

IN THIS ISSUE

Open Data: Copyright,
Community, and Cost 1
ECLS-B Seminar 2
ICPSR Updates 2
ANES and NLS
Collaboration 3
Perils of Poorly-Used
Statistics 3
Crossroads Corner 4

.....

DPLS 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
Contributors: Cindy Severt, Senior Special Librarian, and Lu Chou, 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: dpls@mailplus.wisc.edu
Web: <http://dpls.dacc.wisc.edu/>
Hours: Monday-Friday 8:30-4:30

.....

This issue of DPLS News is online at:
<http://dpls.dacc.wisc.edu/pubs/Newsletters/nov06news.html>

Please Note

DPLS will be closed:
Thu./Fri. Nov. 23-24 - Thanksgiving
Mon./Tue. Dec. 25-26 - Christmas
Mon. Jan. 1 - New Year's Day
Mon. Jan. 15 - Martin Luther King, Jr. Day

OPEN DATA: COPYRIGHT, COMMUNITY, AND COST

by Cindy Severt

One of the newest trends in the world of social science data is the concept of Open Data – a concept that is still being defined even as it raises issues more thought-provoking than the concept itself. Elementally, “Open Data is a philosophy and practice requiring that certain data are freely available to everyone, without restrictions from copyright, patents or other mechanisms of control” (Wikipedia entry on Open Data, 11/13/2006).

Presuming that the data in question does not jeopardize respondent confidentiality or other ethical concerns, it appears that copyright is at the core of the Open Data discussion. Open Data challenges the very notion of copyright which allows the author to retain the rights to their intellectual property. But does traditional protection of intellectual property have a place in a world where that property can be easily duplicated and distributed

Is “data” intellectual property, or is it an agglomeration of common property (facts)?

regardless of whether or not the property is in the public domain? For that matter, just because intellectual property can be made freely and openly available and accessible, should it be? Furthermore, is “data” intellectual property, or is it an agglomeration of common property

(facts) that one individual or entity has taken the time and trouble to collect, albeit with an ulterior motive? Has technology so outpaced copyright law that it has been rendered obsolete?

The strongest argument for Open Data is a communal one. Not surprisingly, the concept of Open Data began in the hard sciences where it can be argued that data belongs to the human race: “The Earth's atmosphere, oceans, and biosphere form an integrated system that transcends national boundaries. To address global issues, it is essential to have global data sets and products derived from these data sets” (*On the Full and Open Exchange of Scientific Data*, Committee on Geophysical and Environmental Data - National Research Council, 1995).

The most compelling argument against Open Data is the cost of management. For all its proponents’ rallying cries that data should be free, there are costs involved. Without a procedure in place to identify, collect, standardize, archive, and disseminate data, it will languish unused and deteriorating on a shelf like a non-circulating library book -- and how open is that? Public goods require an investment, at the very least, to cover the operation costs of maintaining open data, but not at the cost of dividing the social science research community into the haves and have-nots.

Continued on p.3

ECLS-B SEMINAR

by Lu Chou

The National Center for Education Statistics (NCES) will sponsor a three-day (January 10-12, 2007) seminar in Washington, DC on the use of the Early Childhood Longitudinal Study Birth Cohort (ECLS-B).

The ECLS-B was designed to look at children's health, development, care and education from birth through kindergarten. Three waves (9-month, 2-year and preschool) of data have been collected from about 14,000 children born in 2001. Their parents, care providers, preschool teachers and school administrators were interviewed. An additional round of data collection is underway for Fall 2006, with a final round to be completed in Fall 2007. The ECLS-B provides a wealth of information for researchers on cognitive, social, emotional and physical development across multiple contexts such as home, nonparental care and school entry for young children.

The seminar is open to advanced graduate students and faculty members, as well as government and private-sector researchers. The seminar is free of charge, and support for travel and lodging will be provided to accepted seminar applicants. The application deadline is November 27, 2006.

For more information, please visit the NCES site, <http://nces.ed.gov/whatsnew/conferences/confinfo.asp?confid=71>.

