

VOLUME 1 NUMBER 2 1999

OF SIGNIFICANCE...

**A TOPICAL JOURNAL OF
THE ASSOCIATION OF PUBLIC DATA USERS**

**APDU99—
RESPONSIBLE DATA USE:
PROTECTING THE FUTURE
OF PUBLIC DATA**

APDU ASSOCIATION OF PUBLIC DATA USERS

Of Significance... represents an effort on the part of the Association of Public Data Users (APDU) to bring together a variety of perspectives on specific issues of importance to the collectors, disseminators, and users of public data. Each issue will focus on a single theme related to the mission and interests of ADPU members.

APDU is a national network...

- that links users, producers, and disseminators of government statistical data

Our members share a vital concern...

- about the collection, dissemination, preservation, and interpretation of public data

As an association, we are committed...

- to helping users identify public data that meet their needs and apply those data in cost-effective and appropriate ways
- to establishing effective two-way communication between data producers and users, and
- to bringing the perspectives and concerns of public data users to bear on important issues of government information and statistical policy

Of Significance...

Of Significance... represents an effort on the part of the Association of Public Data Users (APDU) to bring together a variety of perspectives on specific issues of importance to the collectors, disseminators, and users of public data. Each issue will focus on a single theme related to the mission and interests of APDU members.

Information to Authors:

Of Significance... is published twice a year. Readers interested in submitting articles for possible publication should consult the APDU web site for specific issue themes and contact person. Inquiries regarding future issue themes and submission requirements can also be obtained by calling or writing APDU Administrative Headquarters.

Price:

Single issue: \$45.00 (U.S.)
Yearly subscription: \$80.00 (U.S.)
(Subscriptions available to U.S. individuals and organizations only)
U.S. individuals and organizations receive *Of Significance...* as a benefit of membership. Member benefits also include a quarterly newsletter, conference discounts, and occasional special publications. Information on member categories and costs can be obtained by contacting APDU Administrative Headquarters or checking the APDU web site.

Back Issues and Claims:

Information about availability of issues may be obtained from APDU Administrative Headquarters. Claims for undelivered copies must be made no later than six months after publication. The publisher will replace missing copies when losses have been sustained in transit and when the reserve stock will permit.

Name of publication: *Of Significance...*
Issue: Volume 1, Number 2, 1999
Frequency: Two issues per year, June and December

APDU Administrative Headquarters
Teresa Hall Allen, Administrator
P.O. Box 12538
Arlington, VA 22219
Phone: (703) 807-2327
FAX (703) 528-2857
E-mail: apdu@apdu.org
URL: www.apdu.org

Guest Editor for Vol. 1:2: Janie Harris
General Editor: Deirdre Gaquin
Production Editor: Teresa Hall Allen

Copyright © 1999 by APDU.
All rights of reproduction in any form reserved.

ISSN 1528-1329

Contents

Introduction to APDU99: Responsible Data Use	1
<i>Stephen Dienstfrey and Lisa Neidert</i>	
Opening Remarks	
Responsible Data Use: Protecting the Future of Public Data	3
<i>Wendy Treadwell</i>	
Keynote Speech	4
<i>Kenneth Prewitt</i>	
The Census Bureau's Integrated Information Solutions	5
<i>Mark E. Wallace</i>	
What Has Been Going on in the Public Data Sector?	6
Developments in the Public Data Infrastructure	6
<i>Katherine Wallman</i>	
Developments in Public Data Policy	8
<i>Edward J. Spar</i>	
Reflecting on the Past and Looking to the Future	10
<i>Judith Rowe</i>	
Making Things Add Up for the End User: Issues in Statistical Literacy	14
<i>Katherine Wallman, Marianne Zawitz, Colleen Blessing, and Wendy Treadwell</i>	
Race and Ethnic Statistics: What Can Users Expect?	17
<i>Clyde Tucker and Jorge del Pinal</i>	
Down the Home Stretch to Census 2000: a Public Policy Perspective	21
<i>Fred T. Asbell, Chris Strobel, Chip Walker, and David McMillen</i>	
Respondent Disclosure in Surveys: Issues and Solutions	24
<i>Alvan Zarate, Patrick Collins, and Lisa Broniszewski</i>	

Local Area Indicators: Are We Really Measuring What We Think We Are?	25	Changes in the MSA Definitions: What is the Impact on Local Users ?	35
<i>Margaret Plantz and Anne E. Dunthorn</i>			
Promoting Public Access to Scientific Data: the Devil is in the Detail	28	Progress Report on the Review of the Metropolitan Area Standards	35
<i>George Leventhal, Mark Frankel, and Gary Bass</i>		<i>James D. Fitzsimmons</i>	
What Are the Effects of Welfare Reform on the Nation's Children?	30	The Good, the Bad and the Ugly: A Look at Counties and Metropolitan Areas	37
<i>Alyssa Wigton and Jeffrey Capizzano</i>		<i>Jacqueline Byers</i>	
Roundtable Sessions: Data Producer and User Forums	32	Post-Mortem Report: Address Review	38
		<i>Patricia C. Becker and Joseph J. Salvo</i>	
Do's and Don'ts of Getting Statistical Policy Across to the Local Public	32	Metadata on Steroids	40
<i>Milo Schield</i>		<i>Ann Green, Wendy Treadwell, and Peter Joftis</i>	
Hands-on Time with the American Community Survey	33	The Vision of Integrated Access to Statistics: the Data Web	42
<i>Nancy Torrieri</i>		<i>Cavan Capps, Ann Green, and Mark Wallace</i>	
UNICON and CPS Utilities	33	Electronic Products for Census 2000	48
<i>Bryan Rickard</i>		<i>Patricia C. Becker and Joseph J. Salvo</i>	
Getting to Know Your Data with PDQ Explore	34	The Overlooked Undercount: Children Missed in the Decennial Census	49
<i>Albert Anderson</i>		<i>William P. O'Hare</i>	
The Nation's Data on DVD: an Introduction to USA on DVD	34	Report on the Panel to Review the 2000 Census: Issues and Objectives	50
<i>Brand Niemann</i>		<i>Janet L. Norwood</i>	
		Acknowledgments	51
		<i>Janie Harris</i>	
		Appendix 1: List of Speakers, APDU99	52

Introduction to APDU99— *Responsible Data Use: Protecting the Future of Public Data*

STEPHEN DIENSTFREY
LISA NEIDERT

Alexandria, Virginia, October 24-27, 1999

APDU99 was the 24th annual meeting of the Association of Public Data Users. As with previous conferences, we tried to present a program that was both topical and informative. To this end, the program contained a variety of sessions that discussed the events of the last year in the public data arena. These included tabulations for the new race and ethnic standards and rules for sharing data as described in Office of Management and Budget (OMB) Circular A-110. Other sessions explored where we stand on the ongoing issues about Census 2000 and the Census Bureau's new Integrated Information Solution Initiative.

As is traditional with APDU, the conference had a strong focus on decennial census issues ranging from what has been learned from the census pre-tests, a report by and reaction to the Census Oversight Committee, and updates and commentary on the census time-line. Many sessions centered on the plans for Census 2000. These included a mild sparring match among the partisan staffs of the Census Advisory Board and the House Committee on Government Operations, a discussion of the experiences with the local review of the master address files by Patty Becker and Joe Salvo, and the keynote address of Bureau Director Kenneth Prewitt. Other census related issues were dissemination plans for census products as well as reports on the American Community Survey.

The conference this year explored several topics related to the broad theme of the future of public data. Continued public support for the collection and dissemination of data requires that emerging issues in respondent disclosure, scientific secrecy, and statistical sampling be addressed so that the underlying data collection, dissemination, and preservation programs continue to receive the support and funding necessary for their operations. Some of these concerns require education of the public and the data producers. An example of this

would be problems relating to statistical literacy. Also related to education and public relations are issues associated with the costs and time involved in data collection and in scientific methods of data collection, including sampling. The subject of confidentiality of data is emerging with the increasingly common practice of matching administrative records to survey records, increased geographic detail on records, the breadth of coverage on longitudinal files, and improvements in technology. All the above increase the likelihood of inadvertent disclosure of confidential information, which has the potential to reduce response rates, to increase the cost of surveys, and to reduce the type and amount of data to be publicly disseminated.

In addition, there were sessions that focused on codebook standards, other federal data products, and recent advances in software. There were several sessions of interest to state and local users, ranging from sessions on socioeconomic status (SES) indicators of local areas, the effect of the new coding for race, more discussion of the effects of welfare reform, and changes in the MSA definitions. It was serendipitous that the week before the conference, the recommendations from the Metropolitan Area Statistics Review Committee were released. This definitely falls under the heading of "it is smarter to be lucky than it is lucky to be smart."

In past years, the APDU conference has been covered in the issue of the Association newsletter that immediately followed the fall meetings. This approach, while helpful, was never quite satisfactory because the coverage often had to be pared down in order to fit into the newsletter format. This was particularly apparent last year when the first draft would have required a 32-page document. With the change in publication policy—the newsletter going to a quarterly schedule and a new more substantive publication appearing—it was felt that the conference could be covered more adequately in the latter, this issue of APDU's *Of Significance...*

A conference of this size cannot be put on without the help of many people. The program

Stephen Dienstfrey and Lisa Neidert are Co-Chairs, APDU Conference Committee. Stephen Dienstfrey is an APDU Past President.

planners are mainly the people who try to coordinate the action of others. We would be remiss if we did not take this opportunity to thank them all. First, a big “Thank you” goes to all of those who organized sessions, which included arranging for the speakers and keeping in touch with them. Second, an even bigger “Thank you” goes to the speakers. Without them, there would be

nothing. They were excellent. And a special “Thank you” goes to the staff of The Harris-Smith Institutes, the new administrative home of APDU. HSI came on board after the conference train was almost at full steam. They took over many of the administrative tasks as if they had been there from the beginning.

As good as this conference was, next year is bound to be better. We hope to see you there.

Opening Remarks

Responsible Data Use: Protecting the Future of Public Data

WENDY TREADWELL

During its existence, APDU has seen massive technological changes in the collection, dissemination and preservation of public data, a move from producer driven to user driven products, the shift from mass production to mass customization, as well as the continuing process of balancing statistical, social and political needs. APDU itself has recently gone through the process of examining its structure and purpose related to the changing environment for public data users. It seems appropriate that as we move towards a new census and a new millennium, APDU will also be celebrating its 25th anniversary. What we have discovered is that the basic issues haven't changed:

- Collection of high quality, appropriate data
- Federal data policy
- Confidentiality and privacy
- Preservation
- Product development and evaluation
- Appropriate usage of federal data
- Statistical literacy

Wendy Treadwell, Board President, APDU

How we address these issues has changed. Print communication is rapidly changing to electronic, and electronic communication and data dissemination are moving to more accessible formats. The changing needs for data information and support, our ability to link our members with federal data producers and policy makers, the changing policy-making environment on the Hill, and the explosion in the number of end-users makes focusing our concerns and our resources a challenging task. APDU's membership is diverse, not only in terms of a member organization's role in the data collection, production, preservation and dissemination process, but also in terms of our individual and organizational interests and priorities. With an ever-expanding set of options, APDU and its individual members need to focus their energies in order to accomplish anything.

We invite our members to look at their individual areas of interest and examine ways in which APDU can help them pursue these interests as well as how they can contribute to APDU and support the activities of their fellow members.

Keynote Speech

KENNETH PREWITT
DIRECTOR
U.S. BUREAU OF THE CENSUS

(Summarized by Wendy Treadwell, President, APDU Board)

Given his demanding schedule, APDU was particularly pleased to have Dr. Kenneth Prewitt, Director of the U.S. Census Bureau, as the Keynote Speaker. As APDU completes its first quarter century, it is interesting to note that Dr. Prewitt's career reflects the range of professions which make up the APDU membership. Currently in government service, his background includes a distinguished career in higher education and public philanthropy as well as extensive involvement in numerous professional organizations.

Dr. Prewitt provided an overview and update of the 2000 Census process. While acknowledging that we will have a two-number census due to court decisions, he noted that the goal of the Bureau, as always, is to provide a progressively more accurate census. A number of steps have been taken to assure getting the most accurate count possible. The Local Update of Census Addresses (LUCA) program has been instrumental in identifying housing units prior to the census. Advertising and local organizing to get out the count is also underway. The big pieces are in place, Prewitt advised: census promotion, census execution, and census coverage improvement and quality control.

However, there are a number of unknowns that will affect the accuracy of the count. Given the solid economy, the availability of an adequate labor pool may cause problems in many areas. With unemployment hovering between two percent and three percent in some parts of the country and the

added need for staff to perform a traditional census, it is possible that there will simply not be enough staff hours available to complete the work within the given time frame. Other major factors that could affect the ability of the Bureau to complete its data collection work are the initial level of non-response and the 2000 budget, although Prewitt has recently received positive budget signals from Congress. In addition, there will be the problems caused by the changing demographic and social factors in the United States at the end of the 20th Century. Higher mobility, larger immigrant groups and the variety of languages spoken, as well as increased political cynicism, will all have an impact on the initial non-response rate and the ability of the Bureau to follow up on these non-respondents. The differential undercount by demographic characteristics and by geographic area is a focus of attention.

Aside from obtaining the most accurate count possible, this census will address a much broader issue: "Can we 'count' our way out of the differential undercount?" Given an adequate budget, the Bureau believes that its operations are sufficiently robust to provide an answer to this question. They believe they will be able to tell the country just how far it can get with a traditional count in the current environment. They also believe they will be able to show the effectiveness of the Post-Enumeration Survey/ Accuracy and Coverage Evaluation (PES/ACE) in covering and identifying the accuracy of the count. In short, we should have the information by 2002 for a data-driven decision on what type of census makes sense.

The Census Bureau's Integrated Information Solutions

MARK E. WALLACE

From its earliest days, the Association has worked to enhance public data dissemination and user access. In the seventies, members focused on the access to federal data stored on magnetic tape files, successfully demonstrating the utility of the files for filling data needs in the academic, local and private sectors. APDU members are currently helping to shape the future of public data delivery in an electronic, networked environment. At APDU99, the Census Bureau's Mark Wallace outlined his agency's plans for a data delivery system, the Integrated Information Solutions, and encouraged members' involvement in its development.

(Summarized by Wendy Treadwell, President, APDU Board)

Mark E. Wallace, Chief of the Economic Planning Staff at the U.S. Census Bureau, has spent the last year and a half leading an effort to launch a new program at the Census Bureau called Integrated Information Solutions (IIS). This involves directing activities to proceed from existing dissemination systems, including the American FactFinder (AFF) and the Federal Electronic Research and Review Extraction Tool (FERRET), to a fully integrated information access and dissemination system. Such a system would allow internal and external users to access a data web of information drawn from all areas of the Census Bureau and other federal agencies.

As the largest data collection agency, the Bureau of the Census collects, processes and disseminates data for federal, state and local agencies. To do this, the IIS builds on the structure of the American FactFinder which will provide future access to the 2000 Census and the Economic Census and the Data FERRET which would act as the distributed data component of IIS. The IIS itself

Mark E. Wallace, IIS Project Director, Bureau of the Census

would provide access to integrated data sets and ongoing customer support. The process would include a product creation system and incorporate policies and standards, new business practices as well as metadata and data support systems. The IIS product process would be customer driven. Product concepts would be generated by customer needs and developed based on past experience, resource and design constraints and market analysis. Customers would receive both product access and customer support from the system.

The major implementation issues center on three main areas: definitions of concepts, geography, and disclosure avoidance. Definitions of concepts vary by agency, time, and application. Finding a means of tracking and relaying this information to end-users must be addressed. Geography is another area affected by time and agency variations. It is particularly important to accurately convey geographical boundaries given the increased use of mapping to display socioeconomic data and the needs of the geographic information system (GIS) software involved in this process. Increased flexibility and increased access will require increased vigilance for disclosure avoidance. Automating this routine and providing realistic expectations and accurate data will continue to be a challenge.

What Has Been Going on in the Public Data Sector ?

A long-standing and much anticipated session at the conference is the annual update on federal data policy provided by Katherine K. Wallman and Edward J. Spar. Their topics alert data users to important developments and areas of concern, and their insights clarify the issues. APDU was fortunate again this year to have Katherine Wallman outline the results of initiatives within the federal statistical system. As Chief Statistician in the Office of Management and Budget, she works across agencies on the reviews and recommendations which affect the collection and dissemination of public data. As Executive Director of the Council of Professional Associations on Federal Statistics, Ed Spar provides national leadership in the improvement of public data. An APDU founder and former Board President, Ed continues his support by sharing invaluable information, assistance, advice and opportunities with data consumers.

Developments in the Public Data Infrastructure

KATHERINE K. WALLMAN

When we gather at APDU each year, there is always considerable attention devoted to the products of the statistical system that are of keenest interest to the data user community. Of course that is not surprising, for in the end, relevant, accurate and timely statistics are what the public demands. But the data we provide to the public cannot keep pace with changes in our society and economy unless we regularly review the concepts and constructs that underlie the collection and presentation of statistical information. This somewhat less visible yet absolutely critical work we undertake in collaboration with the statistical agencies has been a focus of our efforts during the past half dozen years. I would like to highlight our work on statistical standards in my remarks to you today.

The first of our efforts to update standards has been in the area of industry classification. Working with our counterparts in Canada and Mexico, we have developed and published the **North American Industry Classification System** (NAICS). With NAICS' introduction, we now have a system that more accurately reflects the current

Katherine K. Wallman, Chief Statistician, Office of Management and Budget

economy and a system that will yield comparable data for the NAFTA countries. NAICS' implementation began with the 1997 data year in Canada and the United States, and the 1998 data year in Mexico. U.S. agencies will implement NAICS from 1999 to 2004. The first major data programs to use the new system were the 1997 Census of Agriculture released in February 1999; the 1997 Economic Census, with advance statistics released in March 1999; and the 1997 Foreign Direct Investment Benchmark Survey that will be published shortly. For most current economic surveys conducted by the Census Bureau, NAICS data will be introduced over several years: for manufacturing data, with the 1998 reference year; for services data, 1999; and for economic indicator data, such as Monthly Retail Sales, 2001. The Bureau of Labor Statistics will begin its conversion to a NAICS basis with its Employment and Wages Report for the 2001 reference year. Data series may not always be revised for years before the respective program's implementation of NAICS United States; instead, bridges will be developed to permit comparisons of pre- and post-NAICS data. Regularly scheduled meetings among the three countries will ensure that there is a smooth transition to NAICS. In addition, NAICS will be

reviewed and updated continuously to ensure that new activities are promptly recognized and to extend NAICS to the 5-digit industry level in all sectors.

