Data Retention and the Panoptic Society: The Social Benefits of Forgetfulness*

Jean-François Blanchette  
Science and Technology Studies, Rensselaer Polytechnic Institute, Troy, New York, USA

Deborah G. Johnson  
School of Engineering and Applied Science, University of Virginia, Charlottesville, Virginia, USA

Modern information systems not only capture a seemingly endless amount of transactional data, but also tend to retain it for indefinite periods of time. We argue that privacy policies must address not only collection and access to transactional information, but also its timely disposal. One unintended side effect of data retention is the disappearance of social forgetfulness, which allows individuals a second chance, the opportunity for a fresh start in life. We examine three domains in which social policy has explicitly recognized the importance of such a principle: bankruptcy law, juvenile crime records, and credit reports. In each case, we frame the issue in terms of the social benefits of forgetfulness, rather than in terms of individual privacy protection. We examine how different policy approaches to privacy might handle the retention of data and propose a comprehensive policy that includes a variety of strategies. The broad conclusion of the article is that data retention and disposal should be addressed as a part of a broader and comprehensive policy approach, rather than in a piecemeal fashion or as an afterthought.

Received 12 October 1998; accepted 19 April 2001. This article greatly benefited from comments of participants at the ACM Policy ’98 Conference, organized by the Association for Computing Machinery, the Computer Ethics: Philosophical Enquiry (CEPE ’98) conference, held at the London School of Economics, and the Graduate Student Conference on Technology and Identity, held at Cornell’s Science and Technology Studies department, as well as from comments by Robert Gellman, David Charmichael, Daniel Poulin, Ann Cavoukian, and three anonymous reviewers. The first author also benefited from the support of a doctoral grant from Canada’s Social Sciences and Humanities Research Council.

Address correspondence to Jean-François Blanchette, Centre de recherche en droit public, Faculté de droit, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal(Qc), Canada, H3C 3J7.

Keywords  data retention, informational privacy, privacy policy, social forgetfulness, social memory, surveillance

It is not enough to keep repeating that memory is socially structured. To have come so far invites a further step. The next thing is to discover what qualities of institutional life have distinctive effects on remembering. (Douglas, 1980, p. 80)

Cheerfulness, the good conscience, the joyful deed, confidence in the future—all of them depend, in the case of the individuals of a nation, on the existence of a line dividing the bright and discernible from the unilluminable and dark; on one’s being just as able to forget at the right time as to remember at the right time; on the possession of a powerful instinct for sensing when it is necessary to feel historically and when unhistorically. This, precisely, is the proposition the reader is invited to meditate upon: the unhistorical and the historical are necessary in equal measure for the health of an individual, of a people and of a culture. (Nietzsche, 1997, p. 63)

On December 28, 1997, Swiss cellular phone users were distraught to learn that the position of their phones (within a few hundred meters) was automatically and continuously registered by their service provider, Swisscom. While this is an inevitable feature of cellular telephony (in order to forward a call to a particular user, service providers must first ascertain the position of the phone with respect to the network), what made this revelation particularly disturbing from the privacy standpoint was the fact that Swisscom retained the data for a duration of 6 months to a year and half (American Press, 1997).

This incident is paradigmatic of a problem that has been largely overlooked in the privacy literature to date: Control over personal information is not only affected through selective access, but also through selective retention of such information. That is, control is not only a question
of who has and who does not have access to personal information (nowadays, seemingly everyone has access but its producer), but who gets to retain or discard it. Most privacy commentators focus on access and control, and address retention only as an afterthought—if at all. A central concern of this article is to make the importance of this component explicit: We argue that data retention must figure as an important element of any comprehensive account of informational privacy.

We begin by framing the data retention issue within broad concerns over the lack of privacy protection in modern democratic societies. Second, we place the issue in the context of a tension between the importance of institutional/public memory and forgetfulness. Once the issue is framed as such, we go on to examine three domains of life in which the idea of the “fresh start” (where individuals move on, leave their past behind them, and begin anew) plays an important role. We conclude that social forgetfulness is best addressed through a comprehensive approach that includes a variety of policy strategies. We describe how data retention can be addressed through the general principles of a code of fair information practices, legislation, self-regulating markets, a property right, and privacy-enhancing technologies.

PUTTING THE DATA RETENTION ISSUE IN CONTEXT

An enormous literature now documents concerns about and threats to personal privacy arising from new information and communication technologies. Concern heightens each time new technologies give rise to new forms of data collection. In the 1990s attention was focused especially on transactional data (web browsing, credit-card use, intelligent highways), in contrast with the 1970s and 1980s when concern was with the scale of record-keeping and collection of personal data. We do not describe the practices or technologies that give rise to such concerns, as an abundant literature already documents this, as well as the privacy policies extant in many countries. The European Union (EU) has become a focus of attention as it struggles with the harmonization of privacy policies of EU countries and with transborder data flows to non-EU countries.1

We agree with others who have suggested that the apparatus of a panoptic society is slowly, but surely, being put into place in the United States (Gandy, 1993). Democracies are generally thought of as societies in which individuals have a high degree of individual liberty and government power is limited and checked. Yet it appears that information and communication technologies are moving us rapidly toward a panoptic society. The panopticon is Bentham’s prison environment, as described by Foucault (1975), in which prison cells are arranged in a large circle with the side facing the inside of the circle open to view. The guard tower is placed in the middle of the circle so that the inside of each cell is in plain view of the guards. The amount of data currently collected as we go about our everyday lives—intelligent highway systems, consumer transactions, traffic patterns on the Internet, medical, educational, financial, and insurance records, and so on—strongly suggests we are moving into a panoptic society. Even if the data are not collected by a single, Orwellian-like entity, but rather by a mixture of public and private institutions, and even if what is observed is not necessarily amalgamated into a single dossier, the possibility of synthesis remains. Clearly, such a panoptic society presents fundamental challenges to the exercise of democratic freedoms and responsibilities.

Again, most of this is not new and we do not belabor the point. Rather we want to draw attention to the fact that most of the work that has been done on this issue has focused almost exclusively on how to control access to data (and the corresponding value of privacy), and neglected retention (and the corresponding value of social forgetfulness). Data protection policies have not proceeded from any comprehensive analysis of the problems occasioned by data retention. Instead, sector by sector, decisions have been made regarding the length of retention of data, with little attention being paid to the cumulative effect of these piecemeal decisions.

