

IMPLEMENTING THE CONFIDENTIAL INFORMATION PROTECTION AND STATISTICAL EFFICIENCY ACT OF 2002

Katherine K. Wallman and Brian A. Harris-Kojetin
U.S. Office of Management and Budget

Keywords: Confidentiality, Data Access, Data Sharing

Concerns about privacy and confidentiality are prominent issues in our society. Television commercials about identity theft and news stories about the use of airline passenger databases to try to identify potential terrorists heighten public awareness and raise fears about how private information can be used in ways potentially harmful to individuals. Against this cultural backdrop, statistical agencies must continue to obtain the cooperation of respondents to provide information critical to the decisions of policymakers, businesses, and individuals.

The ability of statistical agencies to ensure that the information they collect will be used only for statistical purposes and will not be used in any way against the individual, business, or institution supplying the information has long been considered critical to the success of their efforts. Recent legislation has provided a uniform high level of protection for all information gathered by Federal agencies under a pledge of confidentiality for exclusively statistical purposes.

Before I proceed, let me take a moment to underscore the importance of distinguishing among the purposes for which various types of data are collected. Much of the data the government has is “administrative” and is collected for programmatic purposes. The data that I will be referring to are collected for statistical purposes, with a pledge of confidentiality. Researchers may well be interested in both types of data, or even linked datasets that combine both types, but it is critical that we keep in mind both the purposes for which and the conditions under which the data were collected when we consider how to protect confidentiality and provide access to the data.

What is CIPSEA?

In November 2002 Congress passed the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) as Title V of the E-Government Act of 2002; the Act was signed into law by President George W. Bush on Dec. 22, 2003.

CIPSEA represents the culmination of literally decades of work to enhance legal provisions that protect the confidentiality of Federal statistical data.

As enacted, CIPSEA has two subtitles. Subtitle A on Confidential Information Protection applies to all Federal agencies that acquire data under a pledge of confidentiality for exclusively statistical purposes. The purposes of this subtitle are:

- to strengthen pledges of confidentiality and foster public trust in the government’s promise to hold data in confidence;
- to ensure that individuals or organizations who supply information under a pledge of confidentiality to agencies for statistical purposes will neither have that information disclosed in identifiable form to any unauthorized person nor have that information used for any purpose other than a statistical purpose;
- to safeguard the confidentiality of individually identifiable information acquired under a pledge of confidentiality for statistical purposes by controlling access to, and uses made of, such information (Section 511(b)); and
- to ensure that information supplied by individuals or organizations to an agency for statistical purposes under a pledge of confidentiality is used exclusively for statistical purposes.

Subtitle B of CIPSEA promotes statistical efficiency through limited sharing of business data among three designated statistical agencies: the Bureau of the Census, the Bureau of Economic Analysis, and the Bureau of Labor Statistics. The purposes of Subtitle B are:

- to reduce the paperwork burdens imposed on businesses that provide requested information to the Federal Government;
- to improve the comparability and accuracy of Federal economic statistics by allowing the Bureau of the Census, the Bureau of Economic Analysis, and the Bureau of Labor Statistics to update sample frames, develop consistent classifications of establishments and companies into industries, improve coverage,

and reconcile significant differences in data produced by the three agencies; and

- to increase understanding of the United States economy, especially for key industry and regional statistics, to develop more accurate measures of the impact of technology on productivity growth, and to enhance the reliability of the Nation's most important economic indicators, such as the National Income and Product Accounts (Section 521(b)).

Why was CIPSEA necessary?

There are more than 70 Federal agencies or organizational units that carry out statistical activities either as their principal mission or in conjunction with other program missions, such as providing services or enforcing regulations. In addition to these 70 agencies, many other Federal agencies or units may collect statistical information to use for specific program needs (U.S. Office of Management and Budget, 2003). Prior to the enactment of CIPSEA, a patchwork of legislative protections governed the confidentiality of data gathered for statistical purposes by the different agencies and units. Some agencies had strong statutory authority to protect the confidentiality of the data they gathered for statistical purposes, while other agencies had relatively weak statutory authority to protect confidentiality.

The results of a "survey" my office completed, prior to the passage of CIPSEA, 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 the unevenness of statutory and other protections for upholding confidentiality. From a statutory perspective, the essence of that patchwork was as follows:

- one agency, the Energy Information Administration, had legislation that actually requires sharing of confidential statistical information;
- two agencies, the Environmental Protection Agency and the Economic Research Service (in Agriculture), had no specific legislation protecting the confidentiality of their statistical information;
- four agencies, the Bureau of Labor Statistics, the Bureau of Transportation Statistics, the National Center for Education Statistics, and the Social Security Administration, had partial protection for confidential statistical information; and
- seven agencies, the Bureau of Economic

Analysis, the Bureau of Justice Statistics, the Census Bureau, the Statistics of Income Division at IRS, the National Agricultural Statistics Service, the National Center for Health Statistics, and the National Science Foundation had broad statutory authority to protect the confidentiality of information they collect.

