Privacy and Computerized Data Bases

by Thomas E. Brown *

In the United States, privacy legislation generally has been limited to government records at the different levels of government, i.e., Federal, state and local. These laws impose requirements and restrictions on how government agencies collect, maintain and use information about individual persons. The United States, with few exceptions, has not adopted legislation to regulate and restrict the collection, maintenance or use of personally identifiable information by non-governmental entities. In other countries, this type of legislation is frequently referred to as "data protection laws." However, if a threat to the personal privacy exists today in the United States, it comes not from the regulated governmental data bases but from the non-regulated ones outside of government buildings.

The information in these data bases has long be available in paper form without threatening individual privacy. But technological developments have altered the situation. First, with the spread of technology, data is increasingly available in electronic form. Second, computer processing speeds have increased. With data mining tools, a data base query that took six minutes in 1994 now takes less than nineteen seconds. Third, with the availability of increased computing speed, it is now easier to combine data from multiple sources and create comprehensive information products. Fourth, the cost of storing electronic information — on-line, near-line and off-line storage — has dropped. Fifth, personal computers are becoming more and more affordable and thus more wide spread. Finally, with the spread of the personal computers, Internet use is becoming commonplace from businesses and homes.[1]

Technology, then, has permitted the growth of what Vice President Gore has called "profiling," or the ability to build dossiers about individuals by aggregating information from a variety of database sources. These dossiers now have detailed information on the vast majority of the American population, including children. For example, Acxiom Corporation has information on 196 million Americans on 700,000 data tapes with 350 trillion bytes. Information America claims that it has employment and other demographic information on 160 million individuals, 92 million households, 71 million telephone numbers, and 40 million deceased persons. Finally, Medical Marketing Service, Inc., offers its "Patient Direct" data base with information on more that 20 million households documenting an individual's age, gender, income, educational status, and health condition, from allergies (4.3 million households) to yeast infections (1.4 million)[2].

The sources of information are varied. First, governmental records provide a wealth of information on individuals and, under a variety of Freedom of Information laws, are readily available. While some invasive records are from the Federal Government, most are seemingly from state and local governments. These records concern real estate transactions and holdings, marriage and divorce, birth certificates, driving records, drivers' licenses, vehicle registrations, civil and criminal court proceedings, paroles, postal service change-of-address forms, voter registrations, bankruptcies and liens, incorporation documents, workers' compensation claims, political contributions, firearm permits, occupational and recreation licenses, and filings with the Securities and Exchange commission. For example, the Federal Aviation Administration has a list of all individuals licensed to fly in the United States which includes certification class, medical certificate and date of last medical examination. Real estate documents describe the property, dates of sales, selling prices, mortgage amounts, lender, and the names of the sellers and purchasers. Social security numbers are readily available as well, most commonly from state departments of motor vehicles, which also provide individuals' name, address, height, weight, gender, eye color, date of birth, and whether or not an organ donor. For example, New York's publicly available drivers' information includes vehicle ownership, accident reports, conviction certificates, police reports, complaints, hearing records, suspension and revocation orders. Although government records are increasingly available in electronic form and electronic FOIA guarantees access to the records in that format, other records must first be digitized.[3]

Non-governmental entities, but still publicly available information, offer additional sources of information. For example, newspaper and magazines identify and provide background information on individuals, and electronic editions are now frequently available from the Internet. Powerful search tools permit people to search these

newspapers and magazines and find all references to a given individual. Many professional and alumni organizations make membership lists available on Web sites. Indeed, many Web sites contain detailed personal information to anyone who clicks on the URL, such as adoption pages where adopted children and birth parents post detailed personal information in hopes of connecting with their blood relatives.[4]

