaire data are supplemented with physical examination details, such as eye and dental exams, and blood and urine analyses.

While the ethnic resolution of some of the data is not very refined, (e.g. most surveys identify White, Black, and Others, only), several ethnic groups of Black, Asian or Hispanic descent can be studied across time, using NCHS surveys. The Center has conducted the HISPANIC HEALTH AND NUTRITION EXAMINATION SURVEY, 1982-1984, which can be used alone or with the NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY. Topics covered include an acculturation scale, medical history, dental exams, dietary habits, and vision and hearing tests. Studies were also made to measure depression, alcohol consumption and drug abuse..

In addition, the NCHS compiles vital statistics on births, marriages, deaths, and divorces. The Vital Statistics program covering natality and mortality provides important data resources used in studying population change, fertility, and the birth and death rates of the population. These files, available in magnetic form since 1968, are tabulated for local areas, states, and larger geographic regions. The most detailed of these files presents race tabulations for eight categories and ancestry for twenty-four categories.

Data related to fertility, birth expectations and family structures are also available for comparative ethnicity research. The NATIONAL SURVEY OF FAMILY GROWTH contains data for the study of these issues and on childbearing, contraception, and maternal and child health. The CURRENT POPULATION SURVEY (CPS) described earlier also collects data in the months of April and June of each year on fertility and birth expectations.

Neighborhoods, Housing, Urbanization, and Crime

Several data sources exist for comparative ethnic research in the areas of housing, neighborhood change and crime. The federal data sources on crime and education are the Bureau of Justice Statistics (BJS) and the National Center for Education Statistics (NCES). The NATIONAL CRIME SURVEYS, conducted by BJS, study personal and household victimization for twenty-six cities in the U.S. and also contain a national sample. NCES has conducted

five waves of the NATIONAL LONGITUDINAL STUDY OF THE CLASS OF 1972, which was designed to follow the educational and occupational careers of young people in relation to high school and family background. Schools in low income areas and those with high Black enrollment were oversampled.

The Bureau of the Census conducts the AMERICAN HOUSING SURVEY to collect data on housing characteristics, adequacy of heating, neighborhood conditions, crime, availability of public services, and gathers demographic details on individual household members. Data can be analyzed at the national and standard metropolitan statistical area (SMSA) level. Hispanic ethnicity and race categories of Black, White and Other are the groups available for comparative analysis.

Public Opinion and Political Data

Thus far, a great deal of attention has been focused on federal data sources. Researchers who are interested in comparative ethnic political attitudes and behavior, as well as trends in public opinion, will have to rely, for the most part, on surveys conducted by non-government organizations. National-level public opinion data are gathered regularly by the Harris and Gallup organizations and by newspapers and television stations. In addition, local and state polls are conducted by such organizations as the Field Research Corporation which produces the California Polls. The largest depository of public opinion polls, however, is the Roper Center, located in Storrs, Connecticut, which also houses public opinion polls from other countries.

The National Election Studies Center at the University of Michigan has been conducting the AMERICAN NATIONAL ELECTION STUDIES (ANES) under the direction of Warren Miller since 1952. The ANES gathers information for a thorough examination of voting behavior and attitudes toward political parties, candidates, and both current and continuing issues.

In 1970, 1978, 1980 and 1986, the ANES contains a black oversample. However the best resource so far for studying black attitudes and behavior on a wide range of social issues is the NATIONAL SURVEY OF BLACK AMERICANS, 1979-1980, conducted by James S. Jackson and Gerald Gurin. The questions in this survey are similar to those asked over time in the ANES and also in the GENERAL SOCIAL SURVEY (GSS), 1972-1987, conducted by the National Opinion Research Center, located at the University of Chicago, under the direction of James A. Davis and Tom W. Smith. An analogous study of Mexican Americans was conducted by Carlos H. Arce called MEXICAN ORIGIN PEOPLE IN THE UNITED STATES, with data collected in 1978 and 1979.

