

Private Archives and Public Needs

Stephen J. Ceci
Elaine Walker

Cornell University
Cornell University

ABSTRACT: Issues related to psychologists' willingness to share data from research projects supported by tax dollars are discussed. Attention is focused on the legal, ethical, and pragmatic reasons given for refusing to share the raw data of one's federally sponsored research with other psychologists. A proposal to mandate data sharing is put forward, with discussion of the technical and ethical costs and benefits it would likely entail. Such a proposal would be congruent with mission statements contained in grant announcements, the public's right to know, and a changing zeitgeist in the field of professional psychology from a Cartesian to a Baconian orientation.

This year the federal government will dispense somewhere between \$700 and \$900 million as part of its long-standing commitment to the biomedical and social sciences (Holt, 1981). Although most of this money is allocated as contract research for the internal benefit of particular agencies, approximately one fourth will go to fund the basic and applied research and training of psychologists. Despite the very real concern among social scientists that certain types of research funding are being discontinued for political reasons, government support of the social sciences, and psychology in particular, still represents a significant resource—one that requires careful management.

Many agencies, departments, and institutes share in dispensing federal funds to social scientists, with the biggest roles being played by the Alcohol, Drug Abuse, and Mental Health Administration; the Department of Education (DOE); the Department of Defense; the National Institutes of Health (NIH); and the National Science Foundation (NSF), listed in descending order according to their direct contributions to psychologists (Holt, 1981). The missions of these agencies in funding social science research and training through tax dollars can be found in the prefaces to their sundry publications. As can be seen in the quotations below, taken from federal research grant announcements, these agencies are primarily mission oriented; that is, their congressional authorizations carefully specify their goals and authorize them to fund only research on topics that are congruent with these goals. In all cases these goals have to do with solving national needs and enhancing the quality of life in society. For example,

in the opening sentence of the NIH (1980) *Extramural Programs Announcement* there is an introductory description stating that "the mission of the National Institutes of Health (NIH) is to improve human health through biomedical research" (p. i). The statement then describes the specific areas of research that will be considered mission appropriate by NIH. It explains that the earliest stage in the process of determining NIH funding is to reject applications that are unrelated to the mandated mission of improving the nation's health. In describing the award mechanisms of each of its institutes, NIH repeatedly echoes the global mission stated above. For instance, the National Institute on Aging "provides support through all award mechanisms to further the goals of understanding the aging process and improving the ability of the individual and health care professional to respond to the diseases and other clinical problems of the aged" (NIH, 1980, p. 11). The Social and Behavioral Research branch "awards grants for research, and research training to develop greater understanding of the social, cultural, economic, and psychological factors that affect . . . people in society" (NIH, 1980, p. 13).

The Department of Education is also straightforward about its mission and the need for researchers to justify proposals in specific areas in terms of mission appropriateness. Whereas NIH's concern is with national health, DOE's is with national literacy. For example, the National Institute of Education's (1980) five-year outlook statement begins: "The fundamental goal of the NIE, derived from its legislative charter, is to aid through support of scientific inquiry into the educational process, the attainment of equality of opportunity to receive an education of high quality" (p. 391). In another DOE announcement (NIE, 1981b) entitled *Teaching and Learning Research Grants Announcement*, one reads:

This organization enables the Institute to carry out its legislative mandate to support research which will improve educational quality and foster equality of educational opportunity . . . by supporting research to (1) enlarge scientific understanding of human learning, and (2) increase knowledge of effective teaching and assessment practices. . . . Congress has emphasized in the mission the importance of research that contributes to the achievement of the national policy of the United States: to provide every person an equal opportunity to receive an education of high quality. (pp. 1-3)

Even at NSF, which unlike other agencies disseminates large amounts of its budget to untargeted topics in basic research to improve our understanding of research, the mission statement is quite unequivocal. The NSF's original mission statement described its purpose as supporting basic research for the advancement of the nation's health, prosperity, and welfare. As was the case with the DOE mission (and that of its predecessor, the Office of Education) of improving educational quality, the NSF mission arose out of congressional concern about national interests. Similar national concerns spurred the creation of the Department of Defense, the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA), the Department of Agriculture, the Department of Energy, the Department of Justice's Law Enforcement Agency Administration (LEAA), and every other agency empowered by Congress to dispense public funds. Each of these agencies was established to bolster the welfare of some aspect of the nation. Their specific announcements typically begin by restating their national interests and then proceed to limit the areas of research that are eligible for funding and finally describe the linkage between these areas and the nation's welfare.

