

IN THIS ISSUE

So Long to the Long Form . . . 1
ARGUS, Statistical Disclosure
Control Software 2
DACC Department Admin-
istrator Retires 2
IASSIST 2005 Conference in
Edinburgh 2
2005 ICPSR Summer Program
. 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 and other library users in 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 Social Science 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/feb05news.html>

SO LONG TO THE LONG FORM

The trial runs started almost a decade ago. The full-scale launch was delayed two years due to budget woes. But finally in January 2005, questionnaires from the American Community Survey (ACS) began arriving in mailboxes in every county of the United States.

The ACS is the U.S. Census Bureau's response to ever-louder demands for timely demographic data. The traditional decennial census collection carried the immense disadvantage of a ten-year time lag. By the time the next decennial survey came around, the previous information had grown unfor-

About one in forty households will receive the American Community Survey in any given year.

tunately stale. The Census Bureau eased some of the wait through its annual Population Estimates program (<http://www.census.gov/popest/estimates.php>). However, the estimates only covered the basics: popu-

lation numbers, sex, age, race and Hispanic origin, and housing units. For the more detailed information sampled via the long form questionnaire, data users simply had to wait.

No longer. On January 10, the Census Bureau began sending out its full production of 250,000 ACS surveys per month, asking the long-form questions in a rolling, random sample of nearly three million households per year. Where about one in six households received a long form in decennial census years past, about one in forty households will now receive the American Community Survey in any given year. The odds of receiving the ACS over a decade are less than one in four, and no household will receive more than one questionnaire every five years.

The five-year time frame is important. It will take five years, given the number of surveys being sent, to collect enough ACS data to be accurate down to the census tract level. This means that the year 2005 was the cutoff to begin full-fledged data collection, in order to have usable data at all levels by 2010 and so to legitimately abandon the decennial long form.

The Census Bureau raised the alarm in October 2004, when a Senate Appropriations Committee approved a bill that slashed over \$100 million from the Bush administration budget request for the ACS, offering only \$65 million instead of \$165 million. At that funding level, the Bureau warned, the ACS would be cut off at the knees - and the 2005/2010 deadline meant that further postponement could not be tolerated.

Continued on p. 3

ARGUS, STATISTICAL DISCLOSURE CONTROL SOFTWARE

Statistical Disclosure Control (SDC) has attracted much attention recently. As personal computers become more powerful, it is easier to conduct complex statistical analysis and to link files from different data sources. This presents increasing chances of breaking confidentiality by re-identifying business entities or individuals from public-use data.

Several statistical agencies in the European Union have taken on the challenges of the legal confidentiality requirements versus the increased data needs and capabilities of researchers and policy-makers. The Computational Aspects of Statistical Confidentiality (CASC) project is the result of this collaboration, which took place from 2001-2003. The CASC project comprises not only statistical theories and methods, but also the development of the ARGUS software, available from the CASC web site at <http://neon.vb.cbs.nl/casc/>.

ARGUS was developed under Windows NT and runs under Windows versions from Windows 95. It was first written in Borland C++ and then converted to Visual C++ with its user interface written in Visual Basic. The stated goal of ARGUS is to “modify unsafe data in such a way that safe (enough) data emerge, with minimum information loss.” In this way, data producers can safely release the data to

Continued on p. 3

DACC DEPARTMENT ADMINISTRATOR RETIRES

Jean Schneider, one of the Department Administrators for the Data & Computation Center, retired in January after 36.5 years of service to the UW. In recognition of Jean's friendship and service, a cake and punch reception was held at DPLS on a snowy January 5th. As we bid Jean a fond farewell, we reminded her that the trouble with retirement is that you never get a day off!



Above: Jean Mindel and Jean Schneider, Department Administrators. *Right:* Jean Schneider displays certificates of appreciation from the UW and the State.



IASSIST 2005 CONFERENCE IN EDINBURGH

The annual conference of the International Association of Social Science Information Service & Technology will take place May 24 -27, 2005 in Edinburgh, Scotland. (<http://datalib.ed.ac.uk/iassist/>) Held in conjunction with IFDO, the International Federation of Data Organizations, the conference will include a day of workshops, and three days of plenary, concurrent, and poster sessions.

Online at <http://www.iassistdata.org/>, IASSIST's 200 members come from a variety of workplaces including data archives, statistical agencies, research centers, libraries, academic departments, government departments, non-profit organizations, and inter-governmental organizations.

IASSIST welcomes new members (student rate is \$25). Benefits include:

- Peer to peer support and involvement in an international network of data professionals.
- Subscription to *IASSIST Quarterly (IQ)*.
- Participation in IASSIST-L, the organization's email discussion list.
- A copy of the IASSIST Membership Directory.
- Participation on committees, action groups, and interest groups.

SO LONG TO THE LONG FORM

Continued from p.1

Fortunately, the omnibus spending bill approved in late November came much closer to the request at \$146 million, sufficient for the ACS to move forward as planned.

For areas with populations of 65,000 or more, ACS annual data will be available beginning in the summer of 2006. For areas with populations between 20,000 and 65,000, annual data will first be released in the summer of 2008 based on rolling 3-year averages. For smaller areas down to the tract level, annual data will debut in the summer of 2010 based on rolling 5-year averages.

