

# MANAGING INFORMATION PRIVACY AND INFORMATION ACCESS IN THE PUBLIC SECTOR

George T. Duncan  
H. John Heinz III School of Public Policy and Management  
Carnegie Mellon University  
Pittsburgh, PA 15213

## ***Abstract***

Government agencies collect and disseminate data that bear on the most important issues of public interest. Advances in information technology, particularly the Internet, have multiplied the tension between demands for evermore comprehensive databases and demands for the shelter of privacy. In mediating between these two conflicting demands, agencies must address a host of difficult problems. These include providing access to information while protecting confidentiality, coping with health information databases, and ensuring consistency with international standards. The policies of agencies are determined by what is right for them to do, what works for them, and what they are required to do by law. They must interpret and respect the ethical imperatives of democratic accountability, constitutional empowerment, and individual autonomy. They must keep pace with technological developments by developing effective measures for making information available to a broad range of users. They must both abide by the mandates of legislation and participate in the process of developing new legislation that is responsive to changes that affect their domain. In managing confidentiality and data access functions, agencies have two basic tools: techniques for disclosure limitation through restricted data and administrative procedures through restricted access. The technical procedures for disclosure limitation involve a range of mathematical and statistical tools. The administrative procedures can be implemented through a variety of institutional mechanisms, ranging from privacy advocates, through internal privacy review boards, to a data and access protection commission

## THE TENSION BETWEEN PRIVATE LIVES AND PUBLIC POLICIES

### ***An Environmental Scan***

In its normal activities, the public sector captures enormous amounts of data, stores it in very large databases, analyzes some of it, and disseminates information products to individuals, governments, businesses, and other organizations. Much of these data are obtained directly from respondents in surveys and censuses or through building systems of administrative records based on a variety of citizen interactions with government. Surveys include:

- Face-to-face interviews, as with the National Longitudinal Surveys of Young Women conducted by the Bureau of Labor Statistics
- Telephone surveys, as with the Behavioral Risk Factor Surveillance System conducted by the Center for Disease Control, which estimates current cigarette smoking and use of smokeless tobacco
- Mail-back responses (including electronic mail), as with the reactions to their web site obtained by Inland Revenue of the Government of New Zealand (see **Error! Bookmark not defined.**)

Administrative records include:

- Employer-furnished data, as with Social Security Administration earnings records
- Licensing data, as with state Departments of Motor Vehicles and local building permits
- Individual and firm submitted data, as with Internal Revenue Service tax returns.

The Internet has accelerated the demand for access to government information services, primarily by broadening the range of potential data users. Access demand is in commensurate tension to concerns about privacy and confidentiality. The National Science Foundation in its Digital Government program announcement (NSF 1998), affirms, "Given the inexorable progress toward faster computer microprocessors, greater network bandwidth, and expanded storage and computing power at the desktop, citizens will expect a government that responds quickly and accurately while ensuring privacy." This article focuses on ways the public sector can resolve the growing tension between the demand for government data and concerns for privacy protection.

Government databases are rich in information and have evident practical utility for planning, marketing, and research. Still, many would-be users complain they cannot obtain the data they need, often thwarted by confidentiality concerns (Smith 1991). On the other hand, privacy advocates warn of the dangers of unfettered data capture and dissemination. Their arguments are ethically based. "Individuals in the Western world are increasingly subject to surveillance through the use of data bases in the public and private sectors, and these developments have negative implications for the quality of life in our societies and for the protection of human rights." (Flaherty 1989: 1) From a more utilitarian standpoint, public policy would be affected by declining survey participation rates. In 1992, for example, 31 percent of Americans refused to answer at least one survey, compared with 15 percent in 1982 (Leftwich 1993; also see Dalenius 1993). Acrimonious public debates rage about proper use of mailing lists, credit records, medical histories, and Social Security Numbers (Flaherty 1989). The media highlights privacy concerns in the use of ever-larger computer databases--those of terabyte size, labeled "data warehouses".

Broad access to data supports democratic decision-making. Access to government statistical information supports public policy formulation in areas ranging through demographics, crime, business regulation and development, education, national defense, energy, environment, health, natural resources, safety, and transportation. Thrust against the evident value of data access is the counter value that private lives are requisite for a free society. This article deals with an important aspect of the tension between information privacy and data access—the proper handling of personal information that is collected by government. Other privacy issues such as video surveillance, telephone interception or 'bugging', Internet censorship, protecting children, encryption policy, and physical intrusion into private spaces are outside my purview.

A variety of governmental agencies have important roles in collecting, storing, analyzing, and disseminating information. Certainly this is the case with the federal statistical agencies, such as the U. S. Census Bureau, National Center for Health Statistics, and the Bureau of Labor Statistics. But it is also true of state public health organizations and functional agencies such as departments of motor vehicles. Each agency is tasked with the dual responsibilities of protecting privacy and confidentiality while disseminating information to client users, often including the general public. (Duncan, Jabine, and de Wolf 1993; Duncan and Pearson 1991).

What are some contentious issues in the clash between demands for information privacy and data access? To address this question we examine a classification of some of the bills submitted in the 105<sup>th</sup> Congress. Besides general privacy and data access concerns, bills were introduced addressing medical record privacy, the sale of consumer records, mandating databases, and copyright protection of databases.

- **General Privacy And Data Access Concerns**

**Error! Bookmark not defined..** Prohibits Federal officers and employees from providing access to social security account statement information, personal earnings and benefits estimate statement information, or tax return information of an individual through the Internet or without the written consent of the individual, and to establish a commission to investigate the protection and privacy afforded to certain Government records.

**Error! Bookmark not defined..** Prohibits Federal agencies from making available through the Internet certain confidential records with respect to individuals, and to provide for remedies in cases in which such records are made available through the Internet.

**Error! Bookmark not defined.** Creates Commission to look at statistical agencies. Commission will look at privacy implications of collection and use of statistical information.