It is apparent that some psychologists have tended to ignore the reality of these national interest clauses found in mission statements, believing them to be palliatives inserted by politicians to appease taxpayers concerned about the deployment of their tax dollars in frivolous research. However, it has become evident that whatever their genesis, these mission statements are now serving as bases for current policy decisions related to the funding of psychological research. For example, Title 1 of the National Research Act of 1974 (Public Law 93-348) mandates that the use of all federal funds (through the Department of Health and Human Services) must be governed by "the nation's need for biomedical and behavioral research . . . in areas to be periodically re-established" (National Research Council, 1981, p. 111). Both of the controversial 1980 and 1981 reports to Congress by the National Research Council's Committee on a Study of National Needs for Biomedical and Behavioral Research Personnel urged a reallocation of training money, specifically in the field of psychology, to conform more closely to national needs and interests. Although it is too early to assess the full impact of these reports, it is clear that they have already catalyzed the way psychologists go about training other psychologists. As just one example, the ratio of predoctoral to postdoctoral support has been recommended at 1 to 3 by 1985. That is, for every training grant to ADAMHA (and soon NIH), a department of psychology will have to agree to provide space and resources for

three postdoctoral fellows for each predoctoral trainee. Such a dramatic break with our traditional method of training is justified in these reports to Congress on the grounds that this shift is best for the nation (NRC, 1981).

An even more recent (and ominous) signal that accountability to taxpayers is being endorsed over and above the scientific merit of one's proposed research is the feud that developed when USDA implemented its notorious loyalty checks to ensure that those reviewing grant proposals embraced the President's values and agendas. In a recent issue of *Science*, John Schrote of the USDA defended this mixing of politics and science:

To suggest that only those scientists who are recognized by the present scientific establishment can make those determinations (distribution of grant funds) is the height of elitism. We reject that notion and we believe this Administration should be responsible to the taxpayers, not a peer-review committee. (Marshall, 1982, p. 1391)

Clearly, the political and social climate of the 1980s is far different from that which existed in the late 1960s when there was money to burn for individual researchers and few overriding considerations of national welfare.

The Problem

Nelkin (1982) has documented the tendency among some researchers who receive federal support to view the products of their research as their "private intellectual property." Some investigators who receive federal aid to conduct their research have refused to share their raw data with other members of the scientific community who requested to see them, or have insisted on favorable publication agreements stipulating that any publications based upon their data include their name as first author. It has been suggested that any manuscript containing someone else's data be sent to them for review (Clubb, Austin, Geda, & Traugott, Note 1).

The issue of proprietary rights over the control of scientific information is the focus of increasingly complex legal and administrative disputes. Cecil and Griffin (Note 2) provide an excellent analysis of recent court decisions in cases in which members of the scientific community were sued for refusing to share their federally supported data with others.

In his discussion of obstacles to data sharing, Hedrick (Note 3) argued that the beneficiaries of data sharing are the scientific community, the data

This research benefited from the extensive comments provided on previous drafts by Richard Rockwell at the Social Science Research Council.

Requests for reprints should be sent to Stephen Ceci, Department of Human Development and Family Studies, Cornell University, MVR Hall, Ithaca, New York 14853.

requestors, and ultimately the general public. It will be argued here that the nation's welfare is best served by a mandatory extension of the public's "right to know" to cases of data collected with public funds (cf. Cecil & Griffin, Note 2). If one accepts the idea that an important, if not principal, justification for government support of social science research is the advancement of the nation's health, prosperity, and welfare, then the tendency of some researchers to view federal funding as a mechanism for realizing personal ambitions is not only misguided but out of tune with a growing zeitgeist of public accountability and the concomitant subjugation of personal goals to societal needs (Bevan, 1976, 1980; Fishman & Neigher, 1982). It is the purpose of this article to illuminate some of the ethical, legal, and pragmatic implications associated with the mandatory sharing of raw social science data in cases collected with tax dollars.

Although some of these arguments are thought to apply to data collected with support from private but tax-exempt foundations, the complexities of discussing them would carry us beyond the scope of this article. The interested reader should consult Cecil and Griffin (Note 2) for an analysis of this issue. Also, in terms of their impact (measured by total funds dispensed), these private tax-exempt foundations dispense less than 10% of funds available to social scientists. Our intentions are to address those issues that are considered critical by legal experts, sociologists, and statisticians and to place them in the context of professional psychology. We begin by discussing reasons for refusing to share data and follow this discussion with a proposal for mandatory data sharing. Finally, we conclude with a cost-benefit analysis of this proposal.

Reasons for Refusal to Share Data

Both Clubb et al. (Note 1) and Hedrick (Note 3) have provided many reasons for a scientist's refusal to share data: (a) administrative reasons (e.g., security reasons for nonrelease, financial costs of duplication, data set inadequacies, poor communication); (b) ethical reasons (e.g., concern about the qualification of data requestors, violations of confidentiality); (c) fear that procedural or computational errors in their analysis will be discovered; and (d) fear that divulging unpublished data could result in research by others that preempts the investigator's subsequent planned publications.