RESOURCE LIMITATIONS FOR MULTI-ETHNIC COMPARATIVE RESEARCH

Research in comparative ethnicity is hampered by significant data collection and tabulation problems. In addition to the lack of original survey research, three problem areas can be identified: lack of cohesive assessment of ethnic identity; lack of statistically reliable or geographically representative samples; and, political and financial influences on data collection policy. There is a lack of cohesion among federal data collection agencies as to the particular ethnic groups to be identified or studied, and specific groups are identified with a variety of terminology. When several ethnicities are identified within a particular

enumeration or survey, the sample sizes are usually not representative of individual groups so that comparative analysis, even for large geographic areas is not possible. Political and budgetary effects on the study of ethnic groups are present throughout the federal data collection system. In the discussion that follows, examples of these problems as evidenced in specific data collections will be presented.

Assessment of Ethnicity and Ethnic Background

The data collections discussed thus far can be used with varying degrees of success in multi-ethnic comparisons. One of the major problems rests with the way in which ethnic groups are identified. Many surveys clump ethnically diverse populations into one category. Often ethnic groups other than Blacks are not identified, or are grouped all together as "Others" (e.g. The AMERICAN HOUSING SURVEY, and some of the surveys conducted by the NCHS). Further, among federal agencies (and other survey organizations as well), there is a lack of consistency in the use of terminology to identify specific groups. Race, ethnicity, ancestry, and national origin are concepts that do not have fixed definitions, and that are used differently depending on who is collecting the data, and who is analyzing the data.

Some of the reasons for a lack of consistency are due to the way in which ethnic data are gathered. Most surveys, and such enumerations as the decennial census, ask respondents to self-identify by race and/or ethnicity. However, attitudes, outside influences, and cultural identity can affect response. Recent immigrants to the United States may have language barriers or conflict in deciding how to answer race and ethnicity questions (*New York Times* 1988). However, the lack of consistency is also caused by the data collection agencies themselves. A close look at the Bureau of Census data collection habits will show this.

The decennial Census has probably received the most attention in relation to the collection of information about different ethnic groups. While recognizing the need for such information, the Census Bureau's definition of "ethnicity" and "race" changes each time the census is taken, making longitudinal analysis difficult. The Census Bureau also varies the degree of detail at which persons can self-identify their ethnicity or race. This is evident in an examination of the questionnaires used over time.

For instance, in the 1980 questionnaire, the Bureau asked all households to identify household members as being members of one of the following groups: White, Black, Japanese, Chinese, Filipino, Korean, Vietnamese, American Indian, Asian Indian, Hawaiian, Guamanian, Samoan, Eskimo, Aleut, and Other. The questionnaire also asked for information on Spanish/Hispanic origin or descent and permitted the identification of Mexican, Mexican American, Puerto Rican, Cuban, and Other Spanish/Hispanic people. These two items were asked of all households.

For the 1990 Census, the Bureau had intended to change the level of detail to be tabulated for race and ethnicity. In the course of writing this paper, decisions on how to identify American Indian tribes and the Asian and Pacific Islander populations changed almost on a daily basis. The initial proposed questionnaire grouped all Asians into one category and asked for specification of the particular groups with which respondents identified. Instead of making this data available for small geographic

areas on a 100% basis, the Bureau intended to publish the data for a sample of the population (U.S. Bureau of the Census 1988; APDU Newsletter 1988).

The proposed 1990 Census questionnaire would also have altered the information collected on those of Spanish/Hispanic origin. There would no longer be a specific category for those who 'considered themselves to be Mexican American (of Mexican descent with long-term residence in the United States). Mexican Americans would be collectively identified along with those who called themselves Mexican or Chicano.

While these differences may be subtle to a data collection agency, they are conspicuous to the respondent and to the researcher. Also, the potential for under-representation and misidentification increases and is a serious issue. The literature on undercount is extensive and will not be addressed here, however, its impact on the reliability of census data should not be ignored. In any case, these details are of concern to the comparative ethnicity researcher who requires data sources that differentiate between the ethnic identity of those who have recently immigrated to the United States, and those who have resided here for one or more generations.

The final proposal for the 1990 Census asks Asian and American Indian respondents to write in a specific tribe or race category, such as Chinese, Japanese, etc. The Bureau will tabulate and will produce printed and machine-readable information for "at least 26 Asian and Pacific Islander groups," and will produce a special report on American Indian tribes. This is a considerable improvement over the initial Census Bureau proposal, and will actually provide more race and ethnicity detail than in previous years. However, there is still no evidence that the Census Bureau will cooperate with other federal data collection agencies in order to reach a consensus on the collection of race and ethnicity data, nor on terminology to be employed. And there is no indication that the 1990 questionnaire is the Bureau's final stand on how ethnic groups are to be identified.