A third type is private proprietary information. Maybe the most well known of this type is the information of the three major credit reporting agencies, Trans Union, Equifax, and Experian. But other even more invasive data bases exist with varying degrees of confidentiality associated with them. More than 3,000 corporations in the United States collect, maintain and sell information from their data bases for marketing purposes. Most of the information is provided by the individuals themselves, often through warranty cards or product registration cards. A careful reading of these forms reveals the depth of personal information people supply. As an example, a recent card for a coffee bean grinder asked about sex and ages of all household members, marital status, occupation, income, educational level, credit cards, home ownership, anticipated changes during the next six or 12 months, and finally a list of sixty interests activities in which the respondent participates in regularly. Another series of incredibly revealing data bases are those maintained by banks and credit card companies which list the individual transactions charged to each credit card. But far more sinister is the spread of supermarket customer cards. Last year, a survey indicated that 60 per cent of the supermarkets in the country had started such a program or intended to do so soon. These cards are used to record in a data base each product which a consumer purchases. One's shopping habits reveal a lot about that individual. The purchases of condoms or large quantities of alcoholic beverages connote a life style. Patent medicines reveal aspects about one's health, and purchases of certain brands and products betray one's value systems and interests. Every time somebody makes a telephone call, it creates a record in a database which offers much information to the phone company and other marketers. Through billing records, local carriers and long-distance carriers learn whom people are calling and when and where calls are made. Finally, the Internet itself is a great collector of personally revealing information. Commercial Web sites collect personal information through a variety of means, including registration forms, user satisfaction surveys, contests, and order forms. For example, an online doctor-referral service asks users for their name, mailing address, e-mail address, insurance company, and whether they want information on a variety of health concerns, such as urinary incontinence, hypertension, cholesterol, prostate cancer, or depression. Another Web site is from a mortgage company for pre-qualification for home mortgages. Potential borrowers provide their names, social security numbers, home and work telephone numbers, e-mail addresses, previous addresses, current and former employers, lengths of employment, income, savings, and credit histories including credit cards. In the Spring of 1998, a quarter-million people completed a detailed survey at the ESPN Web site to enter a lottery for tickets to the NCC Final Four tournament. Even without telling the users, Web vendors can collect detailed personal information as they can identify which pages the user visited, what the user bought, where the user linked from and where the user went on the "Net" when he or she left the site, and in some cases, how long an individual was at the site and on each page of the site. Recently reported, some of the largest commercial sites including Lycos-Tripod and Geocities, had begun providing users' reading, shopping, and entertainment habits to a centralized system which links that information the user's age, income, ZIP code, number of children, and information obtained from on-line forms. To protect privacy, the system uses a unique identifying number associated with the hard drive of the computer accessing the sites. Thus when the user connects to a participating site which recognizes the number on the drive, targeted sales messages will be sent. But already, entrepreneurs are trying to link the preferences in compiled computerized data bases with mailing lists of traditional direct marketers. The owners of the system have announced that it will not collect information about sexual interests or health related topics. But with money on the line, such a voluntary restriction is probably not universal or permanent. For example, some Internet sites devoted to specific diseases have solicited data from site visitors and then sold that information, either directly or indirectly through data intermediaries, to companies marketing drugs or other therapies for the specific disease. As Nat Goldhaber, chair of a Web vendor called Cybergold, commented, "What the Internet has done is make explicit what used to be implicit — namely that there dossiers on you than can be built up with great granularity."[5]

As June 24, 1997, there were fifty-one database vendors and information bureaus which will sell detailed personal information on the Internet about telephone use, assets, criminal histories, vehicle and driving records, aircraft, boat and gun ownership and usage, business materials, marriages and divorces, current and previous addresses, and information on neighbors and relatives. Some have unexceptional names, such as Discreet Research whose slogan is "When you need to know." Others are more interesting, such as Dig Dirt, Inc., whose saying is "Because what you don't know does hurt you."[6]

In the collection, maintenance, and use of this information, the United States has adopted as essentially laissez faire approach. Current American privacy law has been described as "sectoral," that is "a handful of disparate statutes directed at specific industries that collect personal data." In fact, the Federal level has only six data protection

laws in place. The first and clearest example is the Fair Credit Reporting Act which guarantees the individual the right to gain access to personal information which the credit reporting agency has and the right to dispute erroneous information and add corrective details. It also generally restricts access to those businesses to which the consumer has applied for credit, insurance, employment, or a lease agreement. The second is Federal Educational Records Privacy Act (FERPA) or more commonly know as the Buckley Amendment. This limits the release of students' educational records to educational personnel and educational institutions. Two other acts are the Cable Communications Policy Act of 1984 which restricts cable television subscriber information and the Telecommunications Act of 1996 which governs customer proprietary network information. The last data protection law in the United States restricts the release of an individual's video tape rentals. Enacted as a reaction to publicizing Judge Bork's video tape rental during his Supreme Court confirmation hearings, the law prompted Secretary of Health and Human Services Donna Shalala to comment, "Our private health information is being shared, collected, analyzed and stored with fewer Federal safeguards than our video store records." In the absence of legislation, case law is working against data protection. In United States v. Miller [425 U.S. 435 (1976)], the Supreme Court ruled that individuals have no constitutional protection of information which that they have voluntarily provided.[7]