Over the years, numerous attempts were made both to shore up legal protection for the confidentiality of statistical information, and to permit some limited sharing of data for statistical purposes. The goal of all of these efforts was to strengthen and standardize statutory protections for the confidentiality of individually identifiable data that are collected for statistical purposes.

What are the benefits of CIPSEA?

By establishing a uniform policy for all Federal statistical collections, CIPSEA stands to reduce public confusion, uncertainty, and concern about the treatment of confidential statistical information by different Federal agencies. By establishing consistent rational principles and processes to buttress confidentiality pledges, the law enables us to harmonize confidentiality claims and set minimum standards for safeguarding confidential statistical information. Such consistent protection of confidential statistical information will, in turn, encourage more efficient working relationships among statistical agencies, resulting in a reduction in both the cost and reporting burden imposed by statistical programs.

The protection of information collected under CIPSEA is supported by a criminal penalty that includes imprisonment for up to five years and fines up to \$250,000 for a knowing and willful disclosure of confidential data (Section 513). Thus, for many agencies this law strengthens the protections afforded to confidential data.

What requirements does CIPSEA impose on agencies?

Subtitle A of CIPSEA establishes a common statutory "floor" for the protection of information that is collected under a pledge of confidentiality solely for statistical purposes. Through this statute, Congress authorizes Federal agencies to give a pledge that provides strong protection of the confidentiality of statistical information. To back up this pledge, an agency will be expected to ensure that it is providing a high level of protection at all times.

In short, an agency must live up to the new authority it has under CIPSEA and maintain the trust the public is placing in the agency to protect information that has been provided under a pledge of confidentiality.

Distinguishing statistical versus non-statistical purposes

Through CIPSEA, the concept of functional separation between statistical and non-statistical uses of information became law. Agencies must ensure that information collected under CIPSEA is used only for statistical purposes. CIPSEA defines a statistical purpose to include the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups. Statistical purposes can also include the development, implementation, or maintenance of methods, technical or administrative procedures, or information resources (Section 502(9)). In contrast, a non-statistical purpose means the use of information in identifiable form for anything other than a statistical purpose, including any administrative, regulatory, law enforcement, adjudicative, or other purpose that affects the rights, privileges or benefits of a particular identifiable respondent. Confidential information covered under CIPSEA is also exempt from release under a FOIA request (Section 502(5)).

Informing the public

Federal agencies are required by the Paperwork Reduction Act (PRA) to submit to the Office of Management and Budget (OMB) for approval any information collection request that will gather information from 10 or more persons or entities. As part of its request to OMB to collect information, the agency must “describe any assurance of confidentiality provided to respondents and the basis for the assurance in statute, regulation, or agency policy” (5 CFR 1320.8(b)(3)).

At the time of the information collection, agencies are required under the PRA to adequately inform potential respondents about the uses of the information they provide. This description must include the reasons the information is planned to be and/or has been collected, the way such information is planned to be and/or has been used to further the proper performance of the functions of the agency, and the nature and extent of confidentiality to be provided, if any (citing authority) (5 CFR 1320.8(b)(3)). When agencies are collecting information under CIPSEA, they are required at the time of collection to pledge to keep the data or

information confidential, and pledge that the information will be used for exclusively statistical purposes (Section 512(a)).

In order to promote greater public understanding of agency pledges of confidentiality for statistical collections, attention needs to be given to the language that different Federal agencies use. Prior to CIPSEA, agencies customized their pledges based on their particular authority to protect the confidentiality of data. To reduce ambiguity and foster understanding of the more uniform protection afforded by CIPSEA, it will be important to make agency pledges more recognizable and consistent.

Because it is generally expected that statistical agencies or organizational units will be collecting data for exclusively statistical purposes under a pledge of confidentiality, CIPSEA imposes special requirements on statistical agencies if they are not collecting information for exclusively statistical purposes. A statistical agency or organizational unit that conducts or sponsors a collection that will not be for exclusively statistical purposes must clearly distinguish any data or information it collects that may be used for nonstatistical purposes (Section 512(c)).

Safeguarding confidential information

In order for CIPSEA to have its intended effect—of reinforcing public confidence in Federal pledges of confidentiality—any Federal agency that makes the CIPSEA pledge must provide a high level of protection for the information collected. Because any inappropriate use or disclosure of CIPSEA-protected information could undermine the trust that the Congress and the public have placed in the data collecting agencies, there is no acceptable level of non-compliance with the CIPSEA pledge.

