

The Eleventh Morris Hansen Lecture Privacy and Confidentiality – A New Era

*Katherine K. Wallman*¹

I am pleased to join this panel today as we examine the ever-growing challenges we face, and must address, as we balance competing interests in providing the broadest possible access to statistics against concerns about protecting data collected under pledges of individual or institutional confidentiality.

The tensions I want to underscore are not new. Almost a decade ago, when I had the honor to address the American Statistical Association as its president, I argued:

“Yet another key issue that demands our attention concerns what I believe is the thorniest – and potentially most consequential – issue facing our national statistical system. This is the growing tension between the need to protect the confidentiality of individual responses, and the desire to provide the broadest possible access to information. Government statistical offices derive their mandate for data collection and dissemination from a citizenry that demands at once both quality information to drive public policy, and protection of the individual respondent from privacy invasion and administrative harm. In response to these cross-pressures, statistical agencies are actively pursuing approaches such as encryption and licensing agreements that extend responsibility for confidentiality protection to the data user. Striking the proper balance between permitting access to accomplish compelling and legitimate research, and incurring the risk, however remote, of inadvertent revelation of individual information is, I believe, a fundamental concern and challenge for official statisticians.”

“Many researchers find access to government data increasingly desirable. The newer data bases are more comprehensive, of better quality, and – with improved data base management techniques – better structured. At the same time, the individuals and institutions that provide the data residing on government data bases – as well as the agencies that sponsor the collection of such information – are well aware that the same technologies that extend analytical capabilities also furnish the tools that threaten the confidentiality of data records. This awareness has the potential to erode – or at least to undermine – our respondents’ confidence in the confidentiality protection for the information they provide.”

That was just over a decade ago. I do not wish to imply that I was prescient in August of 1992 – even I could not have imagined how those words would come to frame so much of my current work.

Let me underscore, as I continue, that I am referring to *individually identifiable* data that are collected from individuals and organizations under a pledge of confidentiality. Such data are frequently, though not always, provided by respondents on a *voluntary* basis – and our continuing ability to gain their cooperation rests to a substantial degree on maintaining their *trust*.

¹ U.S. Office of Management and Budget, Washington, DC, U.S.A.

Absent that trust, and the cooperation it engenders, our statistical system would be hard pressed to deliver the Consumer Price Index, monthly unemployment numbers, measures of educational participation and attainment, indicators of health status and disease prevalence, and a literal host of other critical economic and social indicators to guide our policy and personal decisions.

Over the decades, our statistical agencies have had an outstanding record of maintaining the confidentiality of individually identifiable data – but that is not to say there have not been those who, for seemingly “sound” reasons, have sought to gain access to such information.

In some cases, agencies have had statutory authority protecting the confidentiality of the data they collect for statistical purposes. More often, the agencies’ “authority” has been vested in longstanding administrative policies – or in some combination of statutory and administrative provisions.

The results of a “survey” my office completed last summer of the 14 statistical agencies that comprise the Interagency Council on Statistical Policy – those that would be viewed as the “most statistical” of government agencies – revealed a patchwork of statutory and other protections for upholding confidentiality. From a statutory perspective, the essence of that patchwork is as follows:

- one agency, the Energy Information Administration, has legislation that actually *requires sharing* of confidential statistical information;
- two agencies, the Environmental Protection Agency and the Economic Research Service (in Agriculture), have *no* specific legislation protecting the confidentiality of their statistical information;
- four agencies, the U.S. Bureau of Labor Statistics, the U.S. Bureau of Transportation Statistics, the National Center for Education Statistics, and the Social Security Administration, have *partial* protection for confidential statistical information; and
- seven agencies, the U.S. Bureau of Economic Analysis, the U.S. Bureau of Justice Statistics, the U.S. Census Bureau, the Statistics of Income Division at the U.S. Internal Revenue Service (IRS), the National Agricultural Statistics Service, the National Center for Health Statistics, and the National Science Foundation have *broad* statutory authority to protect the confidentiality of information they collect.

In addition, agencies have derived some protection for statistical information through provisions of the Trade Secrets Act, the Privacy Act, and the Freedom of Information Act.

