

APPROACHES TO INFORMATION ABOUT
CENSUS BUREAU MACHINE-READABLE
DATA FILES (1)

Barbara Aldrich
Data User Services Division
Bureau of the Census
U.S. Dept. of Commerce

During the last several years, the community of Census Bureau data file users has expanded greatly and the number of files available has increased dramatically. This phenomenon will continue when files are released from the 1977 Economic Censuses and the Census of Governments and the 1980 Census of Population and Housing.

As the number of data files distributed by the Census Bureau increases, there is need for a systematic method of providing information about existing holdings of data files and updating that information as new files are released.

While current methods of announcing new file releases reach a very large segment of the user community, the need exists for a single reference publication containing information about the complete collection of files available from the Bureau. This is the need the Bureau proposes to meet with a new publication now available.

In this paper, the current means of data information dissemination will be discussed, followed by a detailed explanation of Directory of Data Files a new Census Bureau publication prepared by the Data User Services Division.

Currently, archivists and researchers obtain initial information about Bureau files from several sources. Data files released during the year are described briefly in Part II of the annual Bureau of the Census Catalog. More current releases are announced in Data User News (published monthly by the Data User Services Division) and in non-Bureau publications such as American Demographics and Review of Public Data Use.

Another less formal means of "spreading the word" about new data files is the data archivist/researcher network. Particularly active at major universities, this informal information network operates outside the constraints of printing and mailing delays.

Although each of these mechanisms for providing information about new files serves a particular audience, none are standardized regarding the level of data information provided or the format in which the information is presented. Nor is there a method of maintaining this information in an easily consulted, cumulative form.

However, this information need will be met by the Directory of Data Files (2), prepared within the Data User Services Division of the Census Bureau. This directory is a cumulative reference guide

-
- (1) Prepared for the 1979 International Association for Social Science Information Service and Technology (IASSIST) Annual Conference, May 7-10, 1979, Ottawa, Ontario. Updated March, 1980.
(2) The Directory of Data Files was prepared by Molly Abramowitz and Barbara Aldrich under the direction of Larry Carbaugh, Chief, Customer Services Branch, Data User Services Division.

for all publicly available machine-readable data files. It serves a two-fold purpose by indicating that a particular file is available and providing brief but concise information regarding its universe, subject-matter, geographic level, and size.

The Directory is a subscription item available through the U.S. Government Printing Office. Subscribers to the basic publication will automatically receive quarterly updates containing information about new files. The Directory is in a loose-leaf format so all update pages can be entered in the appropriate place.

This publication does not replace any of the current systems of data information dissemination discussed earlier nor is it intended to replace the technical documentation as the source of detailed information about individual files.

An abstract is written for each file in the Directory. Each abstract provides enough information so that the user can determine if the file is of interest or if the more detailed technical documentation for the file is needed.

The Directory reflects the efforts of the Bureau to provide authority titles for files and a standardized framework for information about them. This effort complements the on-going development of standards for the Bureau's technical documentation.

The Directory begins with a general introduction to census data and a review of file ordering procedures. The next chapter is divided into program areas: agriculture, economics, general, geographic reference, government, population and housing, and software. A general overview is provided for each. Following this, individual censuses and surveys are outlined, followed by abstracts of the available files produced in these efforts. For example, the population and housing section contains a general outline of the Bureau's role in collection of population and housing data. This is followed by an overview of the Census of Population and Housing and abstracts of the files that are products of that Census. These abstracts are a major portion of the Directory.

Information provided in the abstract includes the bibliographic entry for the file, type of file, descriptions of universe, subject-matter, and geographic coverage, file size, related reference materials, printed reports and machine-readable data, and file availability.

The BIBLIOGRAPHIC ENTRY for the file contains the authority title, the statement of responsibility, place of production, producer and distributor, and the date of production.

EXAMPLE:

Current Population Survey, October 1968-1978 / conducted by the Bureau of the Census for the Bureau of Labor Statistics. - Washington : The Bureau, 1978.

(All subsequent examples are based on this file.)

The bibliographic entry is followed by a statement of TYPE OF FILE and the UNIT OF OBSERVATION. Within the framework of Census Bureau data, the options for type of file are microdata, summary statistics, geographic reference data and software.

TYPE OF FILE:

Microdata, unit of observation is individuals within housing units.

