
The Depository Distribution of CD-ROMs: A Review of the First Year

by Juri Stratford¹
University of California, Davis

Introduction

A large part of the work of depository librarians is providing public access to the vast number of statistical publications produced by various agencies in the Federal government including most notably the Census Bureau and the Bureau of Labor Statistics. Until 1989, the distribution of Federal statistical data in machine-readable form was limited to programs administered by individual agencies, e.g. the Census Bureau's State data center program and the Bureau of Economic Analysis' local area data program, and institutions acquiring data directly from agencies for purchase or through consortiums and archives such as ICPSR.

The first CD-ROM to be distributed to depository libraries was **Test Disc 2**. **Test Disc 2** included state and county data from the 1982 Census of Agriculture and zip code data from the 1982 Census of Retail Trade. This was distributed to a few libraries on an experimental basis in 1988 and was made available through regular depository distribution in 1989 shortly followed by distribution of the **City and County Data Book** CD-ROM in 1990.

Since then the Census Bureau has distributed a number of CD-ROM products through the depository system including data from the 1987 Economic Censuses, the 1987 Census of Agriculture, and the reapportionment data from the 1990 Census of Population and Housing. Other significant CD-ROM products distributed to depositories include the **National Health Interview Survey** produced by the National Center on Health Statistics, the **Toxic Release Inventory** produced by the Environmental Protection Agency, and the **National Trade Data Bank** produced by the Commerce Department. A number of other agencies including the Department of Defense, NOAA, and the Geological Survey are also beginning to distribute CD-ROMs through the depository system.

CD-ROM and Papercopy Distribution

In some cases, the depository distribution of CD-ROMs complements the depository distribution of paper or microfiche products. For example, the EPA's **Toxic Release Inventory** was made available simultaneously to depositories on CD-ROM and on microfiche. The CD-ROM distribution of the 1987 Census of Agriculture and

the 1987 Economic Censuses followed the distribution of these publications to depositories in paper copy. This has also been the case, so far, with the CD-ROM distribution of the **City and County Data Book** and **County Business Patterns**.

In other instances, the CD-ROM distribution provides depositories with materials that they might not have had otherwise. For example, the CD-ROM distribution of the **National Health Interview Survey** data, the PL94 data from the 1990 Census of Population and Housing, and the zip code data from the 1982 and 1987 Economic Censuses represent data not available in paper copy.

A final area, which should concern the depository community, is the replacement of the depository distribution of paper copy or microfiche with CD-ROM. For example, the monthly import and export data from the Census Bureau is now only being distributed to depositories on CD-ROM. Also, the 1985 Congressional Record CD-ROM was recently distributed to depositories on a trial basis as a replacement for the bound edition.

Advantages and Disadvantages of Data on CD-ROM CD-ROMs offer some advantages both to end users and data producers. The electronic distribution of data provides the potential to enhance user access. Libraries can provide access to a vast quantity of Federal data in machine readable form allowing end users the ability to work with the data on their own microcomputers. Where in the past researchers might have had to extract data subsets from tape, or to key in data by hand from printed sources, they can now copy the files directly from the CD-ROMs to diskette. Data from these CD-ROMs are available free of charge and without copyright restrictions.

CD-ROM also offers some advantages to data producers. In many instances, it is less expensive to produce and distribute CD-ROMs than paper copy, and Congress is anticipating a cost-saving through CD-ROM. In a statement before the American Library Association Legislation Committee in January, Robert W. Houk, the Public Printer reported that GPO requested a fifteen percent increase for salaries and expenses primarily associated with the distribution of 1990 Census publications. However, Congress reduced the request from

\$27.9 million to \$26.5 million, projecting that the Census Bureau would distribute a greater proportion of documents in CD-ROM formats, rather than in microfiche as originally anticipated².

When materials in paper copy or microfiche are replaced by CD-ROM, the CD-ROM distribution may be viewed as shifting the expense of production from the data producer to the data user. While many depository institutions are facing budget reductions, they are now also faced with the added expense of devoting CD-ROM stations to provide public access to the Federal data. Other expenses will include acquiring the proper software to work with the data and providing paper for printouts.

Data Format and Software Issues

The Census Bureau data are being distributed in dBase format. For the demographic data, the Census Bureau is distributing software to display tables, usually for specific geographic areas, to the screen or to a printer. The software for the foreign trade data displays data for the current month and beginning of current year to date for the most specific commodity code. The Census Bureau is also distributing programs such as Extract to create small data subsets for output as ASCII files, dBase files, or Lotus worksheet files. Extract can also print tables from the data sets.

