
Distribution Of Census Data On CD-ROM To Depository Libraries

by *Juri Stratford*¹
Documents Librarian
Shields Library
University of California, Davis

INTRODUCTION

The Depository Library Program (DLP) was established by Congress to inform the public on the policies and programs of the Federal government. Through the DLP, the Government Printing Office (GPO) distributes publications to designated libraries. While government documents received through depository distribution remain the property of the Federal government, depository institutions are responsible for the maintenance of, and providing public access to, the documents. This usually involves committing public service and technical service staff, and the purchase of bibliographic tools in addition to committing space.

In the past, depository libraries have provided public access to the printed Census publications while access to Census machine-readable datafiles (i.e. tapes) has been available through State data centers and data archives. The Census Bureau is now beginning to distribute machine-readable datafiles on CD-ROM to depository libraries. This affords depository libraries both new opportunities and new responsibilities.

THE DEPOSITORY LIBRARY PROGRAM

There are almost 1,400 depository libraries. They are located in each state and Congressional district in order to make government publications widely available. These government publications are available for the free use of the general public. For the purpose of depository distribution, a government publication is defined as "informational matter which is published as an individual document at Government expense, or as required by law." [44 USC 1901]

The origins of the DLP can be traced back to the 1790s when the State Department distributed acts of Congress to State governments and newspapers. Funding was sought on an ad hoc basis until 1813 when Congress passed a resolution authorizing "every future Congress" to print additional copies of Congressional publications for this purpose. In 1814, the American Antiquarian Society was designated the first depository library. Responsibility for the program shifted between various agencies and departments, mainly the Department of State and the Department of Interior, throughout 19th Century. Congressional resolutions in 1857 and 1858 affirmed the distribution of congressional materials to

institutions such as libraries and colleges, and other organizations designated by Members of Congress. In 1895, a new printing act was passed, incorporating the old legislation and placing responsibility for the DLP in the office of the Superintendent of Documents at GPO; the act also specified that certain executive materials were to be included.²

The present law, the Federal Depository Act of 1962, increased the number of possible depository libraries; established a system of regional depository libraries which were to maintain a permanent collection, and provide interlibrary loan and reference service; expanded the variety of government documents available for distribution; and established a reporting mechanism, the Biennial Survey, to ascertain the libraries' condition. The 1962 Act has been amended twice: in 1972 to exempt the highest appellate court of each State from the requirement of public access; and in 1978 to extend depository eligibility to law schools.³

DEPOSITORY DISTRIBUTION OF MACHINE-READABLE DATAFILES

GPO has reversed its position on depository distribution of machine-readable datafiles since the early 1980s. At their Fall 1981 meeting, the Depository Library Council (DLC), an advisory body to the Public Printer and the Superintendent of Documents, passed a resolution regarding the feasibility of the GPO providing free access for depository libraries to unclassified bibliographic data bases belonging to Federal agencies. In response to the resolution, GPO general counsel Garrett Brown determined that "...the Depository Library Act of 1962 does not direct the Superintendent of Documents to make published documents available in all possible formats to the libraries. It was the intent of Congress that only printed publications be made available to depositories."⁴

Following the Census Bureau's plans to distribute data from the 1982 Census of Agriculture and the 1982 Census of Retail Trade through the depository library program on CD-ROM as Test Disc 2, the Public Printer requested approval through the Joint Committee on Printing (JCP). In a March 25, 1988 letter to the Public Printer, Congressman Frank Annunzio, Chairman of the JCP, affirmed the Committee's support of the Census project and the GPO's authority to produce and distribute

Government publications in electronic formats.⁵

In 1989, GPO asked its General Counsel Grant G. Moy, Jr. to review the 1982 opinion. He concluded that "the earlier question presented to the General Counsel concerned only the issue of access to unpublished information in a computer data base," and, this was still the case; but "the specific statement in the General Counsel's 1982 opinion, limiting depository distribution to printed products, was disapproved."⁶

Test Disc 2 was distributed to 173 depository libraries as a pilot project in September 1988. The CD was available through regular depository distribution in 1989. At the March 1989 Depository Library Council meeting, Jan Erickson of the Government Printing Office reported on the initial distribution of Test Disc 2, and indicated that it was the Census Bureau's intent to distribute future CD-ROM products through the depository system.⁷ A second Census CD, The City and County Data Book, was distributed to depository libraries in Spring 1990. Software accompanying early shipments of the City and County Data Book were "infected" with the Jerusalem virus.

TEST DISC 2

The Census Bureau's CD-ROM Test Disc 2 contains data from the 1982 Census of Retail Trade and the 1982 Census of Agriculture on a single compact disk. The files are in dBase III format.

