Privacy and Advances in Social and Policy Sciences: Balancing Present Costs and Future Gains

Paul D. Reynolds

Abstract: Individuals have a right to an efficient, effective, and just society. Social science and policy research has provided considerable information about and understanding of some of the most critical issues in modern societies. This has often required extensive, sensitive data on individuals and organizations. As research progresses, continued scientific progress will require more complete and diverse data to further advance understanding of critical societal issues: health, crime, education, economic growth, etc. Increasing concerns regarding individual privacy and organizational confidentiality may lead to restrictions on access to sensitive, critical information on individuals and organizations as well as the capacity to assemble information from diverse public data sets. These dilemmas will be a continuing issue: only temporary compromises can be expected. It is proposed that the optimal strategy for the social science community is multi-faceted: clearly demonstrate respect for individuals and organizations through appropriate handling of private and sensitive information while simultaneously making clear the societal benefits of an enhanced understanding of basic phenomena.

Key words: Ethics; confidentiality.

1. Introduction

A 20 year longitudinal study of a single cohort of males (from ages 10 to 30) and their contacts with the police has found that a minority (18%) were responsible for a majority (72%) of all delinquent and criminal acts. Most (54%) had no contacts of any kind. This suggests that persistent criminality is a major factor affecting the prevalence of victimization (Wolfgang, Thornberry, and Figlio 1987).

While the character and effect of Acquired Immune Deficiency Syndrome (AIDS) is clearly a topic for biological and medical science, its spread throughout the population is related to the sexual practices of distinct subgroups as well as the patterns of illegal drug use. A substantial amount of carefully conducted survey research has provided detailed documentation of the paths of transmission (Feldman and Johnson 1986).

The five Negative Income Tax field experiments involved random assignment of low income and welfare households to different treatments. Some were guaranteed an annual income regardless of whether they worked or not. Changes in
decisions to work before and after participation were compared to changes in households without guaranteed income. There were small reductions in the hours worked for the primary male wage-earner, larger reductions for wives and dependents (Brown 1980, pp. 75–90).

A representative sample of 5,000 U.S. households was tracked annually from 1968 to 1973. It was found that household economic well-being (income/needs ratio) varied dramatically from year to year for a large proportion of households. Only a small proportion of households (9%) are in the lowest fifth of the income distribution for five consecutive years. The major factors affecting household economic well-being are changes in household composition and labor market conditions outside the control of the principal wage-earners (Morgan et al. 1974).

A nationwide study of the variation in public school resources and the educational attainment of students finds that the effect of schools is relatively small compared to the socio-economic background of the students. Further, students from disadvantaged backgrounds tend to perform better when enrolled in public schools where the majority of students are from advantaged backgrounds (Coleman 1966).

Using data collected by a commercial credit rating firm and edited and processed by the Small Business Administration, the role of new firm births and small firm expansions in subsequent regional economic growth appears to be positive; the more firm foundings and expansions, the greater the economic growth in the next two years. However, the higher the rate of firm death and contraction the greater the subsequent economic growth. Firm births and expansions tend to occur in the same regions with higher rates of firm death and contractions (Reynolds and Maki 1990).

* * * *

These six examples of social science research have a number of common features. First, they deal with a variety of topics with major practical and public policy significance: major features affecting criminal victimization, spread of a deadly incurable disease, factors affecting labor force participation, the extent of a permanent "poverty underclass," factors affecting educational success, and the role of entrepreneurship in economic growth. All of these topics continue to be of major significance in modern societies and are a central focus in domestic political discourse. Substantial monies and legislation are often directed toward these issues: crime, health, welfare, education, and economic development.

Second, this research has dealt with some of the most sensitive topics associated with the human experience: illegal and criminal behavior; the complete range of sexual practices; the systematic, communal use of illegal drugs; household income and expenditures; individual labor force participation; academic performance related to ethnic and socio-economic class identification; and the founding, success, and failure of individual businesses.

Third, the research could not have been conducted without substantial information on the specific units of analysis: individuals, households, school districts, and businesses. Five of these six research programs involved longitudinal data collection, necessary to develop inferences about causal relationships. But this required maintaining identity of specific individuals, households, and business firms for repeated data collection over a number of years. Several of the projects involve

2 Unless indicated otherwise, all references to research, customs, regulations, and laws are for the United States.
analysis of substantial, existing administrative data sets; others involved major efforts to collect original data.

Fourth, this research has a major relevance to conceptions and understanding of basic processes. Theories and conceptualizations developed in the social science enterprise were affected by these research programs. Some of these studies provided major new insights into fundamental processes.

Fifth, none of these examples is marred by disclosures that have harmed the participants or invasions of privacy that have caused distress, even though there is a substantial potential for such problems. Because the investigators and agencies responsible for this research completed the work with care and professionalism, no such problems developed. In fact, a 1988 summary of the contributions of the social and behavioral sciences noted that "...past research access to many privileged, confidential, and anonymous data sets in the United States has not produced a single case of harm involving a breach of the anonymity, confidence, or privacy of any respondent (Gersten, Luce, Smelser, and Sperlich 1988, p. 223)."

"Not a single case" refers to the absence of any apparent public controversy in the mass media or the courts. In the context of a 1978 national survey that found that 81% of the respondents had "never been victims of what they felt were improper invasions of privacy" (Louis Harris & Associates and Westin 1981, p. 18) – it is not surprising. This is to be contrasted with hundreds of years of case law that have developed around the use of human subjects in medical research (Katz 1972). Many studies led to severe negative consequences, including death, for the research participants. The use of human subjects in medical research is now so complex that most medical schools have "ethical experts" to guide physician-investigators as they attempt to balance research and subject-patient interests.

In short, these examples represent the types of contributions responsible social science can make to society. There continues to be a substantial potential for making major policy and conceptual contributions with positive effects on the lives of citizens and the efficacy of public organizations. The issue for attention is the future capacity to continue to provide these intellectual contributions.

2. The Major Issue

The major issue has developed as the confluence of several trends. First, there is an increasing public concern regarding the ability to retain a sphere of privacy and avoid surveillance by both public and private corporate actors. Second, the major problems of advanced societies are becoming recognized as complex and multifaceted, often involving the interrelations among a number of diverse constituencies. Optimal solutions will be enhanced by more sophisticated and complete social science analysis. Third, as more is known about these major societal phenomena, it is clear that they involve a wide range of diverse processes. Further advances in understanding will be possible only if more diverse and complete information can be assembled. Existing public data sets are major, low cost sources of such information.

Will significant social science contributions be possible in the future? Two issues are relevant: the availability of resources and legal or legislative constraints on the access of data. Both are affected by the public image – reflected in the public consciousness and conceptions of the elected officials – of the social science enterprise and by their
potential for making important contributions to the understanding of critical issues. What is the best strategy for the social science enterprise to take in response to the current challenges associated with increased concerns over individual and organizational privacy?

There are two relevant dimensions in the characterization of social science. Considering the magnitude of risk for participants (moderate versus none) and the consequences (trivial versus significant) leads to four possible cases. No risk, significant consequence would be the ideal strategy for characterizing social science. The promise of something for nothing would have universal appeal. Unfortunately, not all social science can be characterized as either no risk or significant. It would seem an unrealistic characterization. Moderate risk, trivial consequences is both inaccurate and unlikely to gain much advantage. To emphasize costs without benefits has little appeal. Two major strategies remain.

No risk, trivial consequences has the value of emphasizing the absence of costs. One way to emphasize the absence of any risks is to focus on the lack of consequences for anything. If the results are unrelated to public policy – changes in tax laws or welfare benefits – then it is easier to promise participants they will be unaffected by the research, now or in the future. Unfortunately, this would make it hard to justify requests for resources to conduct the research.

Efforts by social science investigators to emphasize techniques that will reduce the risk of privacy invasions – ideally to acceptably low levels – but fail to emphasize the positive benefits of social science research contributions will reinforce this perspective. If social research is seen as having little practical consequences for the host society that supports it, there is not much reason to expect that it will be taken seriously. If it has little to offer the host society, it is unlikely that exceptions for social science will be made to the increases in the scope and impermeability of legal and technical privacy "shields." Social science research may be treated as an avocation pursued to satisfy the curiosity of the investigators and, therefore, unworthy of special consideration.

Moderate risk, significant consequences has the benefit of emphasizing the payoff from conducting the research. A general perception that the social science enterprise has a major contribution to make to the public discourse on important topics may outweigh concerns regarding disclosure or misuse of private or confidential information. If, in addition, direct harm or embarrassment is very rare, there is even a more favorable balance supporting data base research. Such a favorable benefit/cost ratio enhances the cases for both special privileges and financial resources to support the research.

The following discussion, then, is presented as background for an appropriate strategy for enhancing the potential for social science advances. It is initiated with a brief review of issues associated with moral analyses with special attention to the development of individual rights and a mechanism for foregoing rights – informed consent. Attention is given to the differential problems related to individuals and to organizations as a special type of social collective. The most widely accepted standard for the incorporation of participants in research in the United States, the Common Federal Policy for the Participation of Human Subjects in Research, is summarized. Emphasis is given to the Common

---

3 The potential for high risk social science research, compared to the risks associated with medical research, is quite small, hence the attention to moderate risks.
Federal Policy application to non-experimental data collection. This is followed by detailed attention to informed consent in non-experimental research. The final section proposes the development of a "Federal Data Base Research Certificates," which may facilitate access to federal data sets for social science research.

3. **Overview of Moral-Ethical Analysis**

The basic element of all moral analysis is a personal, individual sensation that something is not right or just or proper. In extreme cases this is a powerful experience associated with moral outrage, as when one hears of a heinous or savage act directed toward innocents. The abuse and murder of young children, for example. One of the major endeavors in philosophical, theological and moral analysis is attempting to determine the basis for emotional reactions to these events. Such arousal may reach maximum strength when it involves the persons themselves. "They can't do that to me!" reflects a common response to personal injustice.

Three legal, moral, or philosophical frameworks have been developed to account for such personal reactions: concern for the integrity of individual rights; inappropriate distributions of costs (harm) and benefits (rewards); and adherence to the appropriate standards for interpersonal conduct.