While a number of problems with consistency and definition of race and ethnicity concepts have been identified with regard to federal agencies, there is no lack of variety with which these concepts are handled by public opinion polling agencies and research organizations. It is also surprising, given the research orientation of many survey groups, that these institutions have not initiated surveys that would provide the kind of information needed for carrying out responsible comparative ethnicity analyses.

Public opinion polls are conducted to obtain quick answers to topical questions, generally political issues. The detail at which issues are addressed makes it difficult to study issue saliency, and there are few demographic variables with which to make comparisons. This is especially true with regard to race and ethnicity, where most polls identify White, Black and Others, sometimes including a question about Hispanic origin if the issue being studied is relevant to this group.

If researchers are to make the best use of secondary analysis techniques for multi-ethnic comparisons, consistent terminology in assessment of ethnicity must be applied. In order to accomplish this, work should be carried out to study how respondents self-identify themselves ethnically. Further research is needed to determine the best methods to be used in data collection and tabulation. This effort will require the collaboration of representatives from ethnic groups, social scientists, and both private and federal data collection agencies.

Sample Size and Geography

Comparative ethnicity research has some specific requirements with respect to sample size and geography. It is in this area that secondary analysis techniques are the most problematic. The comparative nature of the research requires sample sizes that are demographically and geographically representative of the groups studied. Data are needed at not only the national and regional level, but also for community level analyses of social, economic, environmental, and political interactions between cultures. Original small area and community studies are lacking, and there are few surveys that permit such analysis. For those collections which are available, problems occur when there is a representative sample of two or more ethnic groups, but the geography is so broad that detailed analysis is not possible. Conversely, a data file may contain geographic identification for very small areas, but the numbers of people for each ethnic group are too few in a given area to make comparative study possible. Some examples of this are evident in a look at the SIPP, the HHANES, and public opinion polls.

In discussing the SIPP data, an additional sampling issue is raised. To answer some comparative ethnic research questions, longitudinal data are needed. Because the SIPP was intended to have 30 months of coverage, it could be expected to have a reliable longitudinal sample. However, a blow to the sample occurred in the elimination of one or more waves of interviews (so that there are now only 24 months of coverage), and cuts in the size of the sample. Some researchers believe that a longitudinal analysis over just a two year period is not sufficient to produce reliable results, and that the number of people in the survey is too small for analysis of subgroups, such as different ethnic groups.

In addition to longitudinal analysis issues, the actual size of the sample for any one ethnic group is also problematic in many data collections. Although the SIPP provides questionnaire entries for approximately 22 categories of race and ethnicity, the sample size for any one of these is too small to be used for representative analysis.

Public opinion poll data are collected according to current events, generally political issues, so this type of information is not always useful for longitudinal analysis. Public opinion data collection agencies also need to be encouraged to design surveys to collect information from representative samples of the ethnic groups identified in surveys, and, where possible, to oversample the groups studied.

Some credit is due to survey groups for their attempts to identify respondents in terms of race and ethnicity. For example, the ANES contains a large number of ethnic categories, but the sample size overall is approximately 2000 so that in any one of these categories there are not enough people to study, even if weights are used. The ANES questionnaire also explores geographic details on mobility and migration, but is reliably used only at the national level.

In light of the renewed salience of ethnic group affiliation in the U.S., it would seem particularly important for research oriented public opinion surveys such as the ANES and GSS to gather information from a representative sample of as many ethnic groups as possible, particularly those who have just arrived in the United States. The study of the changes in attitudes and behavior socially, economically, and politically that immigrants will experience over time will provide a powerful look at the

dynamic nature of public opinion. It also seems important that detailed research and replications of classic studies of White respondents, such as was done for Black Americans and Mexican Americans, should be carried out for other ethnic groups. These surveys could be used collectively to conduct comparative analyses due to the large sample sizes and the questionnaire similarities between them.