**Error! Bookmark not defined.** Criminalizes "browsing" of tax records by IRS employees.

**Error! Bookmark not defined.** Creates penalties for abuse of information in New Hires Database. Requires data be deleted after 24 months.

- **Medical Records Privacy** (see also, Duncan 1997)

**Error! Bookmark not defined.** Sets limits on disclosure and use of genetic information in connection with group health plans and health insurance coverage, prohibits employment discrimination on the basis of genetic information and genetic testing.

**Error! Bookmark not defined.** Creates "centrally located" national database of health insurance information for processing all claims and outcomes. Gives grants to states to create and operate health care cost containment and quality information systems that contain clinical and treatment data on patients.

**Error! Bookmark not defined.** Defines the circumstances under which DNA samples may be collected, stored, and analyzed, and genetic information may be collected, stored, analyzed, and disclosed, to define the rights of individuals and persons with respect to genetic information, to define the responsibilities of persons with respect to genetic information, to protect individuals and families from genetic discrimination, to establish uniform rules that protect individual genetic privacy, and to establish effective mechanisms to enforce the rights.

**Error! Bookmark not defined.** Requires managed care group health plans to establish written policies and procedures for the handling of medical records; ensure the confidentiality of specified enrollee information; and prevent release of any individual patient record information, unless such a release is authorized in writing by the enrollee or otherwise required by law.

- **Sale Of Consumer Records**

**Error! Bookmark not defined.** to prevent the United States Postal Service from disclosing the names or addresses of any postal patrons.

- **Mandating Databases**

**Error! Bookmark not defined.** A bill to require the national instant criminal background check system to be established and used in connection with firearms transfers.

**Error! Bookmark not defined.** Requires President to recommend creation of national workplace verification system within 90 days.

- **Copyright Protection of Databases.**

**Error! Bookmark not defined.** Implements WIPO (World Intellectual Property Organization) Treaty. Ensures "fair use" of copyrighted materials.

It is evident from this listing that a variety of issues are currently on the forefront of public debate. (Such legislative activity is monitored for example by the Electronic Privacy Information Center (EPIC); see **Error! Bookmark not defined.**). The key issues include:

- restricting access to database information to those with a need to know
- struggles with how to provide access to information through the Internet while protecting confidentiality
- dealing with so-called universal identifiers, especially the Social Security Number
- abuse of databases by government employees
- establishing health information databases
- special problems about genetic information
- ensuring consistency with international standards.

While federal government agencies and certain other private sector organizations operate under federal legislative constraints related to privacy, that is not the case for other major record systems. No federal law ensures the confidentiality of medical records, and state laws vary. Similarly unprotected are

insurance files, credit card transactions, most state government records, criminal records, employment records, and phone bills. The legislative environment for privacy and information issues is currently in flux, and likely will change. Organizations that monitor legislation in this area suggest that some one thousand bills were introduced in this area at the federal and state level in 1993 alone. In the next section we examine legislation that is already on the books that both attempts to protect government data and constrains its use by agencies.

### **Relevant Legislation**

Some agencies are guided by specific legislation. The U.S. Census Bureau, for example, is governed by Title 13 of the U.S. Code, which provides for tight controls on individually identifiable data. In particular, Section 9 provides that

(a) Neither the Secretary, nor any other officer or employee of the Department of Commerce or bureau or agency thereof, or local government census liaison, may, except as provided in section **Error! Bookmark not defined.** or **Error! Bookmark not defined.** or chapter 10 of this title - (1) use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied; or (2) make any publication whereby the data furnished by any particular establishment or individual under this title can be identified; or (3) permit anyone other than the sworn officers and employees of the Department or bureau or agency thereof to examine the individual reports. No department, bureau, agency, officer, or employee of the Government, except the Secretary in carrying out the purposes of this title, shall require, for any reason, copies of census reports which have been retained by any such establishment or individual. Copies of census reports which have been so retained shall be immune from legal process, and shall not, without the consent of the individual or establishment concerned, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceeding.

Similarly for the Internal Revenue Service, Section 6108(c) of the U.S. Internal Revenue Code of 1986 stipulates that

No publication or other disclosure of statistics or other information required or authorized by subsection (a) or special statistical study authorized by subsection (b) shall in any manner permit the statistics, study, or any information so published, furnished, or otherwise disclosed to be associated with, or otherwise identify, directly or indirectly, a particular taxpayer.

The U.S. Social Security Administration enjoys comparable legislative protection of its data through Section 1106 of the Social Security Act.

On the other hand, the Bureau of Labor Statistics has no specific statutory protection to preserve the confidentiality of identifiable information. Instead, the Bureau's confidentiality policy is established by Commissioner's Order 3-93, *Confidential Nature of BLS Records*, which in Section 7(a) states that data "...collected or maintained by, or under the auspices of, BLS under a pledge of confidentiality shall be treated in a manner that will ensure that individually identifiable data will be used only for statistical purposes and will be accessible only to authorized persons." (See de Wolf (1995) for discussion of BLS confidentiality policy.)

Chapter 5 of Duncan, Jabine, and de Wolf (1993) presents a comprehensive view of the legislative environment of federal statistical agencies. Under current legislation, the degree of data protection depends on the agency that holds it without regard to the sensitivity of the information. Also, data sharing among agencies is difficult because agencies with a high degree of legislative protection of their data are reluctant to share with those with a low degree of protection. For example, the National Agricultural Statistics Service (NASS) has had a complicated relationship involving the sharing of lists of farms for the

Census of Agriculture, which is conducted every five years. The Census of Agriculture was officially moved from the Census Bureau to NASS on December 31, 1997. In conducting the 1997 Agriculture Census, the Census Bureau provided NASS employees (many of whom were Census Bureau employees) with farm list information. To do this they were made Census special sworn employees. Restricted access is provided through the Census Bureau's Bowie computer center to NASS headquarters and the regional centers. In 1992 the Census Bureau swore in a limited number of NASS employees to see data collected under the auspices of Title 13 but they were required to come to the Census Bureau's headquarters in Suitland, Maryland to see the data. The official transfer of the program was delayed until the end of 1997 so that the Census Bureau would continue to have the authority to get IRS tax return data to construct the list. (NASS had no such authority). Beginning in 2002, NASS will have to work with IRS to get access by amending the tax code or through proposed data sharing legislation if it becomes law.

