

University of San Diego

Digital USD

Copley Library: Faculty Scholarship

Copley Library

Spring 3-1-1991

Using U.S. Bureau of the Census CD-ROM Test Disc 2: A note

Steven W. Staninger

University of San Diego, ssan@san Diego.edu

Follow this and additional works at: https://digital.sandiego.edu/library_facpub



Part of the [Library and Information Science Commons](#)

Digital USD Citation

Staninger, Steven W., "Using U.S. Bureau of the Census CD-ROM Test Disc 2: A note" (1991). *Copley Library: Faculty Scholarship*. 20.

https://digital.sandiego.edu/library_facpub/20

This Article is brought to you for free and open access by the Copley Library at Digital USD. It has been accepted for inclusion in Copley Library: Faculty Scholarship by an authorized administrator of Digital USD. For more information, please contact digital@san Diego.edu.

Using U.S. Bureau of the Census CD-ROM Test Disc 2: A note

Description, Abstract, or Artist's Statement

Data on U.S. Census Bureau CD-ROM Test Disc 2 is accessible using dBASE III Plus®. Different computer hardware configurations are used to determine the most efficient method of manipulating the data. Various software options for accessing Census CD-ROMs are also mentioned. Using the same dBASE commands with each system, comparisons are made between 8088 and 80386 configurations. The 80386 is found to be essential to the efficient searching of large census files. Libraries will need to invest in advanced computing capabilities and staff training to respond to the GPO's commitment to CD-ROM technology.

Disciplines

Library and Information Science

USING THE U.S. BUREAU OF THE CENSUS CD-ROM TEST DISC 2: A NOTE*

STEVEN W. STANINGER**

Research Services Department, Central University Library, University of California, San Diego,
La Jolla, CA 92093, USA

Abstract — Data on U.S. Census Bureau CD-ROM Test Disc 2 is accessible using dBASE III Plus®. Different computer hardware configurations are used to determine the most efficient method of manipulating the data. Various software options for accessing Census CD-ROMs are also mentioned. Using the same dBASE commands with each system, comparisons are made between 8088 and 80386 configurations. The 80386 is found to be essential to the efficient searching of large census files. Libraries will need to invest in advanced computing capabilities and staff training to respond to the GPO's commitment to CD-ROM technology.

INTRODUCTION

The Bureau of the Census released data on compact disc to depository libraries throughout the United States in early 1989. Test Disc 2 contains data from the 1982 Census of Retail Trade and the 1982 Census of Agriculture. By releasing this data in CD-ROM format the Census Bureau has posed a challenge to libraries to provide access to machine-readable data. Much has been written about the increase in non-print format government information and the problems it presents. Although there is an awareness of the ethical and practical issues presented by the Government Printing Office's (GPO) move towards non-print materials, this paper will focus primarily on solving the technical problems associated with gaining access to data on CD-ROM.

The GPO is committed to decreasing drastically the amount of paper documents distributed to depository libraries. This was made clear at a Census Bureau training course offered at University of California at Berkeley in August 1989 [1]. The Census Bureau representative stressed the commitment to CD-ROM and provided many documents detailing the reasons behind this policy decision. In short, the representative emphasized that the Bureau was providing the data required of it by law, and access to it was the responsibility of the depository library. A comment by one of the members of the audience that this trend towards more and more machine-readable data will require a great deal of time spent by librarians conducting mediated searches was met with the response, "That's something you're going to have to work out."

A report issued by the Office of Technology Assessment [2] shows that the future promises a dramatic increase in machine-readable data and a decrease in paper and microfiche documents.

*The author would like to acknowledge the valuable assistance of Jim Jacobs, UC, San Diego, Central University Library.

**Steven W. Staninger is involved with computer-assisted reference service in the Research Services Department of the Central University Library at the University of California, San Diego. He has previously worked in the Government Documents Department at the Library.

Several congressional committee reports are also heralding the move towards machine-readable data [3]. In order to provide access to these potentially powerful data sources, it is crucial that the librarian and the libraries be current with the necessary technologies.

CENSUS TEST DISC 2

Test Disc 2 comes with software and technical documentation [4]. The files contained on the disc are in dBASE III Plus® format [5]. The purpose of this study was to ascertain if custom reports could be efficiently created by accessing the CD-ROM disc directly through the dBASE III Plus® program without using the software provided by the Census Bureau. The test was initially conducted using an IBM PC XT [6] with an 8088 processor and a 20 MB hard drive ported to a Philips CM 100 compact disc reader [7]. This is a standard configuration for accessing most CD-ROM data.

