IN THIS ISSUE

. . . .

DISC News contains articles about local, national and international data issues. It is published twice a semester by the library staff and sent to faculty, graduate students, and librarians, with a focus on the Social Science departments.

Editor: Joanne Juhnke, Special Librarian
Staff contributors: Janet Eisenhauer Smith, Data Analyst/Archivist; Cindy Severt, Senior Special Librarian

. . . .

. . . .

This issue of DISC News is online at:
**http://www.disc.wisc.edu/pubs/
Newsletters/feb08news.html**

RESTRICTED DATA: HOW DISC CAN HELP

by Janet Eisenhauer Smith & Joanne Juhnke

Last February in this newsletter, we announced the merger of the former Data & Program Library Service (DPLS) with the data services operations of the Center for Demography of Health and Aging (CDHA) and the Center for Demography and Ecology (CDE). This February, we highlight the new and enhanced services for restricted data that DISC is now able to make available to UW-Madison: facilities for secure data storage and analysis, and

> The process of applying for restricted-data is often complex, but DISC stands ready, willing and able to help.

expanded assistance in navigating the application process for restricted data.

Survey and administrative data produced with public funds, either by government or at universities and public research centers, is more widely available for public download and use than ever before. At the same time, however, much socio-behavioral data collected by such researchers cannot be released as-is to the public because of risks to study participants and the confidentiality promised to them. Survey researchers who might have been willing to openly share their data with colleagues after simply removing direct identifiers are now facing increasingly conservative rules for defining conditions under which sensitive confidential data may be shared.

In this context, it is no surprise that requests for restricted data have been on the rise at DISC. The process of applying for restricted-data is often complex, but DISC stands ready, willing and able to help.

Gaining access to restricted data at UW-Madison generally involves several components:
- A human subjects protocol application to one of the UW-Madison Institutional Review Boards (IRB).
- An agency-specific application for a license to use the restricted data, which must include a data protection plan and explain why any available public-use data is not sufficient for the research.
- Contract approval of the data license through Research and Sponsored Programs (RSP) at UW-Madison.

Data Analyst/Archivist Janet Eisenhauer Smith has extensive experience assisting researchers who are creating or using sensitive confidential data, in the context of CDHA. DISC Director Jack Solock has also coordinated complex restricted data contracts in his work with CDE. Their experience

ICPSR SUMMER PROGRAM 2008

by Joanne Juhnke

The ICPSR Summer Program in Quantitative Methods of Social Research is now accepting applications for summer 2008, for four-week courses and 3- to 5-day workshops in Ann Arbor, Michigan. The program focuses on research design, statistics, data analysis, and social science methodology. Four-week sessions run June 23 through July 18, and July 21 through August 15. A detailed syllabus and online registration are available at **http://www.icpsr.umich.edu/sumprog/**.

When considering ICPSR Summer Program fees, please note that although UW-Madison students have in the past been able to attend as CIC Traveling Scholars, that option is no longer available this year due to financial pressures. More information about tuition and fees can be found at **http://www.icpsr.umich.edu/sumprog/2008/tuition.html**.

One travel stipend—approximately the cost of round-trip airfare from Madison to Detroit—is available through DISC for defraying the cost of one individual traveling to Ann Arbor for the ICPSR Summer Program. The stipend is limited to UW-Madison, and preference will be given to students. To be considered for the travel stipend, contact your ICPSR Official Representative at DISC, Cindy Severt: **cdsevert@wisc.edu** or 262-0750.

A BLIZZARD OF DATA FROM ICPSR

by Cindy Severt

From stem cell research, to volunteering for the Peace Corps, to a sea of red on Badger football days, UW-Madison has always done things in a big way—and the same holds true of ICPSR data downloads! From July 2006 to June 2007 the UW-Madison community downloaded 25,280 individual data files from the ICPSR archive at **http://www.icpsr.umich.edu**, totaling 290.679GB. Compared with the median of 3,710 individual files totaling 56.129GB for similar institutions, UW-Madison is accessing nearly seven times the number of files.

What exactly is being downloaded? The National Survey of Midlife Development in the United States (MIDUS), 1995-1996 leads the way:

- National Survey of Midlife Development in the United States (MIDUS), 1995-1996: 29 downloads
- National Assessment of Educational Progress [United States], 1970-1980: 20 downloads
- General Social Surveys 1972-2004 (Cumulative File): 18 downloads
- American National Election Studies Cumulative Data File, 1948-2004: 13 downloads
- National Longitudinal Study of the Class of 1972: 13 downloads
- United States Historical Election Returns, 1824-1968: 12 downloads
- Public Health Impact of Direct-to-Consumer Advertising of Prescription Drugs, July 2001-January 2002: [United States]: 11 downloads
- Current Population Survey, January 2004: Displaced Workers, Employee Tenure, and Occupational Mobility Supplement: 10 downloads
- American Citizen Participation Study, 1990: 9 downloads
- American National Election Study, 2002: Pre- and Post-Election Survey: 9 downloads

Though the majority of ICPSR downloads originate from departments in the social sciences, DISC's users are literally all over the campus map. Here are some diverse examples from 2007:

Department	# of Files Downloaded
Business Administration	327
Psychology	185
History	71
Computer Science	15
Medicine/Dentistry	13
Nursing	10
Library & Information Studies	7
Law/Legal Services	4
Humanities	2

RESTRICTED DATA: HOW DISC CAN HELP

Continued from p. 1

of working with sensitive data and data licensing, added to the assistance already offered at DPLS, is now available to DISC users from across campus.