Individual reactions to violations of basic human rights are frequently quite dramatic. Many have very strong reactions to the most serious instances, such as the capricious detention and torture of political opponents. But milder reactions occur to other situations, such as those where individuals are denied the right to political self-determination. Reactions may even occur in some research contexts, as when unsuspecting individuals are exposed to a research risk without their knowledge or consent.

Moral disquiet may also result from situations where there is an uneven distribution of costs and rewards. The presence of hunger in situations of surplus food often raises concern that there is an inappropriate distribution of benefits. Concerns develop when some may suffer unduly for the benefit of others, as when the poor and disadvantaged are used as research subjects in testing medical therapies that will benefit the elite.

Moral unease may also result from situations where one treats another inappropriately. It may arise from seeing customers stepping over the hungry homeless on the sidewalk to enter a restaurant. Aren't people supposed to show more charity? It may also occur in research settings where humans are exposed to various experimental treatments where they are not fully aware of all purposes and procedures—treated like "rats in a lab."

Moral outrage may reach its peak when all three elements are present. The Nuremberg Code was developed in reaction to medical experiments with concentration camp inmates in Nazi Germany. For example, in order to document the details of the failure of critical biological systems, normal adults were placed in ice-water until they died. As the research subjects were capriciously chosen from the inmate population and given no choice in the matter, they had no opportunity to exercise any rights. Their sacrifice was related to efforts to facilitate the survival of German fighter pilots downed in the North Sea, a rather extreme case of unequal distribution of costs and benefits. Finally, for one person to deliberately freeze another to death is not considered an appropriate interper-

---

4 This material is covered in more detail in Reynolds (1979, 1982).
sonal treatment. One of the major reasons the concentration camp medical research caused such a strong reaction was that all bases for moral outrage were present.

These underlying issues are briefly sketched in Figure 1. The focus of the analysis is the presence of individual moral or ethical arousal. The sources of such concerns are indicated in the left side of Figure 1. They are divided into the immediate source of the emotional response and the underlying issues or conceptualizations that provide an understanding of the basis for the moral reaction.

The right side of Figure 1 focuses on the next stage of the process, the actions that might be taken to ameliorate the moral uneasiness. This is a partial list, ready for augmentation. Some actions may have a direct relation to the source of the moral outrage. For example, if concern reflects the uneven distribution of benefits, a person may take action to correct this imbalance through direct action, such as distributing food to the poor. The person providing the response may feel much better about themselves, but often the problem continues to persist.

Major changes in the underlying situation will often require concerted collective action. Such is the case when new laws or government programs are initiated. The organized public response to concerns regarding individual privacy is an excellent example of a collective response to a widely perceived moral problem. In this case, the suspicion that private information might be used to the disadvantage of those who were the source of the information, either deliberately or through carelessness, has been a central issue. Much of the writing on privacy threats reflects efforts to develop a special interest group that can be effective in changing administrative procedures and laws related to privacy.

The writings on these issues are extensive; they commence with the earliest of efforts of theologians and philosophers. For the purposes of this analysis—oriented towards one individual right (of privacy) and the potential contributions of the social science enterprise—the emphasis will be on the issues underlying the first analysis. This includes attention to the idea of rights, a recent development in the evolution of social contracts between rulers and the ruled.

However, the other strategies cannot be ignored. One of the basic justifications for the social science enterprise is that it provides collective benefits that outweigh the costs borne by the research participants, an application of cost/benefit analysis. Furthermore, the interpersonal norm that one should not invade the privacy of others is violated in research on sensitive personal topics. As professional snoops, social scientists must live with the notion they are not always “nice people” when they study important topics.

4. The Social Contract and Rights

Much of the history of societal development reflects the continuing conflict between the individual and the host society—the state. These conflicts produce dilemmas which may be resolved informally in small scale pre-literate cultures. But in complex and multi-faceted societies, the solutions of these dilemmas become, as well, complex and multi-faceted. The following reviews three critical elements of this relationship: the development of social (political) contracts, rights, and informed consent.

A critical issue in socio-political systems is assigning responsibility for control over different aspects of individual and social life. All societies have one or more individuals responsible for decisions regarding common or group interests. Such interests
Fig. 1. Overview of relationship among moral analysis, outrage, and action

<table>
<thead>
<tr>
<th>ETHICAL ANALYSIS</th>
<th>ETHICAL RESPONSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underlying Legal, Moral</td>
<td>What to do? How to relieve moral</td>
</tr>
<tr>
<td>or Philosophical Issues</td>
<td>outrage?</td>
</tr>
<tr>
<td>1. Justification for</td>
<td>Rights violated or ignored without</td>
</tr>
<tr>
<td>formation of a political</td>
<td>obtaining informed consent. (How to</td>
</tr>
<tr>
<td>system. Give leaders</td>
<td>infer true voluntarism if behavior is</td>
</tr>
<tr>
<td>responsibility for</td>
<td>ignored?)</td>
</tr>
<tr>
<td>selected tasks (foreign</td>
<td>Uneven distribution of rewards and</td>
</tr>
<tr>
<td>policy, control of crime)</td>
<td>costs? (If they can be measured</td>
</tr>
<tr>
<td>as long as individual</td>
<td>with a common metric.)</td>
</tr>
<tr>
<td>rights not abridged.</td>
<td>MORAL OUTRAGE OR ETHICAL CONCERNS</td>
</tr>
<tr>
<td></td>
<td>(Queasy feelings)</td>
</tr>
<tr>
<td>2. Criteria for evaluation of public programs and policies? Select those that maximize the ratio of total benefits/costs. (Should disadvantaged get special treatment? How much?)</td>
<td></td>
</tr>
<tr>
<td>3. Strategy for being a good person; standards for personal treatment of others.</td>
<td></td>
</tr>
<tr>
<td>- Follow “rules,” e.g., golden.</td>
<td></td>
</tr>
<tr>
<td>- Situation ethics.</td>
<td></td>
</tr>
<tr>
<td>- Do what feels “just.”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A. Report as “illegal” (Only if there is consensus on what is wrong will laws exist.)</td>
</tr>
<tr>
<td></td>
<td>B. Publicize issues or problems; hoping for public support and/or political pressure.</td>
</tr>
<tr>
<td></td>
<td>C. Confrontation with responsible authorities.</td>
</tr>
<tr>
<td></td>
<td>D. Personal action or charity toward those cheated to compensated for injustices.</td>
</tr>
<tr>
<td></td>
<td>E. Change the “system” to redistribute resources and/or rights.</td>
</tr>
<tr>
<td></td>
<td>F. Personal influence, persuasion to change the habits or behavior or offenders.</td>
</tr>
<tr>
<td></td>
<td>G. Organize peers and professional groups; develop codes of ethics; establish review committees.</td>
</tr>
</tbody>
</table>
may coincide or conflict with the interests of specific individuals. The controversy over the allocation of domains of influence between the rulers and the ruled has been a central feature of the development of the modern political state (Gough 1957).

The emergence of the concept of "right"—as something one is entitled to—was a key feature of the development of modern social contracts. Two different strategies are possible for assigning rights to rulers and the ruled. Traditional rulers, who do not expect explicit consent of the governed to hold their position, tended to assume they controlled or owned everything, which they could then dispense to others. For example, a king might transfer the responsibility to govern a providence and collect taxes to a loyal subordinate. One of the most important examples of a breakdown in the absolute control of all rights by the king was the development of the Magna Carta in England in 1215, where the people—actually a group representing the elite—circumscribed the decision-making domains of the king, making clear that influence in a number of specific domains were outside his control. The king was expected to adhere to this social contract. An explicit statement of an agreement developed between a king and his people.

Rights, then, emerged as a way of constraining the power of a ruler. Rights reflected the allocation of topics on which control was to be exercised by rulers or the ruled. By the time the U.S. Constitution was being drafted, the situation had changed substantially. By then it was common to assume that the people retained control of all domains of individual and societal life. Rulers were to be allowed control over certain aspects, aspects deemed suitable for centralized decision-making (e.g., common defense, control of interstate commerce, etc.). This shift in philosophy is well represented in the Tenth Amendment of the U.S. Constitution, which reserves all rights not given to the government for the people.

Just how certain rights are explicated for this bargaining process, and the arguments used to justify the establishment of rights—religious or philosophical—is a more complex issue. It is clear that the concept is a straightforward aspect of discussions related to the allocation of decision-making privileges among those capable of negotiating a social contract. Substantial complications arise when the rights of those who cannot actively participate in such negotiations are introduced: animals, fetuses, those hopelessly comatose, the mentally incapacitated, or future generations (not yet conceived).

One aspect of this conception of rights, as a key feature of the social contract, is the negotiation between the people and the rulers (or the state). Once the idea that rights are to be allocated between different parties is accepted, the inventory of "rights for allocation" becomes an important topic for debate. Much of the early discussion centered on more immediate and pressing problems—the right to negotiate foreign policy and provide for the common defense, the right to define and control crime and violence within the society, the right to collect revenues to support the central government, the right to adjudicate civil disputes, the right to private property, the right to travel, the right to due process in denial of rights, and so on. Negotiations over the right to privacy or access to information about the individuals in modern societies is a relatively recent development (Wacks 1989). The first serious attention in the U.S. legal system was produced in 1890 (Warren and Brandeis 1890). Apparently this was a response to mass media intrusion into the social events organized
by the elite of Boston, Massachusetts (Prosser 1960).

An important distinction has recently developed between positive and negative rights. This refers to the responsibility of the state in relation to individuals. A "positive" (or active) right is one that the state is expected to actively enforce. For example, the state is expected to take actions to ensure a right to be free from criminal acts – offenses against a person or his or her property. "Negative" (or passive) rights are those which the state will support, but will take no active role to ensure. For example, a person has a right not to be slandered by another. If one person wishes to charge another with slander, the state will, through the court system, make a decision on the legitimacy of the charge and support a valid claim for recompense. However, the state will not take an active role in identifying cases of slander and initiating sanctions. Most cases of privacy infringement of one individual by another would be considered as infringement of a negative or passive right. Individuals have a right to personally defend their privacy interests, but not a right to have the state enforce a general right of privacy. Currently the United States federal government takes an active stance to ensure privacy only for United States federal data, as reflected in the Privacy Act of 1974.

