. . . .

. . . .

. . . .

PDFs in Perpetuity?

In the dozen years since Adobe introduced the Portable Document Format (PDF), its popularity has skyrocketed. PDF files brought to the World Wide Web a quality that HTML files were lacking: the ability to display a document with high-quality, predictable layout.

PDF has other advantages as well. The Portable in PDF means that, as an FAQ on **PDFzone.com** put it, "you can read a PDF document in Windows that was created on a Macintosh that you downloaded from a Web site running Unix." One doesn't have to have software that matches the software on the author's machine, either: just a free PDF-reader program downloaded from the Adobe web site.

> Preservation in the digital world has its own unique challenges.

However, PDF has not been considered a suitable format for digital preservation, or storing a document for the long term. Digital preservation is of vital interest for an organization like DPLS, whose mission includes preserving data and documentation for current and future researchers. Data for secondary analysis is inherently digital, and modern documentation is "born digital" as well. Preservation in the digital world has its own unique challenges, though. Digital formats and media become obsolete with depressing speed and regularity. It is important to make a careful choice for the long haul!

To be a worthy choice for preservation, a digital format must be widely-used, not proprietary to a single company, and its specifications must be openly available. ASCII text, for example, the now-standard computer representation of Roman alphabet and numbers, is a format of choice for long-term storage of social science data. ASCII is very basic, widely used, not associated with any particular company, and its coding is no secret.

PDF in its current form comes close to the criteria, but not without concerns. It is certainly widely-used. It is also currently, and perhaps surprisingly, an open-source format: Adobe has published the PDF specifications, leading to an explosion of PDF readers and writers from companies other than Adobe. However, Adobe is still at the helm of the PDF market. A new version of PDF still has to come from Adobe to have any hope of broad acceptance, and this dynamic makes PDF still an essentially-proprietary format. Also, the very flexibility of PDF raises certain roadblocks. It is possible, for example, to store ("embed") within a PDF document the fonts used in its creation – but it is also possible to create a PDF document

PDFs in Perpetuity?

without embedding the fonts. If the fonts are not embedded, and the end-user of the document does not happen to have the fonts on his or her machine, the document will not retain its original look. Also, PDF documents may be created with security settings such as password protection, which can prevent others from opening the document.

Several organizations are partnering to overcome the preservation concerns about PDF documents. The National Archives and Records Administration (NARA), the National Information Standards Organization (NISO) and Adobe are working to create an archival version of PDF, called PDF/A. The PDF/A standard is slated to include a core set of PDF attributes that Adobe agrees to maintain in future PDF versions, while disallowing the more problematic features. Future PDF-writing programs would then ideally incorporate the PDF/A standard, so that the document's creator could simply choose to save the document in PDF/A.

As the proposed PDF/A standard wends its way through the standards process, with a publication target date of 2006, it is worth noting that sometimes the old ways still provide an additional measure of safety. Both at DPLS and ICPSR (Inter-university Consortium for Political and Social Research), the final archival resort for codebooks is still… the print-on-paper copy!

2004 National Election Studies

A Full Release of the 2004 National Election Studies time-series study is now available from the NES web site, **http://www.umich.edu/~nes/studypages/2004prepost/2004prepost.htm**. Its sample consists of a new cross-section of respondents. There were 1,212 face-to-face interviews in the pre-election study, with 1,066 of those respondents later providing another face-to-face interview in the post-election study.

In addition to content on electoral participation, voting behavior, and public opinion, the 2004 NES contains questions in other areas such as media exposure, cognitive style, and values and predispositions. Special-interest and topical content provided significant coverage of foreign policy, including the war on terrorism and the war in Iraq. In addition, the study carried expanded instrumentation on unemployment and inflation, gender politics, and gay and lesbian politics. An exceptional feature is the insertion of the comprehensive module on representation and accountability, Module 2, from the Comparative Study of Electoral Systems (CSES), at the end of the post-election interview.

CELF Examination

Using vanguard research methods and an interdisciplinary approach, "scientists at UCLA have spent the past four years observing 32 Los Angeles families in a study of how working America gets it done. Day after day." (*Wisconsin State Journal*, 3/21/05).

Conducted by UCLA's Center on Everyday Lives of Families (CELF) the study is one of six long-term interdisciplinary projects sponsored by the Alfred P. Sloan Foundation examining the intersection between family life and work.

Actual field work consisted of videotaping in 32 households with two parents who work outside the home, pay a mortgage, and have two or three school aged children. Taping in each case spanned seven days.

The data consists of over 1500 hours of videotape, and the project is cross disciplinary (cultural anthropology, linguistic anthropology, archaeology, biological anthropology, applied linguistics, education, psychology). A second phase devoted to analysis is about to begin, but trends are already emerging. The reality of both parents working outside the home is proving to have a major impact on family dynamics. According to the *WSJ* article, "it means parents and children live virtually apart at least five days a week, reuniting for a few hours at night. Playtime, conversation, courtesy, and

FULLY-FUNDED NCES SEMINARS

The National Center for Education Statistics (NCES) is once again sponsoring advanced-studies summer seminars on the use of NCES data. Seminars are designed for researchers such as faculty members, advanced graduate students, and other data analysts with statistical knowledge and proficiency in the use of SAS or SPSS, who plan to use NCES data for their research. Each multi-day seminar will be held in the Washington, D.C. metro area, and will be fully funded by NCES (fees, transportation, lodging, meals) for accepted applicants. The primary purpose of these seminars is to demonstrate the richness of NCES datasets and provide instruction on how to use the data properly and effectively.