Many psychologists, even if they are under no formal obligations from their funding agencies, feel that releasing their raw data could constitute an ethical impropriety. Their subjects consented to participate under specific conditions that may not have included serving as the basis for others' studies, whose aims could actually be objectionable to them.

Even in cases in which the investigator has no qualms in this regard (i.e., believes that the aims of the new study are congruent with the original solicitation agreement), data sharing may present a security problem. Can one be sure that the data requestor will safeguard the subjects' confidentiality? (Barnes, 1979; Boruch & Cecil, 1979). This illustrates a dilemma. On the one hand, scientists are expected to provide adequate protection for their subjects; on the other hand, a hallmark of the scientific enterprise is its open nature (Hedrick, Note 3). Coding data prior to its release to disguise all personally identifiable information is not always a solution to this problem for two reasons.

First, testing of the new investigator's hypotheses may require identification of the subjects. An actual example will illustrate this. A colleague of ours was interested in the impact of urbanization during adolescence on the social and emotional development into adulthood. He proposed to assess this by examining the attitudes of adolescents who were reared in rural areas that were undergoing dramatic urbanization during the early 1960s and comparing them with the attitudes of those living in areas where urbanization either always or never existed. What impact might this have on these individuals' later attitudes about marriage, employment, physical health, and so on? Several survey studies of adolescents' attitudes had been carried out by other investigators in rural areas of West Virginia during the 1960s. These subjects are now in their 30s and 40s. Our colleague wished to interview them to assess the stability of their attitudes. As can be seen, this type of historical (within-cohort) analysis requires the identity of the individual subjects if for no other reason than to find their addresses so that they can be interviewed.

Even when the identity of an individual is not supplied, it may sometimes be accurately determined through a process known as triangulation (Clubb et al., Note 1). For example, one may determine the identities of specific individuals by checking the small area portion of the U.S. Census figures against area motor vehicle registration records. Neither of these sources by itself identifies individuals, but together they may.

A variety of proposals have been put forward to obviate these problems of confidentiality. For instance, it has been suggested that investigators or funding agencies might interject a small amount of random error into their data before sharing it with data requestors (Clubb et al., Note 1). This would result in researchers being uncertain about the verity of any individual case record but still confident in the reliability of their aggregate findings. Another suggestion has been to reindex data differently or to group it into larger units than the single subject.

(However, this might devalue the data for certain purposes, e.g., where the aim is to discriminate individuals with rare traits.) Finally, it has been suggested that prior to sharing data with others, one might remove any potentially offensive pieces of information entirely.

The above arguments are unsatisfactory to some. Decisions to delete or reindex data have a way of returning to haunt researchers (Mason, Taeuber, & Winsborough, 1977). As an alternative it has been proposed that individuals requesting the use of another's raw data be obliged to sign a legal statement endorsing confidentiality (see Greguras, 1979). The Department of Justice's LEAA has apparently operated quite successfully with this procedure.

Aside from these confidentiality-related issues, an experimental psychologist, whose view of data is confined to the modal experimental outcome, may refuse to share data out of a fear of embarrassment should the individual requesting it discover procedural irregularities or computational errors (Hedrick, Note 3). The typical funded study of an experimental psychologist yields a corpus of data that can be easily analyzed, in light of the investigator's hypotheses, within a few years of its collection, if not much sooner. This is not to claim that these data could not conceivably be analyzed further at a later time but merely that the investigator's proposed hypotheses, for which he or she was given public funds to test, could be tested with these data in the years immediately following their collection. Requests for one's raw data are usually made following their publication, that is, quite some time following their collection. Although requests for data are made for a number of benign reasons, such as the testing of hypotheses that were not of interest to the original investigator, the fear of embarrassment is a potent deterrent to sharing data (Hedrick, Note 3). In psychology there have been many cases where an investigator has provided his or her raw data to someone who in turn published procedural or computational errors made by the original investigator in previously published accounts of the data (e.g., Boruch & Cordray, 1980; Clark, 1973; Wolins, 1962, 1981; Cordray & Orwin, Note 4). In discussing this matter with psychologists who refuse to share their data, it is apparent that the fear of public embarrassment is never far from their minds when rejecting such requests.¹

A related argument for refusing to share one's raw data is that the inherent reward structure in science encourages data requestors to be undeservedly negative in their evaluations of the original investigator's data or otherwise risk being unable to publish what essentially amounts to a favorable replication of the original investigator's conclusions. These concerns over "temptations to hypercritical-

ity" have been dealt with at length by Clubb et al. (Note 1) and are probably overstated on the part of researchers who refuse to share their data. The peer review process, despite its substantial problems, does provide a means of countering undeservedly critical evaluations.