In a February 1999 *Federal Register* notice, the Office of Management and Budget (OMB) proposed the development of a comprehensive classification system for products produced by NAICS industries: the **North American Product Classification System**. The long-term objective of the joint effort between Mexico, Canada, and the United States is to develop a market-oriented/demand-based classification system for products that is not industry-of-origin based but can be linked to the NAICS industry structure and to international product classification systems, such as the Central Product Classification System of the United Nations. Given the dynamic and intangible nature of many service products, this initiative will be implemented in two phases. Phase 1 will develop preliminary product classifications for a subset of NAICS service industries and will be completed during 2000. This phase will be confined to identifying and classifying the products produced by the industries in four selected NAICS service sectors: Information (sector 51); Finance and Insurance (sector 52), except Insurance (subsector 524); Professional, Scientific, and Technical Services (sector 54); and Administrative and Support, Waste Management and Redemption Services (sector 56). These results will be incorporated in the 2002 Economic Census and related programs. Phase 2 will be launched after the 2002 Economic Census and will develop a complete and fully integrated product classification system that extends to all NAICS industries, including the full range of service sectors. The results of Phase 2 are expected to be incorporated in the 2007 Economic Census and related programs.

On September 30 of this year (1999), OMB announced the new **Standard Occupational Classification (SOC)** that replaced the 1980 system. The revision is intended to produce a pragmatic occupational classification system that will support economic analysis, strengthen the ties between education and work force data, unify federal agency occupational classification usage, and foster international comparability. All federal agencies that collect occupational data will use this new system, while governments at other levels are strongly encouraged to use the new system as well. The new SOC will be used for Census 2000 and will be the framework for the information being gathered through the Department of Labor's Occupational

Information Network, which will replace the *Dictionary of Occupational Titles*. OMB expects to publish the new *1998 Standard Occupational Classification Manual* next year.

OMB has issued geographic definitions of metropolitan areas for use in collecting, tabulating, and publishing federal statistics since 1950. Proposals for revising the standards for use in defining areas after Census 2000 were published for comment on October 20. Comments are due by December 20, and a decision on the revised standards is scheduled for March 2000. (This topic is covered in depth in another session reported in this issue.)

Members of APDU are aware of the revisions that have been made to the standards for collecting and presenting data on race and ethnicity. The major change from the 1977 standard is the opportunity for respondents to federal data collections to select more than one racial category if they wish to do so. OMB expects to issue provisional guidance for the tabulation of data on race and ethnicity early next year. (This topic is covered in depth in another session reported in this issue.)

In addition to our work on statistical standards, I want to mention two other initiatives that should have considerable benefits for the data user community: legislation to even out statutory confidentiality protections for statistical data and permit limited sharing of such data for statistical purposes, and "one-stop shopping" for products of our decentralized federal statistical system.

The House recently passed the **Statistical Efficiency Act of 1999** (H.R. 2885). This legislation will establish consistent and uniform application of the confidentiality privilege and permit limited sharing of data among eight designated federal statistical agencies: the Bureau of Economic Analysis, Bureau of the Census, Bureau of Labor Statistics, National Agricultural Statistics Service, National Center for Education Statistics, National Center for Health Statistics, the Energy Consumption Division of the Energy Information Administration, and the Science Resources Studies Division of the National Science Foundation. During the coming year, attention will be focused on attaining Senate passage of this bill, and on Congressional action on a companion legislative proposal that would make changes in the "Statistical Use" section of the Internal Revenue Code. These changes would reduce the amount of sensitive tax information that will change hands to support statistical programs while substantially

increasing the effectiveness of that support.

Last but not least — certainly for data users — a coordinated interagency effort to provide broader and simpler electronic access to the full range of federal statistics came to fruition in 1997 with the introduction of **FedStats** <www.fedstats.gov>. FedStats is designed to be the initial point of contact to data produced by more than 70 Federal agencies. To date, statistics from 40 of these agencies have been indexed. With more than 300 entries in its A to Z index, plus a search engine for user-defined queries, FedStats has had over 2.6 million visitors in its first two years of operation. Efforts are currently under way to add the remaining 30 agencies. In addition, the search and data access

tools are being enhanced, including the point and click “MapStats” that will zero in on statistics about a state or county. Finally, FedStats serves as a gateway for the teaching of statistical literacy through its **Kid’s Page**. Some 40 statistical agencies provide special web pages for students in elementary through high school. The web pages include fun facts, project ideas, and other information both from the traditional statistical agencies such as Census, BLS, and EIA, as well as from many organizations less often thought of in our statistical agency framework, such as the CIA.

I look forward to our continuing work with APDU as we define and address these and other challenges confronting the federal statistical system.

Developments in Public Data Policy

EDWARD J. SPAR

As of October 28 (1999), the federal statistical agencies (along with most of the rest of the government) were still operating under stopgap measures, and Congress was still attempting to find a way to pay for FY 2000 spending. Agencies including the Bureau of Labor Statistics and the National Centers for Health and Education Statistics among others, may find themselves far short of their requested expenditures. Perhaps the only expenditure immune to the variations being considered on the Hill is the decennial census. However, if the final budget compromise calls for an across the board percentage decrease, even the funds for the 2000 Census may be reduced. Given that the president has threatened to veto any significant reductions in spending for education and health related areas, the budget process may go on for quite some time. As the process drags on there is more likelihood of our seeing another omnibus spending bill.

On October 20, the Office of Management and Budget (OMB) released the recommendations from the Metropolitan Area Standards Review Committee (MASRC) for changes to OMB’s standards for defining metropolitan and

nonmetropolitan areas. This is the first major revision of these concepts since 1970, when OMB developed new areas such as Primary Metropolitan Statistical Areas (PMSA’s) and Metropolitan Statistical Areas (MSA’s). MASRC has recommended a *Core-Based Statistical Areas* (CBSA’s) classification to replace the current Metropolitan Area (MA) classification. The cores (densely settled concentrations of population) for this classification would be Census Bureau defined urbanized areas and smaller densely settled “settlement clusters.” The settlement clusters are new areas to be identified for the 2000 Census. CBSA’s would be defined around these cores. This CBSA classification has three types of areas based on the *total population of all cores* in the CBSA: 1) Megapolitan Areas defined around cores of at least 1,000,000 population; 2) Macropolitan Areas defined around cores of 50,000 to 999,999 population; and 3) Micropolitan Areas defined around cores of 10,000 to 49,999 population. Those counties containing the cores, would become the central counties of the CBSA’s. Territory outside of Megapolitan, Macropolitan and Micropolitan Areas would be termed “Outside CBSAs.” This new approach is a radical change from the previous concept of “metropolitan.” Users should take a very close look at the potential changes and be sure

Edward J. Spar, Executive Director, Council of Professional Associations on Federal Statistics (COPAFS)

to comment by December 20, 1999. The *Federal Register* notice can be obtained via the COPAFS home page at: <<http://members.aol.com/copafs>>.

Just take a look at "What's New." the House has passed H.R.2885, the *Statistical Efficiency Act*. This is essentially the second part (Title II) of an earlier bill to establish a Federal Commission on Statistical Policy to study the reorganization of the federal statistical system and at the same time provide safeguards for confidentiality in data sharing. The current bill, which no longer addresses the issue of federal statistical agency consolidation, would establish Statistical Data Centers for the purpose of sharing of information under strict confidentiality. These would include the Bureau of Economic Analysis, Bureau of the Census, Bureau of Labor Statistics, National Agricultural Statistical Service, National Centers for Education and Health Statistics, the Energy End Use and Integrated Statistical Division of the Energy Information Administration, and the Division of Science Resources Studies in the National Science Foundation. Under the strict confidentiality rules, the Centers could enter into

joint statistical projects to improve the quality and reduce the cost of statistical programs, and identify opportunities to eliminate duplication and other costs by sharing information for *exclusively* statistical purposes. For analysts in the areas of health, education and labor, for example, this bill could yield substantial benefits through the development of more detailed data series.

The Census Bureau is well along the way in determining what the products for the 2000 Census will look like. Data will be available in the form of paper, CD-ROM and on the Internet via the American FactFinder (AFF) system. There are still some questions that users should be asking. For example, what will the AFF system cost? How much data will be available on the AFF? Has the Census Bureau thought through the need for ongoing maintenance? Will there be user input in the design of the AFF? Regarding CD-ROM, will there be new software more advanced than the GO software? Finally, there is some concern that the Bureau is not publishing enough information on paper. Further discussion is needed about the possibility of producing more socioeconomic and race-ethnicity data for local areas.

Reflecting on the Past and Looking to the Future

JUDITH ROWE

Judith Rowe recently retired as the Princeton University Data Archivist. Data users owe Judith a debt of gratitude for her long national service in the improvement of data access and policy. For many years she also hosted the APDU headquarters in her department, and she served a term as Board President and 1990 Census Products Committee Co-Chair. In recognition of the significant role Judith has played in public data service, APDU presented her with a lifetime membership. We again called upon her, asking her to address APDU99. At Monday's luncheon, she educated and entertained us with these reflections on an illustrious career—which is now entering a new stage of activity and accomplishment.

Like most people I'm better at reflecting on the past than on looking to the future. For example, in the 1980's when we were planning for the 1990 data products, I never predicted CD's, the Internet, or even wide use of PC's...but few people did. The one thing I have predicted successfully is that paper is not going away. If people doing reference don't get their data products in paper form, they print the data out and create their own paper products. Witness the number of stores devoted almost entirely to printer products and paper. People doing research are less dependent on paper. In any case I'm not sure where the technology of data collection or data delivery is going, but I have a pretty good idea where we've been and certainly where I've been...and what the policy issues are likely to be in the future... But let's save that until later.

I guess it all began for me when I first started dating a man whose family was in the Jacquard weaving business. Shortly after that I started Yale as a graduate student in Sociology where I was introduced to the numeric keypunch, the sorter and the punched card. We heard rumor that there was a machine called a collator in one of the administrative offices, but we never saw it. We matched cards by holding them up to the light and looking at the ID numbers, a very time-consuming process.

More than ten years went by during which I was a marketing researcher and then a full-time mother and a part-time researcher. By the end of that period my husband, then the IBM salesman at Princeton, announced that Fred Stephan, who had been a Census colleague of Deming and Hansen, among others, was having difficulty filling a position. He was looking for a social scientist with knowledge of data processing to work at the Office

for Survey Research and Statistical Studies. Stephan had also served on the Social Science Research Council Committee to analyze the 1948 polls that predicted Dewey over Truman. Clearly this was the job for me. When I started working at Princeton, we had alpha-numeric keypunches, more advanced sorters and reproducers. The 101 had a more impressive board. This was a tabulating machine, widely used by people in marketing and social research. We never had one at Princeton but they had one at Columbia among other places. In addition to our unit record equipment, we also had Friden and Monroe calculating machines so that we could compute chi-squares and do percentages. Although I had no contact with it, Carl Helm, to whom I reported initially, reminded me that he and John Tukey among others were writing programs in machine and assembly language for a 650 at that time.

Down the hall from my office there was a project under the direction of an economist named Bill Bowen, later to become president of Princeton and now head of the Mellon Foundation. It was using boxes and boxes of punched cards containing data from the 1960 Census to do the research which resulted in the classic Bowen and Finegan volume, *The Economics of Labor Force Participation*, which was published 30 years ago.

There was talk of 1970 Census products and Fred was eager to have us participate. I was less eager but Princeton was already involved with data service as an early member of ICPSR, the Inter-university Consortium for Political and Social Research. I kept our entire data holdings listed on one sheet of 8 1/2 x 11 paper. We received our data on 200bpi tapes and then later on 556bpi tapes and

processed them on the IBM7090 and then on the 7094 using a homegrown program called Survey and then later Data-Text, which had been written at Harvard in assembly language. Similar programs, usually written in FORTRAN were being developed everywhere but they were not designed to process either aggregate data or the hierarchical data that has characterized census microdata products.

However, Jack Beresford had organized DUALabs with financial support from the Ford Foundation and with staff that included Connie Citro and others. DUALabs members included Columbia, Florida, UCLA, Wisconsin, Yale, a partnership of Princeton and Rutgers, as well as Market Statistics then represented by Ed Spar, the Detroit and NYC Planning Departments and the Rand Corporation. By squeezing out the blanks and zeroes and filling up the tapes, the DUALabs data products were compressed to occupy one-tenth the number of tapes as those being distributed by Census. Under the direction of Jack Barrett, software was written to retrieve and display the data. By the time we acquired these tapes, Princeton had an IBM 360/91 for batch processing. A 370/168 soon followed. Although data were increasingly being transported on round reels with standard labels written at 800bpi and then at 1600bpi, we were still reading data into the computer by using the 1401 card reader to input punched cards.

Princeton's first external data customer was the local League of Women Voters. After several years of providing both internal and external data service, we were approached by Applied Urbanetics and Claritas, both of whom had contracts with federal agencies to use 1970 Fourth Count census data for the creation of special extracts. In my naiveté I suggested that this should entitle them to get these data from the Census Bureau without charge. It was then that I first understood that the United States had a federal government and that Princeton could provide the tapes more cheaply than Census. But Princeton's administrators were sure it was illegal for me to resell census data. Fortunately Princeton's legal officer had been at the Ford Foundation when the complicated arrangement involving DUALabs, the Center for Research Libraries and the assorted participants was consummated, and he understood it all perfectly so the sale went forward.

By 1974 the START community, the original DUALabs members, had organized into APDU, the Association of Public Data Users. Within the same period IASSIST, the International Association for

Social Science Information Service and Technology, and GODORT, the Government Documents Round Table of the American Library Association, ALA, were born and the Federal Statistics Users Group (FSUC) which lacked the focus on machine-readable data products was waning.

At the behest of Sue Dodd of UNC and others, IASSIST seriously addressed the issue of cataloging machine-readable data files or MRDF. This interest coincided with the appointment of an ALA committee to recommend such rules for AACR II, Anglo-American Cataloguing Rules II, which was soon to be published. The first chairman of that committee was John Byrum, now at the Library of Congress (LC), but then head of cataloging at Princeton. John and I had written what was probably the only seminal paper I ever wrote (it appeared in a publication then called *Library Research and Technical Services* or LRTS), and John asked me to join ALA so that he could appoint me to the committee. LC's liaison to that committee was Henriette Avram, the mother of the MARC record. The committee's first meeting was mid-winter at the Palmer House. The weather was bitter and although the committee met almost constantly, I did have some free time. I wandered around and discovered two meetings, which changed my professional life. GODORT was just being organized and I convinced them that they needed a task force on MRDF. I became coordinator of that task force. Then I found the census subcommittee of the now defunct ALA public documents committee. It seemed to be doing nothing. The rationale for this was that since they were not an official advisory committee they couldn't offer Census any advice and Census could not seek their advice. Well, I won't go into that long story but I soon became chairman of a very hard working census subcommittee. Librarians were being recognized as important information intermediaries and on November 11th of 1974, Director Vincent Barabba welcomed to the Census Bureau the members of the Pilot Seminar for Librarians. The subcommittee reviewed numerous publications, and with the blessing of David Kaplan made substantial contributions to the design of the 1980 printed products.

During the 1970's, meetings were held everywhere to explain the 1970 Census products and to plan for the 1980 products. Other agencies as well as Census were exploring new formats for data delivery. Against the advice of a special advisory committee that urged Census to concentrate on doing what they could do best,

Census developed the short-lived software called CENSPAC. On the plus side, however, they developed the standard geographic record and a whole new structure for organizing the aggregate data, both of which were also used with the 1990 data.

The 1980 census data products were distributed by Census in a much more usable form to a much wider audience than had been the case a decade before. In academia the NSF funded an ICPSR proposal submitted by Jerry Clubb and Eric Austin which increased the number of users still further. I served on the committee that ICPSR appointed to set policy for census data acquisition and distribution.

During this period, statistics, particularly in machine-readable form, was becoming a sexy topic. Everyone wanted to get into the act. NTIS, the National Technical Information Service, became the data distributor for NCHS, the National Center for Health Statistics, and the Federal Reserve among others. OMB, the federal Office of Management and Budget, and the congressional Office of Technology and Assessment, OTA, and the Joint Committee on Printing, JCP, were the most notable groups that addressed data policy issues. National Archives increased their data holdings, particularly of census data products and contributed substantially to our understanding of archival procedures such as appraisal.

COPAFS, the Council of Professional Associations on Federal Statistics, was born. Kathy Wallman was the first Executive Director. I was the first chairman and Ken Prewitt coordinated the evaluation committee that recommended that COPAFS become a permanent organization. APDU has been an active participant in COPAFS since its initial organization. One of COPAFS' early efforts was the organization of congressional hearings for the Garcia Committee, which then had Census oversight responsibility and wanted to hear who used the census data. Kathy and Helen Peck worked very hard. It was Ed Spar's first experience giving testimony. He was very good but very nervous.

The Joint Committee on Printing appointed a subcommittee, which was organized by Bernadine Hoduski. I represented the American Library Association on this subcommittee, whose mission was to recommend ways in which depository libraries could receive computer products as well as paper products from the Government Printing Office. Bernadine was already sold on CD's and in these pre-Internet days one of the pilot projects, which

grew out of the Committee's recommendations, was one which provided free access to an online service containing federal data.

In the meantime more and more data librarians were appearing in traditional libraries and more and more libraries were acquiring microcomputers. I was still at Princeton's CIT and I remember our first microcomputer. It was in a locked room with a burglar alarm. I'm not sure to what use other people put it but my primary use was CONFER which several of us were using to plan an IASSIST meeting. It never occurred to me to use it for census data although we soon had PC's which rivaled the 370's of the previous decade with price tags in the thousands instead of in the millions.

By the time the 1990 Census was completed, PC's and CD's were everywhere and mainframes were almost nowhere. The paper maps of the 70's had been replaced by microfiche maps in the 80's and now by GIS systems in the 90's. The Depository Library Council of the Government Printing Office was increasingly involved in the distribution of documents in CD form and had begun to address issues of archival storage. Princeton still has its mainframe...although not for long...so we continued to get our data in magnetic form although the library was getting much of the same data on CD's. Gopher, bulletin boards, listservs and e-mail were the new technologies; the former two now things of the past.

By the mid-1990's our recently rechristened Data and Statistical Services had been transferred from CIT to Firestone Library's Social Science Reference Center. Although the professional librarians provided first-line reference, the DSS staff was responsible for data retrieval and data analysis as well as for the location of non-standard data items. An increasing amount of public data was now available on the web for both reference and research. The statistical consultants soon became web developers as well and an increasing amount of data become available online including the 1970 Fourth Count which Doug Mills had removed from DUALabs compressed form several years before. Is it still possible to save the rest of these older data? Who knows. Some of it is clearly gone. Are archival procedures for the future well in place? What of National Technical Information System (NTIS) data holdings? The subcommittees of the Federal Committee on Statistical Methodology and other related groups are quite sensibly focusing on privacy, access, data sharing and data definitions, federal-state cooperation and other information policy issues. Those involved with the information

infrastructure are also looking at copyright, information ownership and Internet charges. Clearly electronic transfer of data saves distributors paper, printing and postage costs but there are other hidden costs to both user and distributor that must be carefully evaluated. According to the most recent CPS, Current Population Survey, report on computer usage, although computer use is up, only about one in five people use the web. Most people are still dependent on printed documents and may

be so for the foreseeable future. Do we have procedures in place to acquire, archive and make available data from multi-national agencies and multi-national corporations? Are we dealing efficiently with issues of privacy and confidentiality?