In addition to this legislation there are a variety of other relevant laws and regulations at the federal level. They include, but are not limited to: Computer Security Act of 1987 (<http://www.epic.org/crypto/csa/>), Copyright Act of 1976 (<http://www.law.cornell.edu/copyright/copyright.table.html>), National Archives and Records Administration Regulations, Freedom of Information Act (<http://www.aclu.org/library/foia.html>), Information Technology Management Reform Act of 1996 (<http://www.itpolicy.gsa.gov/mke/capplan/cohen.htm>), Paperwork Reduction Act of 1995 (<http://www.law.vill.edu/chron/articles/ombdon.htm>), and the Privacy Act ([http://www.eff.org/pub/Legislation/privacy\\_act\\_74\\_5usc\\_s552a.law](http://www.eff.org/pub/Legislation/privacy_act_74_5usc_s552a.law)). Discussion of each of these can be found on the web at the indicated URLs.

## WHAT PRINCIPLES SHOULD GUIDE DATA STEWARDSHIP?

The principles set forth here for data stewardship derive from Duncan, Jabine, and de Wolf (1993). The United States, and a growing list of other countries, embraces a freedom that recognizes pluralism, public decision making based on representative democracy, and a market-oriented economy. Consistent with this ethos an ethics of information can be built on three principles: democratic accountability, constitutional empowerment, and individual autonomy. These principles can provide a useful guide for assessing the societal impacts of information policies of any organization, whether in the public sector or not. They have particular interpretations for government agencies that are explored in Section 3.

**Democratic accountability** is the assurance through institutional mechanisms, culture, and practice that the public obtains comprehensive information on the effectiveness of government policies. Prewitt (1985) explored this concept. The technology of the web is the most exciting implement for fostering democratic accountability. A quick click to the Social Security Administration's web site at **Error! Bookmark not defined.** yields SSA's Accountability Report for FY 1997. It provides full disclosure of SSA's financial and programmatic operations. This web presentation gave the agency the right to assert, "With its publication on November 21, 1997—less than 2 months after the close of the fiscal year—SSA became the first Federal agency to publish its FY 1997 Accountability Report. FY 1997 marks the eleventh year that SSA has published audited financial statements, the fourth year that SSA has received an unqualified opinion on its financial statements and the third year that SSA has been authorized by the Office of Management and Budget to streamline and consolidate statutorily required financial reporting into a single Accountability Report."

**Constitutional empowerment** is the capability of citizens to make informed decisions about political, economic, and social questions. Constitutional practice emphasizes restraints on executive excess and broad access to the political process through the direct election of representatives, as well as through separation and balance of power. Many government agencies have seized upon the web as a vehicle for providing information broadly to the citizenry. A prominent development in this regard is the plan announced on June 25, 1998 by Bruce A. Lehman, Commissioner of Patents and Trademarks, to create the largest Government database on the Internet. More than two million patents will be searchable by key

word. Including trademark information the database will comprise more than 21 million documents and require 1.3 terabytes of storage.

**Individual autonomy** is the capacity of the individual to function in society as an individual, uncoerced and cloaked by privacy. Individual autonomy is compromised by the excessive surveillance sometimes used to build databases (Flaherty 1989), unwitting dispersion of data, and a willingness by those who collect data for administrative purposes to make them available in personally identifiable form. Government agencies have both ethical and pragmatic reasons to be concerned about individual autonomy. Ethically, agencies ought to respect individual dignity and protect the personal information entrusted to them. Pragmatically, without attention to individual autonomy agencies will find it difficult to enlist the voluntary cooperation that smoothes operations.

## **THE SPECIAL ROLES OF GOVERNMENT AGENCIES: *FUNCTIONAL SEPARATION***

In implementing the three fundamental principles of democratic accountability, constitutional empowerment, and individual autonomy, a government agency should affirm a policy of *functional separation*. This policy makes a distinction between administrative data and statistical data. The distinction is on the basis of use:

- administrative data are used so that data on an individual has a direct impact on that individual
- statistical data are used to create aggregate measures that have an impact on individuals only through substantial group membership

Thus Joe Brown's liquor license application is initially part of administrative data since it is used to determine whether to issue Joe a license. When a database of such applications is used to determine whether females are issued liquor licenses as frequently as males, this constitutes a statistical use. Conceivably, such a study might affect administrative practice about license issuance in which case it might affect the chances of Joe Brown's subsequent application. This impact is due solely to Joe being male and is not determined by his particular data.

Agencies accept responsibility for protecting privacy and confidentiality for several ethical and pragmatic reasons. First, it is the right thing to do. Generally accepted ethical standards in the profession of data collection require attention to privacy and confidentiality (International Statistical Institute 1986, American Statistical Association 1989). Second, they get better data this way. Professionals believe that confidentiality pledges lower nonresponse rates and improve the quality of responses (Singer 1993). Third, the law requires it. Privacy and confidentiality protection is often mandated by legislation or regulation (Duncan, Jabine, and de Wolf 1993).

The Privacy Protection Study Commission (1977, 574) proposed,

That the Congress provide by statute that no record or information contained therein collected for a research or statistical purpose under Federal authority or with Federal funds may be used in individually identifiable form to make any decision or take any action directly affecting the individual to whom the record pertains, except within the context of the research plan or protocol, or with the specific authorization of the individual.