The upcoming seminars and their application due dates are as follows (all dates are 2005):

Topic	Application Due	Seminar Date	Contact
NHES	May 2	June 15-17	Joy Butler
ECLS-K	May 6	June 27-30	Christine Forest
SASS	May 6	June 27-30	Joy Butler
NAEP	May 23	July 5-8	Christine Forest
NELS:88 / ELS:2002 II	June 6	July 20-22	Joy Butler

Application information is posted on the NCES web site at **http://nces.ed.gov/conferences/**. If you have any questions regarding these seminars, please contact Joy Butler at (703) 807-2315 (**joyb@smdi.com**), or Christine Forest at (703) 516-8873 (**christinef@smdi.com**).

2005 GSOEP-CNEF WORKSHOP

Cornell University will hold a workshop September 8-10, 2005, to introduce researchers to the German Socio-Economic Panel (GSOEP) and the Cross-National Equivalent Files (CNEF). The CNEF promotes cross-national comparative research using a subset of data from the GSOEP and three other countries' panel studies: the British Household Panel Study (BHPS) from Great Britain, the Survey of Labour and Income Dynamics (SLID) from Canada, and the Panel Study of Income Dynamics (PSID) from the United States.

The workshop fee is $200. Scholarships are available to partially offset the cost of attending the workshop. Scholarship applications must be postmarked by July 15, 2005, and workshop applications must be postmarked by August 1, 2005. For more information and a link to the online application form, visit **http://www.human.cornell.edu/pam/gsoep/2005conf/**.

CELF EXAMINATION

intimacy are falling by the wayside, and children are driving the minivan as most family decisions and purchases are geared toward the kids' activities. Few people have unstructured time. 'We've scheduled and outsourced a lot of our relationships,' says the study's director, Elinor Ochs, a linguistic anthropologist. 'There isn't much room for the flow of life, those little moments when things happen spontaneously. And we're moving from a child-centered society to a child-dominated society.'"

For more on the project's goals and eventual results, see **http://www.celf.ucla.edu/**.

NEW STUDIES AT DPLS

- Crash outcome data evaluation system (CODES): public use data files, Wisconsin, 1992-2002.
- Cross-national equivalent file, 1980-2002, (BHPS-GSOEP-PSID-SLID).
- Data from a 1997 insurer study of auto injury closed claims.
- German socio-economic panel (GSOEP) 1984-2002.
- Government finance statistics, 1971-1997.
- National Election Pool poll, 2004: Democratic primary election day exit polls (various states).
- National election study (NES), 2004.
- State Politics & Policy Quarterly Data Resource, state data, 1974-2001.

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 DPLS web site.

National Priorities Project Database

The National Priorities Project (NPP) is a United-States-based non-partisan education and advocacy group that examines the community-level impacts of federal tax and spending policies. The NPP takes a particular interest in the trade-offs between military spending and tax breaks versus social spending.

The NPP Database, at **http://database. nationalpriorities.org/**, contains state and county level data on U.S. federal spending in the areas of hunger, military, income & poverty, housing, education, and labor, in addition to basic demographics. Users can select up to five "datasets" (i.e., specific federal programs or demographic attributes) to create an HTML table covering multiple states or multiple counties within one state, with information as far back as 1983. Visitors are requested to complete a free registration to use the database after their first visit.

State Politics and Policy Quarterly - Data Resource

The SPPQ Data Resource compiles 50 variables from various sources for the United States by state, covering the following areas: population and vital statistics; politics; education; crime; and business & economy. The list includes such items as divorce rate, total number of police officers, gross state product, and per-pupil education expenditures. Most variables are annual figures, and some go back as far as 1975.

Visitors to the site, at **http://www.unl.edu/ SPPQ/datasets.html**, can download the entire set or a single subject area, in Excel for-

mat. A separate codebook in Word lists variable descriptions and sources.

Pew Hispanic Center

The Pew Hispanic Center, supported by Pew Charitable Trusts, was founded in 2001 to study the U.S. Hispanic population and the impact of the Latino community within the United States as a whole. Online at **http://pewhispanic.org/**, the Center commissions studies on topics such as education, immigration, labor, and economics, including some public opinion surveys.

Several datasets are available for download on the Pew Hispanic Center web site, including the 2004 National Survey of Latinos: Politics and Civic Participation. Research reports back to 2002 are also available.

National Women's Health Indicators Database

The National Women's Health Indicators Database made its debut carrying women's health data from the year 2000 for the United States, and the Office on Women's Health of the U.S. Department of Health & Human Services intends to update the database annually. Data can be retrieved for national, regional, state and county geographic areas, and can be broken out by race/ethnicity and age. Indicators cover mortality, access to care, infections and chronic disease, reproductive health, maternal health, mental health, prevention, and violence and abuse.

Access to the database at **http://www.health status2000.com/owh/** is free, and visitors can use the drop-down menu interface to make their own tables and graphs. The site also incorporates a mapping function using ArcView GIS.