It is certainly understandable that fear of embarrassment might raise doubts about sharing data. Still, it is important to the welfare of both the nation and our science that accuracy, as a principle, override personal factors. Fishman and Neigher (1982) have cautioned psychologists to cease functioning "within a culture of individual professional freedom" (p. 544). In emphasizing psychology's shrinking resources and greater public accountability, these authors have reminded us of the need to search for more cost-effective means to pursue research. This is bound to bring frustrations and embarrassments resulting in part from the way we, as psychologists, have been socialized through our own training. Although these potential personal embarrassments are understandable, they need to be considered against the backdrop of what resources society is committing to psychology and what return is expected on the investment.

Less understandable than withholding data for fear of embarrassment is the consignment of one's data to private archives ad infinitum and the refusal to share the data for fear that they could someday be of further research interest to the investigator. In some fields of social science research, most notably macrosociology and demography, the extended use of very large data sets is common. These data sets are pregnant with possibilities, as evidenced by the very fact that macrosociologists often purchase sets collected years ago and mine them productively for many additional years. For example, it is common to see many researchers independently working with the same data sets to test different models. Often

¹ In a recent study by Peters and Ceci (1982), the validity and reliability of editorial review practices were examined by submitting typed manuscripts of previously published articles to the same journals that had published them. To test certain hypotheses resulting from reviewers' statistical criticisms, it was necessary to examine the raw data from several of these studies. Numerous authors declined to provide copies of these data because of fear that discoveries might be damaging to their professional reputations, despite promises of strict confidentiality. In a related vein, Leon Kamin has recently tried repeatedly (without success) to obtain copies of a well-known researcher's raw data from an interview study establishing the connection between genotype and schizophrenia. Despite the fact that the researcher in question was supported by federal tax dollars to collect these data, he has not made them available to Kamin, even though the latter offered to cover all costs involved. Given Kamin's well-known reputation (Kamin, 1974) for scrutinizing other researchers' data, it is possible that the motive behind the researcher's refusal to provide his raw data was a fear of possible embarrassment.

these data have been collected by private or semi-private groups and made available to researchers for a price, for example, the National Public Opinion surveys and the Roper Institute's data sets. (Such commercial data are often protected by copyright against duplication.) Sometimes, however, investigators receive public funds to amass large data sets. The Berkeley Guidance Study and the Oakland Growth Study are cases in point. These enormous data sets contain information about individuals born around the time of the Great Depression who have been followed longitudinally up to the present. As such, they represent unique opportunities for researchers interested in issues having to do with transmission of values, life transitions, and the impact of economic and social milestones (e.g., military service, divorce, loss of income) on development. Unfortunately, it is often the case that large, publicly funded data sets of this sort are kept from interested social scientists by the original investigator or some committee acting on behalf of the investigator's institution. Requests for copies of portions of these data, accompanied by offers to reimburse the costs of duplication, are reviewed by the investigator (or the appointed institutional committee) for their merit, the degree to which similar research is being done elsewhere, and so on.² In other words, these investigators or committees have set themselves up as peer reviewers. Underlying such behavior is the view that these data are theirs, not ours.

Clubb et al. (Note 1) give a balanced treatment of the costs and benefits involved in ad hoc data sharing versus intermediary repository facilities for data sharing. The former refers to a mandate that investigators assume personal responsibility for sharing their data with those who request it. This method is inexpensive and nonbureaucratic and, most importantly, keeps the data in the hands of those who know it most intimately—the original investigators. The negative aspects of ad hoc data

sharing are that the burdens of storage and duplication fall on the individual investigators, and it may not be clear from where or whom to request data, which is not the case with an intermediary repository. For their part, intermediary repositories (e.g., NSF's data archives) have their own costs and benefits. However, Clubb and his colleagues believe that these intermediary data banks constitute the most efficient means of using existing data.

Proposal: An Open Data Bank

It is proposed that in the long term, a national data bank be created or extant operations at some agency already versed in the problems of archiving large amounts of data in easily accessible form (e.g., the Census Bureau, the Library of Congress, or the National Archives' Division of Machine-Readable Data) be expanded to accommodate the routine cataloging and dissemination of raw data from all publicly sponsored investigations. This would encompass all research grants, federal cooperative arrangements, decennial censuses, and contracts in which the classification of data is not clearly in the interest of national security or the protection of human subjects. The technical capability for coding and coordinating such a massive data base is already available, although various implementation proposals differ in their administrative and institutional details (e.g., David, 1980; Duncan, Beresford, Chartrand, Day, & Kasputys, 1980). Congress, primarily concerned about the inefficient use of extant statistical data for policy decisions, recently created the Office of Federal Information Policy to oversee the better management and access to data (H.R. Bill 6410 and counterpart Senate Bill 2608). Numerous professional groups, such as the American Statistical Association, have put forward proposals governing issues ranging from the use of mechanical data coding and access systems to the need for confidentiality (American Statistical Association, Committee on Privacy and Confidentiality, 1980; Bonnen et al., 1981).