Policies and procedures. Those are the issues that we must address in the immediate future. The rest will take care of itself.

Making Things Add Up For the End User: Issues in Statistical Literacy

KATHERINE K. WALLMAN
MARIANNE W. ZAWITZ
COLLEEN BLESSING
WENDY TREADWELL

With numbers joining dollars as the currency of policy debates in society, media and government, public understanding of quantitative information has become as important as public access to the numeric data. The first issue of APDU's *Of Significance...* featured statistical literacy, the topic of this conference session and an important concern of the APDU leadership. The Association is working to focus attention and expertise on public education. Panelist and APDU Board President Wendy Treadwell was joined by federal agency representatives to share their assessment of the problem, the results of their efforts towards statistical literacy, and their recommendations.

(Summarized by Jocelyn Tipton, Yale University Library)

This session focused on the importance of statistical literacy today, what two federal agencies are doing to address this topic, and what APDU members can do to help promote it. Jocelyn Tipton, session moderator, reminded the audience that the first issue of the APDU journal *Of Significance...* was devoted entirely to the topic of statistical literacy and that four of the contributors would be participating in the session. APDU, as an advocate of broad public access to statistical data, shares both a concern for the development of a statistically literate populace and a responsibility for helping to ensure that development. Given the promise of increased access to an ever-growing amount of federal data and increased public involvement in the discussion of federal data collection processes, it is time we took this responsibility seriously. As data producers, intermediaries and users it is our role to make sure that statistics are presented and interpreted appropriately within our own constituencies.

Katherine Wallman, OMB, began the session by laying the groundwork for the other panelists by providing an overview of statistical literacy and stressing how important a topic it is for APDU to deal with. She stressed that citizens encounter statistics at every turn in their daily lives yet are

Katherine K. Wallman, Chief Statistician, Office of Management and Budget. Marianne W. Zawitz, Bureau of Justice Statistics. Colleen Blessing, Energy Information Administration. Wendy Treadwell, President, APDU Board, and Coordinator, Machine Readable Data Center, University of Minnesota

often unequipped to evaluate the information presented to them. While more data are available and the ease with which people can access these data is improving, the gap between computer literacy and statistical literacy remains significant. As statistical methods become an object for political debate, citizens are left in a quandary concerning the scientific versus the political aspects of arguments presented to them. She finished her talk by challenging the audience to heed the needs of our consumers in the larger society we hope to serve (government, industry, education institutions, or media) by working to improve statistical literacy among these groups.

Marianne Zawitz, Bureau of Justice Statistics, discussed how data providers have an obligation to present data clearly and accurately and that data presentation is a partnership between data providers and data users with a goal of statistical literacy. Quality presentations of data take advantage of how we already process information, reduce the number of processes required to understand the data and tear down some fundamental obstacles to understanding. She also explained that processing numbers is different than processing words. Numbers have properties that provide additional display methods, numbers have an architecture and even without values, we know what is more or less. Statistics are presented as text, graphics and tables. Quoting Edward Tufte, she said that statistical graphics should have "simplicity of design and complexity of data." She suggested that graphics be used when there are more than 10 data points and when there is a need to show a big picture rather than fine data. Tables should be used when you need exact numeric values and for localized

comparisons. Statistical graphics show the big picture, are paragraphs of data and are best when constructed to convey one finding or concept. The purpose is to show the data, therefore, you should maximize data density and data ink.

Zawitz then discussed William S. Cleveland's hierarchy of graphical perception that deals with position along a common scale, position along identical nonaligned scales, length, angle/slope, area, volume, and color. Several examples were provided to illustrate each of these points. She recommends that pie charts, stacked bar graphs and stacked line graphs be avoided as graphic styles. To show data, avoid non-data ink and chart junk. By reducing non-data ink, such as grids, ticks, frames, and anything that does not convey data, you highlight the data and reduce the number of tasks the user must go through. Chart junk also obscures the data. To prevent this, she suggests using solid lines and avoiding markers on lines. Fill patterns are another important issue. Hatched fill patterns cause a moiré effect which is hard to look at but can be avoided by keeping the space between the bars to a minimum. She then offered tips for making verbal tasks easy to understand including avoiding abbreviations and acronyms, writing labels left to right, using proper grammar and limiting the use of legends only for maps. Lastly, to avoid optical illusions and graphical puzzles, use solids for line styles and fills, avoid data point markers on lines, use the appropriate aspect ratio, start the scale at zero, use only one unit of measurement per graph and avoid using three dimensional designs for two dimensional data. In conclusion, two basic elements of successful data presentations are an appreciation for the user and an interest in showing the data. Public statistical agencies are training their staffs and issuing style guides on good practices as well as designing new presentations of complex statistical concepts.

Colleen Blessing's talk focused on using cognitive interviewing to improve data products. At the Energy Information Administration cognitive techniques are being used to improve web site usability and to change the survey form design for the Manufacturing Energy Consumption Survey (MECS) and the Natural Gas Supply and Disposition Survey. Changing the format of these forms helped to improve data quality, reduce respondent burden, and encourage higher response rates. The cognitive interviews collect information about customers' thinking processes, problem solving approaches, understanding of the product, understanding of the instructions and feelings about

the process. As a result of these interviews the MECS form was updated and improved. The new form incorporates instructions and definitions into the form, provides a simpler page design and uses color to guide respondents through the form. Although the new form is longer, customers seem to like the new layout, and it appears that the data quality in the test was improved and the response rates are encouraging. The Natural Gas Form was used for a training exercise. Four cognitive interviews were conducted with respondents. The EIA was looking to see if the form uses terms and concepts that the respondents can understand, the form asks for data in the same time frame in which records are kept, and the reporting units requested are the same as the units used in their record keeping. During the interviews the testers observed the respondents' reactions and navigation issues, i.e. are they going where they should, and asked the respondent questions about thoughts and feelings as well as about how the respondent would calculate answers. The findings showed that the instructions were too complicated and that the terms, units of measure and calculations were not always understood. The Administration has also done web site cognitive usability testing. This testing was conducted by having one-on-one sessions with customers on the web and having them go through a series of timed exercises; additional demographic data on the users were also collected. The findings from these tests showed that the site contained too much jargon, the search engine was frustrating, there were too many buttons and the format provided too many choices. The demographic data showed low correlations between the number of correct answers and the familiarity of the user with energy terms, the site itself, or experience with the Internet. The EIA plans to use the information gathered in these tests to hire a web design contractor to implement some of the suggestions, then they will do cognitive tests on alternative designs and then implement the successful new ideas. Blessing concluded that by watching customers use the agency's products, whether on a web page or a survey form, you can see where they are becoming frustrated or confused and that by using this information, you can ensure appropriate use of data and accuracy filling out the survey forms.

APDU Board President Wendy Treadwell wrapped up the session by describing ways that we can bring statistical literacy initiatives to the local level and stressing that APDU should continue to be involved in promoting this. For a number of years

APDU has been sneaking the subject of statistical literacy into its conferences. These discussions have been hidden within sessions covering the presentation of statistical information in the press, instructional or explanatory material in data presentations, and statistical methods. Training programs dealing with statistical literacy have focused on training of statisticians, training for journalists, training of staff who present statistics and training in K-12 programs. With the increased use of the Internet and the acknowledgment of the value of information, the discussion of the need for an information literate public has grown. Information literacy is being discussed as a requirement for an informed and active citizenry and also discussed are the means of developing this skill within the general population. Statistical literacy is broader than specific knowledge of basic statistical methodologies and the ability to recognize a 'good' number from a 'bad' one, and a questionable citation from one that provides sufficient information for the reader to verify the source. It is a conceptual understanding of probability, of the

meaning of statistical variance or standard errors and why they are important, and of the implication of using various sampling processes on the accuracy of results.

Treadwell continued saying that as an advocate of broad public access to statistical data, APDU shares both a concern for the development of a statistically literate populace and a responsibility for helping to ensure that development. The problem is no longer the availability of statistics, but the mixed quality of the statistics available and how they are used. We need to actively address how well the data we present are being interpreted and what we, as individuals, in institutions at a local level and as a national organization, can do to address the problem of statistical literacy within the general populace. This is an area of activity that APDU needs to explore more fully. She concluded the session by encouraging us as individuals concerned with the appropriate application of public data for decision making and planning, to consider how various approaches to developing statistical literacy apply within our local environments.

Race and Ethnic Statistics: What Can Users Expect ?

CLYDE TUCKER
JORGE DEL PINAL

Revisions to the Office of Management and Budget (OMB) Statistical Policy Directive No. 15, the federal requirements for collecting and classifying data on race and ethnicity mean changes in the way the data will be reported, and will ultimately impact the research and analysis and comparability of the data. In this session, representatives from the Bureau of Labor Statistics and the Bureau of the Census covered the major changes to the classification system and OMB interagency efforts to develop implementation guidelines, and reported on the application of the new standards in three dress rehearsal sites for Census 2000.

(Summarized by Deborah Gona, Vice President, APDU Board)

Clyde Tucker of the Bureau of Labor Statistics began the session by outlining the primary changes to the race and ethnicity standards presented in Statistical Policy Directive No. 15, the Race and Ethnic Standards for Federal Statistics and Administrative Reporting. These new standards are to be used in the 2000 Decennial Census, with other federal agencies directed to begin adopting them for use in household surveys, administrative forms and records, and other data collection, as soon as possible, but no later than January 1, 2003. The OMB approved the use of the new standards by the Bureau of the Census in its dress rehearsal for Census 2000.

Tucker reminded the audience that the changes to the standards include: 1) separation of the "Asian and Pacific Islander" race category into "Asian" and "Native Hawaiian or other Pacific Islander" categories; 2) allowance for the self-selection of one or more races; 3) prompting for Hispanic ethnicity separately and before prompting for race; and 4) use of the terminology, "Hispanic or Latino" and "Black or African-American." With these revisions, Tucker noted, efforts have been under way by OMB to provide some implementation guidelines and suggest possible different ways of looking at the data. He said that guidelines for implementation ultimately must meet the needs of at least two groups — those involved with carrying out legislative and constitutional mandates and those monitoring and

N. Clyde Tucker, Senior Statistician, Bureau of Labor Statistics. Jorge H. del Pinal, Assistant Division Chief, Special Population Statistics, Population Division, Bureau of the Census

analyzing social, economic and health trends.

To carry out the task of developing these implementation guidelines, Tucker reported that OMB created three interagency groups for implementation. The first, a policy group, was charged with considering the implications of the new standards on statistical policy and with reviewing the work of the other groups. The second, a procedural design group, was charged with developing appropriate question formats, instructions for interviewers and respondents, methods for reporting aggregate data, and training recommendations for interviewers. The third group, on tabulation, was charged with developing methods for bridging to the past and for future tabulation under the new standards.

With the classification revisions and implementation, Tucker noted that there will be some changes in data series, including the demographic characteristics of Hispanics, and the counts and demographic and substantive characteristics of races, especially for smaller racial categories, such as the American Indian and Alaskan Native population.

Regarding the OMB interagency group charged with cognitive testing for implementing the new classification standards, Tucker reported that the group began by developing guidelines for race and ethnicity data collection — both the self-identification questions across various survey modes and the aggregate forms. In testing the methods for self-identification questions, the group conducted phase I in the Washington DC metro area, and phase II in other geographic areas (New York City; Tulsa, Oklahoma; and Sacramento, California). The result of the work was a set of draft guidelines for self-

identification questions. Those draft guidelines are:

- Use an instruction that allows multiple responses to the race question, but does not require or appear to command to do so.
- Consider using an instruction to answer both ethnicity and race.
- Consider providing definitions.
- Tailor wording and format according to mode, remembering especially the limitations of the telephone.
- Use follow-up questions or write-in options when detailed information is sought.
- Follow OMB's Oct. 30, 1997 revised standards.

Tucker noted that while the work has gone well on the self-identification side, the testing of aggregate reporting has been more difficult. In testing the aggregate forms, the group convened an expert panel to review and discuss draft aggregate reporting forms, recruited subjects experienced with administrative data reporting, conducted cognitive laboratory interviews to test draft forms using dummy records, and conducted on-site establishment interviews to test draft forms using agency data. But there are methodological complications in testing aggregate forms. The response categories and data collection methods used for individual source data are different than the revised standards. Significant variation among similar agencies, organizations and establishments makes it hard to develop a prototype to test. And testing requires careful forms design which may be contrary to a "draft" stage of pretesting.

The work thus far, though, has resulted in a set of draft guidelines for aggregate reporting of race and ethnicity data. Those draft guidelines are:

- Provide definitions that distinguish concepts (single versus multiple race reports and ethnicity versus race).
- If possible, allow for the reporting of every combination of multiple race responses. If multiple responses are to be collapsed, specific instructions must be included.
- Include on the form a method to report individuals with missing race information and/or missing Hispanic or Latino origin information.
- Develop, test and incorporate instructions.
- Professionally design the reporting form.

Tucker reported that the next steps for this group are to test the self-identification questions in other geographic locations, to redesign and test aggregate forms with a variety of agencies and establishments, and to develop and test interviewer training and instructions. He said that most of the work was expected to be done by the end of 1999.

Regarding the work of the OMB interagency group charged with tabulations — both methods for bridging to the past and for future tabulations — Tucker reported the guiding principle: if the data collected under the old and new standards from the same people at the same time can be equated using a bridging method, then the only difference existing between the data collected under the old and new standards at separate points in time can be attributed to true change (as opposed to methodologically induced change) when bridging is used.

Tucker gave an overview of the framework for historical bridge tabulation methods; this framework takes into account how responses are assigned to a category — by a fixed rule (deterministic) or by a probability method (probabilistic) — and whether responses are assigned to more than one category — whole or fractional assignment. Bridging methods for "deterministic whole assignment" are "smallest group" (the respondent who selects two groups would be assigned to the group with the smallest size); "largest racial group named other than white;" "largest group named (including white);" and "plurality," based on the respondents' choice of main race in particular combinations from the National Health Interview Survey (NHIS). Bridging methods for "deterministic fractional assignment," are "equal fractions," or the assignment of multiple responses in equal fractions to each race; and "NHIS fractions," where there is fractional assignment of multiple responses to each race based on fractions from respondents in NHIS combinations (for example, 60 percent of Black respondents are assigned to "black" and 40 percent to "white").

Tucker relayed that in evaluating the bridging methods, measuring change over time is the most important criterion; the major purpose for any historical bridge will be to measure true change over time, as distinct from methodologically-induced change. The ideal bridging method would be one that matches how the respondent would have responded under the old standards had that been possible. The purpose of another criterion, minimizing disruptions to the single race distribution, is to consider how different the

resulting bridge distribution is from the single-race distribution for detailed race under the new standards.

In evaluating both bridging and future tabulations, Tucker said the hope is that whatever is decided in the way of procedures can be used in a wide range of programs and in varied contexts. It also is essential that tabulations meet the confidentiality standards of the statistical organization while producing reliable estimates and following recognized statistical practices, and that the tabulation procedures can be implemented with a minimum of operational difficulty, by individuals with relatively little statistical knowledge, and can be easily explained to the public. Finally, he noted a criterion in future tabulations — because of changes in the categories and the respondent instructions accompanying the question on race (allowing more than one category to be selected) — is that the underlying logic of the tabulation procedures must reflect, to the greatest extent possible, the full detail of race reporting.

Tucker reported on the results of evaluation analyses of three data sources: the 1993-1995 National Health Interview Survey, the May 1995 Current Population Survey (CPS) Supplement on Race and Ethnicity, and the Washington State Population Survey. The evaluations were in terms of match to the reference distribution, the misclassification of race, and preliminary outcomes assessment.

The results? On the criterion for measuring change over time, regarding the comparison to the reference distribution, the closest matches were for “largest group,” “plurality” and “fractional assignment” methods. Overall misclassification was greatest for “smallest group” and “largest group other than white” methods, except for certain categories (especially American Indian / Alaskan Native). With regard to outcomes, health outcomes, such as lack of insurance or poor or fair health, were largely unaffected by the bridging method, except, again, for the American Indian / Alaskan Native category on the question of no insurance. The evaluations did suggest possible effects on the American Indian / Alaskan Native category for unemployment and labor force participation, and that ultimately, the effects may depend upon the outcome being measured.

On the criterion for minimizing disruptions, Tucker reported, the “fractional methods,” “plurality” and “largest group” were closer to single race proportions under the new standards. Other whole assignment methods showed substantial

differences for the American Indian/Alaska Native category.

In terms of the range of applicability, the “largest group” and “equal fractions” methods were least sensitive to context; the “smallest group” and “largest other than white” methods varied by geography; and the “plurality” and “NHIS fractions” were more problematic because only a national distribution was available.

On the criterion of confidentiality, there was no greater problem since racial categories used before were being reproduced. Reliability, on the other hand, was good for the “largest group” method, but was a problem for other methods, particularly at smaller geographic levels.

With regard to statistical defensibility, all methods other than “equal fractions” were based on an observed distribution; the “smallest group” and “largest other than white” were based on historical record, while others were based on the minimization of error.

In terms of ease of use, “fractional assignment” requires duplication of records or multiple variables; “equal fractions” is independent of the distribution; and other methods were distributions dependent at different geographic levels.

As far as skill required, “deterministic whole assignment” required relatively little skill compared to “fractional” methods, and all is relative. Regarding understandability and communicability, “fractional assignment” may be difficult to accept.

Tucker said that further work on bridging will involve 1) comparing previous combined distribution to bridges back to that distribution; 2) conducting the same analyses with dress rehearsal data; and 3) providing suggested strategies depending upon what is going to be measured.

The next speaker, Jorge del Pinal, Population Division, Bureau of the Census, reported on what the Bureau found with multiple race reporting in three sites from the 2000 Census dress rehearsal in 1998: Sacramento, California, an urban center; South Carolina (Columbia City and Irmo Town and 11 counties), an urban/rural setting; and Menominee County, Wisconsin, an American Indian reservation. del Pinal cautioned that these were just the results from the dress rehearsal — with a limited number of sites, lack of representativeness, processing errors, and some results based on unedited data — and that the results could vary in 2000 as a result of changes in edits and allocations and the method for adjusting for coverage. Nevertheless, the Bureau looked for the effects of distribution by race (sequencing

Hispanic origin before race), multiple race reporting, common combinations, and non-response.

del Pinal reminded the audience that the potential combinations of the six new OMB race categories of “American Indian or Alaskan Native,” “Asian,” “Black or African American,” “Native Hawaiian or Other Pacific Islander,” “White,” and “Some other race,” a category which the OMB allowed the Census Bureau to add for Census 2000, can result in 63 groups. Respondent reporting of two or more races refers to combinations of two or more of the aforementioned race categories. For this purpose, a respondent reporting “White” and “Asian” would count as two or more races, but “Japanese” and “Chinese” would not.

In the three dress rehearsal sites under consideration—Sacramento, South Carolina, and Menominee—the percent of respondents reporting that they were two or more races was 5.4 percent; 0.8 percent; and 1.2 percent, respectively. In these sites, del Pinal noted, Hispanics were more likely to report two or more races, and about 9 of 10 of those respondents who reported two or more races, reported only two.