But currently, some 80 bills to strengthen the data protection are pending in the Congress. Some would restrict the dissemination or use of the social security numbers; others would allow individuals to stop the post office from selling their change of address requests; and at least one would establish an independent regulatory commission to control the collection, maintenance, and use of individually identifiable information. According to one observer, the New York Times' Nina Bernstein, "But with few exceptions, the proposals seem to be going nowhere. Beneath the surface of their popular appeal, most are mired in unresolved conflicts over contradictory goals: on the one hand, preserving personal privacy, and on the other, the advantages of quick, computerized access to personal information for fighting crime, fraud and waste, or promoting growth in the information economy."[8]

The one exception will be data protection of health information. The Health Insurance and Portability and Accountability Act of 1996, commonly known as Kennedy-Kassebaum Act, required the Clinton Administration to propose legislation on the creation, maintenance and use of medical information on individuals by September 30, 1997. [Ironically, this same legislation also required that the Administration create a standard medical identifier so individuals' medical records could be widely available, regardless of the health care program, from a database of health information.] If the Congress does not enact legislation by August 1999, then the Administration must issue regulations. So on September 11, 1997, Secretary Shalala proposed legislation. But the Administration's proposal is only one of several laying in the legislative hoppers. Reaching a consensus will not be easy in the tug-of-war between consumer groups, law enforcement agencies, and health care professionals. Should law enforcement officials need a warrant or court order to get medical records? Can records obtained investigating insurance fraud be used for other criminal prosecutions, such as information about illegal drug use lead to prosecution? If consumers have the right to change or delete medical information, then some argue that this could endanger a vital resource needed for medical research, for public health analyses, and for improved medical care.[9]

If a legislative consensus is difficult, self-regulation is an option which conforms to the laissez faire approach and has an honored tradition in the country's history. Last December, the Federal Trade Commission announced an agreement amounting to self-policing by "individual reference services," that are businesses which have been selling personal information to the general public. By the end of this year, fourteen of the largest of these organizations, including the three major crediting reporting agencies and the largest direct mail marketers, announced that they would no longer provide information to the general public. They would also limit the types of personal information they would provide to commercial users like marketers, banks, lawyers and journalists. Yet they would still allow unrestricted access by law enforcement personnel, licensed private investigators, and corporate security staff. According to The New York Times, this agreement embodied the Administration's strategy of self-regulation. "The agreement sets in motion the first meaningful trial of the Clinton Administration's privacy policy, the stated goal of which is to protect individual privacy in the Internet age without resorting to new laws and regulations." Privacy proponents have objected to the agreement as too little. While individuals can request that their records be erased from some data bases, they cannot access all of the information being maintained and disseminated about them, and to have themselves removed from selected data bases, individuals would have to contact each of the fourteen reference services separately. An obvious loophole will be for an individual to hire an attorney or private detective to serve as an intermediary.[10]

In this same vein of self-regulation, the 3000-member Direct Marketing Association has issued "Guidelines for Personal Information Protection " which stipulates that personal data collected for marketing purposes should be only for that purpose. It further maintains to its Committee on Ethical Business Practices to investigate the misuse of marketing information. The Association has also announced that it will require, beginning next year, its

members to disclose publicly how they gather and use data. The Council of Better Business Bureaus is working on a model for self-regulation that would impose sanctions on businesses that fail to follow a code of conduct to protect people's privacy. A recent innovation to self-regulation is incorporating a seal onto a commercial Web site to indicate that the site follows an established code of conduct regarding privacy. A nonprofit group called "TRUSTe" already has a system in place for about 120 companies. In July, another group, Online Privacy Alliance, emerged with the same intent.[11]