Our approach to data retention begins from the insight that the endurance of data is a feature that has invisibly but powerfully changed with the shift from paper-and-ink to electronic systems of record-keeping. In the paper-and-ink world, the sheer cumbersomeness of archiving and later finding information often promoted a form of institutional forgetfulness—a situation with parallels to human memory.2 The forgetfulness of the paper-and-ink world was implicit in the material being of institutions, the available storage space, the budget for file cabinets, etc. Often the institution’s memory/forgetfulness was not even recognized as a policy issue but dealt with as a matter of physical facilities.3 In many cases, as storage technologies have gained in practicality, ease of remote access, and lowered in price, the shift to an electronic medium changed the default position from one of forgetfulness to one of memory.4

Whether the paper-and-ink environment or the electronic environment favors data retention, the point remains that decisions about length of retention of data may be made unintentionally or in an ad hoc manner, rather than with an eye to privacy policy or institutional memory per se.5 We find ourselves in a world that captures endless data on us and then decides (sometimes by failing to decide) how long to retain this data. When data are lost or deleted, our behavior is forgotten. When data endure, our behavior is not forgotten and some important values may shrink with it—values that are fundamental to democratic
society. In other words, we must ask, what are the social implications of a lack of institutional forgetfulness?

We begin our investigation of this question within the U.S. context, for several reasons. First, there is a general consensus that in the United States too little is being done to stop the onslaught of personal data collection. There is even, to some extent, a consensus on the nature of the problem in the United States. It is that privacy protection policy has been ad hoc and piecemeal, rather than comprehensive (Regan, 1995; Gellman, 1997). At the same time (and perhaps ironically), the United States has traditionally understood itself to be a place where individuals could get a “second chance.” The idea that an American citizen can sometimes “wipe the slate clean” and start anew is, no doubt, tied to the immigrant, pioneer histories of so many Americans. Whatever its origins, the idea is in tension with current U.S. data collection and retention policies.

The idea that Americans value the opportunity for a “fresh start” was recognized in the early literature on privacy, and periodically recurs in current literature. Westin and Baker (1972), in their seminal work, Databanks in a Free Society, understood that this value was perceived to be under siege because of computers:

Many citizens assume, out of a variety of religious, humanistic, and psychiatric orientations, that it is socially beneficial to encourage individuals to reform their lives, a process that is impeded when individuals know (or feel) that they will automatically be barred by their past “mistakes” at each of the later “gate-keeping” points of social and economic life. Because the computer is assumed not to lose records, to forward them efficiently to new places and organizations, and to create an appetite in organizations for historically complete records, the computer is seen as threatening this forgiveness principle. (Westin & Baker, 1972, p. 267)

Interestingly enough, Westin and Baker went on to point out that the key question about erasure or noncirculation of derogatory information was not a technical matter in the organizations they visited. It was an issue of social policy, on which society has to choose between the “forgive-and-forget” and “preserve but evaluate” theories of record-keeping in each substantive area (p. 268). In his study of police surveillance practices, Gary Marx has underlined how surveillance information “transcends time”—that is, “it is available for analysis many years after the fact, and in totally different interpretive contexts” (Marx, 1986, p. 150). He remarks that this threatens to undermine some basic American values:

The idea of “starting over” or moving to a new frontier is a powerful concept in American culture. The beliefs that once a debt has been paid to society it is forgotten and that people can change are important American traditions. Americans pride themselves on looking at what a person is today rather than what he may have been in the past. Devices, such as sealed or destroyed records, prohibitions on certain kinds of record keeping, and consent requirements for the release of information, reflect these concerns. However, with the mass of easily accessible files, one’s past is always present, for erroneous or sabotaged data, as well as for debts that have been paid. This can create a class of permanently stigmatized persons. (Marx, 1988, p. 223)

Of course, the extent to which Americans truly have valued, or continue to value, the opportunity to move on beyond one’s past (especially when it is weighed against other goods, such as law enforcement) is an open question. By contrast with Westin and Baker, and Marx, Gandy (1993) more recently articulated the value of forgetfulness, but with a more defensive thrust. Referring to “the right to be forgotten” as one of the fundamental principles of data protection identified by Flaherty (1989) in his study of privacy policies in Western industrialized societies, Gandy explained:

The right to be forgotten, to become anonymous, and to make a fresh start by destroying almost all personal information, is as intriguing as it is extreme. It should be possible to call for and to develop relationships in which identification is not required and in which records are not generated. For a variety of reasons, people have left home, changed their identities, and begun their lives again. If the purpose is non-fraudulent, is not an attempt to escape legitimate debts and responsibilities, then the formation of new identities is perfectly consistent with the notions of autonomy I have discussed. (Gandy, 1993, p. 285)

But while Westin and Baker, Marx, Gandy, and yet others have drawn attention to the value of starting over, of having a portion of the past forgotten, the issue has been cast, implicitly or explicitly, as one involving a tension between personal or individual privacy and social goods. They have portrayed the issue as a matter of balancing individual privacy against such social goods as law enforcement, government efficiency, or national security. Yet there is reason to believe that this framing of the problem is inaccurate and biased against individual privacy.

The lesson of the 1980s and early 1990s is that when personal privacy is put into a cost-benefit analysis, it generally loses. The needs of government agencies and private organizations or institutions for more accurate and efficient information systems so as to further their goals (law enforcement, national security, administrative efficiency) overpower the desire (need, interest, or right) of individuals to have information about them kept private. Regan (1995) describes how this framing of the issue has led to the loss of privacy protection in several major public policy contexts. She argues against such a reductive framing of privacy on grounds that it does not recognize the social importance of personal privacy. Hence, in our analysis of institutional forgetfulness, we want to argue for forgetfulness as a social good, not just an individual good.
THE VALUE OF SOCIAL FORGETFULNESS

Privacy as an individual good and privacy as a social good are inextricably tied together. To see this, one need only appreciate that the kind of world we live in makes us into certain kinds of beings and certain kinds of beings are essential for a certain kind of world. Democracy depends on individual citizens who are capable of formulating plans for their lives, taking action, thinking critically, and making decisions. Yet individuals of this kind can not develop in an environment of constant surveillance. The problem is not just that democracy is squelched when individuals live in fear of repercussions for any nonconforming behavior; it is also that the mere fact that one is being watched changes the way one behaves, as Bentham and Foucault have taught us. Individuals change their behavior when they believe they are being watched, and come to see themselves as they believe they are seen by their watcher. The very nature of self and the kinds of personalities that develop in a surveillance society are different.7

