
Sharing Statistical Information for Statistical Purposes

Katherine K. Wallman and Jerry L. Coffey, Office of Management and Budget

Congress has recognized that a confidential relationship between statistical agencies and their respondents is essential for the effective conduct of statistical programs. However, the specific statutory formulas devised to implement this principle in different agencies have created difficult barriers to effective working relationships among these agencies. The development of mechanisms to establish a uniform confidentiality policy that substantially eliminates the risks associated with sharing confidential data will permit significant improvements in data used for both public and private decisions without compromising public confidence in the security of information respondents provide to the Federal government.

Initiatives of the Statistical Policy Office to enhance public confidence in the stewardship of sensitive data and to permit limited sharing of confidential data for exclusively statistical purposes received a substantial impetus in the 1995 reauthorization of the Paperwork Reduction Act. The Act strongly endorses the principles embodied in statistical confidentiality pledges and charges the Office of Management and Budget (OMB) to promote sharing of data for statistical purposes within a strong confidentiality framework.

This paper discusses the history, the promise, and the current status of initiatives to strengthen and improve data protection while promoting expanded data sharing for statistical purposes. The most recent efforts include the OMB Federal Statistical Confidentiality Order, the Statistical Confidentiality Act (SCA), and companion legislation to the SCA that would make complementary changes to the Internal Revenue Code.

■ Introduction

A promising initiative to improve the quality and efficiency of Federal statistical programs is a legislative proposal that would allow the sharing of confidential data among statistical agencies under strict safeguards. The development of this approach has been a painstaking,

careful process that has been supported and nurtured by Administrations of both parties over many years.

The Administration's Statistical Confidentiality Act and two companion initiatives--the OMB Federal Statistical Confidentiality Order and an amendment to the Internal Revenue Code--address two issues that are vital to ensuring the integrity and efficiency of Federal statistical programs and, ultimately, the quality of Federal statistics. These are:

- the unevenness of current *statutory* protections for the confidential treatment of information provided to statistical agencies for exclusively statistical purposes; and
- the barriers to effective working relationships among the statistical agencies that stem from slightly different statutory formulas devised to implement the principle of confidentiality for statistical data in different agencies.

The proposed legislation would establish policies and procedures to guarantee the consistent and uniform application of the confidentiality privilege, and would authorize the limited sharing of information among designated statistical agencies for exclusively statistical purposes.

■ Initiatives Span More Than Two Decades

Efforts to address confidentiality concerns with respect to Federal statistical data have a history that extends for more than 25 years. Such efforts have been endorsed on both sides of the aisle in Congress. The roots of the policies in the Administration's current Statistical Confidentiality Act 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 July 1977, the Privacy Protection Study Commission stated that "no record or information... collected or maintained for a research or statistical purpose 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..." Later, in October of that 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 (PRA), which largely addressed the *efficiency* recommendations of the Paperwork Commission. The legislative history of that Act recognized the unfinished work of fitting the concept of functional separation of statistical information into the overall scheme.

The first attempt to deal with issues related to confidentiality and sharing of statistical data was made by the Carter Administration's Statistical Reorganization Project (popularly known as the "Bonnen Commission.") This effort paralleled the legislative development work by OMB that became the Paperwork Reduction Act. The

initiative identified a group of statistical agencies that could serve as protected environments--or *enclaves*--for confidential data and attempted to create a harmonized confidentiality policy by synthesizing the several prescriptions in existing laws. The initiative was left behind by the fast-track PRA for two reasons:

- First, each new prescription to solve problems in one agency raised new questions in other agencies, so that objections to the language increased as the draft legislation became longer and more complex.
- Second, the approach failed to appreciate that some large data bases--e.g., Census and tax files--represented more significant risks and, thus, needed more elaborate confidentiality protection than other files.

During the first Reagan Administration, this prescriptive formula became more and more complex, as attempts were made to incorporate comments from both statistical and nonstatistical agencies. The draft proposal was eventually withdrawn when it became apparent that almost no one could understand how all the myriad definitions and exceptions fit together.