The question, del Pinal said, is how to portray this information: single race or “lower-bound” or all-inclusive or “upper-bound.” In a single race format, each category would be shown individually: “White” alone, “Black or African American” alone, “American Indian/Alaskan Native” alone, “Asian” alone, “Native Hawaiian and Other Pacific Islander” alone, some other race alone, or two or more races. In an all-inclusive race format, each classification would be shown alone or in combination with one or more other races. But ultimately what difference would it make? del Pinal showed the range between numbers for the upper- and lower- bound distributions. In all presentations for each site, the upper-bound distribution showed slightly higher percentages.

del Pinal then reported on how moving the Hispanic origin question before race affected the answer to the race question. Only Sacramento showed a higher percentage of respondents not answering the race question (in 1990, 2.7 percent, and in the 1998 dress rehearsal, 5.2 percent). Analyzing this by origin revealed that in all three sites, Hispanics appeared to have difficulty (for example, in Sacramento, 21.6 percent of those reporting Hispanic origin did not answer the race question; in South Carolina, 12.8 percent; and in Menominee, 8.8 percent).

del Pinal reported that of the three sites, Sacramento had a higher percentage (20.9 percent) reporting they were Hispanic, and of those reporting Hispanic origin, 80.3 percent reported they were Mexican/Mexican American/Chicano. Interestingly, even though respondents were not invited to do so, about 1 percent of the respondents reported two or more origins. In Sacramento, 58 percent of the respondents reporting two or more origins gave Hispanic and non-Hispanic combinations.

In closing, del Pinal reviewed the findings of non-response to the Hispanic origin question. In Sacramento, 4.9 percent; South Carolina, 5.2 percent; and Menominee, 4.6 percent of the respondents did not answer the origin question. In Sacramento, of those who gave no response to the Hispanic origin question, 9.4 percent were American Indian/Alaskan Native; 9.2 percent were Native Hawaiian or Other Pacific Islander; 7.7 percent were Asian; 6.8 percent were Black; and 2.7 percent were White. del Pinal noted that the Hispanic origin question was moved ahead of the race question to increase Hispanic response; when it was placed after race, respondents were less likely to answer the Hispanic origin question.

After the formal remarks of each of the speakers, audience members had an opportunity to pose questions. In response to the question of whether it would be left to the discretion of the agencies to determine the methods for reporting the data, Tucker noted that the purpose of bridging was for trend analysis over time; however, there will be suggested strategies for reporting the data according to the new standards.

In terms of how much race detail will appear in electronic versus paper format, representatives of the Bureau of the Census responded that the Bureau was looking at the addition of paper products. On paper, though, users likely would only see counts for race/origin as opposed to detail.

The moderator posed a question to the audience to consider: When the results become available and the categories are reported, there will be a need to educate the public. How can APDU help us get ourselves ready and help the public to deal with this? One APDU member noted that there will be a need for a simple explanation of how to compare this census to the last. Another noted that bridging analysis work is underway, one example being Professor Stephen K. Doig, Arizona State University, conducting research under a grant from the Russell Sage Foundation.

Down the Home Stretch to Census 2000: a Public Policy Perspective

FRED T. ASBELL
CHRIS STROBEL
CHIP WALKER
DAVID McMILLEN

After nearly a decade of research, planning, and preparation, the Census Bureau is launching Census 2000. Through the nineties, key issues such as funding, counting methods, and content have been the focus of much attention and discussion. In November 1997, Congress established the eight member US Census Monitoring Board, with four members appointed by Congress and four by the President. For APDU99, representatives of the U.S. Census Monitoring Board and of congressional oversight committees agreed to share updates on their current activities and accomplishments.

FRED T. ASBELL
U.S. CENSUS MONITORING BOARD
CONGRESSIONAL MEMBER

WHAT WE DID

We conducted a block-by-block analysis of all the populated areas surveyed in the 1990 Post-Enumeration Survey (PES). We compared the undercount in each block before and after adjustment. We looked at the undercount: how many people were missed in the census. Then, we looked at the adjustment: how many people would have been added or subtracted from statistical adjustment. We documented our method and findings in a September 30 report to Congress: "Unkept Promise: Statistical Adjustment Fails to Eliminate Local Undercounts, as Revealed by Evaluation of Severely Undercounted Blocks From the 1990 Census Plan," which we've distributed. [Available online at <<http://www.cmbc.gov>>]

WHY WE DID IT

We did this analysis for three primary reasons:

1) Full Disclosure: This analysis is an attempt at full disclosure of what adjustment will and will not accomplish at the local level, so local and national decision-makers can make informed decisions. It is not a call to scuttle sampling. The decision of whether or not to adjust the census, or to use adjusted numbers, is not ours to make. Our job is to review the process, ask hard questions where needed, and report our findings to Congress and the communities they represent, so they can make those decisions.

2) Local Data: Obviously, block- and tract-

level data are vital to redistricting. Those data are also used to determine Title 1 funding for schools, Community Development Block Grants, HUD housing grants, and all manner of educational and community planning. It is vital that the people who use these data know what they can expect from the census and adjustment.

3) Statute: The law that founded the Board, Public Law 105-119, charges us to review how the Bureau's plan "shall achieve maximum possible accuracy at every level of geography."

WHAT WE FOUND

We had four major findings. They're listed on page 1 of 28 in the report.

1) Statistical adjustment will not correct large undercounts or overcounts in small areas such as blocks and neighborhoods.

2) Heavily undercounted areas will remain heavily undercounted.

3) Statistical adjustment will add people to many overcounted areas (areas where the Bureau mistakenly counts too many people).

4) Until the census is improved in local areas that are heavily undercounted, the differential undercount will persist at the local level.

The real danger of adjustment is twofold:

1) People may rely on adjustment to fix large undercounts, instead of taking the necessary steps to get a full head count; and

2) Adjustment will appear to fix the

undercount for the entire state—even as heavily undercounted neighborhoods remain heavily undercounted.

Since adjustment will add people, on average, to undercounted and overcounted neighborhoods throughout every state, enough people will be

added to the state's total population to make it look like the undercount is fixed. But people in heavily undercounted neighborhoods still won't get their fair share, and legitimate complaints about their undercounts would be contradicted by the notion that the undercount had been officially "fixed."

CHRISTOPHER STROBEL
U.S. CENSUS MONITORING BOARD STAFF
PRESIDENTIAL MEMBER

Summary of Key Points

Past Board Activities:

We issued two joint reports in 1999 (April 1 and October 1). The April 1 report covered operational aspects of the 2000 Census. The October 1 report was a bipartisan endorsement of the Bureau's planned paid advertising campaign for 2000. [Available online at <<http://www.cmbp.gov>>]

Upcoming Board Activities:

At our Board meeting on August 3, the Board voted to pursue a joint project to carry out case studies of the enumeration planning process in two Census Bureau regions. The studies will look at the strategies the Bureau will use to reach hard to enumerate areas and populations, and how those strategies are developed at the regional and local levels. After deliberation, the Board selected the Dallas and New York regions for the study.

CHIP WALKER
MAJORITY (REPUBLICAN) STAFF
COMMITTEE ON GOVERNMENT REFORM
U.S. HOUSE OF REPRESENTATIVES

The debate over the use of statistical estimation has subsided in Congress, not so much because the Republicans and Democrats have reached agreement on the issue, but both sides have recognized political realities. The Supreme Court ruled in January 1999 that statistical estimation could not be used for the purposes of apportionment in the House of Representatives. Republicans and Democrats disagree on whether that ruling also applies to redistricting.

In the wake of that decision the Administration has decided to conduct both a census using statistical estimation and a full enumeration. Republicans do not have the votes to override a presidential veto and therefore cannot prevent the use of estimation. On the other hand Democrats do not have the votes to change the law that prohibits the use of estimation for apportionment which was the basis of the Supreme Court's decision.

Both Republicans and Democrats want a successful census. Both would like to see the differential undercount eliminated. For these reasons there have been several areas where both sides have been able to join forces with the Census Bureau. One such area is the first-of-its-kind

advertising campaign. Both Chairman Miller and Ranking Member Maloney are very supportive of this campaign and particularly an offshoot of the ad campaign - the Census in the Schools Program. In fact, in October both Chairman Miller and Ms. Maloney spoke at the kick off event at the invitation of the Census Bureau.

Chairman Miller and Ms. Maloney have also joined forces to promote the Congressional/Census Bureau partnership program. Recently a resolution supporting the partnership program was passed unanimously by the House.

H. Con. Res. 193

- Recognizes the importance of achieving a successful census.
- Encourages groups to continue to work toward a successful census.
- Reaffirms our spirit of cooperation with the Census Bureau and asserts a public partnership between Congress and the Bureau of the Census.

Congress has also appropriated all the money that the Census Bureau requested for operations relating to the census, including the \$1.7 billion additional request in the wake of the Supreme Court decision. Chairman Miller and Ms. Maloney have

worked hard to ensure that during the budget negotiations the Census Bureau would continue to be funded at the levels needed to conduct their ad buys and other expensive operations.

While there are areas of cooperation between

Republicans, Democrats and the Census Bureau, Chairman Miller will remain vigilant on how the Bureau spends its \$4.5 billion 2000 appropriation and all aspects of the 2000 Census operations.

DAVID B. McMILLEN
MINORITY (DEMOCRATIC) STAFF
COMMITTEE ON GOVERNMENT REFORM
U.S. HOUSE OF REPRESENTATIVES

(Summarized by Janie Harris, Cornell University Library)

As he has at many previous APDU conferences, David McMillen shared his perspective on congressional actions concerning the census. He reminded attendees of two considerations: that a decennial census requires good fieldwork and that it always attracts political interest because of its effect on reapportionment. With each successive decennial count, public interest has grown along with costs and publicity about the census. As has its predecessors, the 2000 Census has attracted partisan politics. Currently, budget negotiations between Congress and the White House weren't yet resolved over the fate of the Social Security trust fund surplus. Both sides were aware of the political implications of their actions on budget bills. At this time, partisan positions on the census were increasingly affected by other government initiatives requiring public funds, by deadline pressures, and by census oversight concerns. As a result, he foresaw accommodations being made on Census Bureau funding measures.

McMillen also discussed other public data issues currently the focus of congressional attention. He was hopeful that Congress would support the Statistical Efficiency Bill of 1999. The bill incorporates decades of work to maximize

efficiencies across agencies in federal data collection. McMillen was pleased with the strong bipartisan support given to public data affected by the Fathers Count Act of 1999. The House Ways and Means Committee had just moved on the bill which includes some funding for the Census Bureau's Survey of Program Dynamics. Amending earlier Welfare Reform legislation, the bill will support keeping participants in the survey panel so that long-term monitoring data are more accurate. McMillen mentioned a recent Senate resolution which criticized the removal from the 2000 Census questionnaire of the question on marital status. Earlier in the decade, Congress had directed the Census Bureau to remove that question from the form because the data fulfilled no federal mandate. And at this late date, the final 2000 questionnaires are literally rolling off the printing presses.

McMillen made an additional point about public data policy. He expressed concern about the public response to proposed changes to Office of Management and Budget OMB Circular A110. He described the mail response as an Internet contest which generated thousands of electronic messages to OMB. We are in a new era, he cautioned the audience. The cycle of data collection, dissemination and preservation was changing with new technologies. He urged us to think about the effects of change as we considered responsible data use.

Respondent Disclosure in Surveys: Issues and Solutions

ALAN O. ZARATE
PATRICK COLLINS
LISA BRONISZEWSKI

Public survey data provides a wealth of information for researchers. Federal agencies, however, have an obligation to balance the legitimate needs of researchers for access against the needs of the respondents for privacy and confidentiality. Technological advances and cooperative initiatives are creating a complex environment for the new policies and programs under development. Some of APDU's members have long been involved in federal and private sector discussions of the issues. This conference session provided an update on these issues and an invitation for wider participation.

(Summarized by Lisa Neidert, APDU Conference Committee Co-Chair)

Respondent disclosure is an increasingly important issue for data producers, intermediaries, and users. This session provided an opportunity to have presentations from a data producer, the National Center for Health Statistics (NCHS), and a data producer-sponsored secure data room, the Census Bureau's Research Data Center at UCLA and the University of California, Berkeley, and another presentation that combined the viewpoints of intermediaries and users, the Population Research Institute at Pennsylvania State University. The session ended with an opportunity for the APDU community to become involved in a committee on restricted data; attendees and interested others should send email to the last speaker.

Alvan Zarate is the confidentiality officer at NCHS and is also very involved in a government-wide confidentiality group. Zarate's presentation covered his general principles in establishing effective confidentiality practices within a data producing organization. He discussed issues of legal authority, responsibility, training, data storage and handling, access to confidential data, security, review of data for public release, and a written statement of policy. Many of these points were covered in greater detail.

The next speaker, Patrick Collins, described the Census Bureau's Research Data Centers. The purpose of these entities is to provide researchers

Alan O. Zarate, Confidentiality Officer, National Center for Health Statistics. Patrick Collins, Executive Director, California Census Research Data Center. Lisa Broniszewski, Data Archivist, Pennsylvania State University.

with access to Census Bureau confidential data files. Collins described the scope of data available through this venue, for example, microdata not released in public use files. He described the procedures for gaining access and the typical conditions of access at the centers. Although the research data centers sponsored by the Census Bureau, and comparable centers sponsored by NCHS, provide researchers with unprecedented access to restricted data, there are downsides to this access that Collins described. Briefly, these are cost, time, and inconvenience. Over time, he said, the centers will need to demonstrate their merit, to offer a return on the investment, and to explore the economies of scale, for example, with other agency data added to their holdings.

The presentation by Lisa Broniszewski of the Penn State Population Research Institute represented intermediaries and users. Initially, she described what are confidential data, why researchers want access to these files, and security issues with which researchers and intermediaries have to deal. A key point reiterated throughout her presentation was the need to strike a balance between protecting a respondent in a survey and providing access to the responses for researchers. She described ways researchers at her organization gain access to confidential data and which methods are most preferable for them. Not surprisingly, because of cost, time, and inconvenience factors, secured data rooms rank low. She concluded her session by discussing other possibilities that depend on computer technology at the researcher's home institution. In closing, Broniszewski suggested that interested APDU members join together to address concerns about restricted data, and she offered herself as a contact person. Her email address is <lisa@pop.psu.edu>.

Local Area Indicators: Are We Really Measuring What We Think We Are?

MARGARET (MEG) PLANTZ
ANNE E. DUNTHORN

Many APDU members find themselves producing, providing, or using public data for local areas. They quickly discover that the importance of small area data can be as great as the potential for its misapplication or misinterpretation by the local community. In this session, professional data users share their experiences working within local communities and their recommendations for responsible use of public data, especially as a measurement or an indicator of change.

(Summarized by Deborah Gona, APDU Board Vice President)

Moderator Deborah Gona, of Gona & Associates, opened the concurrent session by remarking that in many ways this would be one where the issues of small area data — broached in an APDU98 session — and statistical literacy — discussed the first day of APDU99 — collide. Gona went on to say that while she had earlier questioned the harshness of the tag line, “Are We Really Measuring What We Think We Are?,” to the session title, she reconsidered after reading a related article in *The Washington Post* just one day prior to the start of APDU99.

The article, she said, described a situation where a Maryland public elementary school was preparing to be honored by the White House as one of the top schools in the nation.¹ The irony was that the same institution simultaneously had been fingered by the local school superintendent as one of the “less productive” public schools in the county, a label based on a newly-implemented, local productivity measure that incorporated the rise and fall of state test scores of a subset of students over several grades. Gona read an excerpt from the article in which a puzzled school official asked, “Which process is real?”

Gona remarked that this was just one example of many across the country, and that especially in recent years, many more communities, community institutions, and organizations have taken up the challenges of measurement — whether measuring progress toward their desired futures, assessing whether expended resources are producing any

benefit, or searching for change and improvement in targeted areas. To do these measurements, she said, they search for indicators, and, increasingly, local, regional and national entities such as the Community Indicators Network, the National Association of Regional Councils, and the National Neighborhood Indicators database of the Urban Institute, are gathering up information about local and regional measures that communities and organizations can review, adopt or adapt for their own circumstances.

The process of searching for indicators that link to desired local futures or that speak to the differences organizations are making within communities is spawning some useful discussions about what it takes to be “effective” or to achieve outcomes, Gona noted. They also beg other issues, including some more technical in nature. For example:

- the choices in indicators;
- the troubles associated with finding indicators related closely enough to the conditions that are of interest and that are meaningful reflections of the changes they expect (a task particularly difficult in the health and human services arena); and
- the availability of this level and specificity of data.

But, she cautioned, some issues go beyond the technical to what many of these projects claim to do or are thought to do. In many instances, we have inadvertently stretched the capabilities and intents of some well-meaning activities, such as community status reports and report cards, and other collections of data that can be otherwise powerful tools for informing action, policymaking, building community, and addressing opportunities for community improvement. Instead, Gona said, we

Margaret (Meg) Plantz, Senior Director, Outcome Measurement, United Way of America; Anne E. Dunthorn, Consultant and Principal, CF Systems.

have used them for purposes for which they were not intended, for example, measuring targeted change — and ultimately assigning blame or credit for progress—within communities or organizations. The speakers at this session were invited to offer their perspectives on these indicator issues.

The first speaker, Dr. Margaret Plantz, Senior Director of Outcome Measurement for United Way of America, told the audience that she had just returned from a meeting of seven local United Way organizations involved in the National Learning Project on Using Program Outcome Data to Create Measurable Change. She pointed out that United Ways and member agencies in many communities have been developing and employing program-level outcome information to track program performance, and more recently, United Ways have begun working on developing viable community-level outcome information. This is becoming increasingly important and of interest, she said, as the United Way system moves more deliberately into community-level problem-solving and looking for ways to impact and add value to the community, along with its traditional fund-raising for local agencies.

Since 1995, Plantz said, United Way agency program-level outcome measures have been used to demonstrate the benefits and changes produced by programs. However, in spite of some efforts to attribute the results, those program-level outcomes cannot be stretched to demonstrate community-level impacts, she said. Some communities and United Ways, for example, have mistakenly assumed that if they could just fund “enough” programs or could just aggregate the outcomes of the programs they fund or could use all of the strategies at their disposal (even in partnership with all of the funders in the community), then they would be able to demonstrate community-wide impact.

In reality, Plantz said, community-level outcomes are benefits or changes for large numbers of community members, and there are not enough programs, with enough clientele, to create the anticipated, measurable change across the community or to hold any individual organization accountable for that level of change. United Ways that have allowed themselves and their communities to think by adopting a worthy, community-wide outcome, such as healthy families, that they are accountable for change in broad indicators, such as a divorce rate, have found themselves looking ineffective when they “fail” to create the change in that indicator. Some community-level indicators are simply beyond local control, but because data are

available, we let them drive what we hold ourselves accountable for, she said.

