

# Introduction to The CSSCR Data Archive

Updated by Tina Tian

Center for Social Science Computation & Research  
145 Savery Hall  
University of Washington  
Seattle WA 98195 U.S.A.  
(206)543-8110

January 2006

<http://julius.csscr.washington.edu/pdf/archive06>

# Introduction to the CSSCR Data Archive

The Center for Social Science Computation and Research (CSSCR) maintains a large electronic data archive applicable to social science research. The data sets at CSSCR are available in many formats. This document describes how to search the data archive and obtain the data sets.

## Finding Data Sets at CSSCR

### *Data sets on the CSSCR Web Site*

If you have Internet access, searching the CSSCR Web page is the most efficient way of finding a data set. A complete list of data sets at CSSCR is available on the CSSCR Web page and much of our data can be easily accessed through a Web browser. The CSSCR archive homepage address is <http://julius.csscr.washington.edu/data.htm>

The data sets on the CSSCR Web page are divided into three categories: ICPSR data, CD-ROM and Census 2000 data. The ICPSR data section includes all online data sets ordered from the Inter-University Consortium for Political and Social Research (which is our Center's major data source), the US Census Bureau and a number of other data sources. The CD-ROM category contains the list of CD-ROMs maintained at CSSCR. Census 2000 section contains Washington State Census 2000 data files and documentations. Clicking on one of these three links will bring you to the ICPSR Resources, the CD-ROM List page or the U.S. Census 2000 files.

On the ICPSR Resources page, you can click on the "Archive Browser" button on the upper left. On the browser page, studies are organized alphabetically by title. Under each title, information such as data source, codename, abstract and storage medium is displayed. To search the holdings, use the "Find in Page" function of your browser. Find is located on the "Edit" menu of Internet Explorer or Firefox. If a study is from ICPSR, you may also view related documentation directly on the ICPSR website through the "Get ICPSR Documentation" link. If you would like to see files associated with any data sets, click "Get Files" on the bottom of each study, that allows you to access and download files from that study. If you already know the codename of a particular study and would like to see or download its associated files, you may simply type the codename in the box on the bottom of the Archive Browser page and press "Submit". Once you are on the "Files Retrieved" page, you can click on the files to read them or save them to your local machine.

For each data set, there is usually more than one file. The following standard extensions are used to distinguish one type of file from another:

<filename>.dat - raw data (ASCII file)  
<filename>.sas - SAS syntax file  
<filename>.ssd - SAS system file  
<filename>.tpt - SAS transport file  
<filename>.sps - SPSS syntax file  
<filename>.svf **or** <filename>.sav - SPSS system file  
<filename>.por **or** <filename>.exp - SPSS portable file  
<filename>.cod - codebook in text format  
<filename>.des - file description in text format  
<filename>.que - questionnaire in text format  
<filename>.doc - documentation in text format  
<filename>.pdf - documentation in PDF format

Another way to look at the Online Archive is go to Home/Data Archive/ICPSR/Online CSSCR Compendium. (The last layer is a text link about a third of the way down the page.) This link puts you into a list view of the CSSCR Data Archive. Use the buttons at the bottom to move through the collection twenty citations at a time, or to the beginning or end of the listings. At the upper right of the title bar is a button that will provide a one-data set-at-a-time form view of each citation. At the bottom of the form view is a button that will create a listing of files associated with the code name you provided in the box. The initial state of the box contains the present code name of the viewed data set. These files are actually links which you can click on to open or save from the Archive to your local hard drive.

If the data set you are interested in is not yet available on-line or is not listed in the database, please contact our data archivist. The data may be available on tape or CD-ROM or through another source; if you don't find it in the Online Compendium, it might still be accessible through CSSCR.

If you do not have access to the CSSCR Web page, you can use the following steps to do a data archive search.

- . **Printed Compendium of data sets:** The printed compendium is same as the version on the CSSCR Web page. It is available at CSSCR main office (Savery 145). The data sets are described in an alphabetical order along with any associated files. When you find a data set that interests you, make a note of the study title and file name.
- . **Codebooks:** Codebooks are maintained on the shelves in Savery 145. Each codebook is labeled by the acronym of a data set title and is arranged alphabetically on the shelves.
- 0. Come to the **data archivist** for more information on how to get access to the data files you are interested in. The archivist can help download the files from CSSCR machines to your diskettes or FTP the files to your mainframe account.

Our ICPSR/Census resource is accessible to University of Washington faculty, staff and students.

#### *Data Sets on CD-ROM*

CSSCR also has an active CD-ROM Archive. More and more data sets are released in CD-ROM format. Most CD-ROM products are for PC-users. The list of CD-ROM titles is posted on the CSSCR Web page at: <http://julius.csscr.washington.edu/cdrom.htm>

Go to the data archivist for the CD-ROMs you want. You are required to present your photo ID and use the CD-ROMs in CSSCR labs only.