The National Research Council sponsored the Panel on Confidentiality and Data Access. In its report, *Private Lives and Public Policies*, the Panel made the following recommendation (Duncan, Jabine, and de Wolf 1993: Recommendation 5.1, p. 134):

Statistical records across all federal agencies should be governed by a consistent set of statutes and regulations meeting standards for the maintenance of such records, including the following features of fair statistical information practices:

- (a) a definition of statistical data that incorporates the principle of functional separation as defined by the Privacy Protection Study Commission,
- (b) a guarantee of confidentiality for the data,
- (c) a requirement of informed consent or informed choice when participation in a survey is voluntary,
- (d) a requirement of strict control on data dissemination,
- (e) a requirement to follow careful rules on disclosure limitation,
- (f) a provision that permits data sharing for statistical purposes under controlled conditions, and
- (g) legal sanctions for those who violate confidentiality requirements.

This recommendation refers to federal agencies. The issues are similar for state agencies and for countries other than the United States. I would, therefore, maintain that these fair information practices are broadly applicable to government operations.

## **PROBLEMS AND OPPORTUNITIES IN ENSURING CONFIDENTIALITY AND DATA ACCESS**

The public sector faces a variety of predicaments as well as opportunities as it seeks to fulfill its responsibilities for both confidentiality and access to data. The problems and possibilities are accentuated by economic and cultural changes, and importantly by developments in information technology. This section will examine these factors as they affect the various stakeholders in the process—the data subjects, the data users, and the agencies that have stewardship of the data.

### ***Data Subjects***

Government agencies depend on individuals, firms, and organizations to provide data that accurately reflect some of the most personal and sensitive aspects of their lives and operations. Some of this data provision is mandated by legislation, as public corporations are required to provide the Securities and Exchange Commission with filing information and individuals must file an income tax return with the Internal Revenue Service. Other data provision is voluntary, as when the Substance Abuse and Mental Health Services Administration interviews people at their residence about use of licit and illicit drugs.

Anecdotal evidence suggests that response rates of federally funded demographic surveys have been declining. To address this issue, the Federal Committee on Statistical Methodology formed a Subcommittee on Nonresponse that collected information on 26 federally-sponsored demographic surveys. Response and refusal rates remained relatively constant but noncontacts fluctuated over the 10-year period 1982-1991. They found a "core" proportion of the population that routinely refuses to participate in federally sponsored surveys.

The key ethical concepts related to data subjects are informed consent and notification. Informed consent is appropriate for truly voluntary surveys, while notification applies otherwise to data collection that is mandatory, such as the Decennial Census, or where benefits hinge on providing information, such as applications for welfare benefits. The Privacy Act requires that each person asked to supply information be informed of (1) the authority under which the information is requested, (2) the principal uses for the information, (3) the "routine uses" that may be made of the information, and (4) the implications, if any, of not providing the information. The National Research Council's Panel on Confidentiality and Data Access (Duncan, Jabine and de Wolf 1993) made recommendations for strengthening these stipulations. Noteworthy among these are requirements that data providers be notified of (1) nonstatistical uses of their data, (2) any anticipated record linkages for statistical purposes, (3) the length of time the information will be retained in identifiable form.

### ***Data Users***

Data users span a diverse range of individuals and organizations. They include academic researchers at Cambridge University, policy analysts for the American Association for Retired Persons and the National

Association of Home Builders, business economists for Wells Fargo bank, and statisticians for the Health Care Financing Administration. They include reporters for the *Toronto Star*, marketing analysts for L. L. Bean, advocates for the National Abortion Rights League, and medical insurance underwriters for Cigna. In general, data users employ the data they obtain for end uses such as policy analysis, commercial and academic research, advocacy, and education. They may also use data for various intermediate purposes such as the development of sampling frames for surveys and the evaluation of the quality of other data.

From a government agency's viewpoint, the primary concern of data users is obtaining access to data. Users desire data that are relevant, accurate, and complete. They also want a usable data format, easy accessibility (low price, little hassle, quick response), timeliness (automatic updates and corrections), and few limitations on use. All of these concerns are legitimate, and they are uncontroversial except for the last, which can raise serious confidentiality concerns.

Often unanticipated are data users who force access through legal action, often as part of a discovery process and involving a court-issued subpoena. As Duncan, Jabine and de Wolf (1993) notes, many statistical agencies lack adequate legal authority to protect identifiable statistical records from mandatory disclosures for nonstatistical uses. An example of this was the ruling that Environmental Protection Agency could not protect company survey responses from the Department of Justice's Antitrust Division for use in compliance activities.

Many data users want to further disseminate the data to other users. This secondary data provision occurs in a variety of contexts: A government agency sponsoring a survey may share the information with another agency. A motor vehicle service may pass on licensed driver information to insurance companies for a fee. This data sharing requires careful managing of the advantages of more efficient data collection against the risks of confidentiality loss. Some laws governing confidentiality of data prohibit or severely limit interagency sharing of data, even for solely statistical purposes.

### ***Organizations representing stakeholder interests***

A number of organizations represent various configurations of the stakeholder interests described above.

- **Association of Public Data Users** (APDU; see <http://www.apdu.org/>) assists users in the identification and application of public data; establish linkages between data producers and users; and bring the perspectives and concerns of public data users to issues of government information and statistical policy.
- **Council of Professional Associations on Federal Statistics** (COPAFS; see <http://members.aol.com/copafs/>) represents academic/professional organizations interested in the production of federal statistical and research data. Member organizations include professional associations, businesses, research institutes, and others interested in Federal statistics. COPAFS seeks to
  - ~ Increase the level and scope of knowledge about developments affecting Federal statistics
  - ~ Encourage discussion within member organizations to respond to important issues in Federal statistics
  - ~ Bring the views of professional associations to bear on decisions affecting Federal statistical programs.
- **Council for Marketing and Opinion Research** (CMOR; see **Error! Bookmark not defined.**) is a non-profit trade association formed to protect the interests of the marketing and opinion research industry. It encourages respondent cooperation and lobbies lawmakers to protect research from restrictive legislation.
- **American Civil Liberties Union** (ACLU; see <http://www.aclu.org/>) affirms both privacy rights and the public's right to know

- **Computer Professionals for Social Responsibility** (CPSR; see <http://www.cpsr.org/>) is a public-interest alliance of computer scientists and others concerned about the impact of computer technology on society.