Each Federal agency remains ultimately responsible, and accountable, for the confidential information that the agency collects under a CIPSEA pledge. The central objective for data protection is to ensure that the pledge of confidentiality for information obtained under CIPSEA is honored by the Federal agency making the promise. Hence, respondents’ trust should be based upon confidence in both the intention and the ability of a Federal agency to keep such a promise. A Federal agency should not make a CIPSEA pledge unless it is fully committed to taking all the actions that are necessary for ensuring that the collected information will be given the highest level of protection. If a Federal agency is not willing or able to provide the protection

required under CIPSEA, then the agency should not make a CIPSEA pledge.

What kinds of data sharing does CIPSEA enable?

Although CIPSEA provides broad confidentiality protection for Federal agencies collecting or acquiring information, its provisions for data sharing among Federal agencies are very specific. Subtitle B of the law, which covers Statistical Efficiency, provides only for the sharing of business data for exclusively statistical purposes among three designated statistical agencies. The three designated statistical agencies permitted to share business data for exclusively statistical purposes are the Bureau of Economic Analysis (BEA), the Bureau of Labor Statistics (BLS), and the Bureau of the Census (Census).

CIPSEA does not alter existing laws that may permit other data sharing among Federal agencies; however, by itself CIPSEA does not authorize the sharing of business data among any Federal agencies other than the three designated statistical agencies, nor does it authorize any sharing of demographic or other types of data among any Federal agencies.

BEA, BLS, and Census are currently working to implement the data sharing provisions of CIPSEA. Prior to sharing business data that were collected under a legal requirement to supply the information—without informing the respondents that the data would be shared with the other designated statistical agencies—agencies are required to publish a notice of any proposed data sharing activity in the *Federal Register*. Such notices must specify the business data to be shared and the statistical purposes for which the business data are to be used, and must allow a minimum of 60 days for public comment.

These agencies are also required to enter into a written agreement before sharing any business data. The written agreement must specify the business data to be shared; the statistical purposes for which the business data are to be used; the officers, employees, and agents authorized to examine the business data to be shared; and the appropriate security procedures to safeguard the confidentiality of the business data.

In June 2003, the first formal proposal for data sharing under CIPSEA was published for comment in the *Federal Register*. BEA and the Census Bureau entered into an arrangement to match data from BEA’s surveys of Foreign Direct Investment in the U.S. and U.S. Direct Investment Abroad with data from the Census Bureau’s Survey of Industrial

Research and Development. The early results from the project confirm the benefits anticipated from data sharing, namely improved sampling frames that translate into increased survey efficiency, potentially reduced costs, and new details on international R&D activities. Based on that match, the Census Bureau was able to add over 500 companies, not previously identified as having R&D activity, to its 2003 R&D survey sample. The match also demonstrated that BEA will be able to augment its existing R&D data with information from the Census R&D Survey, identify data quality issues arising from reporting differences in the respective surveys, and improve its survey sample frames.

Several other projects are in various stages of planning. For example, the Census Bureau and BLS plan research on identifying differences in their business establishment lists. The work will involve matching files and profiling differences, particularly with respect to industry codes, an area long identified as crucial. The IRS is concurring with the Memorandum of Understanding (MOU) that will govern this project.

BLS and BEA are planning to share company and revenue data from several BEA surveys with the BLS. The BLS International Price Program would use the BEA data to study the feasibility of developing a new price index that measures royalties and license fees in international transactions. BEA would benefit by being able to compare its sample frame with the one compiled by the International Price Program and thereby capture any omissions.

Other potential data sharing arrangements that are being discussed include a plan for BLS and the Census Bureau to use records from the Census Business Register to facilitate more efficient sampling for the BLS Producer Price Index (PPI) program, and BEA and BLS are exploring the development of experimental inter-area price indexes to permit comparison of price levels across geographic areas. Clearly, we have much to look forward to in this arena.

Implementing CIPSEA

What are the implications for Federal agencies?

For agencies that have historically had strong legislation to protect the confidentiality of their statistical data, CIPSEA will mean very few changes in their culture or operations. For agencies that had not previously had strong legislative protection, CIPSEA will offer a new world of protection for data gathered for exclusively statistical purposes under a

pledge of confidentiality. This new level of protection brings with it corresponding responsibilities to ensure that the information is adequately protected from disclosure. Agencies must protect the confidentiality of their information through adequate security for IT systems and physical security for information storage. Employees with access to confidential information must also be given adequate training on protecting the information.

In addition, agencies need the infrastructure to provide access to data in a secure environment that will maintain the confidentiality of the data and provide a means for the agency to control and supervise access to and use of the data. They must devise a process for evaluating and selecting proposals for data access, determine effective means by which to exercise supervision and control, and obtain resources to create environments for data access or monitor areas where access has been granted. The implications of CIPSEA for each of the statistical agencies, and how they can provide access to their data, are currently being carefully considered.