Plantz noted that a related assumption has been that community indicator reports by themselves can meaningfully measure the community-level impact of United Way efforts. Unfortunately, she said, we have realized that indicator-based “report cards” are not appropriate ways of measuring the impact of community problem-solving efforts. If done well, Plantz said, report cards can offer excellent measurement and understanding of the status or conditions of a community, but they are inadequate for demonstrating change in conditions that are occurring as the result of specific, deliberate, targeted actions.

Plantz noted that the United Way system distinguishes between “Community Status Reports” and “Targeted Community Interventions” and related indicators² in their purpose, focus of accountability, starting point, key tasks, and uses of the results. The selection of targeted indicators and a measurement plan is a final step, coming only after an intended outcome (or outcomes) is selected from within a broad community goal, and specific strategies and action plans are developed to achieve and implement the outcomes. Plantz cited an example of one community’s winnowing process, narrowing the potential outcomes to one that the United Way and community partners could reasonably impact. Many of the most successful, she said, have strategies that focus on system changes to fix the source of the problem, rather than relying solely on changing individuals. They think through the logic and look for factors that can realistically be affected by the available resources and partners in the community.

She cautioned that appropriate indicators for measuring success may have to be brand new, as many available indicator reports measure too high a level of indicators. Targeted indicators must focus specifically on the factors that the action plan is designed to affect. More often than not, Plantz remarked, these are narrower and more focused than the original community status indicators that gave rise to the concern about the problem.

The second speaker, Anne Dunthorn, consultant and co-owner of CF Systems in Oak Ridge, Tennessee, echoed many of the same sentiments in describing her community’s work on report cards and indicator development. Dunthorn recounted that her community, Oak Ridge, was founded in the 1940s as part of the Manhattan Project to develop materials for atomic bombs.

Most of those recruited to the area back then were young and highly educated or trained in technical work. Now, she said, these original residents are at or reaching retirement age, and without school-age children. The local school district thinks that it needs to measure and sell academic “success” in order to continue high levels of public support and funding for the school system. The question, she said, is what is “success”? How much of it is genetic, how much is due to family resources and socioeconomic status, and how much is the result of schooling? What are they really measuring, Dunthorn asked, and how much is based on perception and what the public wants to hear?

She went on to describe an effort that resulted in the release last year of a “Community Report Card” for Anderson County, Tennessee.³ Dunthorn said that the participants moved away from targeting a single issue to taking on a broad range — widening from health issues to those of safety, housing, environment and others as more stakeholders became involved. For example, she said, both the city and county governments — different agendas and concerns — got involved because they wanted a tool for planning in this area which has an affluent city in the midst of rural poverty, and county rural residents cut off from quick, efficient services by the geography.

The process of developing this report card brought together groups who decided to work cooperatively and use their resources more efficiently, Dunthorn pointed out. The United Way in Anderson County, one of the project sponsors,

realized it lacked resources within the community to solve the identified problems and worked proactively with other groups to leverage state and federal public resources, as well as some from the private sector. The community report card, she said, helped to identify funding “holes” and guide a strategy for securing the necessary funds. Moreover, the community will be using the data reported by social service agencies to various funders as a planning and evaluation tool. Data reported to the state and federal level will be sought and “recaptured” for local use.

Dunthorn closed by reminding the audience that perception is an important part of how we use data. There is always a question of whether we are measuring what we think we are and whether we are even looking at the same thing. But even if we cannot easily compare them to other communities, she said, we must create, in our local areas, the data that are necessary to monitor success.

¹ Brigid Schulte, “Discrepancies Mar New School Gauge: Montgomery Stresses Gains, Not Just State Test Results,” *The Washington Post*, Oct. 24, 1999, A1.

² “Community Status Reports and Targeted Community Interventions: Drawing a Distinction. A White Paper on Change in the United Way Movement.” United Way of America, 1999.

³ “Anderson County (Tenn.) Community Report Card: Building a Healthy Community 1999.” Sponsored by Anderson County Health Council, Concerned Citizens, Covenant Health, and United Way of Anderson County.

Promoting Public Access to Scientific Data: the Devil is in the Detail

GEORGE LEVENTHAL
MARK FRANKEL
GARY BASS

Public data have long been the focus of attention because of their use in congressional reapportionment, cost of living adjustments, interest rates, etc., and in policy formulation. Research data generated in scientific laboratories and on academic computers are now inviting public scrutiny because of their potential to affect public policy. Recent federal regulations opened public access to research data produced with federal funds, and this access raises concerns ranging from cost and competition to confidentiality. APDU's conference session brought together representatives from the government, the research community, and a watchdog agency to share their perspectives on and suggest alternatives to the new federal requirements.

(Summarized by Ted Hull, APDU Board Member)

Ted Hull, session organizer and moderator, introduced this session on the revisions to the Office of Management and Budget (OMB) Circular A-110. An archivist with the Center for Electronic Records of the National Archives and Records Administration, Hull started by discussing the implications of the revision for the federal record status of data collected under grants and contracts. The revision, by using the term "awards" and invoking provisions of the Freedom of Information Act (FOIA) for access to data will, in effect, cause all data created under grant or contract to be federal records. Previously, only those data received by an agency from the grantee become federal records; otherwise data retained by the grantee are not managed as such. Data collected under contract and received by the agency as a deliverable of that contract are federal records; if not specified as a deliverable, they are not federal records. Therefore, by suggesting that all data, regardless of the source of the funding, are subject to the FOIA, the revision effectively changes the federal record status of all data regardless of whether they are under the federal government's physical control or not. Hull went on to introduce the distinguished panelists.

George Leventhal, Association of American Universities (AAU), discussed the history of the revision. The Shelby Amendment, introduced by Senator Richard C. Shelby (R-AL) arose from a US

George Leventhal, Senior Federal Relations Officer, Association of American Universities. Mark Frankel, American Association for the Advancement of Science. Gary Bass, Executive Director, OMB Watch

Environmental Protection Agency ruling on a particulate emission standard based on a publicly funded Harvard University research study. One of the companies affected by the standard sought to replicate the study and appealed to Senator Shelby to assist in obtaining the data from Harvard. They argued that the weather and other medical and survey data used by Harvard were all public data. Harvard refused to turn the data over to the requester or Senator Shelby's office. This resulted in Senator Shelby introducing an amendment to the FY1998 omnibus OMB spending bill requesting revision to OMB Circular A-110 subjecting all data collected under a federal award to the provisions of FOIA.

Leventhal recounted how the university community expressed great concern over the FOIA provision and the effect it might have on maintaining the confidentiality of persons participating as subjects in a research project. This is of particular concern where there are medical proprietary interests and when the private sector is a partial sponsor of the research. Leventhal suggested that if university scientific research is for the essential benefit of science, then the trade secrets exemption of the FOIA would not apply to that scientific research data, and it might be subject to disclosure. He then provided an example of the R.J. Reynolds Tobacco Company suing a Medical College of Georgia researcher for the identities of subjects in a study of the awareness of the cartoon "Joe Camel" among young children. The researcher lost the case and had to identify the subjects under the Georgia Open Records Law.

Leventhal continued his discussion by

providing additional background on the legislative action surrounding the revision. An OMB notice issued in February 1999 indicated that the FOIA would apply only to cases where the data resulted in a federal law, but that notice also broadly applied to all data collected under an award. A subsequent proposed revision to the amendment by Representative Price and Representative Walsh limiting the scope further, lost in committee. However, a second ruling by OMB on the revision issued in August 1999, significantly narrowed the scope of the final revision. Senator Shelby thought the OMB second ruling gave too much discretion to researchers. As a result of the compromise, on October 8, 1999, OMB issued the final ruling on the revision, which made some accommodations to Senator Shelby, by applying OMB Circular A-110 to data used in “any federal action that has the effect of law.” The revision is effective November 8, 1999, and agencies now have to implement the revision. Currently under question is whether the time span of Senator Shelby’s initial amendment expired at the end of FY1999, as it was initially attached to an annual appropriations bill.

Mark Frankel, American Association for the Advancement of Science (AAAS), suggested alternative approaches to data sharing other than formal revisions to the OMB Circular. He recounted how the scientific community has had a long history of incorporating data sharing into its ethical practice. Sharing of data is essential to the progress of science, and researchers routinely share data for the benefit of all. Frankel listed some of the benefits of research data sharing including : resolution of disputes in science; exposure of errors and fraud; verification of original findings; generation of new research and promotion of new analyses leading to new knowledge and insights; provision of a resource for training in research methods; and avoidance of wasteful duplication. In all, one of the goals in data sharing based on these principles is to improve public policy.

Frankel also listed some of the obstacles to data sharing: technical constraints; poor data documentation; contamination of data; loss of control over the data; risks to privacy and confidentiality of research subjects; potential profits; and national security. Therefore, he said, any policy on data sharing needs to consider both the benefits and the obstacles. One of the things the Shelby

Amendment did not consider is that the scientific community has long recognized that peer review is essential for valid science and is not possible without sharing. The amendment also did not recognize existing provisions of the National Science Foundation and National Institutes of Health, which expect funded researchers to make their data available to others. Frankel suggested that current sharing practices could be expanded by methods other than the FOIA.

Frankel then suggested one of the ways data sharing could be promoted would be through a federally funded “virtual archives” of scientific data. The AAAS is currently exploring the feasibility of the development of such a virtual archives. Scientific data, developed with federal funding and not easily replicated by other scientists, would be electronically available via the archives. The public could use this mechanism to identify and locate specific data sets. No award would be given without satisfactory data sharing provisions, given the restrictions of human privacy, intellectual property, and national security. According to Frankel, this plan affirms the ethics of data sharing.

Gary Bass, OMB Watch, asserted that the Shelby amendment is “crazy.” He listed three reasons for this: 1) that FOIA is the wrong method; the amendment flies in the face of all existing FOIA case law; 2) that the scope is inappropriate and not a coincidence that it applies only to awards and 3) that the reason for the amendment focused on benefiting special interests over enforcing regulatory protection. The amendment is not about good science but rather about controlling public policy. Bass does not think OMB has done a good enough job limiting the scope of the amendment. Bass then went on to raise a number of questions about the final OMB issuance. First, to what will the FOIA really apply and how will that affect all non-profits and universities who receive grants? Second, who will absorb the costs for responding to the FOIA and what if an agency doesn’t or can’t force a fee on the requester? Bass thinks that the revision does not address the question of what is good access to public data. He concluded by remarking that the Shelby amendment is a side show, and that attacks will be on procedural implementation of the revision and will not result in what is really needed—an affirmative policy on access and a definition of public access.

What Are the Effects of Welfare Reform on the Nation's Children?

ALYSSA WIGTON
JEFFREY CAPIZZANO

Several years into welfare reform, policy makers and the public are attempting to identify, measure, and evaluate the results. Various indicators of well-being are being compared across the fifty states. The Urban Institute has targeted its resources on providing new, timely and reliable data for such comparisons. Institute data which document the impact of welfare reform on children and their families were the topic of this APDU conference session.

(Summarized by Jocelyn Tipton, Yale University Library)

Sarah Breshears, APDU Board Secretary and a Senior Research Specialist at the University of Arkansas, Little Rock, was the moderator of this session. She introduced the topic by explaining that welfare reform is driving users of public data and that child care issues directly relate to the welfare to work program.

Alyssa Wigton of the Urban Institute provided an overview of the data available from "Assessing the New Federalism," a multi-year Urban Institute research project to analyze the devolution of responsibility for social programs from the federal government to the states. It includes studying the well-being of children and families. She gave brief descriptions of the content and data collection for three data sources available from the project's web site: the State Database, the Welfare Rules Database, and the National Survey of America's Families.

The State Database includes information on the fifty states and the District of Columbia in areas such as income security, health, child and youth well-being, demographic, fiscal and political conditions, and social services. It can be used to compare states with each other or with the nation as a whole, to track trends over time, and to analyze relationships among different state characteristics. This database currently has data from 1993 to the present and is updated twice a year. The data can be used online or downloaded. More information can be found at the Institute's web site: <<http://newfederalism.urban.org/nfdb/index.htm>>.

The Welfare Rules Database (WRD) provides a longitudinal account of the changes in state

welfare rules. The WRD organizes the detailed information on welfare rules across states, time and geographic areas within states and different types of assistance units. The information in this database is being compiled from caseworker manuals and Temporary Assistance to Needy Families (TANF) regulations. This database is still being developed but a beta version is available at: <<http://newfederalism.urban.org/wrd>>.

The National Survey of America's Families conducted in 1997 included over 44,000 households yielding information on over 100,000 people and was repeated again in 1999. The survey focused on economic, health and social characteristics of children, adults and families. The survey has the ability to measure differences among states. The public use files are available on the Institute's web site: <<http://newfederalism.urban.org/nsaf/cpuf/index.htm>>.

There are currently two files available with more in development. The Child Public Use File provides information relating to their family settings and the adults who took care of them. The Most Knowledgeable Adult (MKA) Public Use File provides information about the adult in the household who knew the most about the children. For each of the public use files there is also information about methodology and the text of the questionnaire. The data are in compressed ASCII format that can be downloaded or can soon be used with Tabulator, a Windows-based program which will enable users to design and generate custom crosstabulations. The web site also provides access to all printed reports available from this survey.

Jeffrey Capizzano, also from the Urban Institute, provided examples of how the National Survey of America's Families is being used and the

Alyssa Wigton, Urban Institute
Jeffrey Capizzano, Urban Institute

type of questions being answered by the data. He also provided some background about why we should care about child care usage patterns. Understanding child care arrangements is important because it is a critical work support mechanism, policy makers care about school readiness, and it is impacted by welfare reform measures. State estimates of child care are also important because it is suspected that states differ demographically and fiscally, the cost and supply of child care vary, and state approaches to child care policy also vary. This survey can help to verify these differences and provide policy makers with state specific information.

Capizzano focused on data regarding primary caretakers of preschool children with employed parents and shared several charts and graphs to illustrate his findings. These contrasted child care arrangements by the age of the children and by low or high family income. In conclusion, he found that there were significant state variations and that state specific data such as that available from the National Survey of America's Families is important. The next step will be to try to understand why these differences among states occur. The public release of these findings should be available by the end of the year from the Institute.

Roundtable Sessions: Data Producer and User Forums

Traditionally, APDU conference Roundtable sessions offer data producers, providers and users the opportunity to meet informally for topical discussions and product demonstrations. At APDU99, producers shared and attendees tested several new products and software systems for accessing public data.

Roundtable 1: The Do's and Don'ts of Getting Statistical Policy Across to the Local Public

MILO SCHIELD

(Summarized by Wendy Treadwell, President, APDU Board)

As a follow-up to his contributions to *Of Significance...* (vol. 1:1), Milo Schield led this roundtable discussion on the interpretation and misinterpretation of published data. Using examples from both federal and local sources, the group examined instances in which statistics were presented in ways that led to multiple interpretations or simple misunderstanding. Common errors identified and discussed included confusing a simple difference with a relative difference, confusing “percent” with “percentage point,” and mixing terminology such as “three times as much” and “200% more than” which can confuse the reader. Another problem, which Professor Schield observed while working with students, was their difficulty in understanding tables that provide percentages. He found that having his students try to explain in simple prose what a particular percentage meant was a good test of understandability. Providing the values along with the percent helped to clarify the information in many cases. For example, a table with the title

Milo Schield, Associate Professor, Business Administration and MIS, Augsburg College

“Percentage of Children in Poverty by Race” could be interpreted as the percentage of those children considered “in poverty” broken down by race OR the percentage of children within each racial category who were considered “in poverty.” Providing the base number of children in each racial category as well as the number of those children who were considered “in poverty” clarified the percentage listed in the table.

Recommendations for improving the clarity of reported data included:

- When comparing two values, give both values being compared along with the comparison.
- Use “times as much as” for simple ratios and “% more than” for relative differences.
- Use “times as much as” when the simple ratio is more than one and “% more than” if the ratio is less than three. This avoids the use of small ratios and large percentages, both of which can be confusing to the reader.
- Have someone try to explain the meaning of the numbers in a table to verify that the language used is not open to multiple interpretations.

Roundtable 2: Hands on Time with the American Community Survey

NANCY TORRIERI

(Summarized by Sue Hoover, APDU Board)

The American Community Survey (ACS) is being developed by the Census Bureau to provide demographic, economic, and housing profiles of America's communities every year. The Census Bureau's Nancy Torrieri, who is heading up the American Community Survey group, and two of her staff members presented an overview of the ACS. Handouts were provided as well as the opportunity to do "hands on" work with preliminary data via a laptop—as though it were "real data."

The Census Bureau's web page provides an entire area dedicated to the ACS. The Census

Nancy Torrieri, Bureau of the Census

Bureau's home page is located at <<http://www.census.gov>>. Under "Special Topics" is a section for the American Community Survey. The ACS Home Page covers a number of questions, for example, "What is it?, Goals, Why?, How Does it Work?" etc. Linked from the "ACS Main" and "About ACS" pages, there is a PowerPoint Slide Show presentation that can be viewed and/or downloaded. The "Notes" pages contain the presentation as well as concise explanations of the entire process and the projected timelines.

This APDU Roundtable presented the opportunity to discuss the process and plans for ACS with knowledgeable Census Bureau staff as well as the chance to work with the proposed data. It was informal yet informative.

Roundtable 3: UNICON and CPS Utilities

BRYAN RICKARD

(Summarized by Lisa Neidert, Co-Chair, APDU Conference Committee)

The format of this Roundtable was a live demonstration of the UNICON CPS Utilities. The CPS Utilities contain data, documentation and Windows software to help researchers find and extract data from the US Census Bureau's Current Population Survey (CPS). In the past, the data and extraction software have been on CD-ROM. However, session participants also got to view a web-based delivery method.

Meeting with UNICON programmer Bryan Rickard, participants were able to get guidance on

Bryan Rickard, Senior Programmer, Unicon Research Corporation

problems they were having, ask question about future directions in the software development, and become beta testers of the web version of the CPS Utilities.

The CPS Utilities developed by UNICON addresses problems users have with CPS data files. Utilizing three basic functions, the software permits the user to locate relevant variables with relative ease, to produce data files by simply naming variables and years, to collect and view in compact form all coding and universe information for a variable across survey years, and to ascertain easily the survey question that led to the variable of interest. Further information is available on the producer's website at <<http://www.unicon.com>>.

Roundtable 4: Getting to Know Your Data with PDQ Explore

ALBERT ANDERSON

*(Summarized by Lisa Neidert, Co-Chair, APDU
Conference Committee)*

PDQ-Explore is an online interactive analytic processing system from Public Data Queries, Inc. PDQ-Explore offers remote access to a variety of large, public data sets. Users are able to extract results from these data in seconds.

The format of this roundtable was a live demonstration of PDQ-Explore followed by hands-on experience by the participants with the software and data. Participants had the opportunity to interact with over 100 years of census data with a

Albert Anderson, Public Data Queries Inc.

concatenated version of the Integrated Public Use Microdata Sample (IPUMS) files. Tabulations and descriptive statistics queries were returned in one to two seconds on a file that has eighteen million records.

Longtime APDU member Albert Anderson headed the demonstration. Session participants were able to ask technical questions of the developers of the software, discuss the feasibility of providing access to other data files, and get practical questions about using the software answered. Participants were given copies of a beta version of the software and a guide to using the software. Information on this software system can be found at the producer's website at <<http://www.pdq.com>>.