ICPSR UPDATES

by Joanne Jubnke

This column presents some recent announcements of note from the Inter-university Consortium of Political and Social Research (ICPSR), the massive archive of digital social science data at the University of Michigan in Ann Arbor. The latest news from ICPSR is always available from their web site at <http://www.icpsr.umich.edu/>. To receive e-mail updates of the latest studies added to the ICPSR archive, one can also subscribe to the Recent-updates-and-additions e-mail list at <http://www.icpsr.umich.edu/org/elists.html>.

Undergraduate Summer Internship

ICPSR is accepting applications for its annual ten-week summer internship program for undergraduates, to take place from June 11, 2007 to August 17, 2007. Summer interns will work in a UNIX and Windows environment, describing and preparing data for archiving and distribution. Interns also attend courses in the ICPSR summer program in quantitative methods of social science research. Requirements include completion of sophomore year in a social science major, a strong academic record, and knowledge of a statistical software package. Interns receive a \$3000 stipend, plus room and board and coverage of coursework costs. Application deadline is January 5, 2007; see <http://www.icpsr.umich.edu/careers/internship.html> for application details.

On-Demand Creation of Setup Files

ICPSR has begun offering a new service to those who wish to use data from the ICPSR archive for which setup files do not currently exist. The service serves two ends: to provide users with the files they need, and to help ICPSR create priorities in the process of eventually retro-fitting the entire collection with setup files. To request setup files for a specific study, send a message to netmail@icpsr.umich.edu. User Support staff will discuss your request with you and then provide an estimate of how long it will take to create the files.

Voting Behavior Teaching Module Wins Awards

At the 2006 American Political Science Association convention, the *Voting Behavior: The 2004 Election* online instructional module hosted by ICPSR received two awards, for Best Instructional Web Site as well as for Innovative Teaching in Political Science. ICPSR and APSA have produced a voting-behavior module using subsets of ICPSR data for presidential election years since 1972; however, the 2004 module, authored by Charles Prysby and Carmine Scavo, is the first one to be made available entirely online. The module, which includes online exercises using the SDA online data analysis system, may be found at <http://www.icpsr.umich.edu/SETUPS/index.html>.

OPEN DATA

Continued from p.1

To stay updated about Open Data, users can join the Scholarly Publishing and Academic Resources Coalition (SPARC) Open Data Email Discussion list at <http://www.arl.org/sparc/opendata/>. DPLS will follow up this article next year after than 2007 IASSIST conference: Building Global Knowledge Communities with Open Data.

ANES AND NLS COLLABORATION

by Lu Chou

Seventeen election-related questions were included in the 2006 questionnaire for the children of the National Longitudinal Survey of Youth 1979 (CNLSY 79). These questions mark the first-ever collaboration between two major US surveys, the American National Election Studies (ANES) and National Longitudinal Studies (NLS), in which the ANES has purchased time on two NLS surveys during the next three years. This collaboration opens up new research opportunities in linking electorally-relevant and socially-relevant responses across generations. ANES users will have access to all of the data collected on these respondents in 2006 and 2008, in addition to all of the data collected on these respondents in all previous years of the study.

The question-selection process was a challenging one. Nearly 100 scholars offered suggestions for the ANES questions, totaling 400 minutes of question time. Because only 4 minutes of time were allocated to ANES questions on the CNLSY 79, the ANES team had to select their questions carefully. The final questions were chosen to address not only the theoretical and empirical aspects of election studies but also cover the intergenerational and longitudinal dynamics of electoral experiences. The questions addressed partisanship, turnout and participation, as well as respondents' perceptions of their parent's partisanship. The questions also addressed perceptions about governmental responsiveness, the role of values and trust, and the extent to which people follow politics in the media.

The questions can be fielded again in 2008 for the CNLSY 79 cohort. Meanwhile, ANES is working on a smaller set, one-minute's worth of questions, to be included in the 2008 survey of the NLSY79 respondents.