Files from the Census of Agriculture contain 1982 data by county, with comparable data for selected items from the 1978 census. The technical documentation for the Census of Agriculture describes the data as a single file with a logical record size of 40,320 characters containing 3,360 data fields. However, the dBase III record structure only allows 128 fields. This large record structure is accommodated by storing the data in 28 separate dBase III compatible files. The first file, AG82_GEO.DBF, provides geographic information for each state and county. This information is a guide to the arrangement of data contained in the 27 numeric datafiles. For example, the geographic area indicated by record #170 in AG82_GEO.DBF is "California." This means that state level data for California in the other 27 files is contained in record #170. Each file is named AG82_NN.DBF where NN = 1 to 27. Except for the last file, which contains the last 32 fields, each file contains 128 fields.

Data from the 1982 Census of Retail Trade are available for each 5-digit zip code, including number of establishments, by kind of business, and basic data for retail trade total. The files for the Census of Retail Trade have a much simpler record structure than the files for the Census of Agriculture. The 1982 Census of Retail Trade data are stored in 51 separate files. Each file is named

RC82_XX.DBF where XX = the state postal abbreviation. Data files for the retail trade data have a record length of 155 characters containing 19 fields. There are two files for each state, a DBF file and an NDX file.⁸

COUNTY AND CITY DATA BOOK

The 1988 County and City Data Book CD-ROM contains the same data as published in the printed volume. These files contain data gathered from a variety of Federal agencies and national associations. The disc includes data for states, counties, cities with a population of 25,000 or more, and places with a population of 2,500 or more. Like the Census of Agriculture files on Test Disc 2, each record represents a geographic area, and subject fields are distributed across a number of datafiles. The datafiles range in size from about 200kb to 1mb.

THE ECONOMIC CENSUSES AND THE CENSUS OF POPULATION AND HOUSING

Almost all data from the Economic Censuses previously available on magnetic tape will be on CD. The Economic Census and Census of Agriculture will be available on 9 CDs. The first disc, Volume 1, release 1A was released in early 1990, but has not yet been distributed to depositories. The disc contains data from the geographic area series for wholesale, retail, and service industries for selected states and includes the same statistics statistics as published in the corresponding report series: Geographic Area Series for 1987 Censuses of Retail Trade (all states), Wholesale Trade (selected states), and Service Industries (selected states); Preliminary Industry Series for the 1987 Census of Manufactures (national with some state totals) and selected historical statistics. Plans for the 1990 Census of Population and Housing call for 20-30 CDs to be released from mid- 1991-1993, including redistricting data and block statistics.⁹

DATA EXTRACTION

Each of these CD-ROM products is distributed with a program to display tables, but software is not provided to copy data subsets. The datafiles range in size from 30kb to 3mb; most of these files are too large to manipulate on a microcomputer without first creating a smaller data subset that can be copied to a floppy disk or hard disk. The texts documenting Test Disc 2 and the City and County Data Book discuss using dBase III to work with the files.

The documentation for the Economic Censuses describes the EXTRACT program. The EXTRACT program is a public domain program that was developed to create subsets from the large CD-ROM databases and save them as files on a floppy or hard disk. Extracted files can be created in dBase format, ASCII fixed field format, or ASCII comma-delimited format. Version 1 of EXTRACT is slow and occasionally crashes. A new version of EXTRACT should be released shortly. EXTRACT is

available from the Center for Electronic Analysis, University of Tennessee. However, it has not been distributed with the CD-ROMs.

CONCLUSION

In their paper, "Government Information in Machine-Readable Data Files: Implications for Libraries and Librarians," Ray Jones and Thomas Kinney examine the requirements for the utilization of machine-readable datafiles in retrieval and reference services. Two of their remarks can be paraphrased to apply to the utilization of the Census CD-ROMs in depository libraries. First, when librarians have the responsibility of retrieving numeric information from CD-ROMs either they must know how to program or work with a colleague who programs; and second, librarians will require the critical judgment to determine when data retrieval from CD-ROMs is needed to answer the patron's need most completely.¹⁰ These skills are not widely held by depository librarians at the present time. This is evidenced by the fact that few depository libraries have successfully integrated these materials into their reference service.¹¹

The Census Bureau is currently reviewing the impact of the CD-ROM distribution. The report, *The Role of Intermediaries in the Interpretation and Dissemination of Census Data Now and in the Future*, by Census statistician Sandra Rowland examines these issues. The study credits the experience of librarians assisting in the understanding and use of census data; however, it concludes that "neither the GPO nor the libraries play a big part in the interpretation of data for users" and argues that role of depository libraries is "unlikely to change in the future unless librarians take a more aggressive role as information technicians." It also states that, while the Regional Depository libraries will acquire and hold all census products including data on high density optical storage devices, most depository libraries will acquire and hold fewer census products in the future than they do now.¹²

At present, the Census Bureau appears to have a strong commitment to the depository distribution of their CD-ROM products, and these materials are available to all depository libraries. The depository library community needs to work closely with data archivists to insure that effective use is made of the Census CD-ROM products. Data archivists might try to meet informally with depository librarians within their own institutions to discuss how access to Census data on tape, CD-ROM, and paper copy could best be coordinated. Data archivists might also consider coordinating presentations with state or national government document groups.