A critical feature of the U.S. Constitution, as with other social contracts, is the presence of two distinct types of actors: the ordinary citizens (or natural persons) and the state, represented by elected officials and administrators. Rights were apportioned between these two entities. Extension of rights to "unnatural persons," or juristic (legally defined) social actors, has been a recent phenomena. In the usual case, the juristic person is an organized group, a collection of coordinated individuals sharing resources and efforts to achieve common objectives. It is not common for juristic persons, such as corporations (or business organizations) and even government agencies, to be treated as social actors with claims to rights approaching that of natural persons.

Individuals that share common characteristics, but have no history of systematic interaction and are not coordinated for any common purpose, may also be considered a social aggregate or group, a category of individuals with common interests. This is quite a different type of social entity. In many cases these social aggregates share only cultural or physical traits – Black-Americans, the handicapped, comatose patients. Efforts to gain recognition for a claim of "group rights" have been thwarted by the absence of collectives or aggregates as participants in the theories of social contracts. Only individuals, for example, are referred to in the Universal Declaration of Human Rights, adopted by the General Assembly of the United Nations in 1948 (United Nations 1949). It was a major issue in the Bakke case, where a white male sued for discrimination when refused admission to a California medical school to which less qualified minorities were admitted. The United States Supreme Court ruled in his favor, emphasizing individual over group rights (Van Alstyne 1978). A proposal was made at one university to prohibit research that might create "social risks" to the reputation of a social group or institution (Hart 1973; Irving 1973). It was not adopted. Rights of social aggregates are developing more slowly than that of natural persons or organized collectives.

5. Informed Consent

There are many situations where individuals may wish to give up certain rights (Faden
and Beauchamp 1986). For example, they may wish to exchange property – cash for real estate, or vice versa. They may wish to restrict their freedom of association – a fundamental feature of a marriage contract is a promise to restrict intimacy to one’s mate. Some occupations or leisure pursuits involve a considerable risk to one’s life and limb; this involves foregoing a right to physical well-being. Observing the individual’s behavior – signing a sales contract, a marriage ceremony, or pursuit of a risky occupation – provides important evidence that the individual wishes to forego a right. Individual behavior, however, has generally NOT been trusted as sufficient evidence. A set of conditions has emerged to facilitate judgements regarding a person’s intent in foregoing a right. These conditions focus on the context in which the decision is made; they are unrelated to the actual behavior.

This set of conditions is referred to as informed consent. Three elements are universal, a fourth is involved in risky research. The first critical element requires that a person is capable of making a rational, mature judgement (children or the mentally deficient are excluded). The second is that the individual is fully informed of the conditions associated with making the decision. The third element is that he or she must be true volunteers, acting on his or her own volition without coercion. These latter two elements give rise to the phrase “informed consent.”

The major criteria for informed consent to participate in research has developed in medical research, usually where some type of treatment is being tried – on an experimental basis – for the benefit of the patient-subject. If the major objective of the physician-investigator is to improve the physical well-being of the patient, then these three elements are considered sufficient. If, however, the investigator has other objectives – objectives that may not coincide with the patient’s well-being – then the fourth element is considered critical. It refers to a full awareness of the potential negative consequences.

Hence, the four features of informed consent: rational adult, full information, voluntary decision, and awareness of consequences. In addition to providing a legally binding relinquishing of an individual right, other advantages have been proposed. They include (Reynolds 1979, p. 87):

1. Promotes respect for the participant’s right to self-determination in control of other basic rights (freedom from bodily harm, privacy, freedom of movement, and the like).
2. Promotes respect of participants as unique individuals, as persons.
3. Minimizes fraud and deceit.
4. Minimizes coercion and duress.
5. Encourages self-scrutiny by investigators.
6. Promotes responsible (legal-rational) decisions by respondents.
7. Reduces possibility of criminal or civil liabilities for the investigator.
8. Promotes an egalitarian relationship between investigator and participant.
9. Allows participants to have an interesting experience.
10. Promotes involvement of the public and increases public support and trust in the research enterprise.

The four criteria, applied to situations where individuals consciously decide to forego a right, are considered as critical for active informed consent. There are situations where individuals are assumed to

---

provide passive informed consent. In these situations, usually involving innocuous behavior, persons are assumed to consent unless they take other action. For example, in many large retail stores signs are posted at the entrance informing the patrons that by entering the store they are assumed to have consented to a search of their belongings at any time by security personnel. Entering the store, a behavior, is then considered as passive consent to a search at the discretion of the store personnel. The same rational is now used for many hospital patients, who are assumed to have consented to contribute to research designed to improve the art and science of medicine. Information derived from their treatment may be used, without their knowledge, for research. Written materials provided to patients, especially at major teaching hospitals, include a discussion of their obligation to contribute to ongoing medical research, just as they have benefitted from the contributions of those participating in previous research. Note that this medical research does not involve experiments; the participants are not directly affected by their contribution to the research.

A slightly different rationale is used for research uses of large scale government-sponsored data bases. It is assumed that individuals have elected representatives to make judgements on their behalf. If elected officials agree that it is in the best interests of their constituents that research be conducted with administrative data bases, they can be assumed to provide “informed consent” on behalf of their constituents, who are also the source of the data.

Both rationales – passive consent by participants and proxy active consent by elected officials – may be assumed in some situations. For example, students in a public school (or their parents) are assumed to allow their performance to be used as an indicator of the performance of the teaching and administrative staff, or of variations in techniques or procedures that may be employed (e.g., different schools may use different textbooks). In the United States, where all secondary schools are controlled by elected school boards, these elected officials may be considered to provide consent on behalf of the parents and students. If additional, sensitive information is to be obtained from the students or parents, it may be appropriate to require active informed consent from the participants.

While the establishment of informed consent has provided a workable solution for a wide range of situations where rational adults wish to forego rights, solutions for other situations have been more problematic. While workable procedures for surrogates to make decisions have developed for children, the mentally incompetent, and unconscious accident victims, no widely accepted solutions are in place for other important categories, particularly fetuses, permanently comatose patients, and future generations.

Future generations are of considerable interest, for they will be dealing with the consequences of present decisions. These decisions may lead to long-term costs to be borne by future generations or an absence of benefits – benefits foregone – if research is not completed. Third parties, often very vocal, claim to speak for those – comatose patients, fetuses, future generations – who cannot provide their own response. Their effectiveness is reduced by the knowledge that they will not directly bear the consequences of the decisions.

6. Individual Right to Privacy

Until a society appears in which every
individual obeys every rule and taboo and there is no ambiguity to create choices and tensions, there will be family heads, group leaders, religious authorities, and tribal-national authorities who will engage in surveillance to see that private conduct stays within a socially determined degree of conformity with the rules and taboos of that culture. (Westin 1967, p. 21)

Humans are both social and individualistic. One aspect of a social life – participation in a social community – is a loss of privacy and the normal pressures toward conformity. In small scale societies – such as a preliterate society or small rural community – where most interaction is personal and control resides in informal influence, surveillance of individuals and pressures toward conformity can be extreme. Indeed, an individual request for privacy is often suspect as a prelude to deviance – violation of accepted norms.

With the development of more complex and formal governing mechanisms, inevitably associated with explicit records and documentation, more attention has been given to the nature of individual privacy and external threats. Threats come not only from government, but from other non-government entities-business organizations, political groups, ordinary citizens. This has led to considerable attention to the nature of individual privacy and its critical aspects.

Individual privacy may be related to several features of an individual’s life in modern society: the inner self or the subconscious aspects of one’s personality, personal tendencies and dispositions, social attributes, and individual behavior.

The inner self, the core of one’s personality, may be considered as the most private:

... a large portion of his inner world ... the person discloses to a few confidential friends and shields against the intrusion of others. Certain (other) matters are kept secret at any price and in regard to anyone. Finally, there are other matters which the person is unable to consider. (Pius XII, 1958)

Ironically, modern social science techniques–use of standardized psychological instruments, talented clinicians – may be able to explore and describe the “matters which the person is unable to consider.” However, this is unlikely to happen without the knowledge of the individual.

“Personal tendencies and dispositions” refers to personality traits, attitudes, and tendencies that normally define the individual as a psychological entity. The extent to which a person has authoritarian or aggressive tendencies, places a high value on ecologically sound policies, or enjoys active sports or pornography would fall in this category. Some are considered more sensitive than others.

“Social attributes” would encompass a wide range of ascribed and achieved characteristics: age, gender, ethnic background, religious preferences, educational attainment, occupational trajectory, marital status, and the like. Many of these matters are considered a matter of public record in most advanced societies (residence, marital status, occupation), a few are–in themselves – among the most sensitive of items.

“Individual behavior” is distinct from the previous items in that it reflects what a person actually does, not socially defined attributes, personal dispositions, or core mental processes. Personal income, credit history, traffic violations, work behavior, performance in educational contexts, voting behavior, participation in religious and other voluntary organizations, military service, purchases of goods and service, drug
use, sexual behavior, and the like would be included as individual behaviors.

There are important variations in the treatment of personal activity in public places. While it is reasonable to assume that most public behavior is not to be considered private, there are exceptions. It is considered "inappropriate" to pay too much attention to some private behavior in public places: lovers in an embrace, a heated argument, or the purchase of an illegal drug. The United States Common Federal Policy on the Participation of Human Subjects in Research, described below, refers to private information as related to "behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place" (45 CFR 46.102). No mention is made as to whether "the context" is a public or private location. In some societies, there are legal protections. German law, for example, prohibits the photographing of strangers – in a public place – without their permission (Westin 1967, p. 29).

Of the major aspects of privacy – inner world, personal dispositions, social attributes, and individual behavior – only the latter two are of major relevance with regards to government data sets. There is virtually no systematic information collected on the personality and inner thoughts, and if judgements are developed about attitudes and personal dispositions, they are largely inferred from other evidence. Hence, most of the research involving government data sets is unrelated to these aspects of individual privacy. The remainder, however, includes substantial information about a person’s social situation or status as well as his or her actual behavior. The sensitivity of both may vary, depending upon the individual's preferences and possible consequences.