The argument for privacy as a social good thus encompasses privacy as an individual good; the argument includes both. Privacy is not just something individuals want because it makes them feel good or is good for them; rather, privacy is good for society insofar as it promotes the development of the kinds of individuals who are essential for democracy. A world in which there is no forgetfulness—a world in which everything one does is recorded and never forgotten—is not a world conducive to the development of democratic citizens. It is a world in which one must hesitate over every act because every act has permanence, may be recalled and come back to haunt one, so to speak. Of course, the opposite is equally true: A world in which individuals are not held accountable over time for the consequences of their actions will not produce the sense of responsibility that is just as necessary to a democratic society. Thus, achieving the appropriate degree of social forgetfulness is a complex balancing act, ever in tension between the need to hold accountable, and the need to grant a “fresh start.”

In order to begin understanding the requirements of retention policies, we examined three policy arenas in which forgetfulness seems to play an important and explicit role: bankruptcy law, juvenile crime records, and credit reporting.8 Bankruptcy law involves civil law, juvenile crime records involve criminal law, and the regulation of credit reporting is more concerned with private institutions. We examined these domains to find out if the apparent forgetfulness in these policies is real, to learn how forgetfulness was understood in the development or implementation of each policy, and to understand how the tension between memory and forgetfulness has been played out in American social policy. We also examined the arguments in these domains with an eye to re-deploying them in other domains and to helping us construct a comprehensive approach to data retention.

Bankruptcy Law

The first thing to note about bankruptcy law is that the discussion surrounding it does, indeed, recognize forgetfulness (and forgiveness) as a social good. In the first pages of a 1989 study of bankruptcy and consumer credit in America, the authors write:

Bankruptcy is a powerful phenomenon. It is financial death and financial rebirth. Bankruptcy laws literally make debts vanish. When a judge signs a paper titled “Discharge,” debts legally disappear. (Sullivan et al., 1989, p. 4)

And later:

At the heart of all bankruptcy law, for individuals and for businesses, is the discharge of debts and other legal obligations, the “fresh start.” The notion of beginning anew, of rebirth, lies near the center of our restless, westward-moving culture and is also the central proposition of its dominant religions. Whether a bankrupt debtor, given more time, can pay in full or can pay little or nothing, the relaxation of strict legal obligations is the indispensable centerpiece of American bankruptcy law. (p. 20)

Of course, the textbooks on bankruptcy law and historical accounts of the development of these laws also make it clear that bankruptcy serves the interests of creditors as well as debtors:

Bankruptcy law is a supercollection device for creditors. Indeed, American bankruptcy law arose from two separate bodies of English law, one designed to protect debtors and the other to aid creditors. . . . Ordinary debt collection law has serious flaws from a creditor’s point of view. Its two most important weaknesses are that it is purely state law, making collection across the country very difficult; and it is competitive, with every creditor for itself. Bankruptcy law immediately captures all the debtor’s assets in one country-wide net after a single filing. It also restrains actions by any individual creditor, permitting creditors to act collectively, often through a trustee, to preserve asset values and to ensure a fair distribution. (p. 20)

While the literature we examined did express the concern for forgiveness for mistakes and the good of letting individuals move on, there are reasons to believe that these values alone would not have led to the forgiveness of bankruptcy, were it not for the fact that creditor interests were also served by the forgiveness. Moreover, government (social) interests were at work insofar as there was a perceived need to respond to periodic national financial crises and to facilitate individuals (especially those involved in business) in getting back into economic activity (Warren, 1935).

The literature on the history of bankruptcy law supports Regan’s idea that when policy debates are framed as a
tension between individual interests and social good, individual interests do not win. In bankruptcy law, the tension between individual and social interests was finally (and perhaps, only) resolved when there was a coming together of institutional interests (creditors’ interest in a noncompetitive way to obtain whatever they could), individual interests in being able to start afresh (having their mistakes forgiven and forgotten), and social interests (in responding to major economic crises and getting entrepreneurs back into the economy).

Our research on bankruptcy law thus supports the idea that Americans recognize a social good of forgetfulness. Moreover, the research supports Regan’s conclusion that arguments in favor of social forgetfulness (and privacy protection in general) are more likely to succeed when they are cast in terms of a social good rather than purely in terms of individual interests.

**Juvenile Crime Records**

Juvenile justice has evolved considerably over the last few centuries, concurrently with changing social conceptions of both children and the role of the state. Although there are many different and competing visions of how the state should intervene with regard to juvenile crime, one prominent train of thought has been the liberal (progressive) view of the state as protector of juveniles. Such a view primarily aims at rehabilitating juveniles through deemphasizing their offenses and highlighting their treatment needs (Guarino-Ghezzi & Loughran, 1996). Judge Mack powerfully echoes the sentiments underlying the liberal view:

> Why is it not the duty of the state, instead of asking merely whether a boy or girl has committed a specific offense, to find out what he is, physically, mentally, morally, and then if it learns that he is treading the path that leads to criminality, to take him in charge, not so much to punish as to reform, not to degrade but to uplift, not to crush but to develop, not to make him a criminal but a worthy citizen. (Mack, 1909, p. 107)

Of course, any such goal of rehabilitation must be carefully reconciled with other principles of justice, such as punishment and offender accountability. Juvenile justice statutes, both in the United States and in England, clearly indicate how the courts are expected to hold a balance between the protection of the public and that of the individual child. Section 1 of the Uniform Act states as one of its goal:

> Consistent with the protection of the public interest, to remove from children committing delinquent acts the taint of criminality and the consequences of criminal behaviour and to substitute therefore a program of treatment, training and rehabilitation. (Parsloe, 1978, p. 182)

However, the public interest is here not only defined in terms of protection from delinquent elements, but also in terms of a “reserve capital,” that is, the need to safeguard society’s future. Not only has society an immediate interest in protecting itself from criminal elements, but in the case of juvenile delinquents, it has a future interest in preventing “the deprived and delinquent children of today from becoming the deprived, inadequate, unstable and criminal citizens of tomorrow” (Bean, 1981, p. 126). Clearly, the state has much to gain in avoiding the huge social and economic costs that follow from committing individuals, from an early age, to a lifelong relationship with criminal justice.