These efforts at self-regulation have been haphazard at best. The Federal Trade Commission tersely concluded earlier this summer, "To date, . . . the Commission has not seen an effective self-regulatory system emerge." The Clinton Administration began to hedge last May when Vice President Gore outlined a new administration initiative on privacy. It consisted of a renewed call for legislation regarding medical records, a Federal Web site to assist consumers in deleting their names from commercial data bases, creation of the privacy officer in every Federal agency, and a conference to address the topic in June 1998. This speech, calling for an "Electronic Bill of Rights," was interpreted as an admission that self-regulation had not been as effective as the Administration had hoped. According to Janlori Goldman, a noted privacy specialist at Georgetown University, "This Administration has been singing the praises of self-regulation for some time now, but this is an acknowledgment that there are significant limits to what the private sector will do on its own." As far as the legislation concerning medical records is concerned, Dr. Goldman opined that the Clinton Administration is "using medical privacy as a signal to the public and a stick to industry to say that we have a history of abuse in this area and the Administration wants to do something about it."[12]

But regardless of the mode of data protection — legislation, regulation or self-policing, discussions of data protection revolve around eight principles of fairness: openness, individual participation, collection limitation, data quality, use limitation, disclosure limitation, security, and accountability. While originally proposed by a United States federal advisory committee, the principles were most clearly codified in the Council of Europe's *Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data*. For the United States to be full economic partners with Europe, data protection efforts will have to conform to the Council of Europe's *Convention*. Two statements in the *Convention* should give pause to archivists, namely:

"Article 5 - Quality of data

Personal data undergoing automatic processing shall be:
b. stored for specified and legitimate purposes and not used in a way incompatible

with those purposes; . . .

e. preserved in a form which permits identification of the data subjects for no longer

than is required for the purpose for which those data are stored."[13]

This concept that personal data can be used only for the reason for which it was collected is fairly common in the discussions of data protection. In the above discussion of the Direct Marketing Association, its "Guidelines for Personal Information Protection" stated that personal data collected for marketing purposes should be only for that purpose. In discussing the Clinton Administration's position on health care, Secretary Shalala stated that personal medical information should be "for health care and health care only" with very few exceptions. But from an archival perspective, this limitation on use has a potential problem. Archival theory discusses the primary value of records as the purpose for which the records were created. This is in contrast to the secondary value of records which is the value of the records to someone other than the record's creator for a reason other than that for which they were created, and it is the secondary value of records that justifies the archival retention of records after the record's creator no longer needs them in the course of business, but if the use of personal data is limited to the reason for which it was collected and if the archives are not exempt from this limitation, then records cannot be retained in archives for their secondary values. Indeed, under the Council of Europe's Convention, the records must be destroyed as soon as their primary value has ended. In terms of medical records, for example, if an exception is not made for archival retention of some personal medical information then the history of medicine, the history of technology and the history of science — all currently viable fields of historical study — will be severely curtailed. Application of the Convention's beyond just medical records will threaten the continued existence of records important to future genealogists and family historians. Obviously, a solution is to outline exceptions to the use of limitation provisions in any data protection effort — whether legislation, regulation, or self-policing. One of these exceptions must be for the archival retention of records with significant secondary values to be released only after the passage of time that would permit access without endangering the privacy of individuals. Unfortunately, the possibility of records having secondary values seldom enters into the debates over data protection, but then it seems that the archival profession has seldom entered in the debates over data protection.[14]

[1]Federal Trade Commission, *Individual Reference Services: A Report to Congress*, December 1997, pp. 3-4.

[2] "Vice President Gore Announcee New Steps Toward an Electronic Bill of Rights," This Week's Press Briefings and Releases, July 31, 1968, http://library.whitehouse.gov, August 3, 1998; Robert O'Harrow, Jr., "Data Firms Getting Too Personal?" *The Washington Post*, March 8, 1998, pp. A1, A18-A19; Federal Trade Commission, *Individual Reference Services*, pp. 36; Sheryl Gay Stolberg, "The Numbering of America: Medical I.D.'s and Privacy (Or What's Left of It)," T*he New York Times*, July 26, 1998, p. WK3.

[3]Federal Trade Commission, *Individual Reference Services*, pp. 4-6, 37; Nina Bernstein, "High-Tech Sleuths Find Private Facts Online," *The New York Times*, September 15, 1997, electronic edition.

[4]Federal Trade Commission, I*ndividual Reference Services*, p. 5.