Concerns about the lack, or unevenness, of protections accorded the confidentiality of individually-identifiable U.S. Federal statistical data have a history that extends for more than 30 years. The roots of the policies in efforts to address these concerns – by every Administration over the past three decades – and in our current initiatives that I will elaborate at the close of my remarks – reflect the work of three Commissions that examined statistical and information issues during the Administrations of Presidents Nixon and Ford.

In 1971, the President’s Commission on Federal Statistics recommended that the term *confidential* should always mean that disclosure of data in a manner that would allow public identification of the respondent or would in any way be harmful to him or her should be prohibited. (This commission also recommended that consideration should be given to providing for interagency transfers of data where confidentiality could be protected.)

In addition, the Commission recommended that a promise to hold data in confidence should not be made unless the agency has legal authority to uphold such a promise, and that legislation should be enacted authorizing agencies collecting data for statistical purposes to promise confidentiality as the term was defined by the Commission.

In July 1977, the Privacy Protection Study Commission stated that “no record or information . . . collected or maintained for research or statistical purposes under Federal authority . . . may be used in individually identifiable form to make any decision or take any action directly affecting the individual to whom the record pertains . . .”

In October of that same year, the President’s Commission on Federal Paperwork endorsed the confidentiality and “functional separation” concepts, but applied them directly and simply to *statistical* programs, saying that:

- Information collected or maintained for *statistical purposes* must never be used for administrative or regulatory purposes or disclosed in identifiable form, except to another statistical agency with assurances that it will be used solely for statistical purposes; and
- Information collected for *administrative and regulatory purposes* must be made available for statistical use, with appropriate confidentiality and security safeguards, when assurances are given that the information will be used solely for statistical purposes.

The policy discussions generated by the three Commissions came together during the Carter Administration in a bipartisan outpouring of support for the Paperwork Reduction Act, which largely addressed the *efficiency* recommendations of the Paperwork Commission. The legislative history of that Act recognized the unfinished work of fitting the “functional separation” of statistical information into the overall scheme.

And so that “unfinished work” remained a challenge to the leaders of the Federal statistical system over the ensuing decades. There have been numerous attempts both to shore up legal protection for the confidentiality of statistical information, and to permit some limited sharing of data for statistical purposes. Today is not the forum to explore that history in detail. But one detail of that history is relevant – that in the efforts undertaken by the Carter, Reagan, Bush I, and Clinton administrations to address this challenge, strengthening and evening out statutory protections for confidentiality of individually identifiable data that are collected for statistical purposes has always been a goal. (And, indeed, some new statutes provided such protection for the National Center for Health Statistics, the National Agricultural Statistics Service, and the National Center for Education Statistics.)

During the past year, as we have pursued this challenge with renewed energy, we are facing a far more complex environment – “researcher” access may now seem a rather modest matter. Since September 11, ever more pressing concerns about national security have raised the level of interest in gaining access to identifiable data statistical agencies (and others) collect. In at least one case, that interest led to the passage of legislation that could reverse a prior pledge given to a respondent. Under the U.S.A. Patriot Act, passed shortly after September 11, law enforcement agencies were given the authority to access identifiable data collected by the National Center for Education Statistics – thereby curtailing the broad protection previously afforded NCES data.

More recently, the House passed H.R. 4598 – the “Homeland Security Information Sharing Act” – which calls for sharing of Federal data with State and local officials. In a “Statement of Administration Policy” on this bill, we argued,

“While the Administration supports the bill’s goals, the definition of homeland security information needs to exclude individually identifiable information that has been collected solely for statistical purposes under a pledge of confidentiality. Unless we preserve the trust and cooperation of respondents to Federal statistical surveys, there will be a significant deterioration in the ability of the Federal statistical system to provide high quality aggregate data to guide critical economic and social policy decisions.”

I am pleased to note that as the Homeland Security bill moved to passage in the House, Representative Chambliss offered an amendment, which was accepted, that requires,

“Nothing in this Act shall be construed as authorizing any department, bureau, agency, officer, or employee of the Federal Government to request, receive, or transmit to any other Government entity or personnel, or transmit to any State or local entity or personnel otherwise authorized by this Act to receive homeland security information, any information collected by the Federal Government solely for statistical purposes in violation of any other provision of law relating to the confidentiality of such information.”