Note that such a rehabilitative program is congruent with a number of different philosophical views on the nature of juvenile crime (and the concomitant views with regard to the most appropriate form of punishment). Whether one holds that a child’s criminal behavior is truly criminal or rather simply “naughty,” whether the child is held competent or not to understand the consequences of his or her actions, it is nevertheless understood that, following a certain purgatory, a young person’s mistakes should not unduly burden his or her future goals: “For those offences that could be called ‘crimes’ a child should not be expected to have a criminal record for behaviour that may be transient or reflect a particular stage of development” (Bean, 1981, p. 131). This is the justification for the special provisions within juvenile crime statutes aimed at removing the stigma of a juvenile court history. For example, the Code of Virginia includes provisions for the automatic expungement of juvenile records, for offences that would be felonies if committed as an adult, at the age of 29. All other offences may be expunged at age 19, if five years have elapsed since the juvenile’s last contact with court. … An individual may petition for expungement of all records pertaining to his/her case after 10 years since the date of the last hearing in juvenile court. (Virginia State Crime Commission, 1996, p. 4)

There is thus recognition of the value of social forgetfulness embodied in policies on juvenile crime records. However, echoing our previous discussion on bankruptcy, it is important to note that the background discussion of these provisions points to a coming together of social and individual interests. Individuals are allowed to move on beyond their juvenile criminal records not just because it is good for them, but also because society has an interest in turning juvenile offenders into law-abiding adults. Social forgetfulness serves individual and social interests.

**Credit Reports**

Consumerism, as a way of life, would be significantly dampened without the availability of consumer credit. Without it, families simply could not afford the houses, cars, appliances, and electronic gadgets nowadays synonymous with the good life. The credit-reporting industry has
grown out of the desire for businesses to maximize opportunities for consumers to acquire such goods and services, while attempting to exclude those likely to default on their loans. As Rule explains, “The art and science of credit management lie in determining, in advance, who will pay and who will not, and in screening credit applicants accordingly” (Rule, 1973, p. 178).

Credit evaluation is based on the simple principle that past actions provide a good indication of future behavior. Credit bureaus thus seek to acquire the most complete information possible on individuals, so that their clients (businesses, credit-lending institutions, insurers) may make the most educated guess possible about whether or not to extend credit to applicants. Far from being limited to financial information, the reports assembled by credit bureaus may contain information relating to convictions, suits, employment history, past addresses, family status, etc. In fact, before regulators stepped in, almost any information that could be legally obtained was seen as fair fodder for the credit bureaus’ files, but most importantly:

Credit bureaus place a special emphasis on seeking unfavourable or ‘derogatory’ information. . . . It is much more efficient to aim at excluding bad risks than at including good ones, and derogatory information is to this extent at a premium. (Rule, 1973, p. 193)

Thus, with regard to our previous discussions of bankruptcy and crime records, credit bureaus’ activities would seem to go directly against the idea of granting the opportunity for a fresh start. Such past blemishes are precisely what the credit bureaus are paid to look for:

Worst of all, in the eyes of the credit grantors, are bankruptcy petitions, since they indicate a desire to shirk all debts, which is the most serious sin of all in an industry which prospers only from willingness to pay. (Rule, 1973, p. 194)

In the 1960s, more and more people availed themselves of the services of credit-reporting agencies, for an ever-widening range of purposes. The potential for abuse grew to the point that Congress felt compelled to regulate this booming industry through the Fair Credit Reporting Act (1971, revised 1997). The act was designed to cover a broad range of issues with regard to the activities of credit bureaus; its stated purpose was to protect individuals from the deleterious effects of credit reports, by establishing precise rules under which personal information can be reported. Most pertinent to our discussion, it defined certain categories of information that are subject to obsolescence: bankruptcies, suits and judgments, paid tax liens, accounts placed for collection or charged to profits or loss, and records relating to a crime. For each category, the act established precise time limits after which information must be deleted from credit reports. The FCRA thus ensured that the social forgetfulness principles established in the case of bankruptcy and juvenile crime records were not overwhelmed by the new data collection and aggregation practices of credit bureaus.

In fact (perhaps inadvertently), the act went even further. It prohibited the reporting of “any other adverse item of information” predating the report by more than seven years. It also omitted to make clear not only what it meant by “item of information,” but also how, and from what point in time, it should be judged “adverse.” This is problematic since, as one analyst noted, ‘Items’ may well be continuing matters, such as divorce proceedings or, in investigative reports, disputes with neighbors or employers” (Willier, 1971, p. 55). The interpretive flexibility afforded by such loose formulation, combined with fears of noncompliance with the act, would seem to naturally force upon credit bureaus a conservative reading of what legislators sought to include within the category of “adverse information”:

Since what may be adverse to one creditor, insurer or employer may not be adverse to another, absent any uniform and objective criteria for judgment, almost any items of information must be treated by the agency as adverse. In the extreme, this includes places and time of residence. . . . In short, a consumer reporting agency should look upon any item of information as adverse for purposes of the seven years rule. (Willier, 1971, p. 55)

That is, except for the special categories already mentioned, the act essentially limited credit bureaus to a memory of 7 years or less. Were it not for the generous conditions under which these obsolescence rules may be altogether skirted, the FCRA might have thus provided for some of the strongest policy in current legislation to implement a right to have certain aspects of one’s life forgotten.

Despite its implementation flaws, the FCRA clearly represents a continuation of the philosophies outlined in the case of bankruptcies and juvenile crime records. If the judicial system has sought to provide individuals with some (if limited) means to unburden themselves from their past, the FCRA extends these policies to the new threats posed by data collection, aggregation, and reporting.