While the proposed approach did not succeed, the effort did draw attention to many subtle weaknesses in existing law and led to new statutes and amendments during the second Reagan Administration. In particular, stronger statutory protections were enacted for the National Center for Health Statistics, the National Agricultural Statistics Service, and the National Center for Education Statistics. At the same time, the concept of a governmentwide law for statistical confidentiality and data sharing received a complete overhaul.

A new strategy was presented to the statistical agencies during the Bush Administration. It had five important features that were missing from earlier efforts:

- It was designed to work with the tools already available in the PRA--promoting data sharing but providing for functional separation to ensure that statistical data are only shared for statistical pur-

poses.

- It was designed to be robust with respect to reorganizations within the statistical system. Since every major statistical agency had been involved in one or more reorganizations since 1970, it became apparent that any successful strategy would have to work well in any reasonable organizational environment.
- It was built around a procedural strategy that gives due deference to the precepts of existing law that are tailored to specific risks and builds on agency experience in implementing that body of law. The idea was to adopt a general confidentiality policy consistent with existing law and provide the tools--data-sharing agreements, coordinated rules, and consistent Freedom of Information Act (FOIA) exemptions--to address those risks.
- It provided a means for the major statistical agencies to work closely with other agencies in their areas of expertise. While only the Statistical Data Centers (SDC's) would have broad access to data, any agency that collects its own statistical data could act as a full partner in improving those data under the terms of a data-sharing agreement.
- It strengthened the Trade Secrets Act. This universal confidentiality statute consolidated provisions of tax law, customs law, and *statistical* law, but the statistical implications had been ignored. The new proposal set uniform policies for *confidential statistical data*, increasing penalties and addressing questions of *agents*.

This fresh start--based on a precedent-setting data-sharing order involving the Internal Revenue Service, the Census Bureau, and the Bureau of Labor Statistics--had strong support within the Administration. But the effort failed to reach closure.

The basic strategy developed during the Bush Administration was later expanded and refined during the first Clinton Administration. Criteria for SDC's were incorporated into the Statistical Confidentiality Act, and

every statistical agency that could meet these tests was added to the list of SDC's--bringing the total from four agencies to eight. The relationship to the PRA was fine-tuned, as well, and this process identified some improvements to the PRA that were adopted in the 1995 amendments to that Act.

The final step in the recent initiative involved negotiating a complementary amendment to the Statistical Use section of the Tax Code [26 USC 6103(j)]. This change actually facilitates increased security for taxpayer information by targeting and, thus, limiting the wholesale disclosures permitted under current law. It permits multi-party sharing agreements, so that specific statistical data sets that include tax data can be shared under IRS security procedures with other SDC's.

■ What Factors Argue for Success Now?

After more than two decades, why should we think that these efforts will be any more successful than those of the past? Perhaps it comes down to what can be called the "Three E's:"

- *Experience*--Over the past 25 years, we have learned a considerable amount. The current proposal builds on the experience OMB and the agencies gained through earlier efforts.
- *Environment*--The Federal statistical system is faced with growing fiscal resource constraints. At the same time, the 1995 Paperwork Reduction Act extends requirements for reducing burdens imposed on respondents to Federal surveys. Yet another factor that has affected agency views is the increasing number of proposals for consolidating statistical agencies.
- *Enthusiasm*--Last but not least, the statistical agencies appear to be in a "can do" mood--enthusiastically supporting the development and passage of legislation that will even out statutory confidentiality protections and permit data sharing for statistical purposes.

Whatever the reasons, the agencies have come together on an Administration proposal now embodied in the

Statistical Confidentiality Act and its companion pieces.

■ The Statistical Confidentiality Act

As the centerpiece of this effort, the Statistical Confidentiality Act has two principal functions:

- To ensure consistent and uniform application of the *confidentiality privilege*; and
- To permit limited *sharing of data* among designated agencies for exclusively statistical purposes.