Perhaps this approach should also be used by federal data collection agencies. One successful example of this is the HHANES conducted by the National Center for Health Statistics. There are ethnic identifiers for Mexican, Mexican American, Chicano, Puerto Rican, Boricua, Cuban, Cuban American, Hispano, Latin American, Spanish American, and Spanish people. However, the HHANES sample was selected in areas of the United States where large numbers of Hispanic people reside, so uniform small area analysis across the nation is not always possible, and sample sizes for those Hispanic groups other than Mexican Americans, Cubans, and Puerto Ricans is small.

The acculturation scale in the HHANES would be a useful feature in other national surveys, especially if similar scales could be used for other ethnic groups in generational studies of immigration to the U.S. The HHANES is comparable to the NHANES I and II, so that research comparing Hispanics and Whites is possible. The HHANES was conducted from 1982-1984, and there are no plans to repeat the survey, so longitudinal analysis will not be possible.

Social, Political and Financial Influences

Government data files are produced in accordance with federal policy to attain certain social, political, or legislative goals. These goals are not necessarily those of the social scientist or policy analyst (Cardenas 1979, p. 55). The work of federal data collection agencies tends to reflect the ebb and flow of the federal budget for such purposes, and assumes the general attitude of an administration towards such activities. The same can be said to be true of other data collection operations, but because there is such reliance on government data as an objective source of information, the effect that finances and attitude can have on the data collection process cannot be ignored. This is particularly true for those involved in research on ethnicity and multicultural comparisons.

For instance, researchers have found that adjustments in sampling methodology, or reductions in sample size, can be so altered to fit financial restraints, or political moods, that the resulting data cannot be considered scientifically reliable. And in such surveys where sampling continues to provide adequate coverage and representativeness to ethnic groups, data items and types of information collected are deleted so that needed information is no longer available.

Some classic examples of this are the budget cuts endured by the SIPP and the NCHS NATIONAL HEALTH AND NUTRITION EXAMINATION SURVEY III. The SIPP has experienced many budget cuts resulting in the reduction of the size of the sample, and in elimination of one or more waves of interviews. Budget cuts to NCHS have meant, in part, that the NHANES III sample will be cut by one third and there will be no detailed data available from this survey on different ethnic groups, nor for those aged 85 and older.

Comparative ethnicity research requires valid data at the regional and national level, but much multi-cultural analysis takes place for smaller

areas such as neighborhoods. Collection of geographic detail can be affected by fiscal and political shifts so that small areas (such as those served by local governments and service agencies), cannot produce reliable information about their populations. This would have been the case had the initial proposals for collecting 1990 census data gone forward. The tabulation of 100% data for Asian and Pacific Islanders and American Indian tribes in the 1990 Census was reinstated largely due to pressure from social scientists and the legislative representatives of these ethnic groups.

There are also some hidden fiscal and political effects on data quality and availability. Researchers can make themselves aware of proposed or actual cuts in sample size, adjustments to questionnaires, and increased gaps of time in data collection. But it is difficult to track changes in how budget cuts affect quality of data processing and collection personnel, or the reliability of the data gathered, without real familiarity with the procedures of a survey operation. Often serious problems do not surface until sometime after the data has been collected and distributed.

CONCLUSION

This essay has attempted to provide an overview of current and publicly available data which can be used in comparative ethnicity research. The limitations in available data are considerable. There is room for improvement in both the private sector and within federal data gathering agencies. Many of the problems stem from fiscal and administrative policy. The most disturbing effect of federal policy upon public data is that so much of current social science research is based on the data that is available, rather than on new data collections. Because of this reliance on previously collected data, the research, and the information available for this research, is in both overt and subtle manifestations, being determined by the federal government (Cardenas 1979, pg. 77).

In his essay on science and politics in ethnic identification, Lowry (1982, pg. 42) asserts that some of the problems with Census data could be alleviated if social scientists were to challenge the Bureau's definitions of race and ethnicity, and the procedures the Bureau uses to collect such information. He further chastises the research community for not providing the Bureau and other federal data collection agencies with a scientifically reliable methodology for measuring and understanding ethnic diversity. While it is certain that some within the federal data collection agencies recognize the need for conducting research within a methodologically sound environment, the agencies as a whole are hampered by political and financial constraints. In order to have reliable information, researchers will have to develop both legislative and scientific strategies.