THE NEW THREATS TO SOCIAL FORGETFULNESS

While these three cases illustrate historical recognition of the social value of forgetfulness, the trend in recent decades has been in the other direction. Nowhere is this more evident than in the case of transaction-generated information (TGI) which records the details of our interactions with organizations and individuals (phone calls, purchases, geographical location, banking transactions), facilitating aggregation and inordinately increasing our capacity for social memory. As is often the case with computerization, there is in principle nothing fundamentally new about TGI; rather, it is both the scope of and the new possibilities offered by the enterprise that promise to alter social memory in both subtle and dramatic ways:
Quantity: As more and more of our activities are taking place over electronic networks, more categories of data are being collected every day. From an initially fairly limited set including phone calls, banking and credit card transactions, the list now includes highway tolls, e-mail, web browsing, cellular phones, grocery shopping, etc.\(^\text{12}\)

Granularity: For each category of transactions, a finer granularity of data collection is possible; a phone call over a cellular network may be recorded in terms of originator, destination, duration, time of day, type of device used for the call, geographical location of device, movement of device during the call, network services used, etc. This increased capacity for precise metering of user’s activities is part of the tremendous attractiveness of TGI for organizations.

Cross-correlation: Once collected, TGI is easily aggregated and correlated with other kinds of data: Web browsing, demographics, credit card transactions, and cellular use together provide a much finer resolution of the digital persona than each can by itself.

Predictive power: Most importantly, quantity plus diversity plus cross-correlation combined lead to the possibility of “discovering” information not (explicitly) present in the data collection process itself. In other words, such data have predictive power. Because data are collected in electronic format, they are easily amenable to a variety of treatments: multidimensional and statistical analysis, neural networks, information discovery systems, all technologies precisely aimed at extracting new information from the vast warehouses of electronic information gathered by organizations. Even when the information is not available in a suitably discrete format, image-analysis software or text-analysis algorithms may be used to extract pertinent data from video or free-flowing texts. Such technologies may be used with regard to marketing, network management, credit-risk analysis, sales productivity, etc., with the hope that they may help discover rules and patterns of behavior, and predict the future with some reasonably good probability.\(^\text{13}\)

While critics of the panoptic society have justly remarked on the ubiquity of data collection practices, we underline how such practices invisibly extend the persistence of social memory and diminish social forgetfulness. What the preceding list points to is a subtle yet dramatic change in the nature of this memory. Human activities and interactions that were, at one time, not part of the public record now have the possibility of being recorded in varying levels of detail. In most cases, however, there seems to be little concern over the effects of data retention. In fact, organizations have come to see and use such transaction-generated information as a legitimate and highly useful competitive asset.

**POLICY STRATEGIES FOR DATA RETENTION**

We have argued, then, that social forgetfulness is an important social value that is quietly slipping away because of the increasing use of increasingly sophisticated personal data together with a neglect of data retention policies. We have also argued that privacy policy debates should not be framed as a matter of balancing the social goods of information against individual rights or interests in privacy. Rather, the issue should be understood as involving tensions between social goods, the social good of privacy (and forgetfulness), and other social goods. When the value of social forgetfulness has been recognized, such as in bankruptcy law, juvenile criminal records, and credit reporting, legislation has been developed to provide a form of forgetfulness.

The question remains as to what can and should be done to more broadly address the loss of social forgetfulness caused by data retention. Of course, we are not arguing that social forgetfulness should trump all other social values. Our claim is only that it should be given proper consideration in information management decisions and practices.

To begin to answer this question about what can and should be done to address the loss of social forgetfulness, we do two things in this section. First, we sketch what we take to be the most promising policy model for data privacy in general. Our claim is that data retention cannot be addressed in isolation. It can only be addressed effectively as part of a broader privacy policy. We argue for a comprehensive approach encompassing a variety of mutually reinforcing strategies. Second, we review several of these strategies and consider the adequacy of each for addressing data retention.

**A Comprehensive Approach**

Our focus is the United States, and it is well recognized that the U.S. approach to data protection has been piecemeal, ad hoc, and reactive. Unlike Europe and the European Union, the United States has resisted comprehensive legislation in favor of a patchwork of national and state laws. The patchwork of laws addresses government and private-sector use separately, as well as discrete domains within each separately. For example, there are separate laws addressing credit records, driver’s license information, family and educational privacy, telephone records, and video rental records. In many domains, such as records of Internet use, there are no laws, in anticipation that the market will take care of privacy.

Of course, privacy policy in the United States is contested and is becoming increasingly visible, in part, at least,
because of the pressure for harmonization of policies that will be needed for an intensely global economy. The public discussion has led to a variety of proposals for addressing data privacy. These proposals include:

- More (and/or improved) legislation aimed at discrete domains, such as medical records.
- Harmonization of U.S. policy with the EU Directive.
- Creation of an information market for secondary use of personal information (Laudon, 1996; Hunter, 1999).
- No action in the private sector, so that the marketplace has a chance to mature and develop responses to consumer interests in privacy.
- Increased use of privacy-enhancing technologies (PETs).

Initially, when addressing the question of what can and should be done to address data retention, we “seem” to be thrust into the heart of an impossible dilemma. It appears that we must choose between the limitations of a piecemeal approach to data protection and the limitations of comprehensive legislation. Since data retention issues arise in such a wide variety of contexts—any time personal information is collected—comprehensive legislation covering the expanse of data collection would seem to be the only viable approach. Without it, we risk creating a patchwork of inconsistencies and we risk missing domains as we react to incidents of abuse. On the other hand, the nature of the personal data, the context in which it is collected, and the values its use can promote all seem to necessitate that distinctions be made in the way various data is treated. Consider, for example, the differences in the appropriate retention period for birth and death records, medical records, purchase records, membership records of political organizations, and records of use of a parking lot. The appropriate retention period for each would seem to vary widely; for example, birth and death records might be kept (effectively) forever, while records of entry into and exit from a parking lot might be kept for only 24 hours. If distinctions must be made, then it would seem that a piecemeal approach is the most feasible strategy.

This impossible dilemma is, however, a false dilemma, and it can be avoided by means of a conceptual shift in understanding what constitutes a comprehensive policy. We propose that a comprehensive data protection policy be thought of not as a single piece of legislation, the “magic bullet” that will apply to all domains and solve all the problems. Rather, a comprehensive policy should be understood as a policy approach that makes use of a variety of policy strategies consistent with one another and mutually reinforcing. In other words, a comprehensive policy is one that begins with a set of general principles defining broad standards for personal data protection.

The general principles are then implemented in a variety of strategies including legislation in specific domains, structured markets, self-regulatory practices, and privacy-enhancing technologies. Our proposal is consistent with Lessig’s (1999) insight that individual behavior is regulated in four ways, by law, norms, technology, and the market. Lessig emphasizes how the four work together in mutually supporting ways.