Such organizations provide consequential input to government agencies in dealing with privacy and information issues.

## MANAGING CONFIDENTIALITY AND DATA ACCESS FUNCTIONS

### *General Issues*

Wide-ranging mechanisms exist to deal with conflicts about the capture and dissemination of data. They span federal legislation, interorganizational contractual arrangements, intraorganizational administrative policies, and ethical codes. They also include technological remedies such as the release of masked data that may satisfy data users needs for statistical information while posing little risk of disclosure of personal information (Duncan and Pearson 1991). In managing confidentiality and data access functions, government agencies have two basic tools for responsible provision of information: restricted data and restricted access. As developed in Duncan, Jabine and de Wolf (1993) these concepts have the following interpretations:

- **Restricted data.** Data are transformed to lower disclosure risk. This is accomplished through disclosure limitation techniques such as (1) release of only a sample of the data, (2) including simulated data, (3) "blurring" of the data by grouping or adding random error, (4) excluding certain attributes, and (5) swapping data by exchanging the values of just certain variables between data subjects.
- **Restricted access.** Administrative procedures imposing conditions on access to data. These conditions may depend on the type of data user; conditions may be different for interagency data sharing than for external data users. An example of an institutional arrangement for restricted access by external data users is the Census Research Data Center at Carnegie Mellon University (see <http://www.heinz.cmu.edu/census/>).

Various restricted access policies (Jabine 1993a,b) have been implemented in the last twenty years. Notable has been the fellowship programs run jointly by the American Statistical Association, the National Science Foundation together with four agencies, the Bureau of Labor Statistics, the Bureau of the Census, the US Department of Agriculture, and the National Institute of Standards and Technology. The Fellowship programs require that specific research projects and their data needs be evaluated. If approved, data users relocate to the agency to gain access to unrestricted data. In some cases of restricted access, for example to the Panel Study of Income Dynamics and the National Longitudinal Survey of Youth (Jabine 1993b), the researcher must post bond. The money will be forfeited if the researcher fails to honor the release agreement, say by unauthorized sharing of the data or performing analyses not specified in the proposal.

The Bureau of the Census has long sought a mechanism by which it could make detailed Census information more readily available to researchers, as well as connect Census data to other important national datasets, such as those housed at the Environmental Protection Agency and the Department of Justice, while maintaining the integrity and confidentiality of that data. With this in mind, the Bureau recently granted Carnegie Mellon University's H. John Heinz III School of Public Policy and Management the only Census Center to be housed at a university. Through access to such valuable data, the Center has attracted nationally renowned scholars to engage in inter-disciplinary, collaborative research on important policy issues.

## ***Institutional Mechanisms***

### **Privacy or Information Advocate**

A *privacy or information advocate* is a one-sided intervenor (Kaufman & Duncan 1988) whose mandate is to counterbalance power and resource inequalities among parties to a data dispute. At the U. S. federal level, the Internal Revenue Service appointed Robert Veeder, a specialist in the Privacy Act and Freedom of Information Act at the Office of Management and Budget, to be the IRS's first privacy advocate. As an example of his activity, he presented a paper entitled, *Making Information Accessible while Protecting Privacy*, with Sara Hamer, Associate Commissioner of the Social Security Administration, to a seminar of the Federal Internet Institute in December 1997.

Privacy or information advocates act to right a power imbalance by championing the position of a weaker party. It is presumably difficult for an advocate to switch gears between privacy advocacy and data access advocacy. They are quite limited in their ability to address privacy and information disputes, as advocacy is their only tool. Access to advocates may be hindered if they are located within a bureaucracy, hence obstructing their visibility.

### **Privacy and Information Clearinghouse**

A *privacy and information clearinghouse* provides a forum for intervention in disputes between information organizations and data providers as well as between information organizations and data users. It provides education and advice to those having questions and concerns about privacy and data access procedures.

An exemplar of such an institutional mechanism is the Privacy Rights Clearinghouse in California. It offers information on how consumers can protect their personal privacy. They provide a web site at <http://www.privacyrights.org/> and also operate a telephone hotline for those who seek information about privacy issues. The Privacy Rights Clearinghouse was established with funding from the Telecommunications Education Trust, a program of the California Public Utilities Commission. Some of their materials were developed through the University of San Diego, Center for Public Interest Law, which administered the PRC from its inception in 1992 to October 1996. Given the need for such a clearinghouse to have a reliable source of funding and appropriate administrative support, its existence is contingent on highly specific circumstances. It may not be possible to replicate these conditions in other states.

### **Ombuds**

Another intercessory mechanism with one-sided characteristics is an *ombuds*. An ombuds works within an agency to deal with complaints by data providers or data users.

The federal government has recently taken a step in the direction of using ombuds mechanisms for information and privacy disputes. Office of Management and Budget Circular A-130, that was revised February 9, 1996 (see <http://www.whitehouse.gov/WH/EOP/OMB/html/circulars/a130/a130.html>), provides uniform government-wide information resources management policies. In Section 9a(10) it provides that an ombuds(man) be designated by each agency. The ombuds(man) is to be a senior agency official charged "to investigate alleged instances of agency failure to adhere to the policies set forth in the Circular and to recommend corrective action as appropriate".