More information about the ANES is available online at <http://www.electionstudies.org/>. Scholars can become involved in the ongoing question-selection process for ANES surveys through the ANES Online Commons at <http://www.electionstudies.org/onlinecommons.htm>.

PERILS OF POORLY-USED STATISTICS

by Joanne Jubnke

I recently received an e-mail message pointing out a web site that purports to estimate the number of people in the US who share any particular combination of first and last names, based on data from the U.S. Census Bureau: a site called How Many of Me? at <http://ww2.howmanyofme.com/>. To anyone at all familiar with census data, the site looks suspicious immediately, despite an attractive interface and well-written text. Anyone who further reads the fine print discovers that in fact the site's authors have their tongue quite firmly in cheek, as they describe their site as "More accurate than a Magic 8-ball. Less accurate than distributing and collecting 300 million surveys." At least the disclaimers are there for those who choose to look!

Unfortunately, very little statistical misuse comes with its own debunking. The peer-review process provides the filter for journal literature, but the news media (who use statistics all the time) have no such rigorous controls. One helpful counterweight to statistical blunders in the media is the Statistical Assessment Service (STATS), <http://www.stats.org/>, at George Mason University. Since its founding in 1994, the non-partisan STATS has worked to monitor "the use and abuse of science and statistics in the media." Recent analyses have included "best college" lists, the benefits of breastfeeding, and the Johns Hopkins study of excess-deaths in Iraq.

Crossroads Corner highlights web sites recently added to the searchable Internet Crossroads in Social Science Data, available on the DPLS web site.

National Annenberg Election Survey (NAES)

In the past two U.S. presidential election cycles (2000 and 2004), the National Annenberg Election Survey has carried out a major cross-sectional survey conducted as daily telephone interviews with samples of 50 to 300 people per day, for a total of more than 100,000 interviews with some respondents interviewed multiple times. In each cycle the interviews started over a year before the general election, thereby covering the primary elections as well, and in 2000 the interviews continued several months past Election Day in order to continue to track issues since the outcome of the presidential election remained so long in doubt. Interview questions focused on perceptions and behaviors relevant to the campaigns as well as the political system generally.

DPLS has acquired the 2000 NAES data, along with the accompanying book *Capturing Campaign Dynamics: The National Annenberg Election Survey* from Oxford University Press. The data is available on the public PC workstations at DPLS. The NAES web site, at <http://www.annenbergpublicpolicycenter.org/naes/index.htm>, provides information about the survey and the investigators, as well as providing press releases and reports from both 2000 and 2004.

Pollster.com

Since 2004, Democratic pollster Mark Blumenthal (aka Mystery Pollster) has been blogging about, as he put it, “demystifying the science and art of political polling.” More recently, he and UW-Madison’s own Charles Franklin co-developed the **Pollster.com** site, through which they have been

analyzing political polling fast and furiously through the recent elections. The site’s twin strengths are the commentary and the graphic display of polling results tracking the various political campaigns. Users will find maps and charts that track the polls for each individual race, with links to the originator of each poll. Some of the originator links lead to newspaper articles or press releases, while some (notably Survey USA) lead to more detailed online tables.

The latest polls reported on the site are now focusing on the 2008 presidential primaries, so watch the site for more to come!

Worldmapper

The Worldmapper site takes a catchy title (“The World as You’ve Never Seen It Before”) and puts it into data-driven action, featuring cartograms that display global regions “re-sized according to the subject of interest.” A total-population world map, for example, displays India and Japan swollen to outsized proportions, while the United States looms large on the map reflecting private spending on health-care, and Southeastern Africa dominates the map of HIV prevalence. Some of the broad topics include health, education, transportation, communication, work, and housing, but the list continues to expand. A new set of health-related maps added on November 13 brings the total up to 227 maps and counting.

Each map comes with a downloadable PDF poster and downloadable data files in Excel and OpenDoc format. The Worldmapper project is a collaboration between the University of Sheffield (UK) which hosts the site, and the University of Michigan.

Worldmapper can be found online at <http://www.sasi.group.shef.ac.uk/worldmapper/index.html>