

Accessing Social Statistics

JOSEPH W. DUNCAN

Introduction

THE FEDERAL STATISTICAL system collects, compiles and publishes a vast amount of statistical information on social characteristics and conditions of the population in the United States. These social data cover such areas as population; health and nutrition; housing and environment; transportation; public safety; education and training; work; social security and welfare; income and productivity; social participation; and culture, leisure and use of time. These data are frequently underutilized because there is a lack of adequate information about these data sources and access to existing information is sometimes difficult.

This article is intended to assist users of federal social data in accessing those data. A general overview of the federal statistical system and the central coordinating office is presented, along with a discussion of its data access policy. Finally, some selected source documents which aid in accessing social statistics produced by the federal statistical agencies are listed.

Overview of the Federal Statistical System

The statistical system of the U.S. government is decentralized, with responsibility and authority for statistical activities divided by subject matter among the agencies. These organizational arrangements for

Joseph W. Duncan is Chief Statistician, Office of Information and Regulatory Affairs, U.S. Office of Management and Budget.

This paper is exempt from copyright.

producing federal statistics include several statistical collection and analytical agencies, many statistical units, and the statistical activities and outputs of major program agencies. Currently, there are over ninety federal agencies authorized to collect, tabulate and disseminate statistical data.

The origin of federal data collection activity, and in particular social statistics, can be traced back to the Constitution of the United States, which required an enumeration of the population within three years after the first meeting of the Congress and every ten years thereafter. Consequently, the first census of population was conducted in 1790. Subsequent legislation requiring the collection of statistics resulted in the formation of new statistical units in the federal government.¹ This division of responsibility for statistical activities necessitates a central agency with responsibility and authority for providing general policy guidance on the development of an integrated statistical system to meet the needs of the federal government policy-makers and other users of federal statistics. The most recent in a series of efforts to coordinate federal statistical activities is the establishment of the Statistical Policy Branch in the Office of Information and Regulatory Affairs (OIRA) in the Office of Management and Budget (OMB).

The statistical policy function was transferred to OMB effective April 1, 1981, under the requirements of the Paperwork Reduction Act of 1980 (P.L. 96-511). Section 3(a) of that act requires the president and the director of OMB to delegate to the administrator of OIRA all their functions, authority and responsibility for statistical policy and coordination under Section 103 of the Budget and Accounting Procedures Act of 1950. These include the responsibilities: "to develop programs and to issue regulations and orders for the improved gathering, compiling, analyzing, publishing, and disseminating of statistical information for any purpose by the various agencies in the executive branch of the Government. Such regulations and orders shall be adhered to by such agencies."² In addition to exercising this general statistical policy authority, Section 3504(d) of the Paperwork Reduction Act specifies that the statistical policy and coordination functions are to include:

- (1) developing long range plans for the improved performance of Federal statistical activities and programs;
- (2) coordinating, through the review of budget proposals and as otherwise provided in this section, the functions of the Federal Government with respect to gathering, interpreting, and disseminating statistics and statistical information;
- (3) developing and implementing Government-wide policies, principles, standards, and guidelines concerning statistical col-

Accessing Social Statistics

lection procedures and methods, statistical data classifications, and statistical information presentation and dissemination; and

- (4) evaluating statistical program performance and agency compliance with Government-wide policies, principles, standards, and guidelines.³

Thus, the government-wide statistical policy function extends to all social and economic statistics and includes activities such as: (1) the planning and coordination of statistical programs, agencies and issues across all departments and all subject areas; (2) the review of statistical forms and reporting plans; (3) the issuance of statistical standards and guidelines for ensuring the quality, comparability, timeliness, and accuracy of federal data; and (4) international coordination between U.S. government agencies and international organizations on statistical matters.