Roundtable 5: The Nation's Data on DVD: an Introduction to USA on DVD

BRAND NIEMANN

*(Summarized by Lisa Neidert, Co-Chair, APDU
Conference Committee)*

USA on DVD is a reference and analysis tool on DVD-ROM that is usable on any Windows-based operating system with a DVD-ROM drive. The information that USA on DVD displays in maps and tables combines Environmental Protection Agency (EPA) databases with geographical features and statistics on demographics and economics from the 1990 Census.

The format of this roundtable was a live

*Brand Niemann, US Environmental Protection
Agency (EPA)*

demonstration of the USA on DVD followed by hands-on experience with the software and data by the session attendees. Participants generated various combinations of data with various multi-layer map plots. The plots ranged from neighborhood maps to state or regional thematic maps. The software was easy to use and made the presentation of geographically-based data reasonable for novices. The amount of information available in both text and fully graphical GIS maps and plots also makes this an attractive product for skilled professionals. Further information is available at the website:
<<http://www.usaondvd.com>>.

Changes in the MSA Definitions: What is the Impact on Local Users ?

The federal government classifies cities into types of metropolitan areas for statistical and programmatic purposes. Statistical purposes for federal agencies include the collection, tabulation and dissemination of public data for geographic areas. Programmatic uses can involve federal and state disbursements for the improvement of health or housing, or for economic development. The private sector also utilizes the classifications, for example, for ranking or targeting urban markets. Achieving metropolitan status has a major impact on a local area, as does its loss. At APDU99, two presentations dealt with this important statistical classification. James D. Fitzsimmons of the Census Bureau reported on the current review of the standards which will define metropolitan areas in the coming years. Jacqueline Byers of the National Association of Counties discussed the local impact of the classifications.

Progress Report on the Review of the Metropolitan Area Standards

JAMES D. FITZSIMMONS

The Metropolitan Area Standards Review Project is approaching its goal of producing standards for defining statistical areas in the coming decade. Underway for several years, the project is an interagency effort conducted under the auspices of the Office of Management and Budget (OMB). The new standards resulting from the project will be applied with 2000 decennial census data to produce updated definitions of areas by mid-2003.

OMB recently published a *Federal Register* notice (October 20, 1999) presenting recommendations from the Federal Interagency Metropolitan Area Standards Review Committee (MASRC) on changes to the current metropolitan area (MA) standards. This notice is the second of three notices related to the review. The first notice was published by OMB in the *Federal Register* of December 21, 1998. OMB expects to publish the final standards in the third notice before census day (April 1) 2000.

Background on the Current Metropolitan Area (MA) Program

For 50 years the MA program has provided standard statistical area definitions at the metropolitan level. During this period, large

numbers of directly comparable MA data products have been made available to government, business, scholars, citizens' organizations, and others interested in analyzing and otherwise using information about the nation's major centers of population and activity.

The general concept of an MA is that of an area containing a large population nucleus and adjacent communities that have a high degree of integration with that nucleus. This general concept has remained essentially the same since MA's were first defined before the 1950 Census. The purpose of MA's also is unchanged from when they were first defined: the classification provides a nationally consistent set of definitions for collecting, tabulating, and publishing federal statistics for geographic areas.

In the years preceding each decennial census, the MA standards have been reviewed and revised to ensure their continued usefulness and relevance. The revised standards then have been applied to the new census data to produce definitions for that decade.

The Current Standards Review

The Metropolitan Area Standards Review Project is the fifth review of the standards since they were first published; it has been an especially

James D. Fitzsimmons, U.S. Census Bureau

thorough review, reflecting as a first priority users' concerns with the conceptual and operational complexity of the standards that have evolved over the decades. Other key concerns behind the particularly thorough nature of the current review have been: (1) whether modifications to the standards over the years have permitted them to stay abreast of changes in population distribution and activity patterns; (2) whether advances in computer applications permit consideration of new approaches to defining areas; and (3) whether there is a practicable way to capture a more complete range of settlement and activity patterns than the current MA standards capture.

Review activities began early in the decade, when the U.S. Census Bureau commissioned studies by four universities to identify specific issues that should be addressed and to propose approaches to those issues. The reports coming out of these studies were assembled and published as a U.S. Census Bureau Population Division working paper, which then became the centerpiece of discussion at an open conference in 1995. Last year's *Federal Register* notice (December 21, 1998) presented alternative approaches to defining metropolitan and nonmetropolitan areas and sought comments on those approaches. A second open conference took place in January, 1999. Comments received in response to the December 1998 notice, views expressed at the two conferences and at dozens of meetings with professional and data user groups in the past few years, and work conducted by staff at the U.S. Census Bureau and other agencies informed MASRC's extensive deliberations this past spring and summer as it prepared its recommendations for OMB on changes to the standards.

MASRC Recommendations

MASRC has recommended a Core-Based Statistical Area (CBSA) classification to replace the current MA classification. This CBSA classification has three types of areas based on the total population of all cores in the CBSA: (1) Megapolitan Areas based on cores of at least 1,000,000 population; (2) Macropolitan Areas based on cores of 50,000 to 999,999 population; and (3) Micropolitan Areas based on cores of 10,000 to 49,999 population. The cores on which the areas are based are U.S. Census Bureau-defined urbanized areas (UA's) and a proposed new geographic entity for Census 2000—U.S. Census Bureau-defined settlement clusters.

Megapolitan Areas would comprise a recognizable set of the largest areas and would account for a substantial share of the nation's total population. The identification of Micropolitan Areas extends concepts underlying the core-based approach to smaller population centers that are not addressed in the current MA program. Territory not included in Megapolitan, Macropolitan, and Micropolitan Areas would be termed "Outside CBSA's."

MASRC has recommended using counties and equivalent entities as building blocks of CBSA's throughout the United States and Puerto Rico. (Minor civil divisions would be used as building blocks for an alternative set of areas in New England only.) Those counties containing cores, MASRC recommends, should form the *central counties* of CBSA's. MASRC also recommends that only commuting data should be used to include counties beyond central counties—the *outlying counties*—to complete definition of CBSA's. MASRC has recommended adoption of a single commuting threshold of 25 percent to establish qualifying linkages between outlying counties and central counties.

Mergers of adjacent CBSA's to form a single CBSA should take place when commuting data indicate that strong ties exist between the two areas' central counties. *Combinations* of adjacent CBSA's should take place when there are weaker but still important commuting ties between entire CBSA's. The CBSA's that are combined should retain separate identities in addition to being recognized as parts of *Combined Areas*.

MASRC recommends identifying the city with the largest population in each CBSA, as well as any additional cities with large population or employment totals, as *principal cities*. The title of each CBSA should include the name of the largest principal city. If there are multiple principal cities in a CBSA, the names of the second-largest and third-largest principal cities should be included in the title, in order of descending population size. Titles of Combined Areas also would use names of principal cities.

The full set of recommendations and the October 20, 1999, *Federal Register* notice of which they are part are featured at OMB's web site at <<http://www.whitehouse.gov/OMB/fedreg/index.html>> or at the U.S. Census Bureau's site at <<http://www.census.gov/population/www/estimates/masrp.html>>.

The Good, the Bad and the Ugly: A Look at Counties and Metropolitan Areas

JACQUELINE BYERS

(Summarized by Jocelyn Tipton, Yale University Library)

Achieving metropolitan status has a major impact on a local area. New programs and opportunities become available once an area is designated as a Metropolitan Statistical Area (MSA). But what happens when an area falls below the magic number and no longer qualifies as an MSA? Jacqueline Byers' presentation looked at the advantages and disadvantages of counties belonging to metropolitan areas.

Ms. Byers discussed several benefits for counties belonging to an MSA or a Metropolitan Area (MA). [For definitions of these classifications, please see the Census Bureau glossary at <<http://www.census.gov/dmd/www/glossary.html>>] Urban identification impacts the workforce availability for business and economic development recruitment. Counties in MSA's are eligible for certain federal funding or higher levels of funding for education, housing and transportation. There are also economic impacts of belonging to an MSA. The concentration of businesses and people in MA's attracts new industry, generates new industries, speeds the expansion of knowledge, increases technological innovation and increases productivity. MA's and MSA's have greater access to labor, capital and customers. Workers are more diverse and better educated because of access to training, colleges and universities. These workers bring more specialized skills and higher production because of the concentration of workers which in

Jacqueline Byers, Director of Research, National Association of Counties

turn brings higher salaries. The final benefit of belonging to an MA or MSA is that passenger, freight, and regional transportation is more concentrated, which lowers costs.

Ms. Byers went on to discuss some of the disadvantages to belonging to MA's or MSA's. The primary example she used was the Dupage County-Chicago/Cook County, Illinois, MSA. The first disadvantage is how the U.S. Department of Housing and Urban Development (HUD) defines market areas for rentals based on MSA/MA fair market value. In DuPage County there are newer suburbs, wealthy families and a high demand for housing which leads to higher rents. HUD uses MA market areas to develop fair market value using older inner city and inner ring rental rates. Rental subsidies are often insufficient to subsidize real rents. Few housing owners in DuPage County will accept certificates, and low income people have to go elsewhere for housing. There is a definite connection in commuting patterns with downtown Chicago, but not a connection in housing markets and patterns. Many new jobs in the contiguous counties are in service industries with lower pay so employees cannot live near their jobs. Transportation is often designed to go into the central city in the morning and out in the afternoon, which does not help people commuting the other way. Also, because of a county's being in an MA, agencies are required to look at all counties in the MA as one ring of economic development and one big market. This is not always true; many contiguous MA counties have become employment centers themselves with economic engines that are separate from the core.

Post-Mortem Report: Address Review

PATRICIA C. BECKER
JOSEPH J. SALVO

Urban based demographers know firsthand how important census results are to local officials for whom the counts tally prospective voters or federal aid recipients. Patty Becker of Detroit and Joe Salvo of New York City are consummate professionals whose public data expertise is respected and sought both locally and at the national level. APDU has long benefited from their work to improve data collection and population measurement, as well as from their support of the Association. At this year's conference, they shared their experiences with the Census Bureau's efforts to improve respondent address lists.

In 1970, the Census Bureau conducted the first mail-out/mail-back census. This meant that, instead of building the address list through the field enumeration, the Bureau had to have an address list, or MAF (Master Address File), *before* the census. The mail-out/mail-back activity covered about 65 percent of the nation in 1970, but has steadily increased to nearly 100 percent for the 2000 Census.

In designing the 1980 Census, the Bureau realized that the MAF that it could establish on its own, using a combination of commercial lists and field work, was not good enough. Accordingly, they instituted the Local Review Program. As designed, this was to provide for local community review of housing unit counts by block both before and after the census. Due to various problems, not unlike what has happened in recent years, the pre-census local review effort was cancelled and the post-census review was only made at the enumeration district level—an assignment area for field enumerators that was approximately the size of a census block group.

Local review was again part of the plan for 1990, and this time it was carried out according to that plan. Local communities had the opportunity for both pre- and post-census reviews using housing unit counts aggregated to a census block. In both 1980 and 1990, the post-census phase also included some population counts.

The 1980 and 1990 review efforts were conducted at the block/enumeration district level because the Census Bureau was prohibited, by law,

Patricia C. Becker, APB Associates and Southeast Michigan Census Council; Joseph J. Salvo, Department of City Planning, New York City

from sharing its actual address list with local governments (or anyone else, for that matter.) A major improvement came about in the Census Address List Improvement Act of 1994 (P.L. 104-130). This bill made two important changes. First, it required the US Postal Service to share its address list with the Census Bureau. This provided a continuing source of new address information to the MAF. Second, it permitted the Census Bureau to share its actual address lists with local governments. This is the provision of law that gave genesis to the LUCA program: Local Update of Census Addresses.

LUCA did not get off to a very good start. Even though the law was passed in 1994, the Census Bureau did not officially notify local governments of their opportunity to participate until February, 1998. One outsider observed that the Census Bureau's Geography Division, which was charged with responsibility for 2000 Address List Development, believed that the Postal Service's Delivery Sequence File (DSF), combined with the 1990 Census address list, would in fact cover the vast majority of addresses and that LUCA would provide only marginal assistance in improving the MAF.

Subsequent events proved this expectation to be false on a number of fronts. The biggest problems were and are in old central cities and in fast growing communities, but virtually every community could improve its MAF by doing LUCA. The full story of LUCA is yet to be told. The process itself is not complete. Significant research efforts will be undertaken over the next year to evaluate and measure what happened.

This effort is especially important because the MAF is intended to be a permanent, continually

updated file. It is scheduled to be the sampling frame for the Census Bureau's American Community Survey and perhaps for other federal surveys as well. It also has very significant potential for use in population estimates. Finally, of course, it will certainly be the basis for the MAF to be used in the 2000 Census.

Bottom line: stay tuned!

Metadata on Steroids: A Report on the Data Documentation Initiative Beta-Tests

ANN GREEN
WENDY TREADWELL
PETER JOFTIS

If users are finding data access easier, it is due to the hard work of data producers, intermediaries and archivists to make it so. Among the APDU members are national experts on metadata, who with their peers are planning and developing an international, interrelated dynamic system of shared standards and infrastructures. This year's panel session on metadata offered a discussion of many new changes for consideration and contacts for follow-up.

(Summarized by Wendy Treadwell, President, APDU Board)

Why is the Data Documentation Initiative (DDI) so important? What has metadata got to do with accessing data? Why should APDU members care? During this 90-minute session, the speakers provided a thumbnail sketch of the complex and interrelated structure which underlies the interface systems that provide access to social science data. The interface may have the flashy bells and whistles and the search systems may have the neat features which prescribe how we locate and access data, but it is the availability and structure of the underlying metadata which both drives and constrains these systems.

This session laid out changes in the current infrastructure and the current state of metadata standards development. The current infrastructure consists of packages of objects such as data files on various media and related documentation or reports, as well as a variety of data distributors and intermediaries. The new infrastructure seeks to deliver discreet products and the ability to provide interactive access within which the user can extract, convert and manipulate data in order to create customized products. While these advanced access systems (interfaces and search engines) have improved the ease of access, in many cases they have not made the data themselves any less complex to use. Users still need help to locate, evaluate, acquire, format and interpret statistical resources. They need people to serve as intermediaries either as interpreters or as interface

Ann Green, Director, Social Science Statistical Laboratory, Yale University; Wendy Treadwell, President, APDU Board, and Coordinator, Machine Readable Data Center, University of Minnesota; and Peter Jofitis, Inter-University Consortium for Political and Social Research

developers. They need real-time analysis, subsetting and extraction tools in order to work with the data more directly. And finally they need "intelligent data," data that knows about itself, how it can be used and how it relates to other data objects. All of these speak to the high priority of developing metadata. According to Jostein Ryssevik of NESSTAR (Networked Social Science Tools and Resources), secondary users must rely on the amount of formal metadata that travels along with the data in order to exploit its full potential. The metadata is the pivotal point for any resource discovery system (paper-based as well as digital). Detailed metadata about the sources as well as variables is needed to increase the precision of the discovery process. Metadata is actually the bridge between the data, the users, and the intellectual process. It is a vehicle in the process of knowledge accumulation.

The primary problem is that we still have barriers between the data users and the metadata. In many cases we still lack documentation of complex instruments that reflect accurate flow patterns. Where we do have electronic documentation, we are often unable to search and manipulate the unstructured codebook information. There may be clumsy or dead links between the descriptive information and the data. We are hindered in providing access to a broad range of data due to the varying structures of metadata repositories that have been designed for producers rather than users, and by the problems of interoperability among data extract and analysis systems. In order to effectively locate and retrieve data, users need both descriptive information about the data as well as the rules, relationships and context of individual data objects. The data must be able to tell a story about itself or provide the ability to go back and find the details a user needs. While

documentation is becoming more fragmented (brief data object descriptions attached to data objects), users still think in terms of The Codebook, a codified collection of supporting documentation for the data. What we need is to provide a mechanism for building linkages among the pieces of documentation in order to reconstruct the context.

What we need to do is to develop the Data Web with metadata at the core. Such a Data Web would integrate access to disparate data sources, provide the ability to analyze more than one study at a time and integrate the access and analysis systems including a wide variety of tools, from geospatial and statistical to qualitative and multimedia tools. To build this Data Web we must address two fundamental problems, a lack of standards and a lack of metadata. Metadata is crucial to the development of an integrated social science infrastructure. For new data, metadata needs to be captured at the point of data collection. We need to build modular online archives of metadata at the study level, file level and data object level. This work will need a unified effort based upon shared standards and goals and grounded on the principles of context, interoperability and preservation.

The Data Documentation Initiative (DDI) addresses these needs by promoting a non-proprietary neutral standard. It is committed to both context and format standards based upon standards in use and in development among data archives and data producers. It is robust and detailed, yet modular and flexible. It meets the requirements of preservation and allows for reconstruction of the full documentation. By acting as an interchange format, it permits the transfer of particular contents to an unlimited array of systems and architectures. Finally, the DDI offers interoperability and compatibility with resource discovery and cataloging systems (Dublin/Warwick Core, MARC, etc.). Developed under the administration of the Inter-University Consortium for Political and Social Research (ICPSR), the DDI reflects the interests and needs of representatives from an international group of archives, data producers, researchers and software developers. All stakeholders are committed to a neutral non-proprietary standard. For additional details on the exact structure and history of the DDI, see the DDI main web site at the Inter-university Consortium for Political and Social Research: <<http://www.icpsr.umich.edu/DDI/>>.

In order to build a digital metadata web based upon the DDI, we must make it easy and

inexpensive to produce standard metadata. The beta-tests showed that the overall structure of the DDI was sound although additional work was needed in some areas to improve the computer readability through some more structured attributes. In addition, better explanations and examples of how and when to use specific element tags or attributes were needed for consistent application. Other comments focused on the need for specific tools to facilitate the broad application of the DDI. This means that the process from data collection to metadata production needs to be streamlined through the provision of specific use authoring tools and software converters. Older documentation currently available only in print needs to be scanned and tagged. Data dictionaries in proprietary formats like SPSS, SAS or OSIRIS need to be converted to DDI compliant tagged documents. All of these documents should integrate library cataloging information and link to the PDF image or other appropriate material. All of this needs to be done in order to make it easy to integrate DDI metadata into systems for resource discovery and statistical analysis.

The value of the DDI is best shown through the web based discovery tools that use or are considering using the DDI Data Type Definition (DTD). This session included a presentation of the beta version of NESSTAR, the Networked Social Science Tools and Resources project of the UK Data Archive, the Norwegian Social Science Data Archive, and the Danish Data Archive <<http://www.nesstar.org/>>. This discovery and analysis tool is driven by the DDI structure and is being developed to access data at a variety of archival sites in Europe and North America. The use of structured metadata means that the user can effectively search through multiple archives at levels below the bibliographic information currently available in other systems, discover what data are available and extract or analyze the data online.

