

# Privacy, Confidentiality, and Data Sharing

David McMillen<sup>1</sup>

Alan Westin points out that privacy policy is a balance of competing pressures of a democratic society. We value and institutionalize privacy. At the same time we ask citizens to voluntarily disclose information about themselves for the good of society, and sometimes even require information to be provided to the government. At times, hopefully only for legitimate and well constrained law enforcement purposes, the government invades an individual's privacy in order to maintain a lawful society. "Managing this tension among privacy, disclosure, and surveillance," Weston said, "is the central challenge of privacy protection and definition." (Alan Westin, *Privacy and Freedom* (1967). That was in 1967, and it is equally true today.

This balancing process is a never ending one, as we have dramatically witnessed over the past two years. Two years ago, the Speaker of the House regularly railed against red light cameras as an invasion of privacy. Today cities across the country turn on thousands of surveillance cameras every time the terrorist threat goes to orange.

Privacy also has a psychological dimension that is often unrecognized in discussions about collecting statistical information. Individuals vary in their need for privacy from the recluse to exhibitionist, and each individual's need varies across time, sometimes seeking more and sometimes less privacy from the world. At one extreme are those who refuse to participate in retail "savings" cards because they do not want their purchasing habits tracked by the merchants. At the other extreme, are those who assume that between businesses and the government nothing is secret, and so these discussions of protecting privacy are empty exercises. Most of us fall somewhere in between, and are inconsistent in our behavior. That inconsistency in part results from the fact that our attitudes towards privacy vary, conditioned by our most recent experiences, and by the situation at hand. We may choose to allow our grocery purchases to be collected because we are willing to trade that breach of privacy for the savings realized. Some grocery stores recognize this logic and reinforce it by printing at the bottom of each receipt how much you have saved each year. That same person, and I speak from personal experience, may well insist that some number other than the Social Security Number be used on his driver's license, knowing full well that every credit reporting agency, financial institution, and many government agencies already have a link between your name, address, and SSN. Statistical

---

<sup>1</sup> An earlier version of this paper was presented at the 2003 Joint Statistical Meetings in San Francisco, CA. Benjamin Chevat was a co-author on that paper. The contents of both papers are solely the opinions of the authors and in no way represent the Congress, the members for whom they work, or any Committee of Congress. David McMillen serves as a professional staff for Rep. Henry Waxman, Ranking Democrat on the House Committee on Government Reform. Benjamin Chevat serves as the Chief of Staff for Rep. Carolyn Maloney. I would like to thank Jay Casselberry who organized the JSM session that pushed me to begin organizing these thoughts, and Gerry Gates, Shelly Wilkie Martinez, Joanne McNabb, and Ed Spar for their helpful comments that have resulted in a substantially improved paper. I would also like to thank Kathy Wallman and Brian Harris-Kojetin both for their helpful comments and for their patience in working through many difficult issues.

agencies must develop policies for the collection and disclosure of information to meet the public needs against this complex background of attitudes and behaviors.

An individual's attitudes about privacy will not necessarily be predictive of that person's behavior at a different point in time, or within a different context. Understanding the relationship between individual thoughts about privacy, and government information collection is made more complex because the information is sometimes collected voluntarily, and sometimes "under penalty of law." To encourage individuals to overcome their own sense of privacy, and thus providing private information to the government, agencies promise that the information will be held in confidence. Put another way, the promise of confidentiality is one of the techniques interviewers use to convert reluctant respondents. The question then, is what does that promise mean? What are the terms of the implied contract between the agency and the respondent, and how well does the respondent understand those terms?

It is useful at the outset to be clear about what is meant by confidentiality for the definition can vary both in how the information is retained, and how it is used. A narrow, or first order, definition of confidentiality means that only agents of the organization requesting the information will use the information, and only for the specific purpose for which it was collected. A second order definition requires that identifiers like name and address be stripped from the rest of the information; however, it is often the case that sufficient detail remains so that individuals can be easily identified. A third order definition removes enough information so that it is not specific enough to uniquely identify a single individual. All three definitions are used by statistical agencies. First order information is rarely shared with other agencies or used within the agency for purposes other than the purpose for which it was collected. Second order information is shared in a variety of ways. The specifics of that kind of sharing will be discussed later. Third order information defines what are commonly thought of as public use data sets.

Against this backdrop of shifting attitudes about privacy, and a variability in the definition of confidentiality it is my fundamental premise that we must, to the greatest extent possible tell the public what we do with the private information they choose to share with us. Vague statements about "used only for statistical purposes" or "combined with data from other government agencies to produce summary information" are not sufficient. I will pick this thread up again at the end of the paper when I talk about guidelines for implementation of Title V of the Electronic Government Act.

Three other issues should be clarified at the outset of this paper so that all may read it with a full understanding of the terminology used. Those three issues are: 1) the extent of data sharing allowed by Title V of the Electronic Government Act; 2) what is meant by data sharing; and 3) the distinction between information collected voluntarily and information collected under penalty of law.

There is some disagreement about the extent of data sharing allowed under Title V of the Electronic Government Act. Subtitle B of Title V, Statistical Efficiency, limits data sharing to

three agencies and to “business data.”<sup>2</sup> There is, however, disagreement of the interpretation of Sec. 512 of this Act.<sup>3</sup> Sec. 512 states “Data or information acquired by an agency under a pledge of confidentiality for exclusively statistical purposes shall not be disclosed by an agency in identifiable form, for any use other than an exclusively statistical purpose, except with the informed consent of the respondent.”<sup>4</sup> It is my contention that this language clearly allows for the disclosure of data or information in identifiable form for exclusively statistical purposes. This interpretation is supported by the language in Sec. 512(b)(2), which states that the disclosure is “authorized” only when “the head of the agency approves such disclosure and the disclosure is not prohibited by any other law.”

A clarification about what is meant by data sharing in this paper is also necessary at the outset to minimize misunderstanding. In this paper I identify data sharing in the context of the first and second order definitions of confidentiality above. This can be most easily illustrated with a few examples. All of the following examples are what I would classify as sharing data regardless of whether or not personal identifiers are shared or stripped from the records:

- The Census Bureau matches survey records to income histories held by the Social Security Administration;
- The National Center for Education Statistics licenses data to researchers;
- An agency agrees to a research project and makes the researcher a sworn employee of the agency and in doing so grant the researcher access to first or second order data;
- The Census Bureau provides access to matched data sets like the SIPP/SSA match to outside researchers either in Suitland or at one of the agencies regional data centers.

There are a variety of ways these sharing arrangement can be established. There are standardized programs like the fellowship program sponsored by the American Statistical Association. There are situations where the agency solicits experts in a specific field to conduct research for the agency. There are situations that arise serendipitously. Depending on the agency, the researcher is made a sworn special employee, or signs an agreement accepting liability for unauthorized disclosure. Regardless of the ways in which these arrangements are constructed, they constitute the use of an individual’s information by someone who is not normally an agency employee, and often for purposes beyond that for which the information was collected.

It is necessary also at the outset to specify that throughout this paper I am discussion only information provided voluntarily to the government. In a voluntary survey we are talking about an implied contract between the agency and the citizen. The citizen gives up his or her private information under the conditions of confidentiality set out by the agency. If they don’t want to accept those terms, the respondent is free to refuse to participate.