7. Privacy and Harm

The extent to which a person is "harmed" by invasions of privacy or disclosures of personal information can be considered in several ways. Harm may be associated with: (1) infringement of the privacy sphere; (2) disclosure of information that is embarrassing or awkward for the individual; (3) disclosure of information that leads to a decrement in the individual's situation – disruption of social relationships, loss of economic advantages, judicial proceedings; and (4) an indirect loss of self-esteem or status associated with a change in the social image of an aggregate or group the individual is associated with.

Much of the attention to the right of privacy is generated by concern that individual information, inaccurate or inappropriate, may be used to the detriment of the person(s) involved. This was the central focus of the initial treatment of this topic, the one that brought the potential for privacy related injury of modern technological advances to public attention (Westin 1967). Surveys of typical adults designed to measure public perceptions find evidence of concerns related to employment, credit, insurance, doctors and hospitals, the news media, the government, and computers. The level of public concern regarding privacy invasions associated with research has been too small to tabulate as a separate category (Louis Harris & Associates and Westin 1981). Much of the most vigorous discussion is related to invasion of a "privacy sphere." For example, an extensive list of "privacy interests" is presented in Figure 2.

The privacy interests in Figure 2 have no official status other than a reflection of the concerns of one informed observer. The

---

6 An earlier work by the same author focused on the appropriate balance between privacy interests and effective utilization of the information contained in data bases (Flaherty 1979). The most recent work, however, emphasizes only the protection of privacy, broadly defined (Flaherty 1989).
Fig. 2. Privacy interests of individuals

- The right to individual autonomy
- The right to be left alone
- The right to a private life
- The right to control information about oneself
- The right to limit accessibility
- The right of exclusive control of access to private realms
- The right to minimize intrusiveness
- The right to expect confidentiality
- The right to enjoy solitude
- The right to enjoy intimacy
- The right to enjoy anonymity
- The right to enjoy reserve
- The right to secrecy


A second level of potential harm may result from private data made public in such a way that an individual is not able to present a desired public image (Griffin 1990). For example, a middle aged man may be describing past athletic achievements in a social gathering, only to be reminded by his spouse of many years that he was actually on the reserve team and not a varsity starter. Embarrassing, yes, but such “harm” requires a number of elements, including the provision of alternative information to the same audience that the individual is trying to impress at the same time and relevant to the same topic.

It would be rare for all these elements to be present for much of the information that may be present in government data sets. The risk of such embarrassment seems small.

Major detriments in a person’s situation are three: change in the person’s personal social relationships, such as divorce; effect on the economic situation, such as the loss of a promotion; or civil or criminal proceedings, such as a challenge regarding the person’s income taxes. While data on certain aspects of a person’s life could lead to such changes, it would require the provision of specific data on specific participants to other individuals that could affect the participant’s life. Knowledge that a person had a sexual liaison may lead to a divorce, but only if the spouse has that information. Knowledge that a person was on parole as a juvenile for delinquent acts could affect an employer’s evaluation, but only if it was available and believable. With appropriate technical and supervisory procedures in data gathering and analysis, it is relatively easy to protect a participant from this type of official harm. When the claim is made that research with massive data sets has not harmed individuals, it is this type of effect that is being considered.
Finally, there is the potential effect on an individual from his or her relationship to a specific group. For example, if research on government data sets finds that those in certain income levels appear to have adequate income to pay a larger deductible with regards to medical insurance payments, then the use of that information may result in a slightly larger out-of-pocket personal cost for medical care. On the other hand, this group of individuals may have benefitted from a more efficient and equitable distribution of resources for medical care. The cost of being in a special income group, identified as worthy of special treatment in research, was offset by the general benefits to all. Without the research, the effect on this group may have been much larger.

A major issue is whether or not the ordinary individual will be satisfied with the knowledge that participants will not be affected by “disclosures of private information for research.” Public reaction to a Swedish project that assembled available government data on all persons born in the greater Stockholm area in 1953 is instructive. “Project Metropolitan” was made possible by the standardized use of personal identification numbers and imposed no additional burden on any informant; all the data was in existing government records. All procedures for access to data and the key identifiers were reviewed and monitored by an impartial Data Inspection Board. When the project was emphasized by mass media coverage, however, there was considerable public discussion and concern. The scientific value of a detailed longitudinal record on a representative sample was not seen as a justification for the collation of personal data leading people to be “treated like rats in a cage.” (Flaherty 1989, pp. 153–155). Similar distrust of governments has led to considerable resistance to official population censuses in the Netherlands and (Western) Germany. The extent to which there is a popular appreciation of the benefits of responsible social science research is an open question.

8. Right of Collectives and Organizations to Confidentiality

There is a wide range of social collectives, including families, organizations, communities, school systems, and ethnic groups (or categories) in developed societies. The basic social contract, however, is related to the allocation of decision-making responsibilities between the individual citizen and the state. The idea of rights was developed to facilitate allocation of decision-making responsibilities to these two parties to the social contract. The idea of individual rights is well established in all societies and cultures of the world. Rights associated with collectives are, however, another matter. As organizations and collectives do not have, in the same sense as natural persons, an inner self, predispositions or personalities, or behaviors, it is awkward to talk of an organizational right to privacy. Organizations may, however, be considered to have a right to preserve confidentiality, hence “confidentiality” will be utilized in this discussion.

One type of collective, the family, has emerged with special status in most societies. Only recently has it become clear that an absolute right of “family autonomy” can lead to abuses within family units that are considered morally unacceptable (child sexual abuse, spouse abuse, abuse of elderly relatives, etc.). An appropriate balance of the “family right to confidentiality” and surveillance to ensure that individual members are not abused has yet to be developed.

A similar problem has developed in pro-
viding rights of autonomy and confidentiality for other types of collectives. One major type of collective, the organization, is a source of most of the major contributions to modern societies—goods and services, justice, education, medical care, research and development, and so forth. Contemporary societies are societies of organizations. As with individuals, a major issue for modern societies is the appropriate balance of surveillance and confidentiality afforded organizations. Two perspectives have developed.

... privacy is a necessary element for the protection of organizational autonomy, gathering of information and advice, preparation of positions, internal decision-making, inter-organizational negotiations, and timing of disclosure. Privacy is thus not a luxury for organizational life; it is a vital lubricant of the organizational system in free societies. (Westin 1967, p. 51)

This first perspective reflects the economic conception of organizations as autonomous agents in competitive markets and emphasizes the need for organizational confidentiality to further "fair" competition (Westin 1967). Organizations, expected to survive in a competitive situation, have some rights to confidentiality regarding their activities. Commercial organizations developing new products or new marketing plans are entitled to some protection. Likewise political organizations, competing for votes or resources or influence, have an expectation of confidentiality. It has been suggested that one check on the legitimate political structure is political groups who should have the right to organize in confidence.

An alternative perspective emphasizes the extent to which the fundamental nature of advanced societies has changed. Modern societies are now seen as societies of organizations, or juristic persons, where most relationships between the individual and the state are mediated by organizations—work organizations, government agencies, political parties, educational and religious institutions, recreational and leisure groups, the legal system, the mass media, and so forth (Coleman 1982). Accepting that modern societies are dominated by these special types of collectives leads to further attention to their existence, structure, and operating procedures. It leads to the suggestion that many aspects of organizations should be available for public scrutiny to ensure that the power and influence of organizations are not used for inappropriate goals.

There is no question that organizations are capable of much damage, to individuals or the host society. Attention to safety in the workplace, where organizational savings are offset by risks borne by individual workers, led to the establishment in the United States of the Occupational Safety and Health Administration and a loss of organizational privacy with regards to facilities, equipment, and work procedures. Widespread concern related to decisions regarding hirings and promotions within business organizations led, in the United States, to the Equal Employment Opportunity Commission, guidelines, and surveillance of employment practices. A recent United States Supreme Court ruling has determined that all materials developed in university tenure decisions are subject to review by the Equal Employment Opportunity Commission if a valid claim of discrimination has been filed (U. of Penn. vs EEOC 1990). While this does not make tenure materials public documents, it does provide for the possibility of an independent external review of tenure decisions.

A major issue is determining the confidentiality domains appropriate for different
types of organizations. A preliminary list of major domains of organizational life, "candidates for confidentiality," is aligned with major organizational types in Figure 3. The organizational types are ordered, from left to right, with regards to the current perspectives in the United States regarding appropriate levels of confidentiality.

At one extreme, government agencies are usually required to provide full public disclosure of almost all aspects of their operations: hirings, budget allocations, purchasing decisions, and procedures to implement policy directives. Many government decisions are made in open meetings with a substantial public presence. The purpose of the Freedom of Information Act is to minimize the "right of confidentiality" for United States federal government operations.

At the opposite extreme are religious organizations. Based on the special status of religion in the U.S. Constitution, there is virtually no government surveillance of the internal workings of religious organizations. While problems are rare in established religions, there has been greater evidence of improprieties in charismatic mass media "religions," where financial irregularities have been so flagrant some have lost their tax exempt status. Rare, dramatic examples of internal problems include the ritualistic sexual or physical abuse of children in some cults and the mass suicide of over a thousand within the Jonestown colony (a United States religious cult that relocated to British Guiana in the late 1970s).

Business organizations clearly have an intermediate status. Some aspects are considered matters for public review, others are considered confidential within the organization. Of some importance is the effect of legal form on the disclosure of information. Publicly held corporations are required to provide substantially more information than privately held corporations, partnerships, or sole proprietorships.

The rows of Figure 3 indicate a tentative list of topics on which disclosure of information may be an issue. The sensitivity of these topics varies for different types of organizations. Most organizations developed for legal purposes are usually attracting clients or participants and would probably not consider their existence or purpose confidential. In contrast, their competitive strategy may be considered highly confidential. There may be diversity on a dimension within an organizational type. Some businesses are quite proud of their size and growth, others are concerned lest it attract competitors to their markets or industry. Internal structures (number of divisions, job specialization) are generally not considered sensitive, but internal processes – related to the allocation of resources, hiring, marketing strategies, and the like – may be seen as very sensitive issues.