Of significant interest to the APDU community is the involvement of the Federal Electronic Research and Review Extraction Tool (FERRET), an analysis and extraction tool of the US Bureau of the Census in collaboration with the Bureau of Labor Statistics and other statistical agencies <<http://ferret.bls.census.gov/>>. Additional sites of interest include: Survey Documentation and Analysis (SDA), Computer Survey Methods Program (SCM), University of California at Berkeley <<http://csa.berkeley.edu>> and Virtual Data Center Project (VDC), Harvard-MIT Data Center <<http://thedata.org>>.

The Vision of Integrated Access to Statistics: the Data Web

CAVAN CAPPS
ANN GREEN
MARK WALLACE

Metadata experts are developing concepts and systems towards a dynamic, web-based linkage between data users and data sources and producers. References to such a “data web” were made at the APDU99 presentation on the Census Bureau’s integrated system and at the panel session on metadata. To provide an update for *Of Significance...* readers, the following paper is included in this issue. APDU is pleased to publish the authors’ latest vision of a Data Web. Cavan Capps is the Project Manager of the Federal Electronic Research and Review Extraction Tool (FERRET) data access system in the Demographic Surveys Division of the US Bureau of the Census. He is currently working with a team at the Centers for Disease Control to explore development of preliminary components of the Data Web. Ann Green is the Director of the Social Science Statistical Laboratory at Yale University and a long-term participant in the Data Documentation Initiative. Mark Wallace directs the Integrated Statistical Solutions project at the US Bureau of the Census and heads a FedStats task force for developing and implementing integrated data product concepts via the Data Web.

Introduction

Traditional modes of distributing statistical data to users (flat or hierarchical files and accompanying documentation) are no longer central to the distribution plan of most data suppliers. Instead, we see a proliferation of interactive web sites offering the ability to produce custom tables, maps, graphs, and subsets of data for downloading. To date, however, most Internet based tools offer only crude forms of access to data collections, and users continue to interact with single surveys or single compilations one at a time. For example, to locate age-race-sex tabulations by county or census tract, users must choose among over a dozen demographic sites offering different methods of drilling through varying amounts of supporting descriptive text—all pulling tabular data from the same summary tape file source. The sites do not allow linkages of data across data sets, nor do they provide full descriptive information regarding the context, utility, and dynamics of the data objects. But, more dynamic systems are in development. What users need is a Data Web, where heterogeneous data sources from multiple databases and from a wide variety of data producers are linked across the web. The Data Web, however, depends

Cavan Capps, Demographic Surveys Division, Bureau of the Census; Ann Green, Director, Social Science Statistical Laboratory, Yale University; Mark Wallace, Integrated Statistical Solutions Project, Bureau of the Census

upon a new level of commitment to the intelligent data object as we move toward a highly distributed, semantic, and machine-oriented data object + metadata model.

What kinds of integration would be possible if data objects from distributed data sites were linked via a Data Web? If the data objects were intelligent, it would be possible to have them interact with other data objects in distributed databases. The user would not need to know how or where the data are physically organized or stored. Users would develop queries based upon the meaning of the data, not the data base structure of the data, or the data base language. Users would be able to query complex hierarchical data sets, reoccurring data sets, complex supplements and topical modules, or longitudinal data sets without ever thinking about how the data are stored. The user should think only about how the data are logically or semantically organized.

The Data Web will become a reality only if data resources and documentation are created with this distributed yet integrated model in mind. An integrated web of data is impossible without metadata, the information that describes the characteristics, values, and rules of data objects. Federal, local, academic, and private data producers should be encouraged to take part in the effort to produce high quality standard metadata. However, conflicting standards are inhibiting the ability to move forward in the creation of metadata content

for historic collections or for systems currently in development. It is important that emerging standards take the concerns of the Data Web into consideration or the goals of linked statistical access and analysis will be further delayed. The stakeholders (producers, analysts, researchers, students, archivists, etc.) need to be involved in the design, implementation, evolution and production of metadata content so that their needs and contributions fit the requirements of integration and interoperability.

The key to the data object + metadata model is the linking of a value to its context. An examination of the history of data access tells us much about what users need to understand about data to exploit their utility. The normal process started with the potential user locating data resources of interest and examining some form of codebook (usually in print) that described the variable content and layout of a file. Users had to glean information from the codebook to understand how their programs needed to be written, to make their way into the file, and to struggle with programming to get some results. At the data object level, there was not much information recorded in the files with the data values; it was stored in the printed codebooks. Only a limited amount of information about specific data objects made its way into machine-readable form. As a consequence, the way that data objects (e.g. variables) are described has grown out of an extremely simplistic descriptive model defined by minimal labeling in statistical packages. The lack of complete metadata at the data object level is ongoing and is becoming critical.

The Great Promise of the Digital Object/Data Web Model

Imagine you are a researcher, a policy maker, a business person, or a middle or high school or a college student, and you want to know what data are available online about a particular subject. You could get on the web, pull up a data browser, and search for all the data bases known to the web with data on that topic. Not only do your findings include large national and international data sets maintained by separate government agencies and data archives, but they also include distributed data bases maintained by individual schools, libraries, universities, townships, cities and port authorities. The data would include the latest data available as well as historical data. But most importantly, the data objects would describe themselves in ways that explain how to use them and how they compare to other data sets maintained elsewhere. Once the data

have been selected, you could use web-based tools to tabulate and visually explore the integrated data in standard yet dynamic ways. You would not need to remember or look up formulas for complex weighting and estimating rates because the data would know how to produce the weighting and rate estimation even if data from different data sets were required.

Once the data are displayed in an Internet viewer such as a spreadsheet, graph or map, you could click on the data to describe itself and to explore other ways of manipulating and displaying the information. For example, you might want to click on a cell in a table and plot the time series for the data element in that cell. This would enable you to view the data over the last ten years to see, for example, if what it is measuring is above or below normal. You might want to then see if other time series data are related to the time series that you just created. You might want to plot the seasonally adjusted series, the inflation adjustment series, or other related series or components of the leading indicators. Not only might you want to explore time series, but also you might want to integrate summary data for various geographic units. You might want to create a table by using sex and race data as row variables and data from both a national data set and a state data set as column variables. Although transparent to you, in each of these cases the data might be created and maintained by separate government agencies or data archives located in different parts of the world. You will be able to tabulate data from different data bases simultaneously in ways that are meaningful to you in real time.

Taking this ideal scenario further, imagine being able to access a web site and quickly and easily get relevant and useful information on whether it makes sense to relocate your business and family to a suburb of New York City. You could request data on important variables of your selection (such as retail sales per capita; restaurant sales as a percentage of disposable household income; various age, race, and gender characteristics; cost of living and crime standards; and recent student-to-teacher ratios and test scores in public schools). You could make educated location decisions by comparing these data to data from other locations nationwide. Or perhaps you might want to create a map or chart with other available data sets to dynamically ascertain other integrated statistical solutions such as: How is my community doing? How is my family doing? How is my business doing? What sales targets should I set

for this county? Are health indicators in my county comparable to those for other areas of the country with similar demographics and incomes? Where are the best places in the nation to retire, given my circumstances? How can my metropolitan area optimally respond to a hurricane or earthquake?

Existing web-based tools, though successful in providing comprehensive access to individual data sets, are not able to provide such a problem- and solution-oriented view of the data. Providing such “integrated statistical solutions” will require extending current capabilities beyond the provision of individualized streams of products and static data pages to interactive tools within an integrated data and metadata environment. In short, there is a need to develop a Data Web to power the provision of integrated statistical solutions to users. This could be achieved through the dynamic integration rules that would be available to the distributed data sources of the Data Web. Providing such integrated statistical solutions from the Data Web is the mission of the growing “Integrated Statistical Solutions” project initiated at the U.S. Census Bureau and now supported by several federal and state agencies.

What kinds of manipulation would be possible if data objects were linked with their matching metadata? Intelligent data objects could change the way that users interact with views of data in tables, graphs and maps. Instead of dumping numbers into stand-alone spreadsheets, tables or maps—or into poorly documented files for downloading—the data in all views would remain tied together with, or in communication with, the underlying use-oriented metadata. The data representation in all data views would have all of the self-knowledge to suggest appropriate behavior, appropriate inferences, and relationships, as well as to explain the origin of that data object. Users would have the ability to reconfigure data on the fly, flip matrices, combine hierarchical records into customized configurations, generate tabular output from multiple microdata files at once, and link data from over a dozen agencies and archives at a time. Users could interact directly with the data instead of manipulating the controls of a program to manipulate numbers, and would be able to launch other tabular, graphic or mapping views. Moving views into analysis packages of the user’s choice would be streamlined as the intelligent object would carry with it the descriptive and behavioral information needed for modeling, spatial analysis, etc. The constraints of the past in terms of data

management and manipulation would no longer be relevant.

New web tools could be created by expert intermediaries to help users navigate through complex data to answer a wide range of typical questions. A “hot web report writer” could be developed to allow non-programmer experts to develop web pages that pull the latest data from different data sites, combine it and display it to answer typical user questions. In this way, less experienced users could be informed by the data, and data literacy could be furthered. Finally, data in the system should have access to web-based geographical information services. Without changing anything, you should be able to find data using a map as a searching device.

Low Cost Distributed Architecture

The survival of the Data Web will depend upon increasing capacity as content and users increase, maintaining reliability of the overall system, and at the same time not increasing costs. The distributed nature of the World Wide Web, which now focuses upon documents, provides a blueprint for constructing a similar World Wide Web for data using the same techniques of distributed hardware, software, content and responsibility. This principle, combined with the ability to inexpensively mirror services throughout the web, will insure that capacity reliably increases as the content and users increase. There must be inexpensive software that allows organizations to publish data quickly and inexpensively on the Data Web. The architecture must support collaboration among data users, data suppliers, and the development community, and be organized for high performance in response to high volume and continuous queries.

The Data Web will depend upon distributed data dissemination metadata engines which will contain use-oriented intelligent data object descriptions about the relationships among data, data use rules, and semantic rules for query construction and manipulation. The metadata will provide the ability to evolve data object properties to propagate appropriate attributes to perhaps hundreds of low-cost ‘distributed dissemination metadata repositories’ (DDMR’s). User-oriented metadata requirements will be developed interactively by behavioral psychologists, data archivists, and other data mediators working with collaborating developers.

Geographical Information Systems (GIS), distributed to ensure capacity and to avoid single points of failure, will allow geographic queries to

drive resource discovery and intelligent reports. A range of services, including multi-layer thematic maps and star-field finders as well as the ability to map multiple points of time and roll up geographies will be developed.

Wrappers, small modules installed on distributed data servers, can allow very different servers and databases to look and act the same over the web. An intelligent database gateway, composed of a wrapper driven by metadata, can be developed to allow queries to access data from multiple databases over the Internet. As a result, all data servers described to the Data Web will have the same ability to perform complex tabulation, to link together aggregate data, to do intelligent graphics and more. This gateway can be constructed by using the common Structured Query Language (SQL) as a standard interface to all commercial and specialized databases. The wrapper allows the database to self-describe the physical layout of the data and allows the local database administrator to organize her database differently for local needs and still present a consistent view of the database to the Data Web network. Intelligent activities such as recoding variables, filtering for confidentiality, and verifying statistical reliability can be available through this wrapper to all databases.

Building the Data Web

Such a Data Web will shift the way that data users currently think about statistics, providing new avenues for data visualization, data interpretation and exploration between the relationships implicit in data. For this new age of data access and statistical literacy to be realized, it is vital that institutional collaboration be developed and supported. Competition among institutions tends to increase the proliferation of disconnected stovepipe sites and makes data discovery and data integration difficult if not impossible. The lack of collaboration and integration among various efforts is also very expensive. The tendency toward centralization can create inefficiencies. Centralized computers do not evolve as quickly in terms of growing, matching capacity and taking advantage of new, cheaper technology. Putting data on a collection of computers allows them to be upgraded continuously; data dissemination should be a natural outgrowth of data collection.

One of the most significant roadblocks to creating a Data Web is the difficulty and cost that data producers face in making their data available to the web. Currently, a myriad of proprietary systems

is being rushed into production and none of them are easily connected to any other system. A chaotic array of expensive stand-alone data access systems is the result. Because all standard databases and hardware platforms could be supported in the Data Web, the Data Web would exploit the technology that is currently in use so that additional training costs and system reconfigurations are not incurred. Because labor is the most significant cost that an organization faces when making data available to the web, data suppliers should not have to learn complex new technology. Low cost hardware/software solutions should be explored and promoted, and additional Data Web software should initially be free to data providers to encourage participation.

Another significant roadblock is the severe lack of quality metadata. Publishing data on the Data Web requires new commitments to the production of metadata in standard formats. However, creating metadata is expensive and the benefits of metadata resources often are not understood by the data provider. Even when available, metadata are obtuse and very hard to use. Metadata are often written by government statisticians for other government statisticians and do not promote a broader use of the data. A minimum set of metadata should be recommended with clear authoring guidelines that show a provider how the metadata will be used in the Data Web. Moreover, tools must be provided to data providers and users to assist them in generating metadata, especially the aforementioned dynamic integration rules and user-oriented metadata. It should be made clear that standardized, complete metadata obviously make data more usable, and that by adding more metadata, the functionality of the data and the potential for reuse of data resources increases. Corporate metadata should indeed be centralized and optimized to support and streamline survey collection and production processing. However, metadata for data access in a large-scale system like a Data Web should, concomitantly, be widely and inexpensively mirrored or replicated. A single point of failure is unacceptable, and for a single machine that will perform well with all web users accessing it, the cost would be high to purchase and maintain. The software for these distributed mirrored systems should be very inexpensive or free and provide automated synchronization, replication and load balancing that will support thousands of simultaneous users accessing the thousands of data sources available in a true Data Web.

Current Efforts

The depiction of a Data Web that provides seamless access to data in the same way that the current web provides access to documents seems a dream. However, a careful examination reveals that many of the pilot efforts to test the concepts are in place or are being explored.

Web tools: Development of a generalized data browsing tool for the Data Web is important. FERRET (Federal Electronic Research and Review Extraction Tool) is a tool that is designed to browse metadata and to support complex tabulation and intelligent data visualization from all distributed data sets known to the Data Web. Building on FERRET technology, more user-friendly data profiles, or "Hot Reports," could be authored using the aforementioned hot web report writer and made available to users on selected geographic areas and/or topics of interest. Moreover, current interagency collaborative efforts are laying the groundwork for the development of productive synergies among existing dissemination systems. Specifically, activities are underway to leverage the Census Bureau's "American FactFinder" tool, the Centers for Disease Control "CDC Wonder," FedStat's "MapStats" tool, and a next generation of FERRET to produce newly developed "Hot Reports" from the Data Web that could augment all agencies' data offerings with the clear beneficiaries being data users. Equally important, steps are being taken by a number of agencies that have received grants under the auspices of the National Science Foundation's Digital Government Initiative to take applied research success stories into an actual production environment that would advance the construction of the Data Web.

Metadata: A considerable amount of work is being accomplished within the national and international standards committees to reach consensus on standardizing metadata and registries for organizing metadata. Several standards currently exist with regard to establishing and developing terminology, thesauri, and data elements. An important standard with regard to data elements is International Organization for Standardization / International Electrotechnical Commission (ISO/IEC) 11179, which includes the specification and standardization of Data Elements into six parts, including a conceptual model for a metadata registry for data elements. Several agencies worldwide have successfully implemented profiles of the ISO/IEC 11179 standard, selecting appropriate attributes for particular metadata tracking communities. For

example, the U.S. Census Bureau's American FactFinder dissemination tool draws metadata from the Census Bureau's "Corporate Metadata Repository," which is based on the ISO /IEC 11179 standard. Another important emerging standard was developed by the Data Documentation Initiative (DDI) Committee established in 1995 by the Inter-University Consortium for Political and Social Research. Version 1 of the DDI Document Type Definition in XML, to be released in early 2000, provides an extremely rich and flexible set of elements for developing layered presentations of metadata to users. The Data Documentation Initiative <<http://www.icpsr.umich.edu/DDI/>> supports a convergence of archive, analysis, discovery, and production requirements and promotes a non-proprietary neutral standard. The DDI provides highly structured and detailed content and format standards for complete study information, from elements describing the focus and methodology of a study down to specific data objects in files or databases. Most notably here, the DDI was designed to be flexible enough to operate with other existing standards and to be integrated into existing practices. For example, at the U.S. Census Bureau, it is important that DDI metadata can coexist and be easily exchanged with metadata residing in the Corporate Metadata Repository. With the establishment of such relationships in other agencies as well, the DDI holds great promise as an interchange format to meet the requirements of metadata transport, permitting transfer of particular contents to an unlimited array of systems and architectures. Finally, it offers interoperability and compatibility with resource discovery and cataloging systems. The producers of the standard are stakeholders from social science data archives and libraries in the USA, Canada, and Europe (including the national archives in Denmark, Germany, Norway, Sweden, and the United Kingdom) as well as from major statistical data producing organizations (e.g. the U.S. Bureau of the Census, the U.S. Bureau of Labor Statistics, Statistics Canada, and Health Canada). An example of an existing data access system in Europe using the DDI as its core metadata is the European Networked Social Science Tools and Resources Project (NESSTAR) <<http://www.nesstar.org/>>.

Architecture: Collaborative multi-agency code development and maintenance is a critical success factor. Developers from the U.S. Census Bureau and the Centers for Disease Control (CDC) are working together to design, develop and maintain the elements of a Data Web system. The effort also

should allow health data maintained at the CDC to be integrated with demographic, housing, education, crime, economic, and other data from the U.S. Census Bureau and other agencies. Rates and other data can be created in real time using the latest data from both systems. Test programs are currently accessing data from both sites transparently, using heterogeneous databases and different computer hardware systems. Further, given the timely provision of adequate resources from various data providers connected with the Data Web, there is an exciting possibility that data from Census 2000 could be released in an enhanced dissemination mode on the Data Web. That is, in disseminating Census 2000 data, there could be a collaborative effort to combine the strengths of dissemination tools, such as the American FactFinder, FERRET, and MapStats, to integrate these data with data from various other federal, state, and local agencies. Coordinating and implementing these activities in ways that can “bring the data home” to data users is, in fact, the essence of the aforementioned Integrated Statistical Solutions project being supported by several federal and state agencies.

Software development: An Internet development and technology testing lab is being established to support remote collaborative software development and research. The site will support the inter-agency development efforts and potential

integration of National Science Foundation (NSF) research efforts. The machines in the lab may be available to remote developers and researchers who are working on different elements or services potentially useful in a Data Web infrastructure.

Conclusion

The depiction of a Data Web that provides seamless access to data in the same way that the current web provides access to documents seems a social science and health science dream. However, a careful examination reveals that many of the pilot efforts to test the concepts are in place or are being explored by various federal, state, and local agencies with the Integrated Statistical Solutions project, the FedStats task force, and the academic and applied research communities under the Digital Government Initiative. Such a Data Web will shift the way that we currently think about statistics, providing new avenues for data visualization, data interpretation and exploration between the relationships implicit in data. For this new age of data access and statistical literacy to be realized, we must expand recently successful inroads made within the institutional collaborative arena. And throughout our work, the effort of data access by data users must be simplified, meaning that a commitment to “the ease of use, and the ease of data publishing” is critical.