The Extract program requires a large number of dictionary files describing the data. The dictionary files for the Economic Census are about four megabytes and these files have been included on the CD-ROMs. However, in other instances these files have been distributed separately on floppy disks, and they must be installed on a hard disk for use with the CD-ROMs. The City and County Data Book files require 640kb of hard disk space, County Business Patterns files require 650kb, and the monthly import and export files require 3.2 megabytes.

As the Census CD-ROM files are in dBase III format, the data can also be accessed directly using third party database management and spreadsheet software. However, many of the Census Bureau files are large and are very difficult to work with on a microcomputer. The City and County Data Book files are small enough, and most of the Economic Census files are less than one megabyte, though a few are as large as two megabytes. However, some of the PL94 files are as large as 300 megabytes; and the foreign trade export files are currently about 258 megabytes while the import files are more than 550 megabytes.

Public Access Issues

Regardless of what software is used, whether a library uses Extract or dBase or some other software, data

extraction from these files requires a lot of processing time. It appears unlikely that Extract or dBase could be used to access these data files in a public reference area. In a recent article in Government Publications Review, Steven Staninger documents the time required to conduct a few simple searches using dBase III Plus with the 1982 Census of Retail Trade data on Census Bureau's Test Disc 2. In one example, it required twelve minutes to execute a search for a single type of business code in the California data file on an IBM PC³.

There appear to be two strategies to the problem of public access to numeric data files on CD-ROM. A combination of both tactics will probably be necessary to provide adequate public access.

The first strategy is to develop the personnel and physical resources necessary to effectively deal with the problem of data extraction. To adequately deal with electronic formats a depository will require at least one staff member with microcomputer expertise and a strong social sciences background. At a minimum, this person will have to be familiar with DOS, database management software and spreadsheets; a familiarity with statistical programs such as SAS and/or SPSS would also be helpful. This person must also be familiar with the printed sources and have a close working relationship with the local data archives facility, if any, in order to know when CD-ROM is more appropriate than paper copy or data on tape.

There must be adequate physical resources as well. There appears to be a well established base of CD-ROM stations in large depository libraries. This is supported by the fact that at least half of all depository libraries have selected some Census data on CD-ROM*. Many of these depositories are also using these CD-ROM stations to provide end-user access to bibliographic files. However, as Staninger's article suggests, the effective use of the Census Bureau CD-ROMs will not lend itself to this setting. Libraries will need to provide CD-ROM stations out of the reference area where either library staff or end-users can extract data from the CD-ROMs. These CD-ROM stations will also need to devote a large amount of hard disk space to work with the depository CD-ROMs. In addition to the requirements outlined above, many of the depository CD-ROM products have their own front ends requiring a large amount of dedicated hard disk space, e.g. the National Health Interview Survey requires about five megabytes and the Toxic Release Inventory requires about seven megabytes of hard disk space.

The second strategy is to look for solutions in the private sector. While there are a few products, such as PC Stars, which are designed to work with the depository CD-ROMs, many commercial software producers are

reselling the Federal data. For example, Space-Time Research's **Supermap**, marketed in the U.S. by Chadwyck-Healey, contains data from the 1980 Census of Population and Housing Summary Tape Files, STF1-C and STF3-C, and additional county and land area data; and **StatMaster** produced by CyberSoft includes data from the **County and City Data Book**. While these commercial products might be expensive, each of these products provides enhanced access to the data.

There appears to be a fear that this could make depository libraries more dependent upon the private sector for access to Census materials; a Census Bureau report notes that librarians are "concerned by the need for user-friendly software to access census data and fear that it may be available from only the private sector at prohibitive cost."⁵ But there is a long established, successful history of private publishers providing bibliographic access to depositories. As Sir Charles Chadwyck-Healey stated in a recent interview published in **Government Publications Review**: "Governments seem to be extraordinarily bad at distributing information efficiently. It is probably inevitable. I am not sure there is going to be an enormous advantage to having very cheap data available from government sources if those sources are not able to disseminate it in an efficient and effective way."⁶

The depository CD-ROMs will have a great impact on library public services. In fact, if widely adopted, the CD-ROM distribution has the potential to transform the typical depository as we know it. Perhaps the greatest impact of CD-ROM in a library is the increase in workload for the public services staff in whose area the CD-ROM is located.

