



IN THIS ISSUE

Data, Data Everywhere, Why
Can't I Stop and Think?.....1
U.S. Census News.....2
News from ICPSR2
Data Products Caught in U.S.
Budget Crunch.....3
Crossroads Corner.....4
.....

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: Cindy Severt, Senior
Special Librarian

University of Wisconsin-Madison
1180 University Drive
3308 Sewell Social Sciences Building
Madison, Wisconsin 53706 USA
Phone: (608) 262-0750
E-mail: disc@mailplus.wisc.edu
Web: <http://www.disc.wisc.edu/>
Hours: Monday-Friday 8:30-4:30

This issue of DISC News is online at:
[http://www.disc.wisc.edu/pubs/
Newsletters/apr11news.html](http://www.disc.wisc.edu/pubs/Newsletters/apr11news.html)

Note:

DISC will be closed

- Mon. May 16: State Furlough Day
- Mon. May 30: Memorial Day
- Mon. July 4: Independence Day
- Mon. September 5: Labor Day

DATA, DATA EVERYWHERE, WHY CAN'T I STOP AND THINK? (RIME OF THE MODERN MARINER)

By Cindy Severt

Those who analyze quantitative numeric data have long known that data are not information; and information is not knowledge. Rather, knowledge is derived from information which itself is derived from the building blocks of data. For generations social scientists have had an accepted definition of what constitutes “data” and its concomitant arcane terminology (metadata, codebook, variable, label). And they have frequently had access to data librarians such as myself who document, accession, and provide storage and migration for large datasets.

Sailing into view this semester is UW-Madison’s Research Data Services (RDS).

However, the last few years have seen an unprecedented amount of data of bewildering variety – textual, visual, biological, chemical, meteorological, to name a few – produced by research. The vast quantities of data produced are largely due to its digital format. But the more data that is produced, the less an individual researcher is able to swim in it without drowning. The ability to generate data exponentially exceeds the ability to manage it.

Sailing into view this semester is UW’s Research Data Services (RDS) a pilot collaboration between UW-Madison Libraries, DoIT, the CIO office, the Graduate School, and the School of Library and Information Studies to assist researchers with data curation needs, <http://researchdata.wisc.edu/>.

The RDS website has advice and suggestions for 1) managing one’s research data including file naming conventions, metadata, security, storage & backup, sustainability; 2) sharing one’s data, including data visualization, and most importantly; 3) developing a data management plan, along with information on what a data management plan is and why it is necessary.

A data management plan (DMP) is now a necessary component of proposals for funding with several federal agencies including the National Science Foundation (NSF), National Institutes of Health (NIH), and the National Aeronautics and Space Administration (NASA). Just as no two research projects are alike, no two data plans are alike, and thus there is no precise blueprint to follow. There are, of course, plenty of guidelines, and this is where the RDS website provides navigation through tricky waters. To

Continued on p. 3

U.S. CENSUS NEWS

by Joanne Jubnke

AMERICAN FACTFINDER REDESIGN

The U.S. Census Bureau has launched a revised version of American FactFinder, its primary online data dissemination site, at <http://factfinder2.census.gov/main.html>. The first data in the new version includes 2000 decennial census data, 2008 population estimates, and the new redistricting data from the 2010 decennial census. Data from the American Community Survey, the Economic Census, and Population Estimates is still on the legacy American FactFinder at <http://factfinder.census.gov/> but will be moved to the new interface in the coming months.

Meanwhile, 1990 decennial census data has been removed from American FactFinder altogether. Here are some online sources for 1990 data still in existence:

Interactive Interfaces

- CenStats 1990 PL-171 (redistricting) data — <http://censtats.census.gov/pl94/pl94.shtml>
- Missouri Census Data Center 1990 Demographic Profile Generator — <http://mcdc2.missouri.edu/websas/xtabs3v2.html>

Download Only

- Census Bureau 1990 FTP Downloads— http://www2.census.gov/census_1990/
- ICPSR – 1990 Census Data— <http://tinyurl.com/Census1990ICPSR>

2010 REDISTRICTING DATA

According to Public Law 94-171, the Census Bureau must provide redistricting data to all 50 states, plus Puerto Rico and the District of Columbia, no later than April 1 of the year following the decennial Census. Data was released state by state on a flow basis on the new American Factfinder interface in February and March. Final state totals were released on March 24, and a national summary file was added on April 14.

The redistricting tables include population counts by race and Hispanic origin, as well as voting-age population and housing unit counts.

NEWS FROM ICPSR

by Joanne Jubnke

NEW ICPSR SEARCH CAPABILITIES

An April Webinar on new search capabilities at the ICPSR website is now available for viewing; see <http://icpsr.blogspot.com/2011/04/webinar-on-new-search-capabilities.html>. The latest version of the search engine, launched early in 2011, queries the full text of all the documentation files, including descriptive text at the variable and study level, along with the related citations. Users can narrow their search using the right-hand column of the search-results page, where results can be narrowed by subject, geography, time-period, author or series.

NEW ICPSR LOGIN OPTIONS

Beginning May 2, ICPSR users will have new options for logging in to download data. Currently, users must set up a MyData account with an e-mail address and password in order to download data. With the new authentication options, one can use an existing Google or Facebook accounts for ICPSR login, without having to remember a separate MyData password. In addition, anyone already logged into their Google or Facebook account when they visit the ICPSR site would already be logged in for download at ICPSR.

UW-Madison ICPSR users will still need to either use a computer on the campus IP network, or use an off-campus proxy login with NetID.