---

<sup>2</sup> The limitation on sharing is set out in Sec. 524, Title V, P.L. 107-347. Business data is defined in Sec. 502(3).

<sup>3</sup> The language from Sec. 512 is provided in Appendix A.

<sup>4</sup> Sec. 512(b)(1), Title V, P.L. 107-347..

When data is collected under mandatory authority, there is no such contract. The citizen is required by law to provide the information. This then changes the constraints on the agency. No longer is it saying to the citizen, if you accept these conditions please give me your information. Rather, the citizen is given no choice.

It is my position that the use of information collected under a mandatory authority must be limited to the intent for which it is collected. Thus, a survey like the American Community Survey can be used to provide general statistics. It must not be used to match to other data sets or administrative records without full disclosure to the respondent of the records to be matched and the purpose for the match, and the opportunity for the citizen to opt out of the match. Similarly, it must not be used as a screener to provide a sampling frame for other surveys. The government cannot collect information from citizens or businesses under the penalty of law, and then use those data to provide access to that individual or business for other surveys or data collections. When an individual or business has no option to opt out of a data collection then the agency must constrain how it handles those data. In deciding whether a survey will be mandatory or voluntary, the agency has to weigh the operational benefits of mandatory authority against the utility of subsequent use possible with voluntary surveys. In OMB's "Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002,"<sup>5</sup> agencies are required to include in their privacy policies instructions on "how to grant consent to use mandatorily-provided information for other than statutorily-mandated use or authorized routine use under the Privacy Act."<sup>6</sup> While these provisions are a part of the privacy policies on agency websites, it signals that OMB interprets the Privacy Act to prohibit the use of mandatorily provided information for uses other than the statutorily mandated use without authorization from the user.

Finally, before discussing the legislative history of the confidentiality and data sharing act, I would like to put that legislation in the context of other legislation on privacy passed during the last third of the 20<sup>th</sup> century.

### **Privacy Legislation 1970 to 2000.**

Personal privacy is considered by most Americans to be a fundamental right. The Supreme Court ruled in 1965 that the Ninth Amendment is the Constitutional source of a right to privacy.<sup>7</sup> Others point to the Third Amendment's prohibition of quartering troops in private homes during peacetime; the Fourth Amendment's prohibition of warrantless search and seizure; and the Fifth Amendment privilege against self-incrimination.<sup>8</sup> Many citizens can cite no source, but firmly believe the right of privacy to exist.

The Twentieth Century saw dramatic changes to the concept of privacy with the introduction of new technologies from the camera to the telephone to the microphone to the computer. In addition, the advent of a credit economy brought about an implicit trade of privacy

---

<sup>5</sup> Memorandum For Heads of Executive Departments and Agencies, From Joshua B. Bolton, Director of the office of Management and Budget (October 2003).

<sup>6</sup> *Op.cit.*, OMB Memorandum for Agency Heads at III. D. iii.

<sup>7</sup> *Griswold v. Connecticut*, 381 U.S. 479 (1965).

<sup>8</sup> Alan Fl Westin, *Privacy and Freedom* (1970).

for credit. These new technologies, whether in the hand of government, private industry, or private citizens altered our perception of the meaning of privacy, and how that right should be protected. It is not surprising then that Congress responded with legislation. What is surprising is that it is concentrated in the last half of the century.

Harold Relyea, in his sweeping review *Personal Privacy Protection: The Legislative Response*<sup>9</sup> summarizes 28 privacy laws passed by Congress between 1970 and 2001 and several commissions. Literally hundreds of bills were introduced in Congress but failed to win sufficient support to become law. Some of the laws passed were narrow in scope and passed in response to specific incidents. For example, Congress passed the Video Privacy Protection Act in 1988 following the confirmation fight over the nomination of Robert H. Bork to the Supreme Court. Others laws, like the Privacy Act of 1974 or the Gramm-Leach-Bliley Act of 1999, established fundamental principles of federal privacy policy. Most bills govern the way the federal government handles the private information it collects from individuals and businesses. However, Congress has, when faced with overwhelming demand, regulated corporate behavior. I will not try to summarize all 28 laws. My intent is to illustrate this long history on privacy, to inform the administration of the Confidentiality and Data Sharing provisions from the Electronic Government Act. It is my contention that we should draw from this history of privacy legislation the guidance necessary to administer this new authority in such a way as to maintain the trust and confidence of the American public.

While the Privacy Act of 1974<sup>10</sup> was the preeminent privacy law of the 1970s, it was preceded by the Fair Credit Reporting Act of 1970<sup>11</sup> and the Crime Control Act of 1973<sup>12</sup>. That decade also saw the passage of the Family Education Rights and Privacy Act of 1974<sup>13</sup> and the Financial Privacy Act of 1978<sup>14</sup>.

The Fair Credit and Reporting Act of 1970 authorized consumers to request from consumer credit reporting agencies the nature and scope of all information regarding that individual, as well as the identity of the sources of the information and the name of any recipient of the information. This Act also grants the consumer the right to correct or amend the credit report by supplying supplemental information. The Family Education Rights and Privacy Act of 1974 is similar in that it grants parents to right to inspect, correct, amend, and control the disclosure of information in the educational records of their children.

The Privacy Act of 1974, which is more familiar to most people, governs the relationship between the government and the public. Citizens and aliens lawfully admitted for permanent residence are given access to information about them that is held by the government, and the right to correct that information. The Privacy Act establishes principles of fair information practice, including conditions for the disclosure of private information, requirement for

---

<sup>9</sup> Congressional Research Service, RL30671 (May 24, 2001).

<sup>10</sup> Sec. 552a, Title V, U.S. Code.

<sup>11</sup> 15 U.S.C. 1681 *et seq.*

<sup>12</sup> 42 U.S.C. 3789g.

<sup>13</sup> 20 U.S.C. 1232g.

<sup>14</sup> 12 U.S.C. 3401 *et seq.*

accounting for disclosure, and the requirement that agencies specify the authority and purpose for collecting personally identifiable information. The Privacy Act discourages the use of secondary sources for information on individuals. Finally, the Privacy Act requires that records of information on individuals must be maintained with accuracy, timeliness, and completeness to assure fair treatment of the individual. It is interesting to note that just this year, the FBI, exempted itself from the provisions of accuracy, timeliness, and completeness for its investigatory records. The FBI claimed that these requirements posed undue burden on the agency.

The 1980s saw the passage of the Cable Communications Policy Act of 1984,<sup>15</sup> the Electronic Communications and Privacy Act of 1986,<sup>16</sup> and the Computer Matching and Privacy Protection Act of 1988.<sup>17</sup> Again, the theme of an individual's right to control information held about him or her pervades these laws. The Cable communication Policy Act of 1984 gives cable subscribers the right of access to all personally identifiable information collected or maintained by the cable company. Again the right of correction is granted to the individual. The Electronic Communications and Privacy Act extends the principles of telephone privacy to cell phones and email, and prohibits unauthorized interception of electronic communications. The Computer Matching and Privacy Protection Act of 1988 amends the Privacy Act to regulate the use of computer matching of information contained in a system of records subject to the Privacy Act.

The most significant privacy provisions passed in the 1990s were those contained in the Gramm-Leach-Bliley Act, also known as the Financial Services Modernization Act.<sup>18</sup> This Act prohibits financial institutions from disclosing nonpublic information to unaffiliated third parties without providing customers the opportunity to decline such disclosure. This Act also is responsible for requiring businesses to disclose their privacy policies in what have now become almost routine privacy flyers that accompany credit card statements and insurance bills.