An ombuds provides an alternative and generally easily identifiable complaint route for those in dispute with an IO. This increases the power of data suppliers and data users. An ombuds can only be responsive to IO-specific disputes. This mechanism is limited in flexibility because the ombuds can only direct and articulate concerns. In particular, the ombuds typically does not have mediation or arbitration powers.

Access may of course be limited if the ombuds is hidden within the bowels of an IO's bureaucracy. It is essential that an ombuds be granted the authority to act with some neutrality.

### **Internal Privacy Review Board**

The function of *internal privacy review boards* is akin to that of institutional review boards (IRBs) in universities. In fact such IRBs could themselves provide "oversight mechanisms, with suitable definition of their scope to cover research uses of federal data sets, [to ensure that] adequate controls are in place to monitor compliance with data protection rules and regulations by users in the research community" (Duncan, Jabine, and de Wolf 1993: 107).

The Bureau of the Census has a Microdata Review Panel that was formally chartered in 1981 (Cox, McDonald and Nelson 1986). It is charged with reviewing policies for the dissemination of microdata files, especially public use data tapes and CD-ROM products. With a similar charge, the National Center for Education Statistics has a Disclosure Review Board which was created in 1989. It is staffed by NCES employees and a Census Bureau representative.

An internal privacy review board can vary in its influence on the balance of power in privacy and information disputes, depending on how it sees its mandate. Some such boards may see their role as simply ensuring that extant administrative rules and legislative requirements are met. Others may be more actively involved in disputes between an information organization and their data providers or their data users. An internal privacy review board can be quite responsive to specific disputes. If it is granted adequate authority, it can employ a wide range of mechanisms to resolve disputes. Unless the board is specially constituted for this purpose, access to an internal privacy review board by data providers and data users may be quite limited.

### **Administrative Review Agency**

An *administrative review agency* would derive its mandate from legislative or executive authority. It would function like the OMB Statistical Policy Office. While not having direct administrative responsibility for federal statistical agencies, the Statistical Policy Office provides long-range planning for statistical programs and coordinates statistical policy within the federal government. The office reviews all data collection requests developed by the Census Bureau and the Bureau of Economic Analysis. Other federal agencies submit their data collection requests, including those for statistical purposes, to OMB clearance officers who are not part of the Statistical Policy Office.

An administrative review agency typically would have as part of its charge the responsibility to ensure an appropriate balance of power among agencies, data providers, and data users. As constrained by its legislative mandate and resources, it could have wide-ranging responsiveness to privacy and information disputes, high flexibility in dealing with them, and potentially high access by concerned parties.

Perhaps consistent with this notion is the announcement on July 31, 1998 by Vice President Gore that "OMB will be given responsibility for coordination of privacy issues, drawing on the expertise and resources of other government agencies. This will help improve the coordination of U.S. privacy policy, which cuts across the jurisdiction of many federal agencies."

### **Data and Access Protection Commission**

Perhaps the most elaborate institutional mechanism is an independent *data and access protection commission*. It would have legislative authority to regulate all stages of the information gathering and dissemination process by promoting accountability and fair information practices. In various ways such commissions have been implemented in Canada and several European countries (see Flaherty (1989) for a

detailed discussion of their operation). Canada has institutionalized a balance between data protection and data access. They have both a privacy commissioner and an information commissioner. The province of British Columbia combines these functions in a privacy and information commissioner. Both Australia and New Zealand have Privacy Commissioners; the United Kingdom has a Data Protection Registrar; Switzerland has a Data Protection Authority; Norway has a Data Inspectorate; and Spain has a Privacy Authority.

In the United States, most drafts of the Privacy Act of 1974 provided for a permanent privacy protection commission, but this provision was deleted before final passage. Recently interest in such a proposal has waxed and waned. In the House, Representative Wise introduced legislation in 1989 and 1991 to establish an independent data protection board. In 1995 Representative Collins introduced H.R.184 to amend the privacy provisions of the Privacy Act to improve the protection of individual information and to reestablish a permanent Privacy Protection Commission as an independent entity in the Federal Government. In the Senate, Paul Simon introduced Senate Bill 1735 in November of 1993 to establish a Privacy Protection Commission. The bill provided for an advisory and independent commission of five members to be appointed by the President, with the consent of the Senate, to serve staggered seven year terms. As of July 1998 no such bill had passed. Nonetheless, the fact of repeated introduction of this provision and its support from various bodies, including the National Academy of Sciences Panel on Confidentiality and Data Access (Duncan, Jabine, and de Wolf 1993 Recommendation 8.5, p. 217). Although it does not appear on the current legislative agenda, there continues to be interest in the concept.

Data Protection Commissioners could also assist in reviewing data provider related disputes. The National Research Council's Panel on Confidentiality and Data Access recommended, "an independent federal advisory board charged with fostering a climate of enhanced protection for all federal data about persons and responsible data dissemination for research and statistical purposes." A data and access protection commission could have a legislative mandate giving it wide-ranging authority. Such authority might charge it with the responsibility to provide for a balance of power among stakeholders, be broadly responsive to disputes about data, to be flexible in employing mechanisms to resolve disputes, and to provide easy access to all disputants. A commission could also serve as a liaison for the negotiation of international agreements regarding privacy and information issues.

### ***Technical Procedures***

Technical procedures for maintaining data confidentiality involve release of restricted data; techniques developed over the past twenty years have been proposed in both the statistical and computer science literature (Duncan and Pearson 1991, Fienberg 1994, Keller-McNulty and Unger 1993). Unlike restricting access, restricting data is a technical device. It involves such methods as removing explicit identifiers and masking the sensitive data, e.g., grouping into categories or adding noise. By implementing data restrictions, agencies have operationalized statistical disclosure limitation practices. Some guidelines used by the European statistical system (Eurostat) are summarized in Manual on Disclosure Control Methods (1996).