Most business organizations seem to be concerned about confidential data with respect to three types of external agents: tax collector(s), regulatory agencies, and competitors. Depending upon the firm and its situation, different information about the organization may be considered more or less sensitive.

The readers are invited to develop their own judgements regarding the cells in Figure 3 and to make inferences about those domains of organizational life considered more or less sensitive for different types of organizations. This may vary for different countries.

9. Organizational Confidentiality and United States Federal Data Sets⁷

There are two basic laws related to the con-

⁷ This is only a brief summary. A more complete and detailed review of the situation is provided in Jabine (1991).
### Types of Organizations

<table>
<thead>
<tr>
<th>Government</th>
<th>Quasi-public (not-for-profit charitable)</th>
<th>Business</th>
<th>Voluntary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidentiality domains</td>
<td></td>
<td></td>
<td>Political Relig-</td>
</tr>
<tr>
<td>Existence</td>
<td></td>
<td></td>
<td>parties</td>
</tr>
<tr>
<td>Purpose/Activities</td>
<td></td>
<td></td>
<td>ous</td>
</tr>
<tr>
<td>competitive strategy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ownership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scope/size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Budget</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sources/amount</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expenditures/amount</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Structure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Division of labor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authority allocations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Processes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision making</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hiring, promotions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marketing strategy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>External Relationships</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment relationships</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creditors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suppliers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Host community</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. *Organizations and privacy issues*

The confidentiality of United States federal government data on "persons." The Privacy Act of 1974 (5 USC 522a) is designed to prevent disclosure of data on persons and to provide a mechanism for persons to verify the accuracy of government records. The Freedom of Information Act (5 USC 522) is designed to facilitate access of individuals to government data, records, and procedures. For both acts (5 USC 551 (2)):

- "persons" includes an individual, partnership, corporation, association, or public or private organization other than an agency [of the federal government].

However, the Privacy Act of 1974 is related to "individuals," defined as "a citizen of the United States or an alien lawfully admitted for permanent residence" (5 USC 552a (a) (2)). The focus of the Privacy Act is on agency records pertaining to "individuals" not "persons;" hence organizations are not covered.

On the other hand, when "person" is used, as in the Freedom of Information Act, it has been interpreted quite widely. Almost any organization would be included. "Person" has been extended to cover a specialized coordinated activity of a group of large organizations, the Boston
Survey Group; this involved the sharing of data on employee salaries and benefits by a group of major employers (527 F. Supp. 1163). However, a household or family would not be covered; participants would need to be covered as separate "individuals."

A number of state and federal agencies in the United States collect substantial data on organizations. These are primarily for tax purposes or to provide systematic descriptive data on important segments of the economy. There is some variation in the anonymity afforded the organizational respondents.

All states require individual work organizations to file unemployment insurance returns and payments. These tax payments are used to provide benefits for those involuntarily unemployed. The quarterly unemployment returns are consolidated into descriptions of the total employment and salary and wages paid by county and industry and size of establishment.

Known as the ES202 files, this data is provided to the Bureau of Labor Statistics and consolidated into a national report on employment and earnings by county, establishment size, and industry (U.S. Small Business Administration 1988).

The extent to which these tax returns are to be treated as confidential varies widely among the states. For example, they are absolutely inviolate in Minnesota. Pennsylvania, in contrast, makes listings of the firms freely available.

The major source of data on tax entities, often treated as a surrogate for business entities, is the business returns assembled by the federal Internal Revenue Service. From the perspective of individual scholars, this information is absolutely inviolate; only employees of the Internal Revenue Service are allowed access to the identity of those filling tax returns. Some arrangements have been made, however, that will allow other agencies to share IRS data for certain projects (Jabine 1993). Scholars without an affiliation with a government agency wishing to study individual tax returns must become employees of the IRS, at least temporarily.

The U.S. Census makes a considerable effort to gather and publish data on business organizations. Annual data on employment and payroll on establishments, by size, industry, and county, is based on Federal Insurance Contributions Act (Social Security) records and distributed as the County Business Patterns. Censuses of business in most industries (except finance, transportation, and agricultural production) are completed every five years and published as Enterprise Statistics (U.S. Small Business Administration 1988). Substantial public funds are spent on creating sampling frames for these censuses of business entities. However, the individual identities of these organizations are not available to anyone who is not an employee of the Census Bureau. Consequently, other agencies or programs are required to develop — independently and duplicating the Census efforts — survey frames or seek other sources of data. A number of procedures have emerged to facilitate research on this data that involves the organization as a unit of analysis. Customized data runs and analyses will be made at the request of individual scholars, but this involves out-of-pocket costs and precludes full familiarity with the data sets. An active program of providing temporary employment to other investigators, allowing them access to the files on individual business organizations, is in place. As with permanent employees, they are subject to fine and imprisonment if they disclose confidential data; forms acknowledging this constraint are signed on a regular basis.
A third source of systematic data on United States business establishments\(^8\) (unique locations) and firms (complete legal entities) are the files maintained by Dun and Bradstreet, a commercial credit rating service. This data is collected with the full understanding it will be made available to others. Information on specific identifiable establishments and firms is sold to anyone. Samples are frequently purchased for surveys and scholarly research. The complete data set for the U.S., a file in excess of five million establishments, is periodically leased by the U.S. Small Business Administration and provides – after extensive editing, corrections, and adjustments – a portrayal of establishment and firm changes for the entire United States. These data can be used to track specific firms, specific industry sectors, specific counties, or any combination thereof. Anonymity of the firm is maintained to conform to the contract developed with Dun and Bradstreet, which precludes uses that would compete with other potential sales of these listings, such as credit evaluations or developing mailing lists. No federal statute prevents disclosure of individual firm identities from the Small Business Data Base (U.S. Small Business Administration 1988).

A recent development has occurred in United States federal law related to the collection of data on a special type of organization. The Hawkins–Stafford Act relates to the collection, analysis, and dissemination of educational statistics. In referring to the duties of the United States Commissioner of Education Statistics, it includes the following statement (20 U.S.C.S. 1221e-1(i) (4) (B) (i)):

\[ \text{The Commissioner shall ensure that all personally identifiable information about students, their educational performance, and their families and that information with respect to individual schools will remain confidential, in accordance with [the Privacy Act of 1974].} \]

Clearly intended to ensure anonymity for key elements, it takes the Privacy Act of 1974 as the point of reference. But this act applies to “individuals” and not “persons,” and hence, does not cover organizations, such as “individual schools.” Hence, although the intent of this statement is clear, the actual legal status is somewhat ambiguous (Newton and Pullin 1990). Ambiguity is increased by a proposed amendment to provide punishments for persons who determine the identity of individuals, but not “individual schools,” from these data sets (United States House of Representatives 1990).

It is ironic that the United States, with the largest government sponsored data collection system in the world, has not developed even the most fundamental information on organizations. Despite the rising prominence of organizations as critical actors in modern societies, there is no standardized list or directory of organizations – of any type. There are accurate counts of persons, households, dwellings, jobs, farms, and tax entities, but the number of business organizations is unknown (U.S. Small Business Administration 1990, pp. 9–19). Recognizing the critical role of business organizations in advanced societies, most other industrialized nations not only have standardized lists of business organizations, but make them publicly available (American Statistical Association 1980; Bureau of Labor Statistics 1987). The most highly developed Standardized Statistical Establishment List (SSEL) is developed by the

---

\(^8\) An establishment is a single physical location where economic activity takes place, such as a plant, site, installation, or retail outlet. A single business firm, as a legal entity, may consist of one or more establishments. The most accurate lists in the U.S. are of establishments.
United States Census Bureau, but is not made available outside the bureau. As a result, the only current, public list of establishments from which firms (single or multi-establishment) can be identified is provided by a commercial credit rating service, Dun and Bradstreet. As of 1991, Dun and Bradstreet will provide relatively complete information on the presence, industry, employment, and sales of any of eight million U.S. business establishments. The first record costs $95.75, second and subsequent records cost $0.75.9

It is becoming clear that the changes and turbulence occurring within the population of business organizations are critical to understanding economic change and growth; this is why the role of new and small firms appears critical (Reynolds and Maki 1990). There is rough agreement of establishment counts from the various sources, suggesting reliability if not accuracy. But, the absence of reliable censuses of firms (as combinations of establishments), severely constrains the potential for understanding the dynamic processes that are crucial features of the economy. The intense competition developing among advanced economies suggests this should be an issue of strong national priority for the United States.


If the efforts to resolve the privacy/research contributions dilemmas associated with large scale federal data sets was the first attempt to consider the role of research participants there would be much to do. How-}

9 Provided by Dun’s Marketing Services as Dun’s Direct Access. A standard MS DOS personal computer and a clean telephone line are the minimal requirements. The $95 is an annual fee for access to the system.

10 These risks are so low that a survey of research effects completed in the middle 1970s involved asking investigators about risks of negative effects, rather than actual incidents of harm. For example for 1% of 136 medical projects investigators reported a very low risk of death; for 99% they reported no risk of death. For about 15% of projects that involved confidential data, investigators estimated some risk to the participants, for 85% they estimated no risk (Institute of Social Research 1976, Table IV. 8, p. 121).
remained the same, important aspects have undergone substantial change over the past several decades.

The major elements of the United States federal system for supervision of research with human subjects are (45 CFR 46): (1) Explication of standards to be used in the conduct of research including definitions of minimal risk and the criteria for informed consent. (2) Each research organization (most are universities or hospitals) is to establish an appropriate Institutional Review Board (IRB) with the authority to review and disapprove all projects conducted within the institution; regardless of the source or amount of financial sponsorship. [In practice, however, the major function of the IRB has been to help investigators redesign research projects to meet the federal guidelines.] (3) A definition of “exempt” research activities, for which prior review is not required, primarily because of the low risk to the participants. Only their occurrence is monitored. (4) A definition of another set of research activities, considered to provide moderate risk, eligible for an “expedited review.” For these projects, a single IRB member may review the research and provide approval. Each research institution is allowed to set up its own procedure, subject to federal guidelines and approval of their mechanism by the federal Office for Protection from Research Risks (OPRR).