While this summary only touches the surface of this 30 years of privacy legislation, it is sufficient to illustrate a few key principles: citizens have the right to know about and decline third party use of their information; citizens have the right to know about the uses of their information; and, citizens have the right to amend or correct private information. In providing information to a government agency or a business the citizen does not give up all rights to that information, nor does he or she cede to the government the free use of that information.

### **A Brief History of the Data Sharing Act**

This drama played out over eight years and four congresses. Fortunately in the fourth act we reached dénouement, since, as we all know, a fifth act would have foretold a tragedy.

#### **Act I. Abolishing the Department of Commerce: The 104<sup>th</sup> Congress.**

---

<sup>15</sup> 98 Stat. 2779.

<sup>16</sup> 100 Stat. 1848.

<sup>17</sup> 102 Stat. 2507.

<sup>18</sup> 113 Stat. 1338.

When the Clinton Administration first brought confidentiality and data sharing to Congress during the 104<sup>th</sup> Congress (1995-96), the reaction was studied indifference. That did not, however, mean that Congress was unaware of the statistical system. The House was focused on streamlining the federal bureaucracy that had been built, they believe, by years of Democratic control of Congress. The first target was the Department of Commerce, although, the departments of Energy and Education were also on the chopping list.

In the process of abolishing the Department of Commerce, a decision had to be made about what to do with the Census Bureau and the Bureau of Economic Analysis. The drafter of that section, a young staffer named Counsel Nedd, knew little of these agencies, and placed the Census Bureau at the Department of Treasury and the Bureau of Economic Analysis at the Federal Reserve System. It was not long before he was made painfully aware of the folly of these moves, and to his credit, Nedd never blamed his moves on bad advice from others. A subsequent version of the bill to abolish the Department of Commerce showed up in a bill to increase the public debt limit. There both agencies were transferred to the Department of Labor.

A second version of moving these agencies showed up in a bill drafted by Rep. Steven Horn who was chairman of the Government Reform and Oversight Subcommittee on Government Management and Information Technology. While that subcommittee did not have jurisdiction over the Census Bureau in the 104<sup>th</sup> Congress,<sup>19</sup> it did have jurisdiction over government reorganization. Thus, the subcommittee staff drafted a bill (H.R. 2521 in the 104<sup>th</sup> Congress) to create a federal statistical service. The bill transferred the Bureau of Labor Statistics, the Bureau of Economic Analysis and the Bureau of the Census into this new federal statistical service.

Abolishing the Department of Commerce and creating a federal statistical service were not the only legislative proposals for statistics in the 104<sup>th</sup> Congress. Senator Moynihan introduced a bill that created a commission to study the federal statistical system. The charge for that Commission included considering the desirability of a consolidated statistical system, a review of agency missions, issues of privacy and confidentiality, and the data needs of the Social Security system.

---

<sup>19</sup> Jurisdiction over the census was given to the Subcommittee on National Security. The logic of this placement was in part a reaction to the organization of the Government Operations Committee under Democratic control, and in part a signal that the census was not high on the Committee's priorities. The Government Operations Committee (the Government Reform and Oversight Committee was formed by merging the Committee on the Post Office and Civil Service and the Committee on the District of Columbia into the Government Operations Committee) was structured in such a way that jurisdiction over all legislation was concentrated in one subcommittee. In reforming the committee structure, Republicans wanted every subcommittee to have some legislative jurisdiction. Since there are few national security issues within the legislative jurisdiction of the committee, it was necessary to give that subcommittee jurisdiction that was not compatible with the subcommittee title. The census was available because no other subcommittee fought for it.

Finally, at the end of the 104<sup>th</sup> Congress, Rep. Horn introduced the administration's confidentiality and data sharing bill. This bill, introduced only a few weeks before Congress was scheduled to adjourn, was introduced as a courtesy to the administration without any intent to act on the bill.

Thus the 104<sup>th</sup> Congress ended with confidentiality ignored and with most of the attention on reorganizing the agencies through some form of consolidation. There were four independent proposals on the table: 1) Abolish the Department of Commerce and move the Census Bureau and the Bureau of Economic Analysis to the Department of Labor; 2) create a federal statistical service and move into it the Bureau of Labor Statistics, the Bureau of Economic Analysis, and the Census Bureau; 3) create a commission to study the organization and mission of the federal statistical system; and 4) create a general law which protects the confidentiality of information collected for statistical purposes under a pledge of confidentiality, and create a mechanism to allow statistical agencies to share individually identified information.

## **Act II. Searching for a Middle Ground: The 105<sup>th</sup> Congress**

As with the 104<sup>th</sup> Congress, confidentiality and data sharing were not the focus of congressional concern. Consolidation was the concept that ruled. However, the Republican majority in the House did make an attempt to take the various proposals on the table and make some sense of it all. The turning point came when Rep. Horn and Senator Moynihan testified together before the Senate Committee on Governmental Affairs.<sup>20</sup> Senator Moynihan complimented Rep. Horn on the wisdom of his proposal to merge the three agencies that create the Nation's economic statistics, and Rep. Horn congratulated Senator Moynihan on the wisdom of his commission. The charge was then set for the staff to merge these two approaches in legislation.

The result was a bill (H.R. 4620 and S. 1404 in the 105<sup>th</sup> Congress) that created a commission, not to study the federal statistical system, but to design a consolidated system modeled on Rep. Horn's federal statistical service. The bill included a charge for the commission, if it concluded that a consolidated system was in the country's best interest, to send Congress draft legislation implementing that consolidation. The legislation would then be considered under a fast track legislative process.<sup>21</sup> The bill also contained, as Title II, the administration's proposal for uniform confidentiality for statistical information, and the data sharing provisions. As before, these provisions were taken as drafted by the administration, and without discussion by the principles. Both Senator Moynihan and Rep. Horn were interested in the structure and mission of the federal statistical system, not the mechanics of how the agencies completed their tasks.

---

<sup>20</sup> On January 21, 1997, at the beginning of the 105<sup>th</sup> Congress, Senator Moynihan introduced his statistical commission bill, S. 144 in the 105<sup>th</sup> Congress.

<sup>21</sup> Fast track legislation generally provides for a bill to be heard within a specific period of time, often without consideration by the committees of jurisdiction, and often without amendment. Thus forcing Congress into a simple yea or nay on the proposal. Fast track procedures are often used on trade agreements.

In addition, the champions of dismantling the Department of Commerce replaced their proposal to shift the Commerce statistical agencies to Labor with Rep. Horn's federal statistical service. It is worth noting that in addition to creating a statistical service, Title IV of H.R. 2667 "To Dismantle the Department of Commerce" called for more authority and staff for the Chief Statistician, uniform confidentiality provisions for statistical data, data sharing, and that the decennial census functions and budgets should be segregated from other Census Bureau activities.

Very near the end of the 105<sup>th</sup> Congress, October 6, 1998, the Senate Governmental Affairs Committee reported to the Senate the bill S. 1404. This version of the bill reflected changes in the fast track provisions to accommodate concerns of the administration and members of the Committee. The full Senate was prepared to pass the bill once it received a signal from the House that it would also pass there. Unfortunately, that signal never came.