Statistical disclosure limitation practices have allowed agencies to provide increasing amounts of data to the research community. Jabine (1993a) gives an excellent summary of statistical disclosure limiting practices for selected US agencies. The techniques proposed depend on the nature of the data, whether in tabular, microdata, or in on-line form.

### **Tabular Data**

From demographic surveys frequency counts of variables such as age, sex, and race of responding individuals are tabled. Respondents can be identified, and so a disclosure occurs, with small counts in the cells of the table. If a table, for example, showed only one Asian female in a census tract and shows her as an orphan, then a disclosure has occurred. From establishment surveys variables such as Standard Industrial Classification Code and salary levels are used to create tables. For establishment data the

disclosure issue is to avoid releasing information that will identify characteristics of particular establishments. As noted by Cox and Zayatz (1995), there are four principal methods for disclosure limitation of tabular data: cell suppression (Willenborg and de Waal 1996), rounding (Cox 1987), perturbation (Duncan and Fienberg 1998), and modification of the underlying microdata (Griffin, Navarro and Flores-Baez 1989).

## Microdata

Microdata are records directly on the unit of analysis, so may involve data about specific individuals or establishments. Because of the demand for more information than can be obtained from tables, public-use microdata files have been developed by some agencies. A public-use microdata file provides unrestricted access to restricted data. Public-use microdata files have been limited to data concerning individuals. Data on organizations tend to be more highly skewed than for individuals. The skewed distributions coupled with the time series and longitudinal nature of organizational data make many data restriction techniques, including top-coding, difficult. Consequently, very few public-use microdata files for organizations have ever been released. An important exception to this is the U.S. Census of Agriculture, which beginning with the 1987 data has released microdata that has been disclosure limited through high levels of geographic aggregation and data categorization as well as being a 5 percent sample of farms (Kirkendall et al. 1994).

## Online Data

Technological advances in computers and communications offer both opportunities and threats: opportunities to capture, analyze, store, and disseminate large databases more efficiently and threats of unauthorized access to individually identifiable data. Full use of the capability of today's information technology involves data access through on-line data query systems (McNulty and Unger 1989). The data user directly requests all statistical analyses of interest. Steel and Zayatz (1998) lay out the technical procedures for disclosure limitation that will be used for the 2000 Census. In order that the released data products may have both higher quality and lower disclosure risk, they propose data swapping procedures that will target the most risky records. To allow broader and easier access to data and to allow users to create their own data products, they are developing the Data Access and Dissemination System (DADS). Users would submit requests electronically. Because of the possibility of substantially increased detail in tabular data with DADS new disclosure limitation techniques will need to be employed.

## Conclusions

Privacy is a basic American value -- in the Information Age, and in every age. And it must be protected. We need an electronic bill of rights for this electronic age. You should have the right to choose whether your personal information is disclosed; you should have the right to know how, when, and how much of that information is being used; and you should have the right to see it yourself, to know if it's accurate.

-- Vice President Al Gore

Agencies in the public sector play a key role in responding to this challenge. They are central in collecting and disseminating data that bear on the most important issues in the public interest. Advances in information technology, particularly the Internet, have multiplied the tension between demands for ever more comprehensive databases and demands for the shelter of privacy. In mediating between these two conflicting demands, agencies must address a host of difficult problems. These include providing access to information while protecting confidentiality, coping with health information databases, and ensuring consistency with international standards. The policies of agencies are determined by what is right for them to do, what works for them, and what they are required to do by law. They must interpret and respect the ethical imperatives of democratic accountability, constitutional empowerment, and individual autonomy. They must keep pace with technological developments by developing effective measures for making information available to a broad range of users. They must both abide by the mandates of legislation and

participate in the process of developing new legislation that is responsive to changes that affect their domain. In managing confidentiality and data access functions, agencies have two basic tools: techniques for disclosure limitation through restricted data and administrative procedures through restricted access. The administrative procedures can be implemented through a variety of institutional mechanisms, spanning privacy advocates, internal privacy review boards, and a data and access protection commission. The technical procedures for disclosure limitation involve a range of mathematical and statistical tools. The challenge developing and implementing these administrative and technical tools is great, and the value to society of the information that agencies can provide is hard to overestimate.

## **References**

- American Statistical Association, Committee on Professional Ethics (1989). *Ethical Guidelines for Statistical Practice*. Alexandria, VA: American Statistical Association.
- Cox, L., McDonald, S-K, and Nelson, D. (1986). Confidentiality issues at the United States Bureau of the Census. *Journal of Official Statistics*, 2, 135-160.
- Cox, L. and Zayatz, L. (1995) An agenda for research in statistical disclosure limitation. Environmental Protection Agency.
- Dalenius, T. (1993). Discussion: Informed consent and notification. *Journal of Official Statistics*, 9, 377-381.
- Drucker, P. F. (1992). The new society of organizations. *Harvard Business Review*, September-October, 95-104.
- Duncan, G. T. (1990a). Disclosure limitation research and practices: A commentary on two agencies' perspectives. *Proceedings of the Seminar on Quality of Federal Data Council of Professional Associations on Federal Statistics*.
- Duncan, G. T. (1997) Data for health: Privacy and access standards for a health care information infrastructure. *Health Care and Information Ethics: Protecting Fundamental Human Rights* (Audrey R. Chapman, ed.), Sheed and Ward, 299-339.
- Duncan, G. T., & Fienberg, S. E. (1998) Obtaining Information while Preserving Privacy: A Markov Perturbation Method for Tabular Data. Eurostat: Statistical Data Protection 98. Lisbon.
- Duncan, G. T., Jabine, T., & de Wolf, V. (1993). *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*. Washington, D.C.: National Academy Press.
- Duncan, G. T., & Lambert, D. (1986). Disclosure-limited data dissemination (with comments). *Journal of the American Statistical Association*, 81, 10-28.
- Duncan, G. T., & Lambert, D. (1989). The risk of disclosure for microdata. *Journal of Business and Economic Statistics*, 7, 207-217.
- Duncan, G. T., & Mukherjee, S. (1991). Microdata disclosure limitation in statistical databases: query size and random sample query control. *Proceedings of the 1991 IEEE Symposium on Research in Security and Privacy*, 20-22, Oakland, California.
- Duncan, G. T., & Pearson, R. W. (1991). Enhancing access to data while protecting confidentiality: prospects for the future. *Statistical Science*, 6, 219-239.