Over the decades of discussion leading to the enactment of CIPSEA, some hoped that there would be possibilities to extend the data sharing to other domains, such as demographic data, and to include other Federal agencies. To further this long term goal, we believe that it is critical for agencies to demonstrate a consistently high level of protection to information collected under the CIPSEA pledge as well as to achieve the efficiencies that are authorized. If the Federal statistical system can build upon its strong record in this regard, a case for an extension of data sharing to other domains and to additional agencies can better be made.

In the meantime, OMB has formed an interagency team to solicit questions about implementing CIPSEA and to help inform the guidance that OMB expects to issue. Through the interagency team, a variety of questions have been raised dealing with the applicability of CIPSEA provisions to different agency circumstances. While agencies are generally embracing CIPSEA for their collections, there are cases where they may not collect data under CIPSEA because there are non-statistical uses that their departments may make or may want to reserve the right to make with the data. Agencies are also examining how they will distinguish data that they will be collecting under CIPSEA from data that they will not be collecting under CIPSEA. We need to make very clear to our respondents when we are fully protecting their

individually identifiable data and when their data do not have the same level of protection. I anticipate that OMB will be issuing draft guidance for public comment later this year.

What are the implications for researchers?

CIPSEA provides that an agency may designate agents, by contract or by entering into a special agreement, who may perform exclusively statistical activities and be subject to the same limitations and penalties as employees of the agency. CIPSEA's definition of agents specifically includes contractors and their employees, as well as individuals who obtain special sworn status from the Census Bureau. Agents may also be researchers affiliated with an institution of higher learning, but the law is clear that agents must have written contracts or agreements with the agency, including a signed statement binding them to the same penalties as employees for disclosure of data. Furthermore, agents must be under the direct control and supervision of the agency.

Some have viewed the language in the law regarding agents as opening up researcher access to all data collected by statistical agencies. While CIPSEA does provide a means for agencies to designate agents, it also imposes stringent requirements on agencies and agents to protect the confidentiality of the data. Furthermore, it is important to remember that CIPSEA does not diminish any protections in existing legislation. Although CIPSEA may allow some agencies to increase researcher access to their data, it will not remove barriers in existing agency legislation that limit access to confidential data. For example, the Census Bureau is not relieved of its legal requirement that any work by outside researchers accessing Census data must benefit the Bureau.

Thus, CIPSEA does not provide a "right of access" to Federal statistical data for external researchers. Researchers who obtain authorization to access confidential data for exclusively statistical purposes will have to share the responsibility to maintain and uphold the confidentiality of any data they access. Because different agencies have data that vary in their sensitivity, some agencies may not be able to provide access to their data, while agencies that can provide access to their data may need to do so in different ways. Researchers seeking access to confidential information will need to conform to specific agency requirements and respect the confidentiality provisions those data carry. Although the limitations and safeguards may be more restrictive than those prevailing in the researchers'

universities and those they have encountered in accessing other datasets, the limitations and safeguards must be honored.

What are the implications for respondents?

Respondents can expect to receive more uniform pledges from Federal statistical agencies and a higher minimum level of legal protection for information they provide. Thus, in the long term, CIPSEA should reduce confusion and enhance respondents' trust.

The principle underlying statistical data collection is that of *informed consent*. Respondents have a right to know what their information will be used for and who will see their information. But a question arises: is telling a respondent that the information will be used "only for statistical purposes" a sufficient explanation for all of the statistical purposes an agency may have for the data? The issue rests with the public's understanding of what may be encompassed by "statistical uses," and what constitutes adequately informing the public about how the agency intends to use the data it is collecting. For example, should agencies specify that authorized personnel from another agency, or outside researchers with an agency-approved project, all of whom are bound by the same legal requirements to maintain confidentiality as agency employees, may be able to use the data for statistical purposes? What additional information should be shared to adequately inform respondents, and what are the implications for their cooperation with our requests for information?

It seems likely that we need more research to understand respondents' perceptions of confidentiality. In order to maintain the public's trust while gaining maximum use of the data collected, we need to understand the optimum balance between providing adequate information about the possible uses of the data and creating unwarranted perceptions with respect to problems of disclosure. This will be a key issue to address in these times of heightened sensitivity. Recent research suggests that statements about confidentiality are delicate instruments of communication with respondents, and are not easily constructed to convey just what we want.

What are the implications for the public?

CIPSEA provides a strong uniform floor of protection for all information collected by Federal agencies for exclusively statistical purposes under a pledge of confidentiality. We hope that the use of a uniform pledge across agencies conducting statistical collections will increase public understanding of confidentiality protections and maintain public trust and cooperation with Federal surveys. At the same time, greater efficiencies achieved through the limited data sharing specifically authorized by CIPSEA should result in higher quality economic information at lower cost and with less burden on the public.