There was consensus among the House staff that in order for the bill to pass they needed both parties to agree. If either party had a serious problem, the bill was dead. It did not take long for such a problem to arise. Shortly after serious discussions of passing the bill began, the AFLCIO raised an objection. The union would not, under any circumstances, endorse a bill that, as this bill did, removed the Bureau of Labor Statistics from the Department of Labor. The Bureau of Labor Statistics was placed in the Department of Labor, the union representative argued, to protect its mission to fairly represent labor. The struggle to protect that agency was as long as the union's history, and not negotiable. The union would write to each and every Congressman and Senator voicing its opposition to the bill as long as the provision to move the Bureau of Labor Statistics remained in the bill.

The 105<sup>th</sup> Congress ended without passage of any bill that moved a statistical agency, established a commission, or provided new language on confidentiality and data sharing.

### **Act III. Reorganization Retreats: The 106<sup>th</sup> Congress**

At the opening of the 106<sup>th</sup> Congress, Senator Moynihan decided he no longer wanted to play in the game of merging his interests with those of Rep. Horn. On January 19, 1999 he introduced S. 205 "To Establish a Federal Commission on Statistical Policy." This bill was a return to his original interest in a commission that would consider the structure of the American statistical system in an international context, consider the issue of consolidation in that light, and examine substantive issues like social security and the cost of living index. Later that year, Reps. Horn and Waxman introduced H.R. 2885 as the "Statistical Efficiency Act of 1999." This was the administration's bill without modification.

In the 105<sup>th</sup> Congress, the process of abolishing the Department of Commerce was pursued to its logical conclusion. Each House committee with jurisdiction over a part of the department marked up its portion of the bill. That included nearly all of the committees in the House. In most cases, each committee, believing that the work done by its part of the agency was vital to the government, created an independent agency to continue that work. The various pieces were then assembled and brought before the Committee on Government Reform. That

bill created several new agencies, each with inspectors general and full blown bureaucracies. In the end, the sum of the parts was greater than the whole, and the bill died after Committee action.

With reorganization of the Department of Commerce off the table, and reorganization of the statistical agencies taken off the table by the AFLCIO, Congress pursued a more modest agenda. H.R. 2885 was reported by the Committee on Government Reform, and passed the House under Suspension of the Rules on October 26, 1999. Now all that was left was for the Senate to pick up where it left off at the end of the 105<sup>th</sup> Congress.

Unfortunately, passing legislation is not that simple. The Senate Committee on Governmental Affairs had to consider the legislation anew, and gain consensus of its members, some of whom had not been through the negotiations over the statistics bill in the 105<sup>th</sup> Congress. Not only had the players in the Senate changed, but so too had those watching the bill from the outside.

It is important to understand a second legislative battle that affects data access before proceeding. In the 106<sup>th</sup> Congress, the Environmental Protection Agency issued regulations on clean air standards. Those regulations were developed, in part, on data collected and analyzed under a grant from EPA to researchers at Harvard. Industry representatives questioned the conclusions drawn from that research, but without access to the data they could not challenge the conclusions. EPA said that it could not provide access because the data belonged to the researchers at Harvard, and those researchers refused to provide access to the data.

The result was language in an appropriations bill that has come to be known as the Shelby Amendment after Sen. Shelby of Alabama. The Shelby Amendment says that any data collected under contract or grant funded by the federal government that is used to develop agency policy or regulations must be made available to the public through the provisions of the Freedom of Information Act. In effect, Congress legislated access to all contract and grant data used for policy development because researchers in this one instance refused to provide access, and the agency claimed it had no control over or access to those data. The practical effect of the language was mitigated in the administrative regulations developed by OMB.

As negotiations over the data sharing bill proceeded, a public policy think tank raised a serious concern. The fear was that the federal government would use this data sharing authority to create data sets in support of government policy or regulation that would be unavailable to the public. For example, the administration could draw upon data from population and business surveys to develop a data set to analyze the effects of raising the minimum wage on the number of jobs available, a topic that regularly draws heated debate and disagreement among economists. Regardless of the conclusions from that analysis, opponents of the conclusion would have no way to challenge the results unless some right of public access like that provided by the Shelby Amendment were included in the data sharing legislation. The Administration refused to include language that paralleled the Shelby Amendment. It did, however, share with representatives of the think tank alternative language that would be acceptable to the Administration. Unfortunately, those negotiations failed to reach consensus before Congress adjourned and the legislation died at the end of the 106<sup>th</sup> Congress.

## Act IV. The 107<sup>th</sup> Congress

- A) **H.R. 2136:** Sawyer and Waxman, June 12, 2001, “Confidential Information Protection Act”
  - 1) This bill creates the uniform confidentiality protection of information collected for statistical purposes under a pledge of confidentiality.
  - 2) One of a series of privacy bills introduced by Rep. Sawyer in the 107<sup>th</sup> Congress.
- B) **H.R. 5215:** Horn, Sawyer, and Maloney, July 25, 2002, “Confidential Information Protection and Statistical Efficiency Act of 2002”
  - 1) A new bill put forward by the administration that has uniform confidentiality as Title I, similar to the Sawyer bill, and data sharing as Title II.
  - 2) Data sharing provisions limited to business data at the request of the Census Bureau.
- C) **H.R. 5215: Reported in the House.** Horn, Sawyer, and Maloney, July 25, 2002, “Confidential Information Protection and Statistical Efficiency Act of 2002” House Report 107-778.  
Signed into law by the President in December 2002 as Title V of the Electronic Government Act.

The landscape of the 107<sup>th</sup> Congress was quite different than preceding Congresses. While Republicans held on to a slim majority in both the House and the Senate, they had captured the presidency. No longer was Congress at odds with the Administration, and the leadership of both houses talked of a legislative agenda that would address the needs of the new administration.<sup>22</sup> Despite this change, at the beginning of the 107<sup>th</sup> Congress, most congressional staff believed that there was little use in pursuing the data sharing legislation. Rep. Horn’s staff was not interested in introducing the bill until there was some indication that passage was possible, and there did not seem to be anyone working to make it possible.

Rep. Sawyer, as a member of the Commerce Committee decided to introduce a series of bills that addressed information privacy in the commercial arena. To complete his package of legislation, he instructed his staff to work with the Government Reform Committee to develop a bill that separated the confidentiality provisions from the data sharing provisions. H.R. 2136, the Confidentiality Information Protection Act, was introduced by Reps. Sawyer and Waxman in June 2001 and referred to the Committee on Government Reform, which subsequently referred the bill to Rep. Horn’s Subcommittee on Government Efficiency. There was little interest in moving the Sawyer/Waxman bill.

The prospect of a data sharing bill passing the 107<sup>th</sup> Congress changed considerably when the Administration delivered a revised version of the bill from the 106<sup>th</sup> Congress in the Summer of 2002. The primary revision to the bill was that the data sharing provisions were limited to three agencies, the Bureau of Economic Analysis, the Bureau of Labor Statistics, and the Census Bureau. In addition, the data sharing provisions were limited to data on businesses and corporations.

---

<sup>22</sup> This harmony was disrupted in the middle of the first session when Sen. Jeffords left the Republican Party, became an Independent, and control of the Senate shifted to the Democrats.