Electronic Products for Census 2000

PATRICIA C. BECKER
JOSEPH J. SALVO

The conference included an update from the APDU Working Group on Census Products. The Group co-chair and coordinator led this session at which they described their goals and efforts to date. They were joined by Louisa Miller, a Census Bureau representative on the Working Group Steering Committee. She provided a brief overview of product plans for the 2000 Census. Those interested in the activities of the Working Group are encouraged to read the following summary and to contact the Association.

The APDU Working Group on Census Products was funded by the Census Bureau for the purpose of soliciting input from expert users regarding the content, form, and delivery media of 2000 Census products. This overview report covers the period from February through September, 1999.

APDU appointed members Joseph Salvo, Department of Planning, New York City, and Edward Spar, Executive Director, Council of Professional Associations of Federal Statistics (COPAFS), as chairs of the Working Group. APDU member Patricia Becker, APB Associates and Southeast Michigan Census Council, was asked to serve as its coordinator. Formal liaison with the Census Bureau was handled by Stephen Dienstfrey, APDU Past President. The leadership's first task was to solicit the cooperation of a group of expert census data users, representing a wide variety of user types and groups. Particular attention was paid to representing the following user segments:

- State Data Centers
- Universities
- Non-profit sector
- Private for-profit sector
- National data users
- Librarians

The concept of "expert user," in general, is defined to mean the following:

- Working with census data for at least two census periods (i.e., 1980 and 1990; several experts have experience back to 1970 or 1960.)

Patricia C. Becker, APB Associates/Southeast Michigan Census Council; Joseph J. Salvo, Department of City Planning, New York City

- Using census data is a significant part of his or her professional work, and
- Serving as an intermediary, helping others with less expertise to gain access to the data.

A group of eighteen census use experts was assembled who, along with the co-chairs and coordinator, would respond to the assignments. The work was organized in terms of assignments. Each assignment was developed jointly by the Census Bureau and the APDU chairs/coordinator, collectively referred to as the Steering Committee. Census Bureau representatives included: Louisa Miller, Sherry Pollock, Campbell Gibson, and occasionally others. After the assignment was written by the project coordinator and reviewed by the others, it was distributed to the Working Group co-chairs, coordinator, and expert census data users. Appropriate Census Bureau documents were distributed at the same time. The recipients were given a deadline for review and response. The project coordinator assembled the responses from the reviewers and prepared an overall response to the Census Bureau.

APDU is in the process of negotiating a new contract for the next phase of the Group's work. New reviewers will be added, and the first set of new assignments will focus on the structure, form and content of the 2000 Census Summary files. Anyone who would like more information about the APDU Working Group may contact the co-chairs, Joe Salvo by e-mail at: <jsalvo@dcplan.ci.nyc.ny.us> or Ed Spar at <copafs@aol.com>.

Copies of the Working Group Final Report are available from the APDU Headquarters.

The Overlooked Undercount: Children Missed in the Decennial Census

WILLIAM P. O'HARE

A project of the Annie E. Casey Foundation, KIDS COUNT monitors the status of children in the United States in order to provide benchmarks of their well-being. William P. O'Hare is coordinator of the Foundation's annual KIDS COUNT project and was invited to APDU99 to share the results of his work. His latest effort focused on the undercount of children in the 1990 Census and on concerns that the media and policy makers overlook them when investigating the undercount.

(Summarized by Janie Harris, Cornell University Library)

Many Americans know that the massive head count undertaken in a decennial census misses some individuals. However, we are surprised to learn that children are missed more often than any other age group. William O'Hare reported that more than two million children were missed in the 1990 Census and accounted for more than half the total net undercounted population. The net undercount rate for children was twice the undercount rate for the total population. Younger children were more likely to be missed in the census than teenagers, and minority non-Hispanic children were missed more often than adults. Children in the South and West were undercounted at twice the rate as those in the Northeast and Midwest. Children living in areas of high poverty also suffered a higher rate of undercount.

O'Hare shared some of his theories on the causes of the undercount of children. He suggested that children living in foster care or with a grandparent were at risk of not being listed on a census questionnaire because the respondents

William P. O'Hare, KIDS COUNT, Annie E. Casey Foundation

might not consider the children as officially in their households. Children in minority households were also at risk of not being counted because of the reasons for non-response among minorities. He underscored the link between being missed in the census and living in poverty, as many children do.

The implications of undercounting children involve their social, physical and educational well-being. Public funds are not adequately allocated to programs serving children in need who were not counted. Private sector and nonprofit organizations underestimate local needs for support in areas suffering an undercount. School planners lack benchmarks for determining educational building and staffing demands in areas where children are undercounted.

Unfortunately, O'Hare anticipates that the 2000 Census will show an undercount of children, possibly at a higher rate than 1990. Conditions affecting the 1990 undercount will still exist because of trends in the social and demographic environment. His intention is to document and publicize the undercount of children and to encourage interest in solving the problem. Those concerned with the "Overlooked Undercount" can obtain his report with that title from the Annie E. Casey Foundation at <<http://www.aecf.org/kidscount/>>.

Report on the Panel to Review the 2000 Census: Issues and Objectives

JANET L. NORWOOD

Presentations by Janet Norwood are among the most valued sessions at the annual APDU conferences. Formerly as Commissioner of the Bureau of Labor Statistics, then as a Senior Fellow at the Urban Institute, and currently as Chair of the Panel to Review the 2000 Census, she always provides insightful commentary on the state of public data. Her book, *Organizing to Count: Change in the Federal Statistical System*, is a primer on the issues. Janet Norwood's efforts towards the improvement of data collection, quality and policy, have earned bipartisan support. The findings and recommendations of the Panel she is leading will be widely read and considered.

(Summarized by Joseph J. Salvo, Department of City Planning, New York City)

The Panel to Review the 2000 Census was created by the Committee on National Statistics (CNSTAT) of the National Research Council to evaluate all aspects of the 2000 Census. It is a multi-disciplinary group that is charged with taking an objective, nonpartisan, and scientific look at various aspects of the next census, including:

- coverage issues, especially the differential undercount by race
- operations, including data collection, capture and processing; hiring; and supervision
- address list issues
- advertising and promotional effectiveness
- quality of the data collected and products produced

The Panel has elicited a high level of cooperation from the Census Bureau, in their quest to obtain data. This is a big point of emphasis in the initial stages of the Panel's existence. Meetings have already been conducted on operations and sampling. Panel members have reviewed operational plans, address list work, changes in

the race question, and the Accuracy and Coverage Evaluation (ACE) survey. The ACE survey is the primary means of adjusting census counts that will be used for purposes other than re-apportionment. Labeled as a priority by Panel members, ACE sample size and stratification were the subject of a Letter Report issued in May of 1999 by the Panel. The group concluded that the sample size decision was adequate for the purposes specified by the Bureau.

A copy of the Letter Report can be accessed from the home page of the Committee on National Statistics at <http://www4.nas.edu/cbsse/cnstat.nsf/>.

Panel members are attempting to get a view of the census operation "on the ground." They talked with people collecting data in the field, at the Sacramento, California, and Columbia, South Carolina Census Dress Rehearsal sites. Trips were also made to the Census Bureau's data capture centers in Baltimore, Maryland and Jeffersonville, Indiana. More trips are planned for the spring of 2000 to observe census operations.

The Panel has raised issues and asked general questions, but has avoided making specific recommendations at this juncture. The focus has been on determining what the Panel will need to do its work. Although interim reports are likely, right now, there is no timetable for their delivery.

*Janet L. Norwood, Chair, Panel to Review
the 2000 Census*

Acknowledgments

JANIE HARRIS

The interactive dynamics of a conference are difficult to communicate through the published proceedings. The annual meeting of the Association of Public Data Users presents a particular challenge because APDU gathers together professionals whose specific interests and expertise in public data encompass a wide range. The APDU conference allows them to share their diverse expertise and perspectives and to coalesce around issues of common concern.

The program sessions at APDU99 represented the variety of professional interests of the Association members and participants as well as their shared concerns about public data. The contributors to this issue have captured the wide range of session content and also have identified the common themes. As presenters or summarizers, they have compiled informative and interesting reports for which they deserve acknowledgment. Also deserving recognition are those who helped in the editorial process for this issue, in particular Sue

Janie Harris, Guest Editor, Of Significance... 1:2

Hoover as Chair of the APDU Publications/Editorial Committee, Wendy Treadwell as APDU Board President, and Deirdre Gaquin as General Editor for *Of Significance...* Teresa Hall Allen at APDU Administrative Headquarters deserves special acknowledgment as Production Editor. A long-time institutional member of APDU, Cornell University and its Library continue to support Association activities—most recently, the guest editorship of this issue.

Participation in APDU99 indicated the speaker's recognized expertise on a specific public data issue as well as his or her support for others concerned about "Responsible Data Use." APDU is a national network, and to facilitate further contact among the speakers, conference attendees, Association members, and *Of Significance...* readers, there is appended a list of the speakers with their affiliations and addresses. This opportunity to network can enhance the dynamic experience provided to those who participated in APDU99 and can draw others into their active communities.

Appendix 1: List of Speakers, APDU99

Albert Anderson Public Data Queries, Inc. 2901 Hubbard Ann Arbor, MI 48105 Phone 734-213-4964 Fax 734-475-8168 EMail: afa@pdq.com URL: http://www.pdq.com	Sarah Breshears State Data Center Institute for Economic Advancement University of Arkansas, Little Rock 2801 South University Little Rock, AR 72204 Phone 501-569-8530 Fax 501-569-8538 EMail: sgbreshears@ualr.edu URL: http://www.aiea.ualr.edu/csdc/default.html	Patrick Collins California Census Research Data Center University of California Berkeley 2538 Channing Way #5100 Berkeley, CA 94720-5100 Phone 510-642-3149 Fax 510-643-8292 EMail: pcollins@ccrdc.ucla.edu URL: http://www.ccrdc.ucla.edu
Fred T. Asbell Census Monitoring Board Bureau of the Census P. O. Box 610 4700 Silver Hill Road Suitland, MD 20752 Phone 301-457-5080 Fax 301-457-5081 Email: asbel@cmbc.gov	Lisa Broniszewski Population Research Institute Pennsylvania State University 601 Oswald Tower University Park, PA 16802 Phone 814-863-8322 Fax 814-863-8342 EMail: lisa@pop.psu.edu URL: http://www.pop.psu.edu/data-archive/archive.html	Anne E. Dunthorn C F Systems 908 West Outer Drive Oak Ridge, TN 37830 Phone 423-483-6296 Fax 423-483-4316 Email: dunthorn@usit.net
Gary Bass OMB Watch 1724 Connecticut Avenue, NW Washington, D.C. 20009 Phone 202-234-8493 Fax 202-234-8584 EMail: gbass@ombwatch.org	Jacqueline Byers National Association of Counties 440 First Street, N.W., Suite 8000 Washington, D.C. 20001 Phone 202-393-2630 Fax 202-393-2630 URL: http://naco.org	Jorge del Pinal Population Division U.S. Bureau of the Census 4700 Silver Hill Road, Rm. 20011-3 Washington, DC 20233-8850 Phone 301-457-4875 Fax 301-457-2644 EMail: Jorge.h.delPinal@census.gov URL: http://www.census.gov
Patty Becker APB Associates 28300 Franklin Road Southfield, MI 48034 Phone 248-354-6520 Fax 248-354-6645 EMail: pbecker@umich.edu	Jeffrey Capizzano Urban Institute 2100 M Street, NW Washington, DC 20037 Phone 202-261-5270 Fax 202-293-1918 EMail: jcapizza@ui.urban.org URL: http://ui.urban.org	Jim Fitzsimmons U. S. Bureau of the Census Population Division 4700 Silver Hill Road Washington, DC 20233-8860 Phone 301-457-2419 EMail: jfitzsim@census.gov URL: http://www.census.gov
Colleen Blessing Energy Information Administration US Department of Energy 1000 Independence Avenue, SW Room 1E226, EI30 Washington, DC 20585 Phone: 202-586-6482 EMail: colleen.blessing@eia.doe.gov URL: http://www.eia.doe.gov		Mark Frankel American Association for the Advancement of Science 1200 New York Avenue, NW Washington, DC 20005 Phone 202-326-6400 EMail: mfrankel@aaas.org

Deborah Gona
Gona and Associates
1528 Green Hills Road
Lexington, KY 40505
Phone 606-299-0704
Fax 606-299-0704
Email: DGONA@prodigy.net

Ann Green
Social Science Statistics Lab
Yale University
P.O. Box 208208
New Haven, CT 06520-8208
Phone 203-432-3277
Fax 203-432-6976
EMail: ann.green@yale.edu
URL: <http://statlab.yale.edu>

Peter Joftis
ICPSR
311 Maynard, No. 81
Ann Arbor, MI 48104
Phone 734-998-9915
Fax 734-998-9905
EMail: pmj@icpsr.umich.edu
URL: <http://www.icpsr.umich.edu>

George Leventhal
Association of American
Universities (AAU)
1200 New York Avenue, NW
Washington, DC 20005
Phone 202-408-7500
EMail: george_leventhal@aau.edu

Jerry McFaul
SIGCAT Foundation
11343 Sunset Hills Road
Reston, VA 20190
Phone 703 435-5200
Fax 703-435-5553
EMail: mssigcat@aol.com
URL: <http://www.sigcat.org>

David McMillen
U.S. House of Representatives
Committee on Government Reform
511 Ford Office Building
Washington, DC 20515
Phone 202-225-5420
Fax 202-226-2508
EMail:
david.mcmillen@mail.house.gov
URL: <http://www.house.gov/reform>

Louisa Miller
U. S. Bureau of the Census
Population Division, 2021-3
4700 Silver Hill Road
Washington, DC 20233
Phone 301-457-2073
Email: lmiller@census.gov
URL: <http://www.census.gov>

Brand Niemann
Environmental Protection
Agency
401 M Street, SW
Washington, D.C. 20460-0003
URL: <http://www.epa.gov>
URL: <http://www.usaondvd.com>

Janet Norwood
National Academy of Sciences
2101 Constitution Avenue, N.W.
Washington, D.C. 20418
URL: <http://www.nas.edu>

William O'Hare
Annie E. Casey Foundation
701 St. Paul Street
Baltimore, MD 21202
Phone 410-547-6600
Fax 410-547-6624
URL: <http://www.aecf.org>

Margeret Plantz
Outcome Measurement
United Way
701 N. Fairfax Street
Alexandria, VA 22314
Phone 703-836-7100
EMail:
meg.plantz@uwa.unitedway.org
URL: <http://www.unitedway.org/outcomes>

Ken Prewitt
U. S. Bureau of the Census
4700 Silver Hill Road, Rm. 2049
Federal Building #3
Washington, DC 20233
Phone 301-457-2135
EMail:
Kenneth.Prewitt@census.gov
URL: <http://www.census.gov>

Bryan Rickard
Unicon Research Corporation
1640 Fifth Street, Suite 100
Santa Monica, CA 90401
Phone 310-393-4636
Fax 310-393-2503
EMail: bwr@unicon.com
URL: <http://www.unicon.com>

Judith Rowe
Retired, Princeton University
301 Western Way
Princeton, NJ 08540
Phone 609-921-7432
EMail: Judith@princeton.edu

Joe Salvo
Population Division
Department of City Planning
22 Reade Street
New York, NY 10007
Phone 212-720-3434
EMail:
jsalvo@dcplan.nycnet.ci.nyc.us

Milo Schield
Business Administration and
Management Information
Systems
Augsburg College
721 – 21st Avenue South
Minneapolis, MN 55454
Phone 612-330-1000
URL: <http://www.augsburg.edu/depts/faculty/schild.htm>

Ed Spar
Council on Professional and
Federal Statistics, COPAFS
1429 Duke Street, Suite 402
Alexandria, VA 22314-3415
Phone 703-836-0404
URL: <http://members.aol.com/COPAFS/>

Christopher Strobel
U. S. Bureau of the Census
Census Monitoring Board
4700 Silver Hill Road
Suite 1250-3
Suitland, MD 20746
Phone 301-457-9900
URL: <http://www.cmbp.gov>

Jocelyn Tipton
Yale University
P.O. Box 208263
140 Prospect Street
New Haven, CT 06520-8263
Phone 203-432-3310
Fax 203-432-8979
EMail: jocelyn.tipton@yale.edu
URL: <http://www.library.yale.edu/socsci>

Nancy Torrieri
U. S. Bureau of the Census,
DSMD
4700 Silver Hill Road, Mailstop
8700
Suitland, MD 20746
Phone 301-457-3602
EMail: nancy.k.torrieri@ccmail.census.gov
URL: <http://www.census.gov>

Wendy Treadwell
Machine Readable Data Center
Wilson Library
University of Minnesota
Minneapolis, MN
Phone 612-624-4389
EMail:
wendy@mrdc.lib.umn.edu
URL: <http://www.lib.umn.edu/mrdc>

Clyde Tucker
Bureau of Labor Statistics
Department of Labor
2 Massachusetts Avenue, NE,
Rm. 4915
Washington, DC 20212
Phone 202-691-7371
EMail: tucker.clyde@bls.gov
URL: <http://www.bls.gov>

Chip Walker
U.S. House of Representatives
Committee on Government
Reform
114 O'Neill House Office
Building
Washington, DC 20515
Phone 202-226-1973
Fax 202-226-3436
EMail:
chipwalker@mail.house.gov
URL: <http://www.house.gov/reform>

Mark E. Wallace
U. S. Bureau of the Census
4700 Silver Hill Road, Rm. 2578-3
Washington, DC 20233-6100
Phone 301-457-2621
EMail: mwallace@census.gov

Katherine Wallman
Office of Management and
Budget
725 – 17th Street, NW, Rm. 10201
Washington, DC 20503
Phone 202-395-3631
EMail: kwallman@omb.eop.gov
URL: <http://www.whitehouse.gov/OMB>

Alyssa Wigton
Urban Institute
2100 M Street, NW 20037
Washington, DC 20037
Phone 202-261-5579
Fax 202-293-1918
EMail: awigton@ui.urban.org
URL: <http://www.urbaninstitute.org>

Alvan O. Zarate
National Center for Health
Statistics
6525 Belcrest Road, Room 1179
Hyattsville, MD 20782-2003
Phone 301-426-3238
Fax 301-436-3568
EMail: aolz1@cdc.gov
URL: <http://www.cdc.gov/nchswww/default.htm>

Marianne W. Zawitz
Bureau of Justice Statistics
U. S. Department of Justice
810 – 17th Street, NW
Washington, DC 20531
Phone 202-616-3499
EMail: zawitzm@ojp.usdoj.gov
URL: <http://www.ojp.usdoj.gov/bjs/>

**For information on the
conference program, contact:**

Lisa Neidert
EMail: lisan@umich.edu
Phone: (734) 998-7153
Stephen Dienstfrey
EMail: s.dienstfrey@srbi.com
Phone: (301) 608-3883