Data Access Policy

Recently the central coordinating unit has assumed an active role in facilitating access to federal data. As part of its effort to develop a coordinated approach to the general problem of data access, an Interagency Committee on Data Access and Use was established in April 1980.⁴ "Data access" is a broad term used by agencies to refer to publication policies and practices; development of general informational and reference materials; handling of general, public inquiries; servicing technically sophisticated data users; providing machine-readable data files (MRDF); developing user training programs; fielding user surveys; and so forth. The data access policy seeks to assist users in finding statistical information in a timely fashion for whatever the intended purpose. Users frequently have had short time horizons for locating the data, and they are unaware of potentially useful data. Their problems are complicated further by the lack of either a common format for data or a translator for the technical specifications on the data. Further, there is a shortage of resources for data analysis, and frequently special-purpose data are not always suitable for general applications. The Interagency Committee on Data Access and Use was established to address these common complaints and problems associated with using federal data.

Recommended Good Practices

The committee has developed the following selected recommended good practices for data access, which agencies are encouraged to observe

and which provide general standards against which agencies can measure their performance.

1. Federal statistical agencies should designate a specific organizational unit...[and personnel, with responsibility for improving data access in that agency].
2. All federal statistical agencies and agencies with major statistical programs should publish brochures and flyers which describe the statistics they collect and publish, and how to access the statistics; these descriptive materials should be issued and updated in a timely manner.
3. All federal statistical agencies and agencies with major statistical programs should prepare and publish reference materials which provide the user with clearly written technical guidance for using the agencies' data. Reference materials should also be regularly updated.
4. Agencies which serve sophisticated user communities should establish newsletters if the size and importance of their user community warrants it.
5. Agencies should devote special attention to the problem of adequate documentation for data files which are not available for public use but to which access is provided on some restricted basis.
6. Agencies should ensure that their programs for dissemination of information about data sources and contacts are systematically thought out and coordinated.
7. Statistical agencies are strongly encouraged to develop catalogs and directories which provide access information on major topics irrespective of which agencies hold data on the topics.
8. All federal statistical agencies should have a public inquiry service which answers queries from the general public. Telephone numbers for public inquiry should be published in agency brochures.
9. Statistical programs in agencies where statistics are embedded in a larger organization [which] does not have an inquiry service of its own should examine the larger [agency] inquiry service to ensure that queries of a statistical nature are well handled.
10. Agencies should study the utility to management of routinely logging public inquiries or at least spot-checking them periodically.
11. Agencies should sensitize all staff to proper attitudes in the handling of public inquiries.
12. Agencies should seek to promote the full range of their data products or services which have general utility....

Accessing Social Statistics

13. Statistical agencies and statistical programs must devote special attention to the problem of adequate abstracting and technical documentation for public use machine-readable data files. In particular,...each agency should establish within-agency standards for technical documentation, and ensure that the standards are enforced.⁵

The committee suggests that federal statistical agencies implement these recommended good practices as a positive step toward making statistical data more available to the public in machine-readable as well as other forms, and toward providing greater services to users of social data as well as other data.

Directory of Federal Statistical Data Files

The latest development in aids to accessing data is the issuance in March 1981 of *A Directory of Federal Statistical Data Files*. The *Directory* was a joint effort of two agencies within the Department of Commerce, the National Technical Information Service and the Office of Federal Statistical Policy and Standards. The latter agency was the predecessor of the Statistical Policy Branch of OMB. The statistical input to the *Directory* was overseen by the Interagency Committee on Data Access and Use. It was developed to assist users in accessing machine-readable data, and comprises abstracts describing statistical and related files produced by the federal government.

Since the early 1970s, federal agencies have prepared catalogs and directories on their data files; however, the scope, content, quality, and format of the individual directories varied greatly. The *Directory* integrates these individual agency sources of information through a centralized source, and represents the first effort to implement a continuing program to describe agency data files. Hence, a more fully coordinated effort involving all federal agencies producing statistical files is now being undertaken by the Statistical Policy staff. Standards are being applied to the descriptive information for the files and the preparation of the abstracts contained in the *Directory*. Each abstract contains the following eleven information items: bibliographic citation, file reference, general description, geographic coverage, time coverage, technical characteristics, reference materials, related reports, related files, contacts, availability. A sample abstract entry can be found in the appendix to this article.