The Administration explained that the concerns over privacy surrounding the 2000 census, and the heightened concerns about privacy in general, had led to the exclusion of household data from the bill. With that exclusion, it believed, passage would be simplified. Unfortunately, the road to passage would be neither simple nor straightforward.

The House introduced the bill with bipartisan support. It was marked up in Committee, and quickly passed the House under suspension of the rules.<sup>23</sup> Unfortunately, there was little interest in the Senate for moving the bill. Senator Lieberman, Chairman of the Senate Committee on Governmental Affairs, had other priorities that demanded the Committee's attention. Earlier in the 107<sup>th</sup> Congress Senator Lieberman introduced a bill to create a Department of Homeland Security, which the Administration initially opposed as unnecessary. However, by the summer of 2002, the Administration had drafted its own proposal for a Homeland Security Department, and that effort was consuming enormous time of both staff and members in the House and the Senate. Senator Lieberman had also been working on a bill on electronic government since the beginning of the 107<sup>th</sup> Congress (S. 803). The energy not consumed by homeland security was being directed to the electronic government initiative.

Lieberman's electronic government faced a somewhat similar fate in the House as the data sharing bill faced in the Senate. As a courtesy to Senator Lieberman, Rep. Jim Turner, ranking Democrat on the Government Reform Technology Subcommittee, introduced a house version of Lieberman's electronic government bill (H.R. 2458); however, there was no attempt to move that bill. Rep. Tom Davis from Virginia had his own version of electronic government legislation in the 106<sup>th</sup> Congress (H.R. 5024). However, with the change in administration, he chose to not reintroduce that bill, and gave every indication that he had no interest in moving Senator Lieberman's bill. As Congress thrashed out the details of creating the Department of Homeland Security, the fate of both Lieberman's electronic government bill, and the data sharing legislation looked grim at best.

During the consideration of the homeland security bill, Congressmen Turner and Davis worked closely together on a number of issues that fell within the jurisdiction of their Subcommittee. After the homeland security legislation cleared the House, Rep. Davis instructed his staff to explore whether or not Rep. Turner was interested in moving his electronic government bill.<sup>24</sup>

---

<sup>23</sup> The House of Representatives has an expedited process for passing bills with strong bipartisan support. Every Tuesday bills are brought before the House after both sides have agreed that the bills are non-controversial. The normal rules for consideration of bills on the floor of the House are suspended, to allow expedited consideration of the business before the House that day. To assure that these bills have strong support, the House requires that they must pass by a vote of two-thirds of the House.

<sup>24</sup> Some have suggested this change of heart was the result of a vote on the House floor during the homeland security bill. Rep. Turner offered an amendment to limit the liability of companies providing homeland security services. The House leadership preferred to limit jury awards instead. Rep. Turner's amendment failed by one vote, and some have suggested that Rep. Davis cast the deciding vote to defeat the amendment. Rep. Davis had co-sponsored the amendment

The final version of the electronic government bill that passed the House looked quite different from the bill developed by Senator Lieberman and his staff. Rep. Davis added procurement legislation of his that had passed the House, but had not been considered by the Governmental Affairs Committee. A bipartisan effort resulted in the inclusion of the Federal Information Security Management Act to restructure the administration of computer security in the government that had received considerable attention by both Rep. Davis' and Rep. Horn's subcommittees. Finally, at the request of Rep. Waxman, the Confidential Information Protection and Statistical Efficiency Act was added as Title V of the Electronic Government Act.

It is worth noting at this point that there was considerable discussion among Democratic staff about whether or not to support the inclusion of the Confidentiality and Data Sharing provisions in the Electronic Government Act. Between the end of the 196<sup>th</sup> Congress and the middle of the second session of the 107<sup>th</sup> Congress, the climate of government secrecy had changed considerably.

The Administration had made a number of changes in information policy to decrease access to government information. Attorney General Ashcroft issued a memorandum that instructed agencies to revise their approach to Freedom of Information requests. The new rule of thumb was to withhold documents if there were any doubt. Similarly, the Executive Order on classification of documents was rewritten. Gone were the instructions when in doubt about classification don't classify, and when in doubt about the level of classification, classify at the lower level. Not surprisingly, this resulted in an increase in the classification of documents, and a decrease in declassification.

In the debate over the Department of Homeland Security, the Administration insisted on a provision giving the Department an exemption from the Freedom of Information Act for dealing with information about critical infrastructures.<sup>25</sup> The Administration also insisted that the Department be exempted from the Federal Advisory Committee Act for most advisory committees. At the same time, Congress and the public found it harder and harder to get information from the administration.

In addition to increased government secrecy, Congress was faced with repeated attempts by agencies to use government and private data for surveillance of citizens and foreigners alike. The Transportation Security Administration (TSA) proposed culling credit card records, medical records, and government records to screen passengers at airports. When faced with objections to this program, TSA revised the program, but refused to identify what data would be scanned and who would have access to those data. The Defense Department proposed a similar data-mining project that was blocked in the appropriations process. The Patriot's Act provided the Attorney

---

when it was accepted in the Government Reform Committee markup of the bill. If Davis' vote was the deciding vote, he may have believed he owed his colleague a favor.

<sup>25</sup> The Department expanded this exemption in regulation applying it not only to information submitted by businesses to the Department of Homeland Security, but also to information provided through other agencies.

General access to survey records at the National Center for Education Statistics for law enforcement purposes.

In the face of increased government secrecy about what it was doing, and increased attempts to use a variety of informational databases for surveillance purposes, Congress had to decide whether or not it was wise to create a new data sharing authority that would permit the compilation of massive databases. In the end, the decision was to go forward. If the decision were to be made today, I doubt the decision would be the same.

Only one roadblock remained to the inclusion of the data sharing legislation in the E-Gov Act -- the Congressional Budget Office wanted legislative language that would assure its access to survey records matched to IRS earnings data held by the Social Security Administration.

Late in the 106<sup>th</sup> Congress, the Congressional Budget Office (CBO) sought legislative language that would give the CBO access to nearly all data collected by the Census Bureau. That effort was stopped primarily because it happened in the middle of the collection of data for the 2000 census. Few in Congress were willing to risk disrupting the census data collection for CBO.

CBO's need for the data stemmed from legislation passed in the 105<sup>th</sup> Congress. That legislation charged CBO with the task of developing a micro-simulation model to be used to model proposed changes in the Social Security System. To complete this task, CBO sought access to a data set held by the Census Bureau which matched data from the Survey of Income and Program Participation with earnings records (property of the IRS) held by the Social Security Administration. This matched data set existed, and had been used by employees both of the Census Bureau and the Social Security Administration. Experts in micro-simulation agreed that if CBO was to undertake this task, the matched data set was essential.

The IRS had agreed that CBO could access the earnings records, but the Census Bureau also had to agree to allow CBO access to the survey data. CBO had sought access to these data through negotiations with the Census Bureau; however the two agencies could not come to an agreement.

As the data sharing legislation approached passage as part of the Electronic Government Act, the issue of CBO access to this matched data set again came before Congress. Nearly two years had passed since the first attempt, and negotiations between the two agencies had showed little progress until just before the discussions surrounding the E-Gov bill when the Census Bureau finally agreed to provide CBO with access to the matched data set. The conditions were that CBO employees became sworn census agents, and that CBO physically perform its analyses at a Social Security site a few blocks from CBO headquarters.