In a recent article, Steven Zink argues that the positive aspects of CD-ROM have tended to overshadow the human resources required for its use. He notes that "a persistent administrative malady is the assumption that technology will decrease, or at least not require additional, demands on staff time." In fact, the opposite tends to be true. Zink explains that "while technological advances may have resulted in personnel reductions in selected technical service areas, the use of automation where the public directly confronts technology has generally increased the need for user assistance."⁷

Depositories have yet to determine how best to deliver the additional assistance that users of electronic formats will require. Generally, document librarians perceive that the data extraction from the depository CD-ROMs will follow the patterns established by mediated online searches. However, as Elizabeth Stephenson observes: while many librarians are experienced in handling bibliographic data on CD-ROM, "few have any experience or training in the manipulation of numeric files." She believes that librarians will have to become familiar

with the hierarchical structure of the files and statistical languages to effectively work with the depository CD-ROM products.⁸

Conclusion

In 1988, Diane Smith of Pennsylvania State University conducted a survey to determine the preparedness of depositories to provide access to electronic data products. She examined the extent to which depositories were experimenting with the provision of electronic services and looked for characteristics common to innovative libraries. Her survey covered plans to include documents in local online public access catalogs; the use of online databases, CD-ROM, statistical software, and expert systems in depositories; and available hardware within reference areas. Smith concluded that depositories were ill-prepared to deal with electronic formats and that there is a definite need for depository libraries to face the training and collection management challenges presented. She cautioned that "if this work is not done, it appears that there will be a major crisis in the ability of libraries to deal with electronic data; a crisis that questions the viability of the present situation."⁹

As early as 1988, Jones and Kinney argued that when document librarians assume the responsibility of retrieving numeric, textual, or bibliographic information from computer tapes, they must either know how to program or work with a colleague who programs¹⁰. It is unlikely that a majority of depositories will be able to provide adequate programming assistance in-house. This means that many depositories may need to establish collaborative relationships with other units to adequately service the depository CD-ROMs. Likely partners include computer centers, data archives and social science research units or teaching departments.

At present, both product development and participation in the distribution of CD-ROMs through the depository program appears to be coordinated at the agency level. For example, within the Commerce Department, the Census Bureau is distributing CD-ROMs through GPO's depository program while the Patent and Trademark Office is only offering their own CD-ROM product, CASSIS, for sale or through their own patent depository program. Each agency is approaching the data format and user-interface issues differently. While the Census Bureau has committed itself to distribute data to depositories in dBase III format for use with privately-produced software, the Office of the Secretary is distributing the National Trade Data Bank CD-ROM with its own unique user-interface and data format. Some CD-ROM products will require that depository libraries acquire privately-produced software to work with the data; others will require that depository staff invest a substantial amount of time mastering the user-interface provided by the agency. While the true magnitude of this trend has yet to

be determined, the impact upon public service in depository libraries will be significant.

¹ Presented at the IASSIST 91 Conference held in Edmonton, Alberta, Canada. May 14 - 17, 1991.

² Robert W. Houk, "Remarks before the American Library Association Midwinter Meeting: Legislation Committee, Information Update," **Administrative Notes** 12 (February 22, 1991): 1-6.

³ Steven W. Staninger, "Using the U.S. Bureau of the Census Test Disc 2: a Note," **Government Publications Review** 18 (March/April 1991): 172-3.

⁴ Peter Hernon and Charles R. McClure, "Electronic Census Products and the Depository Library Program: Future Issues and Trends," **Government Information Quarterly** 8 (1991): 61.

⁵ Sandra Rowland, "The Role of Intermediaries in the Interpretation and Dissemination of Census Data Now and in the Future," reprinted in **Documents to the People** 18 (June 1990): 81.

⁶ Jean Slemmons Stratford, Juri Stratford and Steven Zink, "Applying an "Entrepreneurial Attitude" to the Dissemination of Government Information. An Interview with Sir Charles Chadwyck-Healey, Chairman of the Chadwyck-Healey Publishing Group," **Government Publications Review** 18 (March/April 1991): 134.

⁷ Steven Zink, "Planning for the Perils of CD-ROM," **Library Journal** (February 1, 1990): 54.

⁸ Elizabeth Stephenson, "Data Archivists: The Intermediaries the Census Bureau Forgot. A Review Essay of "The Role of Intermediaries in the Interpretation and Dissemination of Census Data Now and in the Future," **Government Publications Review** 17 (September/October 1990): 443.

⁹ Diane H. Smith, "Depository Libraries in the 1990s: Wither or Wither Depositories?," **Government Publications Review** 17 (July/August 1990): 312.

¹⁰ Ray Jones and Thomas Kinney, "Government Information in Machine-Readable Data Files: Implications for Libraries and Librarians," **Government Publications Review** 15 (January/February 1988): 30.