The *Directory* is organized into three sections: (1) scope and use of the *Directory*, (2) file abstracts (organized by agency), and (3) appendixes. Users of the *Directory* wishing to review all files available from a particular agency should first examine the table of contents to deter-

mine if files from that agency are contained in the *Directory*. The second step, for agencies with large file holdings, is to refer to its Appendix I, which is a listing of all file titles included in the *Directory*.

An alternative way to use the *Directory* is through the subject matter indexes. Two subject matter indexes are included in the appendixes. Appendix II contains an index arranged in alphabetical order by subject matter key-word or key-word phrase. The subject matter key-words contain descriptors for the types of data items contained in the file and related identifier. Appendix III contains basically the same set of keywords as contained in Appendix II, but is arranged in alphabetical order by agency by file. Appendix III permits the user to determine which key-word phrases were associated with which files.

The ultimate objective of the *Directory* is to include all federal statistical data files which are available to the public or may be processed by federal agencies at a user's request, providing data generated from the file would meet agency releasability standards. The *Directory* focuses on major statistical data files. Major files include all federal statistical files developed expressly for public dissemination, developed at significant cost, of significant value to the public in general, or of considerable importance due to the level of user demand and the use of the file. The preparation and updating of the *Directory* is a joint effort of federal statistical and information-producing and -disseminating agencies, with continuing coordination provided by the Statistical Policy Division of OMB.

Effect of Technology on Data Use

Two recent efforts underline the importance of newly emerging technology. These are the National Indicators System (NIS) and the Decision Information Display System (DIDS). Under the auspices of the White House Office of Planning and Evaluation, and coordinated by the Statistical Policy Branch, these developments in computer technology seek to make statistical information more useful.

The National Indicators System is a program of briefings designed for systematically informing the president, vice-president and senior White House staff on social, demographic and economic trends in the United States in a policy-relevant format. The objective is to assist the president in making responsible decisions by providing periodic briefings which give an objective background description of the conditions in America. The intention is to develop a communication system which draws on the enormous statistical resources of the federal government to

Accessing Social Statistics

describe national conditions with data that relate to the policies currently pending before the president, cabinet, cabinet councils, or that are anticipated to go before them. The system is designed to be an honest broker of information linked to policy planning, but it is not part of the regular policy advocacy process. The responsibility for developing the briefing materials is assumed by the participating federal agencies under the general coordination of the Statistical Policy Branch.

The Decision Information Display System was conceived as a technique for displaying statistical information on a geographic basis so that White House, executive branch and congressional staff could see the impact of issues across states, counties and other political jurisdictions in the United States. It is designed for rapid graphic presentation of social and economic variables about various geographic regions. DIDS is a cooperative, interagency-funded program for the application of information technology in the statistical community. DIDS currently has approximately 3500 data sets immediately accessible to meet demands for data in a short time frame. The data bases are accessible to the user in a menu-driven approach, so that by entering a simple numeric code from a list of data bases, information on a particular topic is available. Since the information display is oriented toward geographically defined information, DIDS forces a common format on the data bases.

Source Documents

As part of its coordinating effort, the Office of Management and Budget's Statistical Policy Branch prepares publications to inform users of developments affecting federal statistics and to aid them in accessing the data. A general overview of federal social statistics can be found in the following publications issued by the Statistical Policy Branch and its predecessor organizations:

Statistical Services of the United States Government, 1975. This document includes brief descriptions of principal economic and social statistics programs, as well as a list of the principal statistical publications prepared by the federal statistical agencies.

A Framework for Planning U.S. Federal Statistics for the 1980's. This publication presents a comprehensive review of all major statistical programs and data series, as well as the publications that are available as a result of these data collection activities.

Statistical Reporter. This monthly publication contains information on newly issued publications and machine-readable data file.

In addition to the above reports, *Social Indicators III*, prepared by the U.S. Bureau of Census, presents a comprehensive variety of statistical information on the current social situation in the United States, along with selected bibliographies. While these publications contain information on accessing social data from the federal statistical system in general, the following is a selected bibliography by agency for users interested in particular subject-matter areas.

Bureau of the Census

Directory of Data Files. This publication contains structured descriptions of public use data files and computer software available from the Census Bureau. Available on a subscription basis, the publication is updated regularly.