Despite this progress, congressional staff, and in particular, administrators at CBO believed that a legislative signal was needed to assure continued access. The immediate reaction from representatives of the Administration was that CBO access should not be a part of the data sharing legislation.

The first line of objection was that there was no need for legislative language because CBO was getting access to the data it needed. Second, the agencies argued, federal statistical agencies have a special relationship with those from whom they collect information. Sharing with other statistical agencies did not breach that relationship, but sharing with CBO did, because CBO had no such special relationship with businesses or individuals, and CBO had no history of protecting confidential information. Third, the agencies argued that CBO was not a statistical agency and to treat it such in legislation was improper.

Congressional staff argued that the very lengthy and protracted process of gaining access to the matched Census/IRS/SSA data set was exactly the reason that legislation was needed. If the current situation allowed for federal agencies to delay access for years, the system needed to be changed. CBO staff pointed out that CBO did collect data from businesses, and had a relationship with them that was not unlike that of the statistical agencies, and that CBO routinely accessed individual and business tax records without a single breach in the protection of that information. Finally, congressional staff pointed out that CBO functions as Congress's statistical agency and not a policy agency. Congress calls upon CBO to develop statistical models and information and then the Congress makes policy based on that information. In fact, the situation that precipitated the conflict was when Congress called upon CBO to develop a statistical model that could be used to illustrate the effects of various changes in the Social Security System so that Congress could decide which policy to adopt.<sup>26</sup>

At an one of the early meetings of principles, Rep. Horn, whose Subcommittee was responsible for the data sharing bill said that if CBO needed the data, we should just write them into the bill and be done with it. He saw no reason to debate the issue further.

A series of drafts of proposed language went back and forth, often raising the heat of the debate rather than furthering the negotiations. The agencies proposed language that said, in effect, nothing in this bill changes current law. Congressional representatives rejected that proposal and responded with language that said that if an agency determined that it could not provide CBO with access to the data it needed to complete analyses required by Congress, then the agency must provide CBO with the analysis needed. That proposal infuriated administration representatives who believed the language subjugated them to CBO. It became clear, however, that if the data sharing language were to become law this issue would have to be resolved. Congress was willing to kill the confidentiality and data sharing provisions if CBO didn't get the access it needed. Similarly, in the 106<sup>th</sup> Congress the administration was willing to forgo passage rather than include Shelby like language. In the end, language was agreed upon, and the bill became law.<sup>27</sup>

---

<sup>26</sup> Others have suggested that CBO is not a statistical agency because it does not fall under the definition of agency as specified by Congress in legislation. Congress defines agencies in nearly every piece of legislation enacted. The most common reference is to Sec. 105 of Title V of the United States Code. If that were the only barrier, it would have been simple to correct that in the legislation.

<sup>27</sup> The final language, Sec. 504(e) reads as follows:

“(e) SECTION 201 OF CONGRESSIONAL BUDGET ACT OF 1974.—

## Some Examples of Data Sharing

A significant amount of data collected by statistical agencies is not held in strict confidence. By strict confidentiality, we mean that the data are held and used only by the agency that collected the data. The Census Bureau has a number of ways in which it shares data. For example, survey data from the Current Population Survey and the Survey of Income and Program Participation (SIPP) are shared with the Social Security Administration. Some of this sharing is for survey administration purposes, like the verification of Social Security numbers, and some is not. However, data from both the CPS and SIPP are matched to Social Security administrative records and historical data on earnings are appended to the survey records. Those historical earnings data cover an individual's annual lifetime earnings covered by Social Security and are technically the property and responsibility of the Internal Revenue Service. Consequently in order to create these matched datasets, the agreement must include three agencies -- the Census Bureau, the Social Security Administration, and the Internal Revenue Service. Those data are matched and used for a variety of research activities, and access to those data is available to researchers at the Census Bureau and the Social Security Administration.

I am not arguing that these matched data sets are not valuable. Indeed, it was just such data that was at the center of the controversy over including access for the Congressional Budget Office in the Statistical Efficiency Act. I am arguing, however, that these sharing arrangements violate a strict notion of confidentiality and that the consequences of these violations may be changing. These data sharing arrangements fall under the second order definition presented above. The data are stripped of identifiers like name, address, and social security number, but contain sufficient detail that individuals could be relatively easily identified. I argue, however, that does not relieve the agency of the responsibility to inform respondents of this use of their information.

Data from the mandatory economic censuses are matched with data from the voluntary economic surveys<sup>28</sup>. These matched data, while held within the confines of the Census Bureau, may violate no rules of confidentiality, although it is unlikely that the responding businesses know of these matches. However, researchers have made wide use of these data under agreements with the Census Bureau and under sworn authority. Again, I argue, swearing in agents, and extensive research proposals does not relieve the agency of the responsibility of informing the respondents of this use of their information.

---

This title, including amendments made by this title, shall not be construed to limit any authorities of the Congressional Budget Office to work (consistent with laws governing the confidentiality of information the disclosure of which would be a violation of law) with databases of Designated Statistical Agencies (as defined in section 522), either separately or, for data that may be shared pursuant to section 524 of this title or other authority, jointly in order to improve the general utility of these databases for the statistical purpose of analyzing pension and health care financing issues.”

<sup>28</sup> One example is known as the Longitudinal Establishment Data held by the Center for Economic Studies at the U.S. Census Bureau.

As recently as 1999, a shudder went through the statistical system when the Internal Revenue System, after auditing the confidentiality procedures used by the Census Bureau, shut down all data sharing that involved the use of IRS data. That, in effect, shut down virtually all Census Bureau data sharing. The charge made by the IRS was that the Census Bureau was sharing information that belonged to the IRS in ways and for purposes that were not allowed under the agreement between the agencies that gave the Census Bureau access to those data. One positive effect of this event is that the Census Bureau had spend a considerable amount of time and energy developing agency wide procedures for handling both requests for access to confidential information, and establishing set procedures for how those requests are reviewed and adjudicated.

The National Center for Education Statistics collects data from children, teachers, and school administrators all under a pledge of confidentiality. Those data are then licensed widely to academics and private research firms for analysis. The agency has stated that numerous agreements that have been made to provide access to confidential data. So much so that it was hard to argue against giving the Attorney General access to education records in the Patriot's Act following the attack on the World Trade Towers. Again, I argue, if agencies are going to license information, the agency has an obligation to clearly inform respondents of these uses.

### **Implementation**

Now that the data sharing bill is law, the difficult task of implementing the law is before us. In the last section of this paper I would like to discuss what I believe are critical questions that must be addressed in developing the implementation policy for Title V of the Electronic Government Act. The concept of privacy is multi-dimensional and transitory, as is the definition of confidentiality. Consequently, the implementation guidelines should be designed to explicate the terms and conditions of the contract between the citizen and the government agency. The government must tell its citizens exactly what will be done with the information the citizens provide to the government. Similarly, the various precedents provided by privacy legislation suggest that agencies should develop policies that explicitly acknowledge that respondents maintain some right of ownership of information provided to the government, even after personal identifiers are stripped. Policies for sharing data should acknowledge and honor that ownership.