- Duncan, G. T., & de Wolf, V. A., (1990). Mediating confidentiality and data access. *Chance* 3, 45-48.
- Duncan, G. T., de Wolf, V. A., Jabine, T. B., and Straf, M. L. (1993) Report of the panel on confidentiality and data access. *Journal of Official Statistics*, 9, 271-274.
- Engelage, C. (1992). Statistical confidentiality in the context of community statistics: the legal framework. Eurostat report, Luxembourg. August 28.
- Federal Committee on Statistical Methodology (1994). Statistical Policy Working Paper 22: Report on Statistical Disclosure limitation Methodology. Washington, DC: U. S. Office of Management and Budget.
- Fienberg, S. E. (1994) Conflicts between the needs for access to statistical information and demands for confidentiality. *Journal of Official Statistics*, 10, 115-132.
- Flaherty, D. H. (1989). *Protecting Privacy in Surveillance Societies*. Chapel Hill: University of North Carolina Press.
- Griffin, R., Navarro, A. and Flores-Baez, L. (1989) Disclosure avoidance for the 1990 Census. Proceedings of the Section on Survey Research Methods, American Statistical Association, 516-521.
- International Statistical Institute (1986). Declaration of Professional Ethics. *International Statistical Review*, 54, 227-242.
- Jabine, T. B. (1993a). Statistical Disclosure Limitation Practices of United States Statistical Agencies. *Journal of Official Statistics*, 9, 427-454.
- Jabine, T. B. (1993b). Procedures for Restricted Data Access. *Journal of Official Statistics*, 9, 537-590.
- Keller-McNulty, S. and Unger, E. A. (1993). Database Systems: Inferential Security. *Journal of Official Statistics*, 9, 475-500.
- Kirkendall, N. J., Arends, W. L., Cox, L. H., de Wolf, V. A., Gilbert, A., Jabine and Zayatz, L. V. (1993). Report of the Subcommittee on Statistical Disclosure Limitation Methodology, Federal Committee on Statistical Methodology, Washington, D.C.
- Lambert, D. (1993) Measures of disclosure risk and harm. *Journal of Official Statistics*, 9, 313-331.
- Leftwich, W. (1993). How researchers can win friends and influence politicians. *American Demographics*, August: 9.
- Manual on Disclosure Control Methods (1996) Eurostat. Luxembourg: Office for Official Publications of the European Communities.
- National Science Foundation (1998) Digital Government Program Announcement. Directorate for Computer and Information Science and Engineering. Washington, D.C. March 15.
- Norwood, J. (1990). Statistics and public policy: Reflections of a changing world. Presidential Address, *Journal of the American Statistical Association*, 85, 1-5.
- Prewitt, K. (1985) Public statistics and democratic politics. In J. J. Smelser and D. R. Gerstein, eds. *Behavioral and Social Science: Fifty Years of Discovery*. Washington, D. C.: National Academy Press.

Regan, P. M. (1984). Personal information policies in the United States and Britain: The dilemma of implementation considerations. *Journal of Public Policy*, 4, 19-38.

Smith, J. P. (1991). Data confidentiality: A researcher's perspective. Panel on Privacy and Confidentiality. Annual Meeting of the American Statistical Association, Anaheim, CA.

Steel, P. and Zayatz, L. (1998) Disclosure limitation for the 2000 Census of Population and Housing. Annual Meeting of the American Statistical Association, Dallas, TX.

Willenborg, Leon and de Waal, Ton (1996) *Statistical Disclosure Control in Practice*. Lecture Notes in Statistics 111. Springer Verlag, New York.

de Wolf, V. A. (1995) Procedures for researcher access to confidential microdata at the Bureau of Labor Statistics. Office of Research and Evaluation, Bureau of Labor Statistics, Washington, D.C.

George T. Duncan is Professor of Statistics in the H. John Heinz III School of Public Policy and Management and the Department of Statistics at Carnegie Mellon University. He was on the faculty of the University of California, Davis (1970-1974), and was a Peace Corps Volunteer in the Philippines (1965-1967), teaching at Mindanao State University. His research centers on information technology and social accountability. He has published more than fifty papers in such journals as *Statistical Science*, *Management Science*, *the Journal of the American Statistical Association*, *Econometrica*, and *Psychometrika*. He has received National Science Foundation research funding and has lectured in Brazil, Italy, Turkey, Ireland, Mexico, Israel and Japan. He chaired the Panel on Confidentiality and Data Access of the National Academy of Sciences (1989-1993), resulting in the book, *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*. He chaired the American Statistical Association's Committee on Privacy and Confidentiality.

He is a Fellow of the American Statistical Association, an elected member of the International Statistical Institute, and a Fellow of the American Association for the Advancement of Science. In 1996 he was elected *Pittsburgh Statistician of the Year* by the American Statistical Association. He has been editor of the Theory and Methods Section of the *Journal of the American Statistical Association*. He received a BS degree (1963) and MS degree (1964) from the University of Chicago and a Ph.D. degree (1970) from the University of Minnesota, all in the field of statistics.

George T. Duncan  
H. John Heinz III School of Public Policy and Management  
Carnegie Mellon University  
Pittsburgh, PA 15213  
Phone/FAX: (412) 268-2172/7036  
E-mail: George.Duncan@cmu.edu