Collectively, social scientists will need to develop a plan of action for the study of the U.S. population. There is a need for both the collection of new data and the improvement of ethnic information now being gathered. A research agenda focusing on comparative ethnicity needs to establish a priority for scientifically reliable sampling and data collection methodologies, and, promote a scientific basis for ethnic concepts and identification procedures in both data collection and analysis. "Both civil rights and science would be better served by a more analytical approach to data collection and dissemination" (Lowry 1982, p. 55). This agenda should be promoted as a standard within data collection

agencies of the government, and the private sector, and should be reviewed as the social network of diverse population groups evolves. In order to compensate for the adjustments agencies make due to anorexic budgets, research funding agencies should be encouraged to support scientific research into the phenomenon of ethnic diversity. Such research should focus on concepts, sampling techniques, questionnaire development, and methods of analysis and interpretation, so that there is a reflection on all aspects of diversity that a multicultural society possesses.

REFERENCES

- APDU Newsletter. 1988. "Race/Spanish Items Set for 1990 Census." *APDU Newsletter* (February): 8.
- Cardenas, Gilbert. 1979. "Critical Issues in Using Government Data Collected Primarily for Non-Research Purposes." *Quantitative Data and Immigration Research* 55-57, Washington, D. C.: Research Institute on Immigration and Ethnic Studies, Smithsonian Institution.
- Chiswick, Barry R. 1979. "Some Data and Methodological Issues Regarding Research on the Economic Progress of Immigrants." *Quantitative Data and Immigration Research* 137-165. Washington, D. C.: Research Institute on Immigration and Ethnic Studies, Smithsonian Institution.
- Committee on Government Operations. 1982. "Hearing Before a Subcommittee of the Committee on Government Operations, House of Representatives." June 3, 1982. *Federal Government Statistics and Statistical Policy*. Washington, D. C.: U. S. Government Printing Office.
- Gerken, Ann. 1988. "What is a Data Archive and What Should the Information Specialist Know About Managing Locally Maintained Data Files?" *Database* 60-65 (August). Weston, CT.: Online, Inc.
- Hakim, Catherine. 1982. *Secondary Analysis in Social Research: A Guide to Data Sources and Methods with Examples*. London: George Allen and Unwin.
- Heim, Kathleen M. 1987. "Social Scientific Information Needs for Numeric Data: The Evolution of the International Data Archive Infrastructure." *Collection Management* 9 (Spring): 1-53.
- Kraly, Ellen Percy. 1979. "Sources of Data for the Study of U. S. Immigration." *Quantitative Data and Immigration Research* 34-53. Washington, D. C.: Research Institute on Immigration and Ethnic Studies, Smithsonian Institution.
- Lowry, Ira S. 1982. "The Science and Politics of Ethnic Enumeration." *Ethnicity and Public Policy* edited by Winston A. Van Horne (Madison, WI.: University of Wisconsin Press), pp. 42-61.
- Momeni, Jamshid. 1984. *Demography of Racial and Ethnic in the United States: An Annotated Bibliography with a Review Essay*. Westport, CT.: Greenwood Press.
- New York Times. 1988. "Congressmen Fear '90 Census Questions Will Skew the Count." *New York Times*. April 15.
- Numbers News. 1988. "Watching Washington." *Numbers News*. April 11.
- Ong, Paul. 1988. Personal communication. UCLA: Graduate School of Architecture and Urban Planning.
- U. S. Bureau of the Census. 1987. *Survey of Income and Program Participation: Users Guide*. July. Washington, D. C.: U. S. Bureau of the Census.
- _____. 1988. "Decision Paper on the 1990 Census Race Item." (May 10). Unpublished electronic mail transmission.

NOTES

1. Researchers and survey groups can pursue additional avenues in order to be informed about data collections, practices, and policies. Many data users, producers and distributors are affiliated with the Association of Public Data Users (APDU), which was organized in 1976 to promote interaction and communication among these groups with regard to federal data. Through its newsletter and annual conference, information about data sources is provided and members can make their data requirements known to federal agencies. The Consortium of Social Science Associations (COSSA) and the Council of Professional Associations on Federal Statistics (COPAFS), monitor and inform members on congressional activity as related to all aspects of social science research.

2. Data files cited in this paper are referenced and described in INDEX OF MACHINE-READABLE FILES FOR USE IN COMPARATIVE ETHNIC RESEARCH. Copies of this index can be requested from the author.