[1*F*]ederal Trade Commission, *Individual Reference Services*, p. 3; O'Harrow, March 8, 1998, p. A18; Nina Bernstein, "Lives on File: Privacy Devalued in Information Economy," *The New York Times*, June 12, 1997, electronic edition; Federal Trade Commission, *Privacy Online: A Report to Congress*, June 1998, p.3, 39; Saul Hansell, "Big Web Sites to Track Steps of Their Users," *The New York Times*, August 16, 1998, pp. 1, 24; Denise Caruso, "Who Knows What About Whom on the Internet," *The New York Times*, April 13, 1998, electronic edition

[5]Federal Trade Commission, *Individual Reference Services*, p. 3; O'Harrow, March 8, 1998, p. A18; Nina Bernstein, "Lives on File: Privacy Devalued in Information Economy," *The New York Times*, June 12, 1997, electronic edition; Federal Trade Commission, *Privacy Online: A Report to Congress*, June 1998, p.3, 39; Saul Hansell, "Big Web Sites to Track Steps of Their Users," *The New York Times*, August 16, 1998, pp. 1, 24; Denise Caruso, "Who Knows What About Whom on the Internet," T*he New York Times*, April 13, 1998, electronic edition.

[6]See <http://www.dresearch.com/> and <http://www.pimall.com/digdirt/moore.htm>.

[7]Federal Trade Commission, *Privacy Online*, p. 62; 15 USC 1681; 20 USC 1232g; 47 USC 551; 47 USC 222; 18 USC 2710; Robert Pear, "Clinton to Back a Law on Patient Privacy," *The New York Times*, August 10, 1997, p. 22.

[8]Peter Maas, "How Private Is Your Life?" *Parade Magazine*, April 19, 1998, p. 5; Nina Bernstein, "Goal Clash in Shielding Privacy," *The New York Times*, October 20, 1997, p. A16.

[9]Steven Findlay, "Prescription for Patient Privacy: Administration today offers plan to ensure confidentiality," *USA Today*, September 11, 1997, pp. 1-2; Editorial, "HHS identifier puts privacy at risk," *Federal Computer Week*, July 20, 1998, p. 24; Arthur Allen, "Exposed," *The Washington Post Magazine*, February 8, 1998, pp. 11-15, 27-28.

[10]Federal Trade Commission, *Individual Reference Services*, passim; Katherine Q. Seelye, "A Plan for Database Privacy, But Public Has to Ask for It," The New York Times, December 18, 1997, pp. A1, A24; John Markoff, "Guidelines Don't End Debate on Internet Privacy," *The New York Times*, December 18, 1997, pp. A24.

[11]Federal Trade Commission, I*ndividual Reference Services*, p. 38; O'Harrow, March 8, 1998, p. A18; Robert O'Harrow, Jr., "White House Effort Addresses Privacy," The Washington Post, May 14, 1998, p. E1, E4; Robert O'Harrow, Jr., "Internet Companies Move to Safeguard Computer Users' Privacy," *The Washington Post*, July 22, 1998, p. A13.

[12]Federal Trade Commission, *Privacy Online*, p. 41; John M. Broder, "Gore to Announce'Electronic Bill of Rights' Aimed at Privacy," *The New York Times*, May 14, 1998, p. A22; O'Harrow, May 14, 1998, p. E4.

[13]Robert Gellam, "Don't fear privacy protection — arm yourself with fairness checks," *Government Computer News*, May 4, 1998, p. 26; Department of Health, Education and Welfare, Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computers, and the Rights of Citizens, (July 1973), Council of Europe, European Treaties, ETS No. 108, Convention for the Protection of Individuals with Regard To Automatic Processing of Personal Data, Strasbourg, 28.I.1981, http://www.coe.fr/eng/legaltxt/108e.htm.

[14]Federal Trade Commission, *Individual Reference Services*, p. 38; Pear, p. 22; T. R. Schellenberg, *Modern Archives: Principles and Techniques* (Chicago, University of Chicago Press, 1956), pp. 28-32; T. R. Schellenberg, *The Appraisal of Modern Public Records*, Bulletins of the National Archives, Number 8 (Washington DC: U.S. Government Printing Office, 1956), p. 6.

* Paper presented at the IASSIST Conference, May 21, 1998, at Yale University, New Haven, Connecticut. Thomas E. Brown, Manager, Archival Services Electronic and Special Media Records Services Division National Archives and Records Administration.