#### *CODEBOOKS*

The Center has a hard-copy codebook for each data set we maintain. The codebook contains information on the study design, sampling method, procedure of data collection, variables and coding. A copy of the questionnaire is also included in the codebook if the data are gathered in a survey. The data files are described in detail, including the data format, file structures, names and types of variables, etc. In short, the codebooks contain the information you need to assess the utility of each data set and to choose any variables you may wish to analyze. You are allowed to check out the codebooks for a short period of time.

### *DATA LOG and CD-ROM LOG*

The data log records the title, source, block size, record length, record numbers, file format, receiving date, and additional comments for each data set. The newest data log is also located in Savery 145 along with the printed compendium. For older tape logs, see the Data Archivist. The CD-ROM log is also kept in Savery 145 along with the data log.

The CSSCR Compendium and CD-ROM list (hard copy and Web version) are normally updated every summer. However, we keep a recent record of any new data sets on the logs, and new acquisitions are announcing in the CSSCR newsletter at the beginning of every quarter.

### **Finding Data Sets at Other Locations**

Most data sets at CSSCR are from ICPSR (Inter-University Consortium for Political and Social Research) and the U.S. Bureau of the Census. CSSCR is a member of ICPSR and one of the coordinating centers of the Washington State Data Center. If you are interested in a data set that is unavailable at CSSCR, you may try the following resources:

- . **ICPSR Guide to Resources and Services** (hard-copy): The Guide is located in Savery 145. All the data sets are generalized into seventeen categories.
- 0. **ICPSR Web site:** The home page address of this site is: <http://www.icpsr.umich.edu>

ICPSR maintains an excellent on-line archive on their Web page. You can perform your data search by study title, abstract, or principle investigator. You can also browse archive holdings by subject, title, principle investigator, or ICPSR study number. For some data sets, machine-readable codebooks are freely available to public. You can read or search for key words or variables by clicking on the uncompressed version. If needed, you can also download the codebook.

You cannot download the data yourself, but once you find a data set you need from the ICPSR archive listings, write down the title and ICPSR study number. You may then ask our data archivist to order it for you. ICPSR member institutions are given FTP access to the ICPSR collections. Ordering from ICPSR is free of charge to UW faculty, staff, and students.

- . **Other Catalogs.** We subscribe to more than ten catalogs that introduce the most recently released data sets (such as *GSS News*, *ICPSR Bulletin*, *NLSY News*, *Data Archive Bulletin*, *Census and You*, etc.). They are all available in Savery 145.

## Other Data Resources at CSSCR

### *DataStream*

DataStream provides access to various global economic databases such as the National Government and OECD series, international monetary funds, company fundamentals, equity and bond indices, exchange and interest rates, economic indicators, spot commodities, fixed income, futures and options, etc. People can use DataStream for economic research, time series analyses, bond, and equity research.

CSSCR shares the DataStream subscription with the Economics Department and Business School. The DataStream index code volumes and reference books are available in Savery 145. Access to the DataStream database is only available through the computer in the archivist's office. The archivist can download the data for you, if you note the index codes, time period and frequency of publication.

### *DRI-WEFA Basic Economics Analytics*

The DRI-WEFA Basic Economics database is available at CSSCR through an econometrics package called E-Views. It is a machine-readable, macroeconomics database. This database contains approximately 7,000 monthly, quarterly, and annual time series that data back to 1946, when available, and end with the latest observations. The data consist of national-level information taken from the U.S. Government, Federal Reserve Board, and other public and private organizations. We keep a reference book in the CSSCR main office; it lists a short description of each series, the units used, time period covered, frequency of publication, retrieval code, and references to additional information about the data and its organization.

### *Geolytics Products*

Our Geolytics products include four demographic and mapping software packages: CensusCD+Map, CensusCD Blocks, CensusCD 1980 and StreetCD 98. With different features contained in each of these packages, they are the most convenient and complete demographic and geographic references in the U.S. market. You can easily view, query, map, and export U.S. population and housing data (including the 1980, 1990 and 2000 U.S. Census, recent estimates and projections) for various geographic contexts, from blocks to the state level. For example, the StreetCD 98 allows you to acquire Census TIGER 98 street data and boundary files easily. You can also view and retrieve geographic data and attributes from the local community to the Nation (such as roads, railroads, hydrography, and landmarks, as well as legal, statistical, and administrative boundaries for census, transportation, health, education and more).

The Geolytics products are accessible from any computers at CSSCR; only CensusCD 2000 are available in the Faculty and Graduate Research room now. Click on the Start button, select Run, and open a related batch file from the O drive to access one of the four software packages. To access:

CensusCD+Map	<b>o:census3.bat</b>
CensusCD Blocks	<b>o:ccdblock.bat</b>
CensusCD 1980	<b>o:ccd1980.bat</b>
StreetCD 98	<b>o:tiger99.bat</b>