The American Community Survey already has considerable data on its web site, <http://www.census.gov/acs/www/>. The site contains both aggregate data and PUMS microdata from all of the testing phases of the survey, back to the first seven test-counties that received the survey in 1996. Since 2000, the ACS has already been operating in 1,239 counties across the nation, gathering a substantial body of data.

The ACS is authorized by the same statutes that govern the decennial census, and those who receive the survey are legally required to respond. The decennial census itself will continue to exist, with the Census Bureau currently preparing for Census 2010. Census 2010 will still feature the short form, to provide the 100 percent count and demographic information necessary for Congressional reapportionment and redistricting. But so long, long form... the American Community Survey is ready for prime-time.

2005 ICPSR SUMMER PROGRAM

Again in 2005 ICPSR will be presenting its Summer Program of 3-5 day workshops and 4-week courses which:

- Offer instruction for the primary development and “upgrading” of quantitative skills by college and university faculty and by nonacademic research scholars.
- Extend the scope and depth of analytic skills for graduate students, college and university faculty, and research scientists from the public sector.
- Furnish training for those individuals who expect to become practicing social methodologists.
- Provide opportunities for social scientists to study those methodologies that seem to have special bearing on specific substantive issues.

For full course listings and application, visit <http://www.icpsr.umich.edu/sumprog/>. It is anticipated that a travel stipend will be available for UW-Madison applicants. Please contact Cindy Severt at cdsevert@wisc.edu for more information.

ARGUS

Continued from p.2

researchers and the public. The ARGUS software has two components for achieving statistical disclosure control. The μ -ARGUS component is for safeguarding microdata, while τ -ARGUS is designed to make tabular data safer.

Using statistical packages like SAS, SPSS and Stata, data producers can apply various statistical methods and risk models to make their data safe. However, it is very time-consuming to do the necessary global recoding and case swapping with any existing statistical packages. In addition, data producers must document any changes they have made to the original data in order to achieve their SDC criteria. For these reasons, producing safe data files using existing statistical packages has been a substantial task.

The ARGUS software is designed to help data producers through these steps. ARGUS allows users to easily apply the various SDC theories and methods incorporated in its features. In addition, it documents any changes made to the original microdata. In the end, ARGUS produces a safe data set and a complete document of the SDC process.

The UW-Madison’s Center for Demography of Health and Aging (CDHA) hosted an SDC workshop last November at the Pyle Center, at which the ARGUS software was demonstrated, and Lu Chou attended on behalf of DPLS.

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

UNESCO Institute for Statistics (UIS)

Hosted by the University of Montreal in Canada, the UNESCO Institute for Statistics (UIS) is the United Nations depository for global statistics in the fields of education, science, culture and communications. The UIS web site, at <http://www.uis.unesco.org/>, provides access to the public data they collect in these fields, going back to their founding in July of 1999. Many of the tables can be manipulated through the Beyond 20/20 web interface, with downloads available in CSV and Excel; others are only available as Excel files. The site also carries questionnaires and manuals for ongoing surveys, and related articles and reports. Featured initiatives include the UIS participation in the UN Millennium Development Goals program; the Education for All movement; and World Education Indicators.

National Survey of Student Engagement (NSSE)

The National Survey of Student Engagement (NSSE) conducts an annual survey for colleges and universities in the United States about student participation in programs and activities that institutions provide for their learning and personal development. The results address the way in which undergraduates spend their time and what they feel they gain from attending college. Participating institutions receive data specific to their school along with peer institution comparisons. The NSSE web site, at <http://www.iub.edu/~nsse/>, carries annual PDF reports back to 2000, summarizing responses from the various types of participating institutions and students. The site also links to related surveys for law schools (LSSSE), high schools (HSSSE), and community colleges (CCSSE).

Over 470 institutions participated in NSSE in

2004, including the UW-Madison. UW-Madison also participated in NSSE in 2001, <http://wiscinfo.doit.wisc.edu/obpa/NSSE/>.

Facster

Fondly known as “The National Data Book,” the *Statistical Abstract of the United States* has just become more readily available, thanks to a new web site called Facster at <http://www.facster.com/>. The Census Bureau publishes the *Statistical Abstract* to serve as a summary of social, political and economic statistics for the US, compiling the tables from both public and private sources. The annual volume is sold in print and CD-ROM format (both of which are in the DPLS collection), and the Census Bureau has put the print volume online as a series of PDF documents at <http://www.census.gov/statab/www/>. However, until Facster came along, one had to use the CD-ROM to get any of the tables in manipulable format.

The Facster web site has taken the tables from the 2000, 2002, and 2003 *Statistical Abstract* CDs, and from the 2003 *California Statistical Abstract*, and created an interface for browsing by category and searching by keyword. The results are presented in HTML tables, but can be cut and pasted into Excel. While the bulk of the *Statistical Abstract* is viewable in Facster, there are still 126 of the 1439 tables in the 2003 *Statistical Abstract* that are under copyright and for which Facster has not (yet) received permission to display the contents. In these cases, a copyright notice appears in the data cells of the table.

One useful Facster feature is the “Select a Subset of this Information” button, that lets one pare down which rows are included in a table. Facster interface quirks include the fact that sections and tables are listed in alphabetical order rather than by the table numbers used in the book. Also, to get the table number and the source of the information, one has to click the “Additional Information” button.