The elements of a comprehensive policy are, essentially, already “on the table” in the United States: The Code of Fair Information Practices and the EU Directive provide a set of model general principles; existent legislation provides models for legislation in specific domains; proposals for addressing secondary use via markets have been put forward; PETs are actively being developed.

We next discuss each of these elements and their potential for effectively addressing data retention and social forgetfulness as part of a comprehensive privacy policy.14

General Principles of Fair Information Practices

The cornerstone of a comprehensive policy is a set of general principles that serve as standards to be followed in implementing practices in various domains and sectors, both public and private. The Code of Fair Information Practices (CFIP, developed and recommended in the 1973 Report of the Secretary of Health, Education, and Welfare’s Advisory Committee on Automated Personal Data Systems) and the EU Directive provide good models for development of general principles. Bennett and Grant (1999) have identified a similar set of principles about which, they claim, there is already a broad international consensus.

The principles include the standard that information collected for one purpose cannot be used for another without consent of the individual; they give individuals a right of access to information about them, a right to contest inaccurate information, and so on. Data retention is not explicitly addressed in the CFIP, though it is addressed in the EU Directive and the principles to which Bennett and Grant refer. The principle in the latter is general in specifying that data should be retained “for no longer than necessary.”

It might be argued that a set of standards of this kind, and especially the data retention standard, is not likely to be effective because it is too general to be effective or enforceable. However, we are not proposing that these standards alone constitute a comprehensive policy. Rather, we propose the standards as part of a set of policy strategies. The standards are the starting place but not the ending place of the policy.

It is important to note that these standards may allow exceptions. There may be certain kinds of records such as criminal records of pedophiles that should be treated differently than all other records. The value of the guidelines is that they provide the backdrop against which exceptions
must be justified. In this way, the standards keep the burden of proof on those who would use data in ways that are nonstandard. In the current environment, the burden of proof is the other way. Personal data are used whatever way possible unless it can be shown that there is a reason to restrict their use.

Consideration of adoption of a broad set of standards is timely since the United States and the EU are currently in a struggle over standards. The European Directive on data protection requires that data flowing out of the EU meet certain data protection standards (European Community, 1995). This means that U.S. companies doing business in Europe must meet those standards. Many U.S. firms have used what leverage they have in Europe to fight against the directive, fearing that it will make it much more difficult and costly for them to do business in EU countries. On the other hand, the EU Directive puts pressure on the United States to develop data-protection principles that are in harmony with those of the EU, for this will make data flows between Europe and the United States seamless. U.S. citizens would have greater data privacy if the EU were to win this struggle.

**Legislation**

The three case studies discussed earlier—bankruptcy law, juvenile criminal records, and credit records—provide examples of legislation that has effectively addressed data retention and social forgetfulness. The legislation we examined was not comprehensive in the broadest sense, but comprehensive within specific domains. Bankruptcy law, juvenile criminal records law, and the Fair Credit Reporting Act define the ground rules and structure practices within a particular domain. Bankruptcy law, in effect, defines the “rules of the game” of financial life. Juvenile criminal records laws specify a limitation on what law enforcement agencies can do in pursuing their goals. Similarly, the Fair Credit Reporting Act specifies the ground rules for engaging in the activity of determining an individual’s creditworthiness based on the person’s credit history.

Legislation of this kind, specifying the treatment of certain kinds of records in certain domains of activity, should be a part of a comprehensive approach to data privacy. Our point is only that it need not be the only strategy for addressing data protection. The sector-by-sector, piecemeal approach has several dangers. When each domain is viewed separately, lengthy data retention practices may seem justifiable, and there is no way to take into account the cumulative effects of decisions made in multiple separate domains. Giving up a little social forgetfulness here and there may seem reasonable until we experience the cumulative outcome of having hardly any whatsoever. Other dangers of the piecemeal approach were mentioned earlier. There is the danger of missing important areas of data collection of retention and the danger of inconsistency from domain to domain. Thus, legislation should be used to protect social forgetfulness, though alone it is not likely to do an effective job.

**The Market and Self-Regulation**

Strong arguments can be made for letting the market take care of data protection (including data retention), though these arguments are generally coupled with the idea that self-regulatory, fair information standards will develop. Culnan and Bies (1999) provide just such an argument emphasizing the importance of trust in long-term marketing relationships. They argue that trust is achieved when companies inform customers about how their personal data will be treated. This information can then be taken into account in the customer’s decision to do business with a company, and with this information in the marketplace, the market will produce greater privacy protection or, at least, the kind of data protection that consumers want. Culnan and Bies argue that fair information practices will emerge because of the importance of trust in consumer transactions.

We are less optimistic about the market yielding significant privacy protection, especially when it comes to secondary use of data and data retention. Among other things, data that are retained (but not legally protected) become extremely vulnerable when companies change ownership and/or change their policies. We are inclined to think that legislation is necessary to require, at a minimum, that companies publish their data retention policies, much as food producers are required to display the ingredients in food containers.

However, our point is not to argue against the use of the market, but rather to emphasize the importance of structuring markets to ensure that they promote data protection standards. In other words, markets alone are not likely to achieve a desirable degree of data protection or social forgetfulness. However, markets together with model standards articulated in the overarching general principles of a code of fair information practices can support markets and facilitate the development of trust between consumers and companies. This was the original intention of the CFIP.

**Personal Data as a Personal Property**

Recently, several analysts of privacy policy have proposed schemata for giving individuals more control of their personal data while at the same time facilitating its exchange. Laudon (1996) proposed a National Information Market (NIM) not unlike the U.S. stock exchange, and more recently Rule and Hunter (1999) proposed a schema involving information agents. In both proposals, individuals own their personal data and can make them available for sale,
through the clearinghouse or through agents. In these proposals, individuals can specify the conditions under which their personal data may be sold, and they receive royalties when the data are sold.

While we will not go into the details of these proposals, systems of this kind would seem to hold great promise for addressing data retention and social forgetfulness. Of course, they could not address all data use and they would not work in isolation from other policy strategies. Their promise lies in eliminating the free, secondary use of personal information. Currently, data about us are obtained from a variety of sources and then bought and sold to a variety of users, especially direct marketers. The new proposals eliminate free use and give individuals control of who obtains access and how it is used, but they allow marketers to continue to acquire personal data.