Data Access Descriptions. Issued irregularly, 1967 to present. Contains user-oriented, topical reports on accessing and using census data and products.

Directory of Federal Statistics for Local Areas: A Guide to Sources, 1976, March 1978. Describes and cites individual tables from reports issued by federal agencies. Subjects and geographic areas smaller than states are displayed in tabular format.

Directory of Federal Statistics for Local Areas: Urban Update, 1977-78, 1981. Similar to the directory immediately above, except that geographic focus is on cities, standard metropolitan statistical areas, labor market areas, etc.

1980 Census Users' Guide (forthcoming). A guide to the 1980 Census of Population and Housing.

Reference Manual on Population and Housing Statistics from the Census Bureau, March 1977. A comprehensive discussion of census and survey reports and computerized products, with attention to limitations as well as strengths of the various data and to skills required to find specific data.

Factfinder for the Nation, Series CFF, issued irregularly. A series of topical brochures describing the range of census materials available on a given subject. In particular, Factfinder No. 5, "Reference Sources," provides a bibliography of Census Bureau data access publications.

Census '80 Introduction to Products and Services, November 1979. A general introduction to the 1980 Census of Population and Housing data.

Housing Data Resources: Indicators and Sources of Data for Analyzing Housing and Neighboring Conditions, 1980.

Accessing Social Statistics

General Accounting Office

Federal Information Sources and Systems 1980. This publication describes federal sources and systems maintained by the executive agencies which contain fiscal, budgetary, and program-related information. The report is prepared as a part of the *Congressional Sourcebook Series* and is an update to a volume prepared in 1977. The report contains descriptions of many statistical files prepared by government agencies.

Bureau of Labor Statistics

BLS Data Bank Files and Statistical Routines. This publication generally describes public use data files and computer software produced and distributed by the agency. Prepared in 1978, the report is now out of print. Updates are now being prepared and a revised report may be released by BLS.

National Archives and Records Service

Catalog of Machine-Readable Records of the National Archives of the United States. This catalog is a second edition publication which describes machine-readable data files accessioned by the National Archives. A new catalog is now being prepared and is expected to be available in 1981. The publication will contain statistical and other types of files.

National Center for Education Statistics

Computer Tapes Available from NCES. This report contains brief summary descriptions of files which are available for public use through the National Center for Education Statistics. The report is updated frequently.

Directory of Federal Agency Education Data Tapes. This publication contains structured descriptions of data files produced by federal agencies which contain education-related data. The report was prepared in 1979.

National Center for Health Statistics

Catalog of Public Use Data Tapes from the National Center for Health Statistics. This report generally describes NCHS statistical programs from which data are produced in machine-readable form, and summarizes the content and structure of public use files prepared. The report is updated annually.

Facts at Your Fingertips: A Guide to Sources of Statistical Information on Major Health Topics, 4th ed., fall 1979. This guide lists sources of

statistical information on over 100 major health topics, and references NCHS data on particular topics and other public and private sources of information.

NCHS Publications on the Health of Minorities, May 1980. An annotated bibliography of selected NCHS reports which present data on health topics of special interest to women.

NCHS Publication on the Health of the Elderly, November 1980. An annotated bibliography of selected NCHS reports which present data on health topics concerning the elderly population.

National Technical Information Service

Directory of Computerized Data Files and Computer Software. This report contains descriptions of data files and computer software distributed by the agency. Originally prepared in 1974, the publication was updated in 1976; it is now out of date,

Office of Federal Statistical Policy and Standards and National Technical Information Service

Directory of Federal Statistical Data Files, March 1981. The publication contains structured descriptions of statistical data files produced by the federal government.

Social Security Administration

Research Publications and Microdata Files: Spring 1980. This report contains a brief summary description for individual microdata files produced by the SSA Office of Research and Statistics. Content, technical features, and availability are described. The catalog is updated frequently.

Lois A. Alexander and Thomas B. Jabine. "Access to Social Security Microdata Files for Research and Statistical Purposes." *Social Security Bulletin*, vol. 41, No. 8, August 1978. This article focuses on the characteristics of SSA microdata files and on the development of a disclosure policy aimed at serving the public interest while protecting the privacy of individuals and the confidentiality of research and statistical information.