Finally, while the Census Bureau assures us that the CD-

ROM production of the 1987 Economic Census and the 1990 Census of Population and Housing will not be produced at the expense of the publication of the tape or paper products, Sandra Rowland's report suggests that we can expect to see fewer paper products and more electronic products for the 2000 Census: "With respect to the year 2000 census, it is very likely that there will be a movement out of printed media and into electronic media for dissemination to the libraries."¹³ While there are certainly instances where researchers' needs would best be served by Census data on CD-ROM, documents librarians and data archivists alike need to monitor the situation to insure that the CD-ROMs are not produced at the expense of other necessary Census products.

¹ Presented at the IASSIST 90 Conference held in Poughkeepsie, N.Y. May 30 - June 2, 1990.

² Hemon, Peter, Charles R. McClure and Gary P. Purcell, *GPS's Depository Library Program: A Descriptive Analysis* (Norwood, New Jersey: Ablex Publishing Company, 1985), pp.3-7.

³ U.S. Congress, Joint Committee on Printing, *A Directory of U.S. Government Depository Libraries* (Washington, D.C.: U.S. Government Printing Office, 1988), pp.1-3.

⁴ U.S. Congress, Joint Committee on Printing, *Provision of Federal Government Publications in Electronic Format to Depository Libraries: Report of the Ad Hoc Committee on Depository Library Access to Federal Automated Data Bases...* (Washington, D.C.: U.S. Government Printing Office, 1984), pp.112-13.

⁵ U.S. Congress, Office of Technology Assessment, *Informing the Nation: Federal Information Dissemination in an Electronic Age* (Washington, D.C.: U.S. Government Printing Office, 1988), p.143.

⁶ *Ibid.*

⁷ "Summary: Spring Meeting, Depository Library Council, Pittsburg, Pennsylvania, March 8-10, 1989," *Administrative Notes* 10 (August 1989): 3.

⁸ Stratford, Juri, review of "CD-ROM Test Disc 2 (machine-readable data file) and CD-ROM Test Disc 2: Technical Documentation," *Government Publications Review*, 16 (1989): 397-398; see also Peter Hemon and Candy Schwartz, "Readers Exchange: Census Product Review," *Administrative Notes* 11 (March 1990): 13-21.

⁹ *Administrative Notes* 10 (August 1989): 3.

¹⁰ Jones, Ray and Thomas Kinney, "Government Information in Machine-Readable Data Files: Implications for Libraries and Librarians," *Government Publications Review*, 15 (1988): 25-32.

¹¹ Diane Smith examines the viability of federal depositories to deal with electronic information in her forthcoming paper, "Depository Libraries in the 1990s: Whither or Wither Depositories?," Government Publications Review, 17 (1990).

¹² Rowland, Sandra, The Role of Intermediaries in the Interpretation and Dissemination of Census Data: Now and in the Future, (Washington, D.C.: U.S. Bureau of the Census, 1989), photocopy, pp.32-33.

¹³ Op cited, p.14.

DATA NEWS

Here are two items that readers of the Quaterly may find of interest. (Submitted by: Jim Jacobs jjacobs@ucsd.edu)¹

1) In the newest edition of the ASIS "Annual Review of Information Science and Technology" (vol. 25, 1990, Martha E. Williams, ed., published for ASIS by Elsevier Science Publishers pp. 3 - 54), Karen J. Sy of the University of Washington and Alice Robbin of the University of Wisconsin have written an article titled: "Federal Statistical Policies and Programs: How Good Are the Numbers?"

In their concluding remarks they say, "...the 1980s saw a serious decline in the quality of our federal statistical system." And, "Throughout the government, but especially in the key coordinating branch of the Office of Management and Budget, we found policy makers who viewed statistics as a burden and not as indicators of who we are as a nation and where we should be going. In sum, the federal statistical system is in serious trouble."

2) The Committee on National Statistics and the Social Science Research Council with support from several agencies, have convened a Panel on Confidentiality and Data Access. The scope of the study includes publicly supported statistical data collection activities. The Panel is in the middle of a two year study and is soliciting short statements from interested parties on the following topics:

Access Problems. (examples of instances where confidentiality law or policies have made it impossible to obtain data.)

Suggestions for Improving Access. (including suggestions for improving access with appropriate safeguards to maintain confidentiality.)

Persons or Businesses Harmed by Disclosure.

You can submit statements to:

George T. Duncan, chair
Panel on Confidentiality and Data Access
Committee on National Statistics
National Academy of Sciences
2101 Constitution Ave, NW
Washington DC 20418

If you have questions, or if you want a more detailed announcement of the charge to the Panel, you may call Virginia de Wolf, Study Director, (202) 334-2550.

¹ Jacobs, Jim. 1991 "Quality of Government Statistics" [computer file]. Edmonton, Alberta: [OR-L](#). Electronic listserv. (LISTSERV@UALTAVM).