Of course, these proposals require two major changes from current practice, changes that could only be implemented through major legislation. The first change that would have to be made would be to declare that all personal data are the property of the individual. The second change would be to prohibit all secondary use of data without the consent of the individual. The second is a corollary of the first, though it is important to mention because there are different kinds of property. The second change specifies that personal information is a kind of property that one does not lose when one sells it to someone else. In effect, the owner licenses the use of the data for a specified purpose and only for the specified purpose.

Thus, these markets in personal data would have to be implemented as part of a comprehensive policy in which the fundamental “rules of the game” of personal data acquisition, use, and retention are specified.

Once these two principles were articulated as part of a set of data protection standards, the best schema for a marketplace in personal data could be debated and chosen. As mentioned earlier, such a market would not involve all personal data. For many activities such as applying for credit or a job or insurance, individuals could be required to release relevant data. Moreover, some personal information such as criminal records would not be in the control of the individual. Use and retention of these records would be addressed by legislation. Nevertheless, the use of data collected for one purpose, in one domain, could not be sold to others without permission from the individual. Indeed, in these secondary market schemata, individuals could decide for themselves how long certain kinds of data remained in the system. For example, if an individual wanted to receive advertisements from financial services but did not want to release his or her financial history, the individual could restrict the sale of his or her personal data as such. If an individual wanted to receive advertising about vacation properties but didn’t want the advertisers to know all the places he or she had lived in the past, the individual could specify this.

To be sure, these proposals for markets in personal data do not address all data protection or data retention issues. Nevertheless, they could be a significant component of a comprehensive policy.

Privacy-Enhancing Technologies

Technology can intervene in two different ways with regard to data retention: It can altogether prevent the collection of identifying data before it accumulates, or it can help anonymize data after collection has taken place. In both cases, the trick lies in cutting off the link between individuals and data.

The first scenario has been extensively explored by cryptologist David Chaum in a series of widely circulated papers (Chaum, 1981, 1985, 1992). Chaum argues for a computerized world in which cryptography plays a central role in providing individuals with some degree of control over their electronic privacy. His vision is highly original in that it posits no fundamental antagonism between two seemingly conflicting concerns: protecting the individual’s privacy, while ensuring organizations of all the expected benefits of computerized bureaucratic rationalization. In fact, Chaum’s work has been a powerful example of how highly original scientific and technological work may flow from a research program articulated around precise societal concerns.15

Although Chaum’s contributions touch many areas within digital security—electronic cash, network anonymity, electronic wallets, signature systems, to name a few—one particular aspect of Chaum’s work is especially relevant to our concern over data retention. Chaum has observed that computerization brings about an important change with regard to the ways in which individuals obtain and present credentials (academic degrees, permits, etc.) to and from organizations. Individuals are less and less involved in the process, and simply do not possess the relevant documents—e.g., universities exchange transcripts directly, without the student’s mediation. Clearly, for Chaum, if information about individuals is stored in remote databases and freely exchanged between organizations, there is no hope for them to regain control of their personal information:

> The trend today is toward taking monitorability and control of the credentials process completely away from individuals by allowing organizations to be the repositories of all credential data. Individuals would merely provide the identifying information that allows linking to their own credentials. (Chaum, 1985, p. 1039)

Not only do organizations bypass individual control by detaining and exchanging personal information, but using universal identifiers (Social Security numbers, SSN)

---

15 Although Chaum’s contributions touch many areas within digital security—electronic cash, network anonymity, electronic wallets, signature systems, to name a few—one particular aspect of Chaum’s work is especially relevant to our concern over data retention. Chaum has observed that computerization brings about an important change with regard to the ways in which individuals obtain and present credentials (academic degrees, permits, etc.) to and from organizations. Individuals are less and less involved in the process, and simply do not possess the relevant documents—e.g., universities exchange transcripts directly, without the student’s mediation. Clearly, for Chaum, if information about individuals is stored in remote databases and freely exchanged between organizations, there is no hope for them to regain control of their personal information:
makes it possible to crosslink credentials between organizations, creating, in effect, a “dossier society.”

Evertse & Chaum (1987) have suggested the use of “digital pseudonyms” whereby each individual is known to an organization by a pseudonym. When an individual receives a credential from an organization, he or she can present it to another organization in order to gain access to some service, but no linkage is possible between the two databases. Because the individual maintains different pseudonyms for each organization he or she interacts with, no crosslinking is possible, and no dossier may be constituted. More relevant to our purpose, Chaum remarks:

> There are additional benefits to changing pseudonyms aside from the weeding out of obsolete information. The periodic reduction to essential information also prevents organizations from gradually accumulating information that might ultimately be used to link pseudonyms. (Chaum, 1985, p. 1042)

That is, the system also provides a structural mechanism by which information linked to individuals can be “forgotten.” A simple change of pseudonym in effect removes any possibility of linking past information to the individual.\(^\text{16}\)

While Chaum’s approach effectively prevents the linking of transactional information to individuals, other approaches attempt to sever the link after data has been collected, by removing all information that can lead to identification. Unfortunately, such scrubbing of data is extremely difficult to achieve in practice. Sweeney (1997) and Schneier (2000) list some of the limitations of securing the privacy of data in this way, showing that for every strategy for removing identifying information, there exists a counterstrategy that can be used to infer identities from contextual information gathered from a group of related records. At a more fundamental level, the concept of individual anonymization may not be sufficient. As Vedder et al. (1998) point out, the privacy of groups is also challenged by the practices of data collection, retention, and mining.

In spite of these difficulties, PETs can and should play an important role within the general framework of a comprehensive policy. Their use would be facilitated by principles that would sketch the contours of the landscape to be achieved through technology, and they, in turn, would support the achievement of those principles.

**CONCLUSION**

In this article we have pointed to the importance of social forgetfulness and explored the relationship between social forgetfulness and information technologies. The nature of public/institutional memory is dramatically changing due to the evolving character of information technologies. While preserving the opportunity for a second chance might have been easily achieved in the past, it has become increasingly difficult today. The ongoing balancing of “discard and forget” and “preserve and evaluate” has been skewed in favor of the latter. Unless data retention issues are addressed explicitly as part of a comprehensive policy approach to personal privacy, we will gradually move to a panoptic society in which there is little social forgetfulness and little, if any, opportunity to move on beyond one’s past and start afresh.