Because this ambitious proposal is unlikely to be implemented in the near future, we propose as a short-term solution that each federal funding agency encourage data sharing by its recipients. Guidelines governing the dissemination of one's data should be negotiated at the time of an award to clarify possible exemptions for reasons such as breach of confidentiality or national security. Transfer agreements and certificates of privacy should specify the responsibilities of both the original investigator and subsequent researchers interested in sharing the data (see Greguras, 1979). Subject solicitation and informed consent procedures should make clear to subjects the possibility that other researchers may be permitted to examine the data,

² For example, the Berkeley and Oakland data sets were collected with federal support to the investigator (see MacFarlane, Jones, & Honzik, 1976). According to the Executive Committee of the Institute of Human Development, University of California, Berkeley, "The Institute is under *tacit* contract [emphasis ours] to see to it that these materials are used in worthwhile research—research that justifies the time and help the participants have granted and have, implicitly agreed to grant in the future. . . . Data in the longitudinal archives are available to any investigator who meets the appropriate criteria as given above. These data may, however, be reserved for the use of the investigator(s) responsible for designing and collecting them or for generating codes and indices based upon them. The period for which data are reserved is determined by the Director(s) in consultation with the investigator(s) concerned and the Executive Committee, is given in writing, and will not generally exceed five years from the date by which the relevant materials are collected and/or generated" (Institute for Human Development, 1981, p. 6).

under strict agreements regarding the protection of their confidentiality. Since this is viewed only as an interim measure, we propose that it be done on an ad hoc rather than intermediary level (Clubb et al., Note 1) to avoid financial complications.

To assuage investigators' fears that under such a completely open policy their proposed hypotheses could be tested (and published) by competitors before they themselves have done so, a limited period of exclusivity might be established. The investigator would be granted exclusive rights to analyze and publish federally supported data for a limited period of time. The length of the exclusivity period could be negotiated as part of the normal peer review process and would be expected to vary as a function of amount of data involved, the number and complexity of hypotheses, and the level of institutional and other support available. Investigators could be asked to sign a statement of assurance (like Greguras's, 1979, information transfer agreement) which acknowledges that their raw data will be made available a given number of months or years after the termination of support to those requesting it. This would give investigators ample time to analyze and report on those aspects of their data that they were specifically funded to gather and analyze. At the same time, it would open new possibilities for interested social scientists to perform secondary data analyses either to replicate earlier reports or to test new hypotheses. To avoid financially burdening the investigator, arrangements regarding the costs of providing data would need to be negotiated (e.g., persons requesting the data would bear the costs of reproduction).

Ethical and Legal Considerations

There is nothing in any of the federal agencies' mission statements that justifies scientists' withholding their raw data from peer scrutiny and use, with the possible exception of research that has a bearing on our national security or results in stress to human subjects. Indeed, as Greguras (1979) and Greguras, Broder, and Zimmerman (1979) have pointed out, human subjects exemptions are already in place at some granting agencies. These were specifically written to facilitate data sharing among researchers for scientific or policy reasons. The foremost purpose, as we have seen, of a federal agency's involvement in social science research is to further its own mission, not the personal research career of the investigator. It is apparent that much of the federally funded data that have been kept from other scientists could further the funding agency's mission were those data made available to others.

Given the foregoing argument about the congressional mandate to funding agencies, the ethical appropriateness of withholding nonclassified

data can hardly be argued. It is easy to see how the nation's welfare is advanced by an investigator's willingness to divulge his or her findings. "The notion of private individual creativity contradicts the concept of open pursuit of replicable and testable knowledge, particularly in the case of costly data collections that can not be readily duplicated" (Clubb et al., Note 1, p. 34).

Many readers probably know, first- or second-hand, about investigators who received tax-supported grants to advance the public welfare and then withheld their data from other researchers who happened to be interested in related issues. From a moral or ethical perspective, other researchers have just as much right to use these data as do the original investigators, provided some agreed-upon period of exclusivity has elapsed and no harm is posed to the original human subjects or to national security. Again, it must be said that the congressionally mandated missions of the funding agencies have to do with national interests and welfare, not private interests (Lowman, 1981). However, as Bevan (1976) aptly observed, "Indeed, we in science have been so concerned during the past several decades with 'getting ahead' that we have not understood that this has increasingly entailed ignoring our responsibility to the larger society for the quality of life in general" (p. 490). Bevan (1980) sounded an eloquent alarm for psychologists to make better use of our limited resources, such as data, especially when these resources are tax supported.