Wendy Alvey. "Policy Analysis with Social Security Research Files." *Statistical Reporter*, May 1980. This article describes SSA's workshop on policy analysis with social security research files held March 1978 in Williamsburg, Virginia.

"Some Statistical Research Resources Available at the Social Security Administration," April 1979. This brochure describes the nature and

Accessing Social Statistics

availability of data from SSA's Lifetime Earnings Records and Continuous Work History Sample.

Bureau of Justice Statistics

Inter-University Consortium for Political and Social Research. *Guide to Resources and Services: 1979-80*. Ann Arbor, Mich.: Institute for Social Research, 1980. This *Guide* provides information on all of the resources and services offered by ICPSR of which the Criminal Justice Archive and Information Network (CJAIN) is a part. These include data archiving, data access functions related to the archives, training in a variety of areas related to ICPSR, and computing assistance.⁶

JOSEPH DUNCAN

Appendix

Sample Abstract From *A Directory of Federal Statistical Data Files*

Census of Population and Housing, 1970 Summary Statistic File 1 (machine-readable data file) prepared by the Bureau of the Census. Washington: The Bureau of the Census (producer and distributor), 1971.

FILE REFERENCE: 101.1010/100180

GENERAL DESCRIPTION:

The Census of Population and Housing, 1970 Summary Statistic File 1 is a series of summary statistic files, each containing detailed characteristics of the U.S. population and housing by geographic area based upon the 1970 census completed count (100%) questionnaires. The files are also known as the 1970 Census First Count Summary tapes. Two types of files—A and B—were prepared separately for each state and include the same subject matter data but contain summary statistics for different levels and types of geography. Each file contains records which correspond to an individual geographic area and contain population data by age, race, sex, and family and household characteristics and housing data descriptive of units and their occupancy.

GEOGRAPHIC COVERAGE:

The geographic coverage differs for each of the file types. File A contains summary statistics for enumeration districts and block groups. File B contains summary statistics for counties, minor civil divisions (or census county divisions), places and congressional districts.

TIME COVERAGE:

Data contained in the files pertain to the date of the census—April 1, 1970.

TECHNICAL CHARACTERISTICS:

Files are packaged and distributed on a state separate basis. Due to the volume and complexity of the structure of these files, please consult the reference materials cited below for further information on size and technical characteristics.

REFERENCE MATERIALS:

1970 Census Users' Guide, Parts I and II, U.S. Bureau of the Census, 1970. See in particular the Part II technical documentation.
Index to the 1970 Census Summary Tapes, U.S. Bureau of the Census.

RELATED REPORTS:

Census of Population, 1970, Volume 1, Characteristics of the Population PC(1), U.S. Bureau of the Census.

Accessing Social Statistics

RELATED FILES:

Census of Population and Housing, 1970.

- a. Other summary statistic files.
- b. Public use sample files.

CONTACTS:

Chief, Customer Services Branch
Data User Services Division
U.S. Bureau of the Census
Washington, D.C. 20233
(301)449-1600

AVAILABILITY:

The files are available from the Bureau of the Census for \$80 per output reel. File order numbers are CSUM70001a(File 1 A) and CSUM70001b(File 1 B).

JOSEPH DUNCAN

References

1. For more specific details on the origins of the federal statistical agencies, see Duncan, Joseph W., and Shelton, William C. *Revolution in United States Government Statistics, 1926-1976*. Washington, D.C.: U.S. Office of Federal Statistical Policy, Department of Commerce, 1976.
2. "Budget and Accounting Procedures Act of 1950." 31 *U.S. Code* §18b.
3. *U.S. Statutes at Large*, vol. 94, P.L. 96-511, §3504d.
4. See Sprehe, J. Timothy. "A Federal Policy for Improving Data Access and User Services." *Statistical Reporter* 81(March 1981):323-44. The appendixes to this article contain lists of statistical data file catalogs and directories, and other data access publications which aid users in accessing federal data.
5. *Ibid.*, pp. 328-32+.
6. *Ibid.*, pp. 341-44.