The following are an example of the kinds of questions that should be addressed:

- 1) What rights do respondents have with regard to the information they share with the government?**
- 2) How does the agency collecting the information inform the public of its rights?**
- 3) If the agency chooses to share information collected from the public under a law such as Title V of the Electronic Government Act, what obligations fall upon the agency?**

- 4) **Once a shared data set is created, presumably for agency purposes, who else can access those matched data?**
- 5) **To what extent does use alter public access?**
- 6) **Who decides who can have access?**
- 7) **What governs user fees charged by agencies for access to those data?**

**1. What rights do respondents have with regard to the information they share with the government?**

The voluntary nature of surveys gives citizens the right to refuse to participate. Interviewers use a variety of explanations to convince reluctant participants including the use of the information to improve government functions, civic duty, and that the data will be held in confidence, viewed only by the employees of the agency collecting the data. In addition, the principles underlying the Privacy Act<sup>29</sup> give the respondent the right to see and correct his or her data. Since the information is given in a question and answer format, this is usually not a problem. However, the possibility that these data might be matched with data from another data collection (administrative or survey data), or that they might be matched with data from another agency puts these privacy rights in a different light. A policy that guides the development of matched data sets should recognize that the match would create data that the respondent never anticipated being combined when responding. Does the respondent have a right to review the matched data? Will the respondent ever know that the matched data exist? Both are questions that should be addressed.

**2. How does the agency collecting the information inform the public of its rights?**

The process of notifying respondents of their rights varies tremendously from survey to survey. This will vary by how the survey is taken, as well as decisions made by those sponsoring the survey. The Current Population Survey tells respondents that their answers will be held in confidence. It does not go beyond that. Interviewers, in trying to convert a reluctant respondent, use confidentiality as a selling point. In neither situation is the respondent advised that his or her answers might be matched with those from other surveys or administrative records.

Despite the fact that the Patriot's Act gives the Attorney General specific privileges to access individual records held by the National Center for Education Statistics, the agency web site is vague about just what that means. Similarly, the agency web site talks about the fact that it licenses the data for research, it is unclear that the average respondent would understand that licensing means that records stripped of personal identifiers, but that often contain sufficient detail to identify individuals, are being shared at universities and colleges around the country.

---

<sup>29</sup> Sec. 552b of Title V, United States Code.

Economic surveys generally notify respondents that the data are held in confidence on the survey form; however, the form does not indicate that those records might be matched to records from the same company provided as part of the mandatory economic censuses.

The communication with respondents during a survey is structured to enlist as much cooperation as possible. That is, of course, the agency's job. This communication does not, however, go on to fully inform the respondent to the ways in which the data might be used by matching to other records. The public policy question that must be addressed is to what extent does an agency have a responsibility to tell respondents how their data will be used? If the respondent is told that the information will be used "only for statistical purposes", is that warning sufficient to permit the agency to share those data with other agencies, academic researchers, or the public policy think tanks?

**3. If the agency chooses to share information collected from the public under a law such as Title V of the Electronic Government Act, what obligations fall upon the agency?**

As statistical agencies begin sharing information collected under a pledge of confidentiality, it is imperative that they provide the public with complete information about the ways in which those data will be used. To do so is in the best interest of the agency and the public. A vaguely worded pledge of confidentiality that simply says that the data will be used only for statistical purposes, does not fully inform the public of how the data will be used. That vagueness can easily leave the respondent with the impression that he or she was misled.

The defense for this sharing is that those outside the government are made sworn employees, or sign an agreement that subjects them to penalties for the breach of respondent confidentiality. While those provisions meet the letter of the law, it is not clear that they meet the spirit of either the Privacy Act or agency language on confidentiality.

The purpose of the data sharing legislation passed as Title V of the Electronic Government Act was to increase the sharing of data across federal agencies. The language gives the administration broad authority to initiate data sharing agreements, and to monitor those agreements. All that is required is a single annual report by OMB. The remainder of how this law is administered is left to the administration to specify in agency guidance.

What then needs to be specified? I would argue that there are two issues that should be addressed at the outside of any guidance. First, what has the agency done, and what will it do, to inform respondents that the information they voluntarily provided the government is being used in new ways. Second, OMB must establish a process where proposals for data sharing are made public, and there is an opportunity for public comment. The process for developing regulations has such a process for notice and comment. The paperwork clearance process has a process for notice and comment. So too should these data sharing agreements receive adequate public scrutiny.

Oversight for the data sharing process falls to OMB. The final decision on whether or not a data-sharing project goes forward falls to OMB. It follows then that OMB should be responsible for the public notice and comment.

**4. Once a shared data set is created, presumably for agency purposes, who else can access those matched data?**

The question of who gets access to shared data is a more complicated one. Current practice is in need of revision. However, developing an appropriate set of principles to guide access is difficult. What is clear, however, is that Congress is willing to step in where need be to alter the rules of access. Three examples illustrate this point: 1) the controversy over CBO access to matched survey and earnings records; 2) the Shelby Amendment, which provides access to data collected under grants from the federal government; and 3) the Sabo bill that prohibits the use of copyright to protect the results of research substantially funded by the federal government.

As noted above, the current system allows for extensive data sharing across agency boundaries. Data are shared with academic institutions and researchers. The issue is not whether or not the data are secure within these sharing arrangements. The issue is how these decisions are made. Currently there is no public notice that this sharing is going on. Access to the data is limited by the agency holding the data. The idiosyncratic nature of this process almost brought the data sharing legislation down before it was passed in the 107<sup>th</sup> Congress.

Earlier in the paper I detailed the debate over providing CBO with access to survey records matched to IRS earnings history data. Those arguing against access by CBO were unable to clearly articulate the basis on which the decision to grant access was made. The inability to describe how access is granted, and why CBO failed to meet those criteria, made the objections to CBO access seem arbitrary. A second example is the data access issue that gave rise to the legislation that is commonly known as the Shelby Amendment, which provided for public access through the Freedom of Information Act to data used to support agency development of regulations.

In the 106<sup>th</sup> Congress, the Environmental Protection Agency issued regulations on clean air standards. Those regulations were developed, in part, on data collected and analyzed under a grant from EPA to researchers at Harvard. Industry representatives questioned the conclusions drawn from that research, but without access to the data they could not challenge the conclusions. EPA said that it could not provide access because the data belonged to the researchers at Harvard, and those researchers refused to provide access to the data.

The result was language in an appropriations bill that has come to be known as the Shelby Amendment after Sen. Shelby of Alabama. The Shelby Amendment says that any data collected under contract or grant funded by the federal government that is used to develop agency policy or regulations must be made available to the public through the provisions of the Freedom of Information Act. In effect, Congress legislated access to all contract and grant data used for policy development because researchers in this one instance refused to provide access,

and the agency claimed it had no control over or access to those data. The practical effect of the language was mitigated in the administrative regulations developed by OMB.

The point I hope to make from these two examples is that in general Congress does not pay a great deal of attention to issues of statistical policy or data access. However, Congress will act swiftly when it sees an institutional need for information. In both cases, one of the underlying, and unspoken issues was reluctance on the part of the executive branch to share information with the legislative branch. When faced with such reticence, Congress will legislate to get what it wants. If the data sharing legislation has within it the seeds to bear the ripe fruit that many have suggested, these conflicts between Congress and the executive branch will arise again. If there is in place a clear and logical set of policies to govern who gets access to these matched data sets, much of the conflict can be avoided.