In advocating a shift in professional attitudes, Bevan and others distinguished between Baconian and Cartesian views of research in psychology. On the one hand there is the

individualistic, esoteric, "dogma eat dogma" . . . nature of the Cartesian view of research associated with the scientific paradigm. (Fishman & Neigher, 1982, p. 540)

[For the Cartesian investigator,] doing science is like running a race, and one's colleagues . . . can therefore only be viewed as strong competitors. . . . The public is perceived as having only one role, that of patron. . . . [In contrast, the Baconian view] is a cooperative activity within a professional community marked by a clear-cut division of labor but bound by a single shared altruistic commitment to the promotion of human welfare. (Bevan, 1980, pp. 780-781)

From a legalistic perspective one might argue that the investigator, by virtue of having developed the grant, has an entitlement to its fruits, including the data. In line with this are the many cases wherein individual investigators are permitted through their grant arrangements to retain patents on the outcome of their government-funded work or to take highly specialized equipment with them in the event they change employment (see National Science Foun-

dation, 1978, p. 17). However, the recipient of a federal grant is nearly always the investigator's institutional affiliation, not the investigator him- or herself. This is done to assure full compliance with statutory requirements as well as to guarantee the financial provisions and scientific resource commitments described in the grant proposal. Thus, legally it is not clear that investigators have any proprietary rights other than those conferred by the granting agency. Ironically, in the case of patents, copyrights, and other forms of ownership, the intention, as stated in federal publications, is not to facilitate exclusivity but just the opposite, to encourage dissemination:

It is the policy that the results of activities supported should be utilized in the manner which would best serve the public interest. . . . The Contractor shall not assert any rights at common law or in equity or establish any claim to statutory copyright in such materials and all such materials shall be made freely available to the Government, the education community, and the general public. . . . Arrangements for marketing or other distribution of materials under copyright, for a limited period of time, may be authorized . . . upon a satisfactory showing that such protection will result in more effective development or distribution or would otherwise be in the public interest. (National Institute of Education, 1981a, p. 5)

The Freedom of Information Act (5 U.S.C. 552) mandated the right of a concerned public to examine unclassified government documents. Although the legal wrinkles are not all pressed, it is clear that this legislation is having an impact on the nation (Cecil & Griffin, Note 2). It has allowed public scrutiny of everything from government-funded economic surveys to Central Intelligence Agency documents, yet in the arena of social science research, the public's right to know within a reasonable period of time appears to be unprotected. Therefore, legally, the practice of regarding publicly financed research as one's own intellectual property is incongruous with both the mission of governmental support of social science research and the right of taxpayers to inspect a product they financed (Fishman & Neigher, 1982; Lowman, 1981; Nelkin, 1982). However, there is no law making it a crime to refuse to share one's data, though the courts have often upheld a plaintiff's request for the raw data of another investigator (Cecil & Griffin, Note 2). Presently, any member of the public has the right to request from a federal agency copies of any grant proposals or final reports, but with the exception of a few agencies (like NSF), investigators are not required to include the raw data in their final report. (Most agencies have created freedom of information offices solely to accommodate this public interest.) Because such requests can come before the investigator has had an opportunity to publish the con-

tents of the final reports, it is customary for freedom of information offices to notify the investigator concerning each request for the proposal and final report. This is viewed as a protection of the investigator's ideas from academic piracy. Even though many find these final reports useful, they do not typically contain raw data (NSF and several others being exceptions); thus they cannot take the place of a well-organized, accessible data bank.

Pragmatic Considerations

To some social scientists, the requirement to deposit raw, and sometimes undigested, data into an open data bank is seen as a discouragement to grant writing. Their logic is that many researchers would forego the tedium and stress involved in grant development if they anticipated that similar data of others would be available for their use. After all, so the reasoning goes, why bother to devote one's limited time and resources to grant writing when the data will be made available to others, especially when similar data may be collected by others and made available to one's self by the same law?

The above argument against data sharing ignores the ecology of research in academia, government, and the public sector. As academic psychologists know all too well, successful grantsmanship brings with it more than the opportunity to test one's hypotheses. Grants enable scientists to travel to conferences; employ research assistants, personal secretaries, and support staff; and purchase equipment that can be used beyond the scope of grant-related research, with the understanding of the agent institution. Grantsmanship frequently serves as a vehicle for career advancement and merit pay awards. Therefore, it is unlikely that the establishment of an open data bank policy would have a negative impact on grant development. The very real rewards associated with being a principal investigator, along with the exclusive use of the data for a limited period, should suffice to maintain the attractiveness of pursuing grants. Clubb et al. (Note 1) have suggested that additional steps might be instituted to reward investigators for sharing data, for example, acknowledging their services on all publications based on data they collected. Therefore, pragmatically, it is not anticipated that the mandatory depositing of raw data would necessarily affect the quantity or quality of social science grant writing.