To begin the process of applying to use restricted data, the first step is identifying and acquiring application materials. The researcher will need to provide an abstract describing the proposed research. Sometimes the description will need to include the research model, which variables will be used and the types of analysis that will be applied, and how results will be reported. If a public-use version of the data exists, the researcher will need to demonstrate why the public-use data is not sufficient for the analysis.

The data license that should result from the application is a contract that defines how the data will be analyzed and describes how the data will be protected from unauthorized use while housed at the licensee's home institution. The steps that will be taken to prevent unauthorized use of licensed data are commonly referred to as the "data protection plan." The contract is usually executed by a formal contracting authority; for the UW-Madison, this authority resides in the office of Research and Sponsored Programs (RSP).

Having DISC's expert help in coordinating the data protection plan with the license application and the RSP approval can save considerable time and follow-up for the researcher. Familiarity with past contracts means that DISC staff know when to recommend adding new projects to existing licenses, or what contractual clauses have proven problematic in the past. Avoiding unnecessary bottlenecks may mean a difference of weeks or even months in obtaining a data license.

To help provide the level of security that data protection plans can require, DISC now offers the "cold rooms," a secure computing enclave in the Sewell Social Science Building. Administered in cooperation with the Social Science Computing Cooperative, each of the two cold rooms employs a system of removable hard-drives and wall safes, each accessible by only one licensee. Each cold room contains a single stand-alone workstation and can be accessed by only one researcher at a time. Only Janet, Jack, and researchers with restricted data licenses are issued cold-room keys; not even the custodial master keys will open the door. This cold-room service is now available to anyone affiliated with the UW-Madison who needs access to such a facility in order to get a restricted data license, with prior approval from DISC and an approved IRB protocol.

As the world of socio-behavioral research navigates the opposing pressures for broader sharing of public data and for more careful human-subjects protections, DISC will make every effort to help our users find their way as well.

NCES INTERNATIONAL DATABASE TRAINING

by Joanne Juhnke

The National Center for Education Statistics (NCES) is sponsoring a 2½-day seminar on the use of NCES International Databases: the Program for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS). The seminar will be held in Washington, DC, May 21-23.

The seminar is for researchers who plan to use data from the NCES International Databases, including advanced graduate students and faculty members, researchers, education practitioners, and policy analysts. Participants attending this seminar should have a solid understanding of statistical methods, be experienced in using personal computers, and be proficient in the use of SPSS or SAS statistical software packages.

NCES will provide training materials as well as computers for hands-on practice. NCES will also pay for transportation, hotel accommodations, and a fixed per diem for meals and incidental expenses during the training seminar.

Applications are due April 4, 2008, and selected candidates will be informed by April 14. For more information or to complete an application, visit the seminar web site at: **http://ies.ed.gov/whatsnew/conferences/?id=309&cid=2**.

C
R
O
S
S
R
O
A
D
S

C
O
R
N
E
R

Crossroads Corner highlights web sites recently added to the searchable Internet Crossroads in Social Science Data, available on the DISC web site at http://www.disc.wisc.edu/newcrossroads/index.asp.

A New Nation Votes: American Election Returns, 1787-1825

An exciting presidential primary season makes a timely backdrop for highlighting an election-data resource that reaches back to the earliest years of American democracy. *A New Nation Votes*, online at **http://elections.lib.tufts.edu/aas_portal/index.xq**, offers data from America's earliest elections, between 1787 and 1825. The scope is much broader than presidential politics; the site covers offices all the way down to Alderman and County Coroner. Around fifteen thousand elections are currently available, about 27 percent of the eventual total that will cover all 25 states that existed during the time frame. The data is the result of decades of research, much of it gleaned from sources such as newspapers and county histories and primarily collected by researcher Philip J. Lampi.

A New Nation Votes is searchable by keyword, state, year, office, candidate, and party. Results are in HTML tables, with links to view PDFs of Lampi's original notebook pages. The entire dataset or data by state may also be downloaded for analysis.

GraphWise

"Charting a world of data" is the tagline at GraphWise, **http://www.graphwise.com/**. This beta-release search engine crawls the web in search of HTML data tables, spreadsheets and downloadable data files, making them available in a searchable index. GraphWise then automatically generates graphs based on what it finds, and allows registered users to manipulate and download custom graphs as well.

As with any search engine, users must check the source site with care, since the crawler has limited means for differentiating between trash and treasure. GraphWise also picks up some items that are laid out in HTML tables but do not have much to do with data, such as course catalog pages (if course-name is in one column and number of credits in another, it looks like tabular data to the crawler).

The GraphWise index is approaching 3 million searchable tables, with a declared goal of growing to 100 million. Among search engines, the GraphWise approach is currently unique, and as the index grows, the searching will likely become increasingly more useful.

UNdata

The United Nations has over the years produced a wide and sprawling array of statistical databases. To bring multiple UN data sources under a single interface, the United Nations Statistics Division has recently announced UNdata, at **http://data.un.org**.

UNdata has begun with 14 databases containing over 55 million data points, covering a range of topics including population, industry, energy, trade and national accounts. The databases are accessible either by keyword searching from a single search page, or through a menu of databases.

UNdata will be replacing the UN Common Database (UNCDB) which is slated to be discontinued in the summer of 2008. Indicators formerly offered through the UN Common Database will be listed under Key Global Indicators, and will be searchable through the main interface as well. However, the trade information in UNdata will not be replacing UN Comtrade, which will continue to cover a deeper and more fully-featured set of merchandise statistics.