A limited number of Federal statistical agencies would be designated as Statistical Data Centers. The eight agencies that currently meet the criteria to become SDC's are the Bureau of Economic Analysis (BEA), Bureau of the Census, Bureau of Labor Statistics (BLS), National Agricultural Statistics Service (NASS), National Center for Education Statistics (NCES), National Center for Health Statistics (NCHS), the Energy End Use and Integrated Statistics Division of the Energy Information Administration (EIA), and the Science Resources Studies Division of the National Science Foundation (NSF).

A key component of the legislation is *functional separation*, whereby data or information acquired by an agency for *purely statistical purposes* can be used only for statistical purposes and cannot be shared in identifiable form for any other purpose without the informed consent of the respondent. If a designated SDC is authorized by statute to collect data or information for any nonstatistical purpose, such data or information must be distinguished by rule from those data collected for strictly statistical purposes.

The procedural strategy for implementing the legislation would be carried out via written data-sharing agreements between or among statistical agencies. The Statistical Data Centers would provide information on actual disclosures and information security to OMB for inclusion in the annual report to Congress on statistical programs. OMB would also review and approve any implementing rules to ensure consistency with the purposes of the SCA and the PRA.

■ Companion Legislation

In addition to the Statistical Confidentiality Act, special amendments have been proposed to the Statistical Use subsection of the Internal Revenue Code--Section 6103 (j). These amendments would authorize limited disclosure of tax data to agencies that have been designated as Statistical Data Centers. In addition, the Research and Statistics Division at the Federal Reserve Board has been added to the group of agencies covered under the IRS companion bill.

The amendments would provide access to tax return information to construct sampling frames and for related statistical purposes as authorized by law. Names, addresses, taxpayer identification numbers, and classifications of other return information in categorical form could be provided for statistical uses. These latter data are not to be used as direct substitutes for statistical program content, but rather can be applied using statistical methods such as imputation to improve the quality of the data. Class sizes or ranges for such data--e.g., for income--will vary by purpose.

The amendment is designed to protect taxpayer rights and maintain proper oversight and control over tax return disclosures, while allowing carefully targeted expansion of access to tax return information for statistical purposes only.

■ The Statistical Confidentiality Order

As an integral step to foster passage of these legislative proposals, OMB felt it was critical to move ahead with efforts to clarify and make consistent government policies protecting the privacy and confidentiality interests of individuals and organizations that provide data for Federal statistical programs. With that aim in mind, OMB developed and sought public comment on an Order that assures respondents who supply statistical information that their responses will be held in confidence and will not be used against them in any government action. The order also gives additional weight and stature to policies that statistical agencies have pursued for decades and includes procedures to resolve a number of ambiguities in existing law. Following the public review

process, the Federal Statistical Confidentiality Order went into effect on June 27, 1977.

■ What Opportunities Will Attend Passage of the Legislation?

For more than a decade, we have worked within the constraints of existing law to make limited comparisons between similar data sets in different agencies. We have set in motion a series of limited exchanges tailored to conform to current law, but they cannot address all of the problems. Moreover, such exchanges could be cut short by an unfavorable interpretation of any one of the dozens of statutes involved. In each of these cases, extraordinary efforts have been required to accomplish even limited data exchanges. Based on these experiences, we believe that even modest exchanges of information could, in the future, unearth and eliminate important errors in existing economic series, enable significant consolidations of overlapping programs (with comparable reductions in costs), and permit substantial reductions in reporting burden imposed on the public.

As the possibility of a law to permit data sharing in a safe environment has become more credible, statistical agencies have begun to identify potential improvements to current operations and programs that this law would permit. These possibilities include the following:

- Integrated data base concepts for information on particular segments of the economy and society, such as educational institutions (NCES, NSF, and Census), health care providers (NCHS, Census, and some program-specific agencies), and agricultural establishments (NASS, Census, and the Economic Research Service of the Department of Agriculture), would improve the consistency and quality of data while reducing current data collection costs.
- Collaboration on sampling frames would improve accuracy and reduce maintenance costs. A more efficient division of labor would make it possible to maintain high-quality frames at minimum cost, both for list frames (Census, BLS, NASS) and for area frames (NASS, Census, NCHS). This approach would avoid duplicate expenditures and

improve quality. Coordination and shared use of relisting information (updates) in large multistage designs could also reduce frame maintenance costs.