Benefits of an Open Data Bank Policy

Minimally, the creation of an open data bank, wherein all government-funded data that are not harmful to national security or human subjects would be deposited and made available after an agreed-upon time has elapsed, is beneficial for four reasons. The first reason concerns its time- and cost-

effectiveness. An open data bank is far more effective in both regards in accomplishing the mission of federal support—enhancing society's welfare (Hedrick, Note 3).

The recent evaluation of the effectiveness of early intervention programs like Head Start (Darlington, Royce, Snipper, Murray, & Lazar, 1980) serves as a good example. This is an issue of large public concern. As parents, educators, and taxpayers, it is in our best interest to learn all there is to know about the effects of early intervention. Darlington et al. found that in evaluating the relative merits of intervention programs, there existed a substantial research literature on individual programs evaluation. However, it was often necessary to go beyond the published summary statistics and examine the raw data and procedures from these different studies. If such data already exist, it seems hardly efficient to fund researchers to recollect them. Although there is something to be said for replicating published work, this is not the issue, because the investigators would probably not have exactly reproduced any of the studies, since replication was not their intention. Moreover, were an investigator simply interested in replication of another investigator's work, the possibility of funding would be slender, given the emphasis on creativity, originality, and so on. Hence, an open data bank policy is both time- and cost-effective in accomplishing the federal mission in social science research. It would result in greater use of data, thus improving the cost-benefit ratio and increasing the likelihood of enhancing the nation's welfare.

The second reason that an open data bank would be beneficial is closely related to the first. In the social sciences we have recently witnessed a profusion of new forms of statistical and technological methodologies. Log Linear, LISREL, variants of time series, and so forth, have been refined, programmed, and made widely available. Such methodological innovations may have been unknown or uninteresting to the original investigator, but they greatly expand the scope of insights obtainable from a data corpus. Large bodies of previously collected data could be productively reanalyzed and modeled with these new forms of analyses if interested scientists only had access to them (see Mason et al., 1977, for a comprehensive listing of descriptions). An excellent example of this is the current practice of program evaluators' undertaking meta-integrative analyses of data from a collection of independent studies and estimating an overall effect size. Wortman's (Note 5) assessment of the impact of busing on school achievement and Smith, Glass, and Miller's (1980) assessment of the effectiveness of psychotherapy are cases in point. At least in Wortman's case, it was necessary to spend a great deal of time

searching for state reports on busing. He reported that these data were sometimes unavailable and other times in poor physical shape. On a far larger and more frightening scale, the existence of entire current population surveys are in doubt, and concerned social science researchers have forcefully argued for the necessity of systematic and centralized archiving of such data in order that they might be available for future scientists working on problems of national interest.³ It is imaginable that methodological breakthroughs will continue as long as research continues. An open data bank with efficient archiving, storage, and retrieval systems permits the most powerful forms of analyses to be carried out on yesterday's data.

So far the first two benefits of an open data bank have had to do with efficiency. It has been argued that it is more cost-effective and less time-consuming to make previously collected data available for new analyses than to fund its recollection. The third benefit concerns the ability to do research that would otherwise be impossible no matter how great the time and financial resources available. An illustration of this benefit is in the area of so-called historical analysis. Historical questions can only be answered with data from the epoch in question. Elder's (1974) analyses of the impact of the Great Depression, service in World War II, and so on, on life-span transitions are historical analyses that were made possible by virtue of his having obtained permission to use data collected as part of the Berkeley Guidance Study. It is impossible to recollect these historical data. There are many data sets from the 1950s and 1960s dealing with childhood and adolescence. Researchers interested in the life courses of these children, who are now in their early to middle adult years, can provide valuable insights about the effects of industrialization, early school leaving, loss of income, divorce, and so forth—provided the researchers are allowed access to these data sets by the original investigators.