That said Congress is not the only stakeholder. Academics, think tanks, trade associations, and public interest groups will all pressure for access. It is clear in a number of arenas from rocket science to public health that there is talent outside the federal government that can be fruitfully be brought to bear on problems of public policy. This is surely true in the realm of statistical information as well. If there is a public policy purpose for developing matched data sets, then we need to develop ways to bring a richer set of talents to bear in analyzing those data sets.

The solution both to sharing data within the government, and bringing public talent to bear on confidential data sets is through public information and education. Blind promises of strict confidentiality must be replaced with accurate and informative descriptions of the ways in which the agency intends to use those data. When unanticipated uses arise, the agency must do everything possible to inform those who voluntarily provided private information and seek permission for new uses of those data.

## **5. To what extent does use alter public access?**

One of the many roadblocks that kept the data sharing legislation from becoming law was the issue of how the matched data sets would be used. As noted above, by enacting the Shelby Amendment Congress has already specified that the way in which data are used can change the right of public access. A similar issue arose as we attempted to pass the data sharing legislation in the 106<sup>th</sup> Congress. Analysts at a public policy think tank raised an objection to the legislation that echoed the concerns of the Shelby Amendment.

The think tank analysts argued that under the provisions of the data sharing language the executive branch could create elaborate data sets, which with the use of sophisticated statistical analyses could be used to inform public policy. For example, the administration could draw upon data from population and business surveys to develop a data set to analyze the effects of raising the minimum wage on the number of jobs available, a topic that regularly draws heated debate and disagreement among economists. Regardless of the conclusions from that analysis, opponents of the conclusion would have no way to challenge the results unless some right of public access like that provided by the Shelby Amendment were included in the data sharing

legislation. The Administration and those representing expanded access were unable to reach agreement and the legislation died in the Senate.

The scenario is not just an abstract argument. If the data sharing legislation creates the potential for creating new and useful data sets for public policy, then inevitably, just such a conflict will arise. Those new and useful data sets will be used for public policy, and the quality of the analysis will be questioned. Science proceeds on the basis of open access and the replication of results, and those principles must be met by federal agencies as well.

## **6. Who decides who can have access?**

The problem then is who should decide who should have access and how should that access be provided. As we saw above in the example of CBO access, if such decisions are left to agencies progress will be slow. Unfortunately, OMB may not be willing to break new ground in defining access to confidential data. However, given the oversight and administrative responsibilities assigned to OMB in the bill that is the logical place for such access rules to be developed. The alternative is legislation.

## **7. What governs user fees charged by agencies for access to those data?**

The last issue to be addressed is whether or not agencies can or should assess fees of those who wish to use these matched data sets. This is not an issue for the agencies participating in the match. They are free to negotiate reciprocal use arrangements as a part of any data sharing agreement. However, agencies should not be left to their own devices to determine access fees for matched data sets. Again, the responsibility logically falls to OMB.

In the world of print, Congress had clearly established that information collected at government expense should be available to the public at a minimal cost. This has historically been defined as the marginal cost of distribution. This principle has not been uniformly accepted. During the early 1990s Congress repeatedly fought with the National Library of Medicine, which was charging licensing fees to a select set of bidders who then provided on a fee basis access to the medical library database of publication summaries and citations. This valuable information set was developed at public expense, but was not available to the public except through expensive subscription services. Fortunately, the technology of the World Wide Web, and enlightened leadership at the National Library of Medicine has reversed this situation. STAT USA, which sells subscriptions to government information, raises similar questions about public access.

The Sabo bill, recently introduced in the House, reiterates this principle applied to the results of research funded by the federal government. The Sabo bill states that copyright protection is not available to the results of research substantially funded by the federal government.

It is my understanding that after all of the negotiations over research content, the swearing in of researchers, and the defining of census value to the research, there is a charge of \$4000 a month to researchers who want to do research using a data set which matches economic

census records to records for the same firms from the economic surveys. This longitudinal data set has a number of valuable applications some of which are rejected. The Census Bureau argues that it can only sanction research that requires access to confidential data if the research has a “census purpose.” If the research has a census purpose, it is difficult to understand why the agency is charging those doing research on the agency’s behalf. Logic would suggest that the agency should be paying the researchers for enhancing the Census Bureau’s understanding of the data.

Clearly this too is an area where strong OMB guidance is needed. It is too easy for the current situation to be misconstrued as a private club for data users where the agency bureaucrats comprise the admissions committee. That, in the long run, will undermine public confidence in both the quality of the analysis and the agencies themselves.

### **Summary**

In summary, it is clear from the legislative history on privacy that when a person provides information to the government, or to a private entity, they do not give up all rights to how those data are used. Conversely, when the government receives information from the public, it is not free to use that information in ways other than for the purpose for which the information was collected.

The central issue then is what are the terms of this implied contract between the citizen and his or her government, and what is the government’s responsibility to make those terms clear and explicit. For information collected under mandatory authority, the OMB guidance makes it clear that there should be no reuse of that information without explicit consent from the individual providing the information. For information collected voluntarily, there are more options, and thus the solution is more complex.

Agencies collecting information from the public should be as clear and detailed as possible in explaining to the respondent how the information will be used once it is in the hands of the government. If the information is going to be matched to IRS records and maintained as a research file that should be stated explicitly. There are two issues at play here. First, even if the information is stored without personal identifiers like name, address, and social security number, there is often sufficient information to identify the individual. Second, after the information is in the government’s hands, the individual retains some rights to how that information is used. The agency collecting the information should explain to the respondent that many uses require sufficient information that complete anonymity is not possible, and respect the rights of the individual. In addition, public notice should be given for any match of information across agencies, not just those conducted under the authority of Subchapter B of Title V of the Electronic Government Act.

Finally, we should keep in mind that our government is based on open access to the citizens it serves. That openness should be one of the principles that guide the development of policies about informing respondents of their rights and responsibilities when asked for information.

## Appendix A

### SEC. 512. LIMITATIONS ON USE AND DISCLOSURE OF DATA AND INFORMATION.

(a) USE OF STATISTICAL DATA OR INFORMATION.—Data or information acquired by an agency under a pledge of confidentiality and for exclusively statistical purposes shall be used by officers, employees, or agents of the agency exclusively for statistical purposes.

(b) DISCLOSURE OF STATISTICAL DATA OR INFORMATION.—

(1) Data or information acquired by an agency under a pledge of confidentiality for exclusively statistical purposes shall not be disclosed by an agency in identifiable form, for any use other than an exclusively statistical purpose, except with the informed consent of the respondent.

(2) A disclosure pursuant to paragraph (1) is authorized only when the head of the agency approves such disclosure and the disclosure is not prohibited by any other law.

(3) This section does not restrict or diminish any confidentiality protections in law that otherwise apply to data or information acquired by an agency under a pledge of confidentiality for exclusively statistical purposes.

(c) RULE FOR USE OF DATA OR INFORMATION FOR NONSTATISTICAL PURPOSES.—A statistical agency or unit shall clearly distinguish any data or information it collects for nonstatistical purposes (as authorized by law) and provide notice to the public, before the data or information is collected, that the data or information could be used for nonstatistical purposes.

(d) DESIGNATION OF AGENTS.—A statistical agency or unit may designate agents, by contract or by entering into a special agreement containing the provisions required under section 502(2) for treatment as an agent under that section, who may perform exclusively statistical activities, subject to the limitations and penalties described in this title.