- Targeted frames--or sample selection services -- from improved master frames could reduce duplicative expenditures in agencies that must currently pay the cost of independently developing these resources for specific surveys.
- Access to specific data details that can resolve uncertainties in particular analyses--e.g., anomalies that arise in the Gross Domestic Product estimation process--would reduce errors in macroeconomic statistics without imposing additional burden.
- Coordination of sample selection across agencies could reduce the total reporting burden that falls on any one household or company (and, thus, improve the level of respondent cooperation).

■ What Systemic Problems Will the Act Address?

- *The Statistical Confidentiality Act creates a credible governmentwide confidentiality umbrella.--The public will know that the entire government stands behind the pledges of statistical confidentiality offered by the SDC's or any agency engaged in joint statistical projects with the SDC's.*
- *The SCA creates the legal presumption that data collected for most purposes may be used in a safe environment for statistical purposes.--This is one of the critical insights of the Privacy and Paperwork Commissions.*
- *The SCA provides consistent FOIA policies for all the SDC's.--This was controversial 15 years ago, but now six of the eight agencies designated as SDC's already have in place statutes that meet the requirements of Section (b)(3) of FOIA.*
- *The SCA permits the data-sharing authorities of the PRA to work without compromising con-*

fidentiality.--By establishing the functional separation principle in law, the SCA facilitates the use of PRA mechanisms to promote and manage data sharing for exclusively statistical uses.

- *The SCA provides a privacy-sensitive alternative to the creation of universal data bases, which each Department has proposed at one time or another to support its own policy interests.*--Statistical methods--particularly sampling--coupled with secure data sharing provide a natural hedge against the big data base (i.e., dossier building) mentality that puts privacy at risk.

In short, the Statistical Confidentiality Act permits the SDC's and their statistical partners to share both expertise and data resources to improve the quality and reduce the burden of statistical programs, while preserving privacy. Moreover, no matter how the organizational boxes for the ideal Federal statistical system may be drawn, this legislation will permit the components of the statistical system to manage their data as if they were a single, functionally integrated organization.

■ **Current Status of the SCA and Related Initiatives**

Culminating efforts that have literally spanned decades, the Statistical Confidentiality Act was initially introduced on a bipartisan basis in the House of Representatives in 1996. Late in 1997, the Administration's proposed legislation was included in a broader bill, S.

1404, introduced on a bipartisan basis in the Senate. With growing bipartisan support in both houses, hopes are high that the SCA will soon become law. The complementary amendment to the Internal Revenue Code is also pending before Congress, with broad bipartisan support. OMB is continuing to work with the House and Senate to attain successful action on the legislation.

In addition to these legislative approaches to foster efficiency and quality in Federal statistical programs, the agencies are actively exploring other means of expanding collaboration to improve the effectiveness of the Federal statistical system. Recently, the Interagency Council on Statistical Policy (ICSP), under the leadership of OMB, has broadened efforts of the principal Federal statistical agencies to coordinate statistical work--particularly in areas where activities and issues overlap and/or cut across agencies. One byproduct of these efforts was the establishment in 1997 of the Interagency Confidentiality and Disclosure Avoidance Group, under the auspices of OMB's Federal Committee on Statistical Methodology. This working group discusses common technical issues involving privacy, confidentiality, and disclosure limitation. The group is currently working on developing a set of generic guidelines for disclosure review, which could be adapted for use by other agencies.

It is our hope and expectation that both the statistical confidentiality legislation and subsequent cooperative efforts will go a long way toward solving some of the challenges the Federal statistical agencies have encountered in a decentralized environment.