ICPSR RESEARCH PAPER COMPETITION

ICPSR has announced six winners in the 2011 Research Paper Competition. The winning papers are available online at <http://www.icpsr.umich.edu/icpsrweb/ICPSR/prize/winners.jsp>.

There will be three competitions for 2012: The ICPSR Research Paper Competition, the Integrated Fertility Survey Series (IFSS) Research Paper Competition, and the Resource Center for Minority Data (RCMD) competition. Look for more information from *DISC News* in the September 2011 issue, or visit <http://www.icpsr.umich.edu/ICPSR/prize>.

DATA, DATA EVERYWHERE, WHY CAN'T I STOP AND THINK?

(RIME OF THE MODERN MARINER)

Continued from p. 1

illustrate, NSF proposals are categorized under four main Directorates: Engineering (ENG), Geological Sciences (GEO), Mathematical & Physical Sciences (MPS), and Social, Behavioral, & Economic Sciences (SBE), <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>—and all require a 2-page maximum data plan. The SBE data management plan consists of five points to be addressed: expected data, period of data retention, data formats & dissemination, data storage and preservation of access, and additional possible data management requirements. All of these are outlined at http://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf.

Feeling overwhelmed by this information overload? Click on the UW Stories link in the top menu-bar at <http://researchdata.wisc.edu/> to be assured you're in good company!

DATA PRODUCTS CAUGHT IN U.S. BUDGET CRUNCH

By Joanne Juhnke

This year's debate over the United States Federal Budget is particularly fraught with proposals for deep spending cuts. Several high-profile data products are among the targets.

The U.S. Census Bureau has proposed the elimination of its Statistical Compendia Branch, which is responsible for the *Statistical Abstract of the United States*, the *State and Metropolitan Area Data Book*, *County and City Data Book*, and more.

DISC has worked to raise the profile of online statistical compendia worldwide through our Guide to Country Statistical Yearbooks at <http://researchguides.library.wisc.edu/yearbooks>, and we make frequent use of the Census Bureau's statistical compendia products as we assist people in their search for the data they need. The *Statistical Abstract of the United States* and its companion products provide an invaluable, easy-to-use first step in navigating sources for United States data.

The American Library Association has issued an Action Alert calling for a restoration of funding for the Statistical Compendia Branch, at <http://capwiz.com/ala/callalert/index.tt?alertid=37054501>.

Also at risk is a suite of data transparency programs initiated by the Obama administration, including Data.gov and USASpending.gov. Initial reports anticipated such severe cuts to the E-Gov fund, which funds the data transparency programs, that sites would begin to be shut down within months. However, Daniel Schuman of the Sunlight Foundation stated on April 18, "Although the future of online transparency is uncertain, what we can know for sure is that our collective efforts averted a disaster. A few programs will survive, with the immediate fate of online transparency in the hands of federal CIO Vivek Kundra and a few key legislators."

The Sunlight Foundation's efforts on behalf of federal data transparency initiatives are online at <http://sunlightfoundation.com/savethedata/>.

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>.

Million Song Dataset

The Million Song Dataset is, in the words of dataset founder Daniel Ellis, “the first and only freely-available, industrial-scale dataset for research on popular music and audio analysis.” Online at Columbia University, <http://labrosa.ee.columbia.edu/millionsong/>, the dataset covers a million contemporary popular music tracks.

The data does not include the audio itself, but rather a set of derived features and metadata including such elements as the song artist and title, the number of beats and bars, loudness measures, and pitch and timbre information. Data files are in HD5 format, a format developed by NASA to handle large, heterogeneous, hierarchical datasets. Wrappers are provided for python, MATLAB, and Java.

The site also includes SecondHandSongs, a database of cover-songs, and the musiXmatch dataset of lyrics to songs found in the Million Song Dataset.

Global Peace Index

The Global Peace Index, online at <http://www.visionofhumanity.org/> and created by the Institute for Economics & Peace, is a measure for ranking countries in order of their peacefulness. The index is made up of 23 indicators, covering both external measures such as weapons exports and number of conflicts fought, and internal measures such as jailed populations and numbers of homicides. Data are compiled from many sources, including the Uppsala Conflict Data Program, the Economist Intelligence Unit, the World Bank, and

the International Institute of Strategic Studies.

The index includes data from 2007 to 2010, for a growing list of countries. In 2010, New Zealand was ranked as the world’s most peaceful country on the Global Peace Index. The United States ranked 96th of 121 countries in 2007, and 85th of 149 in 2010.

The site also recently launched the U.S. Peace Index, a comparison among U.S. states using a combination of 5 indicators: homicide rates, violent crimes, percentage of the population in jail, number of police officers and availability of small arms.

Health.Data.gov

In Nov. 2009, *Crossroads Corner* highlighted Data.gov, an open-data initiative from the U.S. government under the auspices of the U.S. Office of Management and Budget. In February 2011, Data.gov introduced a new component of its growing collection: a site devoted to health data, called **Health.Data.gov**. The Health.Data.gov site uses the same interface as Data.gov, and as of mid-March carried over 200 health-related “datasets and tools” in its catalog.

The site also features a portal called Health Indicators Warehouse, at <http://www.healthindicators.gov/>, sponsored by the U.S. Department of Health and Human Services. The Warehouse organizes its contents by indicators rather than by datasets, allowing users to browse and select indicators by topic, by geography, or by initiative (e.g., Healthy People 2020).

At its launch, the Health Indicators Warehouse held around 1200 health indicators from 170 different data sources.

Note: Health.Data.gov is among the resources threatened by potential budget cuts (see p. 3).