As initiatives to ensure “Homeland Security” proceed in the Senate, and in the Administration, we will continue our efforts to preserve respondent trust and cooperation, and ultimately to ensure the continued availability of critically needed statistical information.

Last, but certainly not least, I want to share the “best news” in our never-ceasing, decades-long efforts to achieve broad statutory protection for individually-identifiable information collected for statistical purposes under a pledge of confidentiality. This year, the Administration developed a new legislative proposal to establish a uniform set of safeguards that would protect the confidentiality of *all* individually-identifiable information collected from individuals, businesses, educational institutions or other organizations – in other words, data collected from any entity – for statistical purposes. This legislation builds upon previous efforts of the Executive and Legislative branches including H.R. 2885 (the Statistical Efficiency Act of 1999, originally offered by Representative Stephen Horn, and unanimously passed by the House of Representatives) and H.R. 2136 (the Confidential Information Protection Act, originally offered by Representative Tom Sawyer in 2001). Most notably, the new bill, given the name “CIPSEA,” will extend statutory protection to *all* information collected for statistical purposes under a pledge of confidentiality – rather than extending such protection to data collected by particular agencies.

CIPSEA, the Confidential Information Protection and Statistical Efficiency Act of 2002,

- provides a clear and consistent standard for the use of confidential statistical information, and prohibits the Federal Government from using confidential statistical information for any nonstatistical purpose (prohibited nonstatistical purposes include use of data in individually-identifiable form for any administrative, regulatory, law enforcement, adjudicative, or other purpose that affects the rights, privileges or benefits of the person or organization supplying the information);
- prohibits disclosure of confidential statistical information under the Freedom of Information Act;

- provides safeguards to ensure that data provided under a pledge of confidentiality are used only for statistical purposes; and
- imposes criminal penalties on Federal employees or agents of Federal agencies who willfully disclose confidential statistical information.

In sum, CIPSEA, without restricting or diminishing any confidentiality protections or penalties for unauthorized disclosure that currently exist, provides one uniform set of protections to replace the current patchwork of provisions, and extends these protections to *all* individually-identifiable data collected for statistical purposes under a pledge of confidentiality.

In a “Dear Colleague” letter seeking support for the 2001 version of this legislation (H.R. 2136), Representative Sawyer wrote:

“I believe that these measures are needed not only to protect the public but also to ensure the public’s continued cooperation and participation in essential government research. Informed public policy relies on it.” (Sawyer, June 28, 2001)

Introducing CIPSEA, H.R. 5215, on July 25, 2002, Representative Horn indicated,

“The bill’s enhanced confidentiality protections will improve the quality of Federal statistics by encouraging greater cooperation on the part of respondents. Even more important, these protections ensure that the Federal Government does not abuse the trust of those who provide data to it under a pledge of confidentiality. . . . the Confidential Information Protection and Statistical Efficiency Act of 2002 makes important, common-sense and long overdue improvements in our Nation’s statistical programs. It is a bipartisan, good Government measure that has the Administration’s strong support. I urge my colleagues to join with us to achieve prompt enactment of the bill.”

Those of us in the Federal statistical system could not agree more. We are thrilled that CIPSEA was unanimously passed by the House of Representatives at 2:50 a.m. on Friday, November 15, as Title V of the E-Government Act of 2002 [H.R. 2458], and subsequently passed unanimously by the Senate sometime after 8:00 p.m. the same day. President Bush’s signature on this legislation will complete the “unfinished work” that has challenged us for decades, and has become ever more important in the current environment.

Postscript: On December 17, 2002, President George W. Bush signed into law the E-Government Act of 2002, Public Law 107–347. CIPSEA appears as Title V in this law. In the coming months, the Office of Management and Budget will be working closely with the statistical agencies to explore the implications of the new legislation for data collection, dissemination, and sharing, and to collaborate in the development of appropriate implementation guidance.

Received February 2003