Both the exempt and expedited categories were developed after it was discovered that almost all social science research and a substantial amount of medical research had little risk of negative consequences for the participants. The most recent development has been the standardization of the policy for virtually all relevant United States federal agencies; a total of sixteen—seventeen if the Central Intelligence Agency is included, and it must be to comply with an Executive Order of the President. There are some slight variations on details, not important to this review, that are related to the special situations of several agencies.

The United States Common Federal Policy provides a well accepted, explicit standard for social science research. For example, the unit of analysis is defined as (45 CFR 46.102):

f. “Human subject” means a living individual about whom an investigator conducting research obtains:

1. Data through intervention or interaction with the individual,
or
2. Identifiable private information.

“Private information” includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record). Private information must be individually identifiable (i.e., the identity

---

11 Coordinated through the Office of Science and Technology Policy, the agencies include the Department of Agriculture, Department of Energy, National Aeronautics and Space Administration; Department of Commerce; Consumer Product Safety Commission; International Development Cooperation Agency (Agency for International Development); Department of Housing and Urban Development; Department of Justice; Department of Defense; Department of Education; Department of Veterans Affairs; Environmental Protection Agency; Department of Health and Human Services (Office of the Secretary; Food and Drug Administration); National Science Foundation; and Department of Transportation. As of January 1991, about half of these agencies had agreed to conform to the Model Federal Policy, approval by the other half were expected in 1991 (Joan Porter, Personal Communication, 22 January 1991).
of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects.

Hence, administrative data coded to mask the identity of the individual before it is provided to the investigator(s) would not be considered research with human subjects and not be subject to the Common Federal Policy. The extent to which the research would be covered if participants could be identified through comparisons of predominate characteristics with other public data sets would depend upon an interpretation of "readily be ascertained."

Most social science research with existing data on human subjects is exempted from any IRB review. The relevant clause is as follows (46 CFR 46.101):

(b) Unless otherwise required by department or agency heads, research activities in which the only involvement of human subjects will be in one or more of the following categories are exempt from this policy:

(4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

No distinction is made as to whether or not these data are in public or private data sets or whether or not the subject had an "expectation" of privacy when the information was provided. Reflecting the concerns with medical and social science experimental research, the Common Federal Policy gives primary emphasis to actual negative consequences for the individual. Access to private information on anonymous and unaware participants is not considered "damage."

Note that the entire Common Federal Policy is oriented toward research with "human" subjects. Research where the unit of analysis is a collective – business organization, public high school, special ethnic category – would not be covered except with respect to the collection of data from or on specific individuals. Collectives, as entities due consideration and protection, have no standing in the Common Federal Policy on the Participation of Human Subjects in Research.

11. Informed Consent in Non-Experimental Research

Data collected in scientific surveys or for legitimate administrative purposes are generally low cost for the participants. The major cost they bear is the time involved in providing the information. The provision of sensitive, private, or embarrassing information can be considered a burden, but it will vary depending upon the nature of the data collection activity and the individual respondent. Ideally, the collection of data would be conducted in such a way that the three major moral perspectives are satisfied: (1) the rights of the participants are respected, (2) some benefits from the research are shared by the participants (or the group they represent), and (3) the procedure reflects respect for the participants as human beings.

Benefits resulting from the research will depend, initially, on the investigators. They determine the research design and its implementation; presumably they are competent and some useful knowledge will result. The use of this knowledge and how it affects the lives of the participants (or the group they represent) may not be under the direct control of the investigators. At a minimum, it would be best if the participants’ lives were not made less pleasant as
a result of their involvement. If the possibility of direct benefits to the participants is remote, it may be appropriate to provide them with a special direct benefit for the contributions; a suitable cash payment is often well received. (If there is too much cash, however, there may be criticism that the informed consent was not truly voluntary.) A summary of the project results is often provided to participants as a demonstration of respect for their contributions.

Respect for the participants as human beings will be demonstrated by the care and consideration shown in incorporating them into the project and obtaining their informed consent. A procedure that is not patronizing and provides a direct, honest response to questions and concerns is usually the most effective way of developing cooperation. Skilled investigators usually treat participants as research partners, rather than merely sources of information. In voluntary survey research, individuals will make the decision to participate almost immediately—seconds into a phone contact, shortly after meeting an interviewer, before finishing the introduction to a mailed questionnaire. The more the approach is individualized, appears serious and carefully prepared, and provides the impression that the data will be used for an important purpose, the greater the chance that the instantaneous decision will be to cooperate.\(^{12}\)

For this discussion it is assumed that the participant is a “rational adult,” fully capable of making an informed judgement. The major issues critical for data collection activities are the nature of the data to be obtained, how it will be used in the analysis, and its eventual disposition. Participants are generally promised that they will remain anonymous, that no person outside the research staff will have access to their identity, and that reports of the research will not allow identification of specific participants. In most cases participants can be assured that there will be no change in their status as a result of participation: no “risk of criminal or civil liability or be damaging to the subject’s financial standing, employability, or reputation” (45 CFR 46.101 (a) (2) (ii)).

While data are usually collected for a specific purpose, unforeseen or future uses of the data are an occasional concern. In an extreme case, data may be destroyed after the project is completed, precluding further, additional analysis. This may be appropriate with extremely sensitive data, such as that related to illegal or extremely embarrassing behavior. In most cases, however, the removal of identifiers and security storage of the master participant list ensures continued anonymity.

High quality data on representative samples can be considered a societal resource. Its value may be substantially enhanced when it is combined with other data or is part of a longitudinal data set. It is not, however, possible to specify or predict future scientific analysis of the data. Hence, it is not possible to obtain informed consent for specific future uses. It is possible to obtain permission to treat the data as a permanent research resource, available for consolidation with other data sets and additional analyses. It is critical, however, to be able to assure participants that future uses or data set consolidations will not place them at risk in the future.

Informed consent with social collectives, such as organizations, is complicated by

---

\(^{12}\) It has been found that most kidney donations are instantaneous decisions; 77% of 30 donors in one study (Fellner and Marshall 1970) and 61% of 114 in another (Simmons, Klein, and Simmons 1977). Most were donating to a relative and made the decision when they first heard of the recipient’s problem, usually in a phone call, long before being officially approached and completing the informed consent procedure.
several features. Unlike individuals, who can normally speak for themselves, it is not always clear who may provide informed consent for an organization. For commercial organizations or government agencies, it is usually the chief administrative officer. For some issues, however, a supervisory authority, board of directors, elected officials or the like may be needed to provide informed consent for the social collective. A second problem develops when specific data are required, for they are usually provided by a specialist in that topic.\textsuperscript{13} If the individual is answering for the organization, then the informed consent may be provided by the chief administrative officer, but if additional information on that organization is obtained from another organizational participant, a second informed consent may be appropriate. In this case, a promise to the participant not to disclose any information they may provide to others in the organization may be critical.

There are situations where individuals may be approached about providing data, and the fact that they were eligible participants would have a substantial potential for a negative effect. In one case administration of a standardized personality inventory was to be given to the children of adults with a known pattern of psychosomatic symptoms, most critical were the patterns associated with schizophrenia. The research objective was to estimate parent–child correlations. In order to mask the identity of the chosen children and their parents, an entire classroom of grade school children was given the personality inventory, even though the results from only one child were of value to the project (Garmezy 1980). In a follow-up study of those who had spent time in Alcatraz (a former United States high-security federal prison) to determine the long term effects of incarceration, contacts were made in such a way that none of the ex-convicts' associates or family members knew of the purpose of the interview. Most associates, even spouses, did not know of these respondents' backgrounds. However, once the private contact was made, cooperation was very good, as the respondents were eager for gossip about former friends from Alcatraz and quite willing to complete the interview (Ward 1988).

A list of the major elements of informed consent is provided in Figure 4. They are separated for individual respondents, organizations, and organizational members. All elements are appropriate for all types of natural persons. Figure 4 is organized around the elements of informed consent provided in the United States Common Federal Policy (45 CFR 46.116-46.117). The parentheses refer to the specific elements in the general requirements. The Common Federal Policy is quite explicit about what may not appear in an informed consent, oral or written. No informed consent "may include any exculpatory language through which the subject ... is made to waive or appear to waive any of the subject’s legal rights, or releases or appears to release the investigator, the sponsor, the institution or its agents from liability for negligence."

Additional informed consent elements are suited for organizational members: (1) freedom from coercion from the organization and (2) a promise not to divulge the

\textsuperscript{13} A mailed questionnaire was sent to 5-8 senior administrators in a wide variety of large organizations in a single urban area. The purpose of the project was to compare the responses of different administrators to a standardized set of questions. In several public agencies, all questionnaires were routed to the same budget office for items related to finances. In one public library, all questionnaires were sent to the person in charge of answering all questionnaires sent to the organization (Reynolds, Knoke, Miller, and Kaufman 1990).
### Elements of informed consent for non-experimental research

<table>
<thead>
<tr>
<th></th>
<th>Individual</th>
<th>Organizational</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Members</td>
<td>Organization</td>
</tr>
<tr>
<td>Participant is a rational adult</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(a, 1) Explain that study is for research</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Explain what data to be collected</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Explain time, resources required of participant(s)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Explain purpose of research</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(a, 2) Describe foreseeable risks</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(a, 3) Describe benefits to participant</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Describe benefits to others</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(a, 4) Alternative treatments</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>(a, 5)Extent of confidentiality, how maintained</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(a, 6) Possible compensation, if any</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(a, 7) Contacts for questions about research, participant rights</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(a, 8) Statement that participation is voluntary</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(b, 1) Statement regarding unforeseen risks of experimental procedure</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>(b, 2) When investigator may terminate participation</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(b, 3) Additional costs of future participation</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(b, 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Individual</td>
<td>Organizational</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Members</td>
</tr>
<tr>
<td>Consequences of participants' withdrawal</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(b, 5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on significance of research results</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>(b, 6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximate number of participants in study</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Other topics:</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Promise no change in participants' situation expected</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Advise of data retained as permanent resource</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Advise that data may be supplemented with other information</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free from organizational coercion</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Promise confidentiality with respect to host organization</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Symbols in parenthesis indicate relevant section in general requirements for informed consent, in the U.S. Common Federal Policy (45 CFR 46.116–46.117)
responses to other organizational members (e.g., coworkers, the boss, subordinates, etc.). Many questions that would be innocuous for a person to answer in his or her home may become quite sensitive in a workplace context. In the first income maintenance experiment in New Jersey, reporters interviewed one of the respondents, who had a low-paying unskilled job in a small manufacturing firm. When his superiors discovered, from the mass media, that he thought he was underpaid, he was harassed and eventually lost his job (Kershaw and Fair 1976, p. 181). Subsequent to this experience, project investigators refused to release participants’ names to the press.