The next element in the abstract is the UNIVERSE DESCRIPTION. If the data are sample data, a brief description of the sampling technique is also provided.

UNIVERSE DESCRIPTION:

The universe consists of persons 3 years old and over who are the civilian non-institutional population of the U.S. living in housing units. A probability sample is used in selecting housing units.

A SUBJECT-MATTER DESCRIPTION follows. This is not intended to be a complete list of the individual variables on the files, but rather to provide an overview of the type of data available on the file.

SUBJECT-MATTER DESCRIPTION:

The basic information collected monthly provides data on labor force activity the week prior to the survey. Comprehensive data are available on the employment status, occupation and industry of persons 14 years and over. Characteristics such as age, sex, race, marital status, household relationship, educational background and Spanish origin are shown for each person 14 and over in the household survey.

In addition, supplemental surveys are sometimes taken at the same time the basic data are gathered. The Current Population Survey, October 1968-78, files contain supplementary data on school enrollment. Items include current school enrollment, type of school in which enrolled and grade attending for all persons 3 and older. Other school enrollment data varies between supplements. The Current Population Survey, October 1974 contains post-secondary school information as well as the school enrollment data.

The GEOGRAPHIC COVERAGE elements provide the information concerning the levels of geographic identification used in the file.

GEOGRAPHIC COVERAGE:

In the years 1968-72, geographic identification includes regions, divisions, 19 States (in the remainder several states are grouped together for identification) and 19 SMSA's. In the years 1973-76 geographic identification includes regions, divisions, 13 States and 34 SMSA's. Beginning in 1977, geographic identification includes regions, divisions, all States and 44 SMSA's.

The FILE SIZE element reports the number of files and logical record count of the file(s). In the case of a series like the CPS, when the size varies from year to year, the approximate range is given.

FILE SIZE:

1 file per year, logical record count ranging from approximately 130,000 to 146,000.

The element RELATED REFERENCE MATERIALS is designed to provide the user with a guide to more information about the files. It references the technical documentation as well as other materials regarding data-collection or sampling. It also provides information on how these materials can be obtained.

REFERENCE MATERIALS:

U.S. Bureau of the Census. The Current Population Survey : Design and Methodology. (Technical Paper 40- U.S. Bureau of the Census) Superintendent of Documents No: C3,212:40. Government Printing Office, Washington, D.C. 20402. Price \$3.25.

This comprehensive document gives detailed information on the CFS program including sample design and rotation, survey operations, and preparation and accuracy of estimates as well as sampling errors. Thirteen appendices are provided which contain very useful information to a CFS user.

"Current Population Survey, October (year) Technical Documentation". Available from the Customer Services Branch.

This is a guide to the machine-readable data file. It has general information about the data, specific content information and a codebook.

A separate abstract element RELATED PRINTED REPORTS cites publications which contain actual data produced from the files being abstracted.

RELATED PRINTED REPORTS:

U.S. Bureau of Labor Statistics. Employment and Earnings, November (year).

The employment information in Section A of this document is derived from the Current Population Survey of the previous month.

Data from the Current Population Survey, October file is usually published in the Current Population Reports P-20 series. The publications for specific years can be found in the Bureau of the Census Catalog. Examples of some of these reports are listed below:

College Plans of High School Seniors: October 1974, P-20 No. 284.

College Plans of High School Seniors: October 1975, P-20 No. 299.

Major Field of Study of College Students: October 1974, P-20 No. 289.

School Enrollment — Social and Economic Characteristics of Students: October 1975, P-20 No. 303.

Nursery School and Kindergarten Enrollment of Children and Labor Force Status of their Mothers: October 1967 to October 1976, P-20 No. 318.

School Enrollment — Social and Economic Characteristics of Students: October 1976, P-20 No. 319.

The last element of the abstract is FILE AVAILABILITY. This section indicates how the file is packaged and sold (by Census division, State, all files as a single unit) and identifies individual file order numbers for each file. Appendix C of the Directory is a chart arranged by file order number which indicates the number of reels required providing several density and blocking options.

Appendix A provides a glossary of Census geographic concepts and Appendix B is a bibliography of Census publications relating to machine-readable data files.

This publication, issued October 1979, is sold by the U.S. Government Printing Office. The cost (which includes the quarterly updates) is \$11. Orders may be placed with Subscriber Services Section (Publications), Bureau of the Census, Washington, D.C. 20233.