To begin, the software accompanying Census Bureau CD-ROM data was examined. The first problem encountered was getting the computer to recognize the disc in the CD-ROM reader. For unknown reasons the software provided by the Census Bureau is designed to look for a CD-ROM drive called drive L:. To get the reader to look for the CD-ROM at the customary drive "D:" was the first order of business. Using Microsoft CD-ROM Extensions Version 2.0 [8] to program the device driver, the following command changes the "L:" designation to the "D:" drive designation:

```
a: MSCDEX D:MDIHS DVR L:D:
```

This done, it was fairly easy to gain access to tables of data if the user follows the menu commands. This method, although slow, allows the user to obtain data traditionally available in print format. The menus, screen displays, and instructions are clear. Any user with basic computing skills should have little trouble using the Census on CD-ROM to find basic data. However, searching for more detailed data is a different proposition.

The files on Test Disc 2 are in dBASE III Plus® format because it is close to being a standard. Other dBASE-compatible programs, such as FoxBASE® [9], can also be used to obtain access to the files. Using dBASE III Plus®, several sample programs were run to test the ease with which the user might extract customized data reports. After loading dBASE III Plus®, the command

```
.set default for D:
```

will allow for direct access to the files on Test Disc 2 by using dBASE commands. The Census of Retail Trade for Arizona is identified as file name RC82_AZ.DBF on Test Disc 2. After loading this file with the command

```
.use rc82.az.dbf
```

the user is ready to manipulate its 5981 records.

The first task is to reveal the structure of the datafile. The 'TAB' fields in the structure are numerical codes for types of tables. The data dictionary to these fields is found in the technical documentation accompanying Test Disc 2 [10]. It is essential that the technical documentation be readily available to the user who is attempting to extract data without using the Census Bureau software. Using the technical documentation to decipher the codes list in the structure of the datafile, the author constructed the following command:

```
.list for business = '016' zip,business,tab_1.1
```

The data dictionary in the technical documentation notes that 'tab_1.1' is number of establish-


```
. list for business = '016' zip, business, tab_1_1
```

Record#	zip	business	tab_1_1
15	00000	016	231
97	85003	016	1
132	85004	016	3
184	85006	016	1
259	85008	016	2
340	85010	016	1
361	85012	016	4
404	85013	016	5
446	85014	016	2
492	85015	016	11
542	85016	016	6
588	85017	016	3
633	85018	016	4
713	85020	016	1
756	85021	016	9
841	85023	016	2
917	85028	016	3
958	85029	016	1
1045	85032	016	2
1088	85033	016	5

Command Line <D> RC82_AZ

Rec: EOF/5981

Num

Enter a dBASE III PLUS command.

Figure 1. Data from U.S. Census Bureau Test Disc 2: Number of "Other Food Stores" in Arizona in Zip Code Order.

ments. The table line stub numbers appendix in the technical documentation reveals that '016' is "other food stores" [11]. Therefore, the above command will list the number of "other food stores" (e.g., convenience and liquor stores) in each zip code in the state of Arizona in zip code order, from smallest to largest zip code number (Figure 1). This has the potential to be valuable information to someone doing marketing research in Arizona. Most patrons, however, would be unable to construct such a search without considerable assistance. Another problem is that this command took 12 minutes to process. Even if the patron was familiar with dBASE commands and had the technical documentation, it is likely that they would think that the search was not working properly.

In an attempt to correct this problem, the Census CD-ROM was loaded on faster hardware. Armed with a Zenith 80386 16 mhz computer [12] ported to a Hitachi CD-ROM drive [13], the author accessed dBASE III Plus® through the Novell Network [14] available in the library. Not surprisingly, the results were dramatically different. The same command for the Arizona file was processed in 2.5 minutes. Thus encouraged, the California file (which, at 40,996 records, is by far the largest) was loaded. The processing took 17 minutes using the 80386 Zenith hardware. Even more speed could be achieved by using an 80386 without going through the network [15].