Because of the wide diversity among research activities, only guidelines can be provided for informed consent procedures. There is no single, universally accepted criterion that can be used for all research to evaluate ethical correctness. Each person has a distinctive moral reaction to any situation with an ethical component. An exchange between individuals – participant and investigator – may result in different ethical reactions for different observers. If informed observers agree that the exchange was morally correct, one can have confidence that an acceptable moral choice was developed.

There is, unfortunately, no way to determine whether a person is a true informed volunteer. Inferences are made about informed consent based on the context in which the behavior occurs. The informed consent guidelines represent, after considerable analysis and exchange, a consensus as to what are the critical features of the context for the decisions. The guidelines represent, therefore, a substitute for a judgement as to whether or not informed consent was ethically obtained. If the guidelines are followed, most observers should conclude that informed consent was ethically obtained – appropriate for the benefits expected from the research and costs (effort, risk) for the participants.

12. Legal Threats to Privacy and Certificates of Confidentiality

A critical feature of the informed consent for research participants is the promise that the information will remain confidential. If the United States data are gathered under the auspices of the Bureau of the Census or in conjunction with a tax filing or other agency-sponsored activities, such promises reflect legal restrictions that can prevent efforts to obtain access to the data through subpoena.

If data are collected through other, especially non-federal, institutions, there is some ambiguity regarding the legal status of the data. There have been some efforts to subpoena federal and non-federal research data. For example, substantial effort was initiated by county attorneys to obtain specific records on specific families participating in the first income maintenance experiment. The families were suspected of “welfare fraud” for accepting both county welfare payments and the income maintenance payments that totaled more than the approved guidelines. The most significant threat to the families was settled with a cash transfer from the project to the county for any “overpayments.” The project administrators also feared public disclosure due to potential subpoena requests from a U.S. Senate committee, which they felt they could not resist (Kershaw and Fair 1976, pp. 177–190). This incident led to a number of analyses of this problem, and a substantial case law tradition has developed in this area (Nejelski and Lerman 1971; Boruch and Cecil 1979, Ch. 8).

For much legitimate research, the United
States courts have ruled that data are immune from subpoena for civil cases, similar to the privilege provided to mass media reporters under First Amendments rights. Disclosure for criminal cases, however, is another matter. While the likelihood that archival type data would have value in a criminal case is extremely small, an absolute guarantee of confidentiality may be unwarranted for data not collected under any legal immunity. While the risk is small, the number of participants in research and administrative data bases, which may number in the tens of millions, suggests this is a problem not to be dismissed for research on sensitive topics.

A United States federal procedure has been established whereby any authorized individual engaged in a pre-approved research project may withhold the names and other identifying characteristics of subjects from any person or authority—regardless of whether or not the research is supported by federal funds. If a Certificate of Confidentiality has been issued, persons authorized to have access to information which would reveal the identity of the subjects "may not, at any time, be compelled in any Federal, State, or local civil, criminal, administrative, legislative, or other proceedings to identify the research subjects ..." (42 CFR 2A.7 (a)). The Certificate of Confidentiality does not prevent disclosure of a specific subject's identity if they agree to do so in writing nor does it prevent voluntary disclosure by the research staff. Confidentiality Certificates are provided, upon application, by the Assistant Secretary for Health and Human Services.

Initially these Certificates were only provided for research on alcohol and drug abuse or mental health, but protection was expanded to include any "biomedical, behavioral, clinical, or other research" by Public Law No. 100-607, par 163 (November 4, 1988). The interim policy statement has interpreted this to include research of a "sensitive nature where the protection is judged necessary to achieve the research objectives." Research is considered to be sensitive if it involves any of the following (DHHS 1989):

a. Information relating to sexual attitudes, preferences, or practices;

b. Information relating to the use of alcohol, drugs, or other addictive products;

c. Information pertaining to illegal conduct;

d. Information that if released could reasonably be damaging to an individual's financial standing, employability, or reputation within the community;

e. Information that would normally be recorded in a patient's medical record, and the disclosure of which could reasonably lead to social stigmatization or discrimination;

f. Information pertaining to an individual's psychological well-being or mental health.

Information in other categories, not listed here, might also be considered sensitive because of specific cultural or other factors, and protection can be granted in such cases upon appropriate justification and explanation.

In short, any endeavour classified as research (a systematic investigation designed to develop or contribute to generalized knowledge) that may involve information that may be damaging to the individual participants is eligible for a Certificate of Confidentiality. Since

there is no restriction on the research technique, this would include any research using data bases – federal, state, local, private, or otherwise. The major restriction relevant to research involving data bases is the emphasis on “individuals” or natural persons, rather than legal “persons,” which would include organizations. Certificates of Confidentiality are not available for research where the units of analysis are organizations, school districts, communities, social aggregates, and the like.

Distinctive features of the procedure include the universal application to all research, regardless of the source of data, and decisions coordinated through one administrative office, the Assistant Secretary of Health and Human Services. While a Certificate of Confidentiality is only provided for a single project, any project related to social science may be covered – including those involving data bases. Such a procedure reflects at attempt to minimize the costs of diverse, numerous, and idiosyncratic procedures for resolving an important issue: insulating social science research from threats to privacy associated with other government activities.

13. The Social Science Enterprise and the Potential for Future Research

A popular Government without popular information or the means of acquiring it, is but a Prologue to a Farce or a Tragedy or perhaps both. Knowledge will forever govern ignorance, and a people who mean to be their own Governors, must arm themselves with the power knowledge gives. (James Madison 1822)

People have an expectation of, perhaps a right to, an efficient, effective and just government. The following elements are all related to the capacity to satisfy this basic right:

1. Social problems are now complex and multifaceted, requiring an enhanced level of sophistication for acceptable solutions.

2. The social and policy science enterprise has much to offer in developing information that can help to solve these problems.

3. The development of a more complete and useful understanding of critical social, human, and economic phenomena will require access to larger, multi-variable, and longitudinal data sets on natural persons and organizations.

4. There is an increasing level of public concern regarding abuses of privacy leading to potential negative effects. Most of this reflects apprehension regarding administrative abuses facilitated by diverse and detailed data sets.

5. No permanent solutions to the privacy-data set access dilemma is likely to be developed; resolution will consist of a sequence of temporary compromises.

The major problem for attention is: How can the potential for significant social and policy science contributions utilizing government data bases be increased?

Several strategies might be used to achieve this end.

Efforts to enhance the availability of government data for research have focused on, first, reducing the risk of unauthorized infringement on privacy and, second, enhancing access to and utility of federal data bases. Both of these efforts have been piecemeal, completed in such a way that it has been difficult for a single person, committee, or agency to balance three elements: (1) the privacy interests of individuals or confidentiality interests of organizations; (2) the operational concerns of specific agencies, administrative and statistical; and (3) the public interest in developing additional knowledge on important topics. This latter, may, of course, be done by
statisticians and analysts within government as well as those in research settings (universities, institutions, medical facilities) outside official government structures.

Attempts to reduce the risk of unauthorized infringement upon privacy have taken several forms. Most ingenious and elaborate have been the various technical solutions related to obtaining sensitive data (such as the randomized response technique), maintaining respondent anonymity in the data sets (using a host of procedures), and insulating investigators from the data sets themselves (Cigrag and Rainwater 1990). These efforts have been quite successful in minimizing the potential for disclosure of individual or organizational identities present in large scale data sets, although they often increase the costs of research or preclude some analyses. Other efforts have focused on the researchers, as reflected in the development of codes of ethics by professional associations (American Statistical Association 1983; Ellenberg 1983; International Statistical Institute 1986; Reynolds 1975). This has done much to generate awareness of the importance of respondent confidentiality and privacy, although professional associations have limited capacity to enforce such codes. A third focus has been on obtaining authorization from respondents for all possible uses of data, through the use of more complete and detailed informed consent procedures (Jabine 1986). Concerns have developed when new uses of the data go beyond the narrow conditions associated with the initial informed consent provided by the respondents (Pearson, Duncan, and Jabine 1989).

The result of these efforts may have contributed to one of the most important patterns associated with research uses of large scale data sets (singly, in combination, or longitudinal): the absence of any public examples of harm resulting from disclosure of privacy. No participant or organization has been disadvantaged from analysis of such data. While there may be some undiscovered examples of abuse, there is no evidence of any systematic patterns of respondent injury. It is probably true that research analysis using large-scale data sets has gone beyond topics initially anticipated by the participants. But since this has not led to disclosures of identities or participant harm, it can hardly be seen as a major threat to the respondents.

Despite all these efforts and a history of "safe stats," there is still public concern that personal privacy may be unnecessarily invaded for social science and policy research (Riche 1990). This may reflect a lack of appreciation of the benefits to policy development of reliable and complete information, which requires the use of extensive data sets. More likely, it reflects a major concern regarding misuse of personal or organizational information by those in positions of authority (Flaherty 1989) and a failure to distinguish research analysis from administrative applications. Nothing is more critical than to distinguish the use of data for legitimate social and policy research from the use of data for routine administrative, monitoring, and compliance objectives. The following focuses on uses for research, where information on specific individuals and organizations is of little value or interest to the investigators.

The value of using large federal data bases for important research questions has been recognized for some time. A large number of such uses have been identified and documented, for example, analysis of individual files, using files from several sources to correct or verify data, and consolidating files from two or more agencies to explore unique or distinctive patterns (Jabine 1993). These efforts, successful and
unsuccessful, have several features in common. The cost, in professional time, energy, and creativity in developing explicit agreements with the agencies responsible for the data has been substantial. The results, however, were unique. There is little "system learning." Agreements developed to facilitate each project were developed on an ad hoc basis. Past agreements have not served, in any systematic way, to facilitate subsequent agreements regarding access to data sets.