Robert Gellman (Personal communication, 1999) warns that there is a trend towards increasing maintenance of data. Even in the three cases we discussed wherein social forgetfulness has been institutionalized, there are signs of this forgetfulness being eroded. More juveniles are being tried as adults; bankruptcy law is being tightened (Johnston, 1998); and limitations on data retention in credit reporting are being undermined by other, nonregulated, information services. Thus, there is need for a reaffirmation of the social value of forgetfulness, along with more extensive study and focus on the topic of data retention, through empirical and cross-cultural studies.

We have argued that data retention cannot be addressed separately from other data protection policies. There are too many kinds of data being used in too many different types of context. Data retention must be addressed as part of a comprehensive data protection policy. We have argued for a comprehensive policy that consistently uses a variety of strategies including an overarching set of standards, legislation in specific sectors, a structured market, and privacy-enhancing technologies. We discussed each of these separately, showing how each contributes to data protection. A market in personal data would work well as part of a comprehensive policy if the market were developed on the premise that individuals own their personal data. A comprehensive policy would counteract the problems of the current American approach, which is piecemeal and ad hoc.

**NOTES**

1. See Schwartz and Reidenberg (1996) for an extensive review and analysis of this question with regard to the United States.

2. This is somewhat echoed by the European Directive on Data Protection (EU Directive), which extends its protection only to cases where “the processing of…data is automated or if the data…are contained…in a filing system structured according to specific criteria relating to individuals, so as to permit easy access to the personal data in question” (European Community, 1995, p. 15).

3. Of course, retention policies are influenced by a variety of factors beyond the availability of archiving technologies, most notably fear of litigation and regulatory requirements—see Grady (1996) and Skupsky (1993) for arguments on how retention of records may both expose to and protect against litigation.

4. In some respects, though, data may well endure longer in paper form than in an electronic environment, depending on a variety of
factors. As David W. Charmichael, county records manager and archivist for Westchester County in New York, testifies, “Westchester County still retains its first book of records from 1684, but its first computer tapes from 1977 are unreadable after just 21 years.” (Charmichael, 1998, Personal communication). In other words, institutional memory can turn out, in an electronic environment, to be a function of how often and what kind of technological changes an institution makes. When new technology is accommodating, data endure and it takes an intentional act to delete them, whereas when new technology is not accommodating, data may become effectively unusable.

5. This is echoed by Schwartz and Reidenberg’s (1996) survey of American data protection law: All requirements for retention of data are requirements of minimum duration, motivated by administrative requirements. In their analysis, Schartz and Reidenberg place great faith in the need for institutions to divest themselves, for reasons of efficiency, of the burden of accumulated data, thus enabling an ad hoc institutional forgetfulness, but also acknowledge that marketing divisions may well wish to keep the data, in order to establish long term consuming patterns [Schwartz & Reidenberg, 1996, sections 10-1(a)(4), 10-2(a)(4), 11-1(a)(4), 11-2(a)(4), 12-1(a)(4), 12-2(a)(4), 13-1(a)(4), and 13-2(a)(4)].

6. This is echoed in Frederick Turner’s classic thesis on the American frontier ideal, The Idea of the Frontier in American History: “In the long run, the effective force behind American democracy was the presence of the practically free land into which man might escape from oppression or inequalities which burdened them in the older settlements” (Turner, 1986, p. 274).

7. See Reiman (1995) for a lucid articulation of this argument.

8. There are of course several other mechanisms within law concerned explicitly with mediating the tension between social justice and the opportunity to start over, such as free pardon, remission of sentencing, amnesty, statutes of limitations, etc. The precise makeup of such devices is naturally highly dependent on the social mores of the times: In France and Britain, for example, free pardon proved a useful mechanism to increase the size of both royal armies and new colonies (Foviau, 1970).

9. See McNamara (1973) for a more complete legislative history of the Act.

10. Even within those rules, credit bureaus found ample room to gnaw at the forgetfulness principle: “In Equifax, Inc., an FTC administrative law judge found that the reporting agency violated the Act by inserting phrases in its reports such as, ‘[i]n compliance with the Fair Credit Reporting Act, no additional information can be reported from this former employer concerning employment experience prior to seven years ago.’ The quoted phrase was inserted in consumer reports only when Equifax believed it had adverse obsolete information” (Sheldon, 1994, p. 160).

11. The rules limiting retention are waved under conditions easily met by almost any substantial credit, job, or insurance application. As a manual from the Associated Credit Bureaus explains, “Congress accepted the argument of some ‘specialty’ consumer reporting companies who make reports on consumers where large sums are involved, and exempted certain reports from the obsolescence section and any adverse item, no matter how old, may be reported if the report is being done for a credit transaction or life insurance policy which will be for at least $50,000; or for employment purposes where the annual salary will be at least $20,000” (Associated Credit Bureaus, 1975, p. 710).

12. Although not yet quite of the same nature, videotaping of public spaces will eventually also fall within this category, especially when coupled with face recognition technology (Thomas, 1998). In the United Kingdom alone, an estimated 200,000 cameras cover public spaces (Davies, 1997).

13. For a more detailed discussion of the technologies of data mining and knowledge discovery, see Mattison (1996).

14. Although we do not discuss them here, systems such as the W3C’s “Platform for Privacy Preferences” (P3P) are part of such a comprehensive privacy policy, insofar as they are “designed to inform users about any secondary use of their data so they can make informed choices about whether or not to provide data that might be used for these purposes” (W3C, 2001, p. 24). Such systems are, however, unable, in and of themselves, to either prevent collection or enforce disposal.

15. Chaum’s highly innovative work has not translated into market share: His celebrated anonymous cash business, Digicash, filed for bankruptcy in 1998. Many of Chaum’s idea about anonymity and pseudonymity have been implemented within Freedom, an “identity-management kit” for the Internet produced by Zero Knowledge. Whatever their commercial success, privacy-enhancing technologies such as those developed by Chaum will have, at the very least, “initiated a shift of imagination” (Agre, 1997).

16. This is precisely why changing personal names is a severely restricted process in some countries—see Lapierre (1995).

REFERENCES


regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Community*, November 23:L: 281.