A recent improvement to accessing Census Bureau CD-ROMs is *Extract* [16]. This is public-domain data extraction software developed specifically for use with Census Bureau files in dBASE III Plus® format. *Extract* can select, display, and extract data from Census Bureau CD-ROM files without using dBASE III Plus®. Providing the user wants only to manipulate the data using these three functions, the menu-driven *Extract* program is an alternative to the often complex dBASE commands.

CONCLUSION

The Census Bureau is committed to issuing data on CD-ROM, a medium the Bureau feels has all the qualities necessary to provide large amounts of data at a reasonable production cost. CD-ROM is also a format they believe will be viable in the future. It is imperative that depository libraries have the equipment necessary to use CD-ROMs. If at all possible, computers running at 80386 speed or faster should be used so that searching the CD can be done in a reasonable amount of time. Equally important is the need for librarians, especially in government documents and reference, to have the training necessary to utilize these emerging technologies. The intentions of the Census Bureau, and ultimately, the GPO, are clear. The debate about the pros and cons of government information in non-print format has become moot. It is time for librarians to prepare themselves, their patrons, and their libraries to provide easy access to machine-readable data.

NOTES

1. "Notes Taken at the Census Bureau Training Course," University of California, Berkeley, August 23, 1989.
2. U.S. Congress, Office of Technology Assessment, *Informing the Nation: Federal Information Dissemination in an Electronic Age* (OTA-CIT-396), (Washington, GPO, 1988.)
3. U.S. Congress, House, Committee on Government Operations, *Electronic Collection and Dissemination of Information by Federal Agencies: Hearings before a Subcommittee of the Committee on Government Operations*, 99th Cong., 1st sess., April 29, June 26, October 18, 1985; U.S. Congress, House, Committee on Post Office and Civil Service, Subcommittee on Census and Population, *Plans and Activities for 1990 Decennial Census: Hearings before the Subcommittee on Census and Population of the Committee on Post Office and Civil Service*, 99th Cong., 1st sess., July 25, 1985; U.S. Congress, Joint Committee on Printing, *An Open Forum on the Provision of Electronic Federal Information to Depository Libraries*, report of the Staff of the Joint Committee on Printing, 99th Cong., 1st sess., June 26, 1985.
4. U.S. Bureau of the Census, *CD-ROM Test Disc 2* (machine-readable data file) Prepared by the Bureau of the Census. (Washington: The Bureau, 1988); U. S. Bureau of the Census, *CD-ROM Test Disc 2 Technical Documentation*, by the Data User Services Division, Data Access and Use Staff (Washington: The Bureau, 1988).
5. *IBM PC XT* (computer hardware), International Business Machines, Inc., Old Orchard Rd., Armonk, NY 10504.
6. *Philips CM 100* (computer hardware), Philips Consumer Electronics Co., P.O. Box 6950, Knoxville, TN 37914.
7. *CD-ROM Extensions* (computer software), Microsoft Corp., 16011 NE 36th Way, P.O. Box 97017, Redmond, WA 98073.
8. *FoxBASE* (computer software), Fox Software Inc., 134 W. South Broadway, Perrysburg, OH 43551.
9. U.S. Bureau of the Census, "CD-ROM Test Disc 2: Census of Retail Trade 1982 Data Dictionary" *CD-ROM Test Disc 2 Technical Documentation*, II-27.
10. U.S. Bureau of the Census, Appendix B: "Table Line Stub Numbers / Kind-of-Business Codes," *CD-ROM Test Disc 2 Technical Documentation*, B-1.
11. *Zenith Data System 80386 16 mhz PC* (computer hardware), Zenith Electronics Corp., 1000 N. Milwaukee Ave., Glenview, IL 60025.
12. *Hitachi CD-ROM Reader* (computer hardware), Hitachi America Ltd., 50 Prospect Ave., Tarrytown, NY 10591.
13. *Novell Network Software* (computer software), Novell, Inc., P.O. Box 5900, Provo, UT 84601.
14. *DESQview 386* (computer software), M & T Publishing, Inc., 501 Galveston Dr., Redwood City, CA 94063. *DESQview 386* has a utility that will show the speed of different hardware configurations relative to the 8088. It indicates that the *Novell Network* is 4.1 times as fast as 8088 and that the 80386 is 8.2 times as fast. A direct 80386 configuration would definitely be the choice for running some of the more time consuming dBASE commands, such as "join."
15. *Extract* (computer software), Center for Electronic Data Analysis, 316 Stokley Management Center, University of Tennessee, Knoxville, TN 37996.