Furthermore, administrators tend to respond, as they should, to the distinctive needs and requirements of their home agencies. Only those responsible for the specific projects speak to the public interest. Unfortunately, their career involvement in the project diminishes their effectiveness as societal spokespersons. In some cases, very worthy projects were terminated when existing laws changed. One example involved the cessation of the use of Internal Revenue Service address lists to locate participants in medical follow-up and epidemiological studies (Jabine 1993). This was a response to a change in confidentiality provisions associated with the Tax Reform Act of 1976 and it ended the practice of assisting the National Institutes of Health in finding, for subsequent interview, those exposed to suspected health hazards. The Internal Revenue Service officials implemented the new provisions and cooperation ceased. Taxpayers gained an increase in privacy; citizens—the same individuals—lost a chance to know more about maintaining good health.

Agency administrators responsible for major data sets have two major concerns, compliance with legislation related to the conduct of their work and the cooperation received from those that provide the data. They are required by law to adhere to relevant legislation; their work performance is evaluated—in part—by the extent to which they can minimize the cost of gathering data from individuals and organizations. Any publicity—accurate or not—that threatens to reduce voluntary compliance will complicate their work. Such projects are likely to be avoided, regardless of the potential societal benefits.

But given that substantial public benefits may be at issue, it is not clear it is in the public interest to have projects thwarted by the insular concerns of agencies focusing upon their current responsibilities. Proposals for additional analysis should be reviewed and evaluated by those with a broader concern for the public interest. Indeed, an agency responsible for data that can provide unanticipated benefits may, in the long run, find its role and budgets enhanced by new application.

The substantial risk of future public benefits forgone suggests that changes should be made in the current system. Several strategies are possible. One would be to work with specific federal agencies to make changes in their legislation, administrative guidelines, and relevant interpretations. Among the dozens of relevant agencies there will probably be some success, some partial success, and some failures. These efforts can be decentralized, with different groups working with different agencies. It is unlikely to result in increased standardization; each agency is likely to retain a distinctive and unique set of criteria and guidelines, reflecting the unique set of laws that authorize its work. Furthermore, it may not do much to facilitate a standardization of procedures for consolidating data across agencies. It may continue to be necessary to develop unique cross-agency arrangements for each project that involves two or more data sets.

A proposal to establish a separate federal agency to develop standardization across federal statistical activities was rejected by
President Carter (Bonnen et al. 1981). It may be time, once again, to propose a federal office of Statistical Policy, one that can focus on the public interest in improving the value and reducing the costs of using the existing massive data sets, developed with public funds.

13.1. **Federal data base research certificates**

Another strategy is to work toward uniform federal standards or procedures that are applicable only to research, taking precedence over individual agency guidelines. While complicated to implement — and potentially taking considerably longer — such a standard would resolve a number of issues with a common procedure, applied in a uniform fashion by a centralized entity. Of substantial importance is to develop a mechanism that allows an independent judgement of the social value of specific projects.

Worthy projects could be provided, upon application, with a “Federal Data Base Research Certificate” (Data Base Certificate). Such a certificate would be accompanied with the privilege of unlimited access to any data maintained by any federal agency (with their approval), immunity from any legal subpoena, and substantial penalties for any investigator or research staff member that divulged the identity of any individual, person, or organization represented in the data set.

There is some diversity among federal agencies regarding the legal and administrative status of their data, administrative and statistical. For this reason, it is proposed that each agency have veto power over contributing its data to such joint endeavors. On the other hand, the conditions associated with the Data Base Certificate are intended to be stringent enough to satisfy the agency officials that “their” respondents (individuals or organizations) will be adequately protected. While there is a remote possibility that some of “their” respondents will be publicly identified, it is more likely that the value of their data will be more widely recognized and, in turn, the agency itself will benefit from participating in a cooperative endeavor.

There would be several elements associated with the Data Base Certificates. A “Federal Data Research Review Board” could evaluate applications and award the certificates, assigning the responsibilities and privileges associated with their use. The Federal Data Research Review Board would be modeled after the Institutional Review Boards associated with approval of research involving human subjects under the Common Federal Policy. It should be designed to represent the major groups that can provide a judgement on the balance between privacy interests and societal benefits. As such, it should represent all major constituencies, including elected officials, representatives of the major agencies, established experts in data analysis and techniques for ensuring data confidentiality, those familiar with major policy issues, private citizens, and representatives of the research participants—individuals and organizations. Consultation with relevant experts should be obtained when appropriate.

The Federal Data Research Review Board should be organized to meet and produce decisions on a periodic basis, four times a year might be sufficient. Projects that were not approved could have the opportunity to revise and resubmit. No funding would be provided by the Federal Data Research Review Board, but funding might be contingent upon obtaining a Data Base Certificate. Adequate funds to ensure that data remained private or anony-
mous could be an issue associated with the Board’s approval of projects. Agency heads could have the right to approve the use of their data, but as the confidentiality standards associated with the Data Base Certificates would meet or exceed that of any existing agency, they would not be expected to consider protection of individual privacy or organizational confidentiality in making decisions regarding approval of the research.

Approval would be provided for specific projects, which could be submitted by investigators, organizations, or agencies. Projects could be for any scientific or policy research that involved data sets, whether or not the participants expected anonymity. Projects that involved single data sets, combinations of two or more existing data sets, or combinations of existing data with new data would be eligible for a Data Base Certificate. The application could include the purpose of the project, significance for the public good, strategy for analysis, and procedures for ensuring respect for the rights and welfare of the units of analysis (individuals or organizations). Applications, as with those submitted for Certificates of Confidentiality, should also specify the individuals (investigators and research staff) that would have access to the identities of individuals or organizations utilized in the analysis.

Some standards used by the Federal Data Research Review Board might be taken from the current Common Federal Policy on the Participation of Human Subjects in Research. However, as all the data will have already been collected, many of the criteria related to selection of participants, informed consent, and the like would not apply. Critical would be the following (Federal Register 10Nov88, 53 (218), p. 45675):

.111 (a) (2) Risks to subjects are reasonable in relation to anticipated benefits, . . . , and the importance of the knowledge that may reasonably be expected to result. . . . The IRB should not consider possible long-range effects of applying knowledge gained in the research (for example, the possible effects of the research on public policy) as among the research risks that fall within the purview of its responsibility (emphasis added).

.111 (a) (7) . . . , there is adequate provisions to protect the privacy of subjects and maintain confidentiality of data.

An important feature of the first consideration is the admonition not to withhold approval based on speculation regarding research outcomes and how this might be used in modifications of existing public policy.

Two criteria developed in the Common Federal Policy could be modified to set a standard for an acceptable risk associated with disclosure of confidential information. The specification of exempt research activities makes reference to “disclosures that could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects’ financial standing, employability, or reputation.” (Federal Register 10Nov88, 53 (218), pp. 45672–45673):

“Minimal risk” means that the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examination or tests.”

Acceptable procedures would be those that reduced the risk of release of data that could subject individuals or organizations, as units of analysis, to a probability of “criminal or civil liability or damage to financial standing, employability, or reputa-
tion" to less than they would "ordinarily encounter in daily life." As the current levels of risks to privacy from use of data sets for research are much lower than the risks associated with private or public administration data sets, this standard could easily be met.

A "Federal Data Base Research Certificate" should provide:

1. Approval for access to any federal data set except those considered critical to national security interests.

2. Complete and total immunity from disclosure of participant identities, whether individuals or organizations, from any legal proceeding or subpoena.

3. Substantial penalties for any member of the research staff that fails to exercise extreme care in the handling of identifying information such that the identity of any individual or organization represented in the data set becomes known to any outside the project staff, even those within the same organization, institution, or agency.

4. Prohibitions against any release of individual identities by the research staff with or without the permission of the individuals or organizations involved.

5. Clear specification of the eventual disposition of any information in the research files that could link individuals or organizations to specific data records at the termination of the project.

6. Full and complete immunity for the identifiers in the data set from any legal subpoena from any source under any conditions – including national security interests.

In short, the tightest possible legal and technical procedures for ensuring privacy, confidentiality, and anonymity for the units of analysis should be associated with the Data Base Certificate. The legal requirements for maintaining confidentiality associated with the Data Base Certificate should equal or exceed that of all existing federal agencies–Bureau of Labor Statistics, Census Bureau, Internal Revenue Service, Social Security, etc.

Such a mechanism would provide for a common procedure and standards for all research involving federal records and other large-scale data sets. Over time the expertise and technology required for such research would be centralized and more easily shared. It should substantially reduce the cost of implementing such projects and enhance the protection afforded individuals and organizations represented in the data sets.

Of major importance, an independent entity would be in a position to evaluate the societal benefits of the research, apart from the costs and complications posed for any specific agency. Furthermore, as the potential for analysis, significant societal issues, and concerns about privacy and confidentiality change – and they are sure to change in the future – this entity could adjust its criteria and provide new compromises to resolve the basic dilemma – the balance of research benefits and privacy risks.

*   *   *   *

Psychological, social, and cultural studies pertain to virtually everything that people treat as a problem in our civilization – violence, theft, pollution, and illness – and nearly everything hailed as a triumph – justice, plenitude, artistry, and freedom.

Modern life has . . . been profoundly changed by an understanding of visual and auditory perception, the aging process, the roles of families and schools in the development of cognitive skills, the causes of gender and racial differences, the dynamics of inflation, and the nature
of poverty and dependency (Gersten et al. 1988, pp. 1–2).

Citizens have a right to an efficient, effective, and just government. Given the current complexity of major social problems, optimal government decisions are not possible without more complete and detailed understanding of these social problems. This understanding will not be forthcoming without research access to large-scale data sets on individuals and organizations. Such access has not produced any problems in the past. The societal benefits of continued advances in social and policy research seem substantial. It is up to the social and policy science enterprise to make its best case for the privilege of making further contributions to society.

14. References


Griffin, W.B. (1990). Letter to George T. Duncan, Chair, Panel on Confidentiality and Data Access, August 10.