This third benefit raises the issue of confidentiality. The subjects in any given study are normally afforded confidentiality, an assurance that is made in the context of public dissemination and discussion of findings. As was pointed out earlier, in the case of historical analyses, it may be necessary for

³ In an effort to obtain an answer to the question of whether national population surveys were indispensable for carrying out basic research on important questions about social and economic change in the U.S., NSF convened a small workshop on public-use samples in 1976. In the conclusion to their investigation, they wrote: "It seems certain that March CPS microdata (i.e., raw, non-aggregated data) do not exist in any form whatever for years prior to 1960. Whether tapes exist for the surveys of the early 1960s, and if so which ones, is unclear. It would be a very early goal to determine exactly which ones have survived" (Mason et al., 1977, p. 140).

a contemporary researcher to know the identity of the original subjects to chart changes in their lives. This is a complex problem needing careful consideration. At issue are the needs of researchers (and society) on the one hand and the participants' right to privacy on the other. It is probably possible to balance these concerns in a way that avoids too much violence to either of them (Hofferbert, 1976; Feinberg, Note 6). One possibility is to place raw data in a coded form on accessible tape. The researcher might be given the names of the subjects but not the scores, outcomes, and life events that are associated with each name. The researcher could then locate these now-mature subjects, explain his or her purposes, and request permission to follow up on the earlier interview. Other possibilities can be imagined. It is interesting to note in this context that from a legal (as opposed to ethical) standpoint, the protection of confidentiality is relative and based on the principle of "reasonable effort." In essence, this holds that the determination of adequate protection of a subject's identity when data are being shared for statistical or research purposes is to be governed by a consideration of the effort required to disambiguate the identity of individual participants (Greguras et al., 1979). Thus, one need not guarantee that a subject's identity would be impossible to discover, only that discovery would require a fair amount of effort.

The fourth benefit of an open data bank is to settle disputes emanating out of previous investigators' works. The modern history of social science, especially psychology, is rife with controversy. There have been notable cases in which the veracity of an investigator's data has been called into question but the investigator was no longer alive and the data could not be located. Leon Kamin's (1974) treatment of the Cyril Burt affair exemplifies this. Had the British government or professional societies that publish journals mandated the routine depositing of raw data, many doubts and speculations (though not all) could have been tested. On a less odious level, there is something to be said for a system of data management that allows for the examination of previously collected data to test new hypotheses, check old claims, and so on. Finally, it needs to be mentioned that psychology lags behind many other fields (e.g., statistics) in its ad hoc handling of data. The American Sociological Association's revised code of ethics for researchers explicitly requires data sharing, even when the data are not collected with public support. Principle A-6 of the code states:

Consistent with the spirit of full disclosure of method and analysis, sociologists should make their data available to other qualified social scientists, at a reasonable cost, after they have completed their own analyses, except in cases where confidentiality or the claims of a fieldworker to the

privacy of personal notes necessarily would be violated in doing so. (1982, p. 9)

The impressive feature of such a code of ethics is its unambiguous intent. It says sociologists "should make their data available . . . at a reasonable cost"; it does not say "whenever possible" or "when convenient."

Conclusion

In the present Baconian proposal, researchers are seen as stewards of society's welfare by virtue of their role in achieving the funding agency's mission. This means the investigator is expected to use public funds in a manner that maximizes society's well-being. We have argued that personal ambitions are, of necessity, secondary to this stewardship. If the short-term and longer term proposals put forth here were to gain acceptance from federal decision makers, researchers would need to be socialized differently. A greater Baconian emphasis on the public well-being would be needed along with a Cartesian de-emphasis on private archives and personal intellectual property. It will become essential in times of projected fiscal constraint for social scientists to learn to share the fruits of the limited research that gets funded (Fishman & Neigher, 1982). From the standpoint of the taxpayers, who pay for this research, it does not matter so much who does what, so long as it gets done by someone, and the more "someones" involved in extracting meaning from a given set of publicly financed data, the better its value to the taxpayer. Keppler-Seid, Windle, and Woy (1980), Parloff (1980), and others have noted the growing perception of the public that government has been unaccountable for its wasteful programs. In this regard, Ohio congressman John A. Ashbrook was quoted in the *New York Times* (August 30, 1981) as describing the social and behavioral sciences as "the foolish fringe folly of researchers who use our tax money like the dilettante squanders his inheritance" (Hill, 1981, cited in Fishman & Neigher, 1982). It is time for social scientists to heed the advice of Holt (1981), which although referring primarily to the NSF, can be extended to all federal agencies, institutes, and departments:

Such concerns about accountability to the public are aimed particularly at the behavioral and social sciences. . . . Some members of Congress think that most research in these disciplines is silly, wasteful, and a poor use of public funds and amendments designed to cut support of the behavioral sciences have been introduced in Congress for the past three years. (p. 29)

Perhaps the creation of an open data bank would be a first step toward such accountability. Instead of duplicative funding of proposals, a few select ones, much larger in scope, could be funded and their outcomes made available for all to analyze. The

first step in implementing any form of data sharing, however, is to convince our colleagues of the correctness of this approach. As Fishman and Neigher (1982) point out, the time has come to put aside Descartes and "bring home the Bacon."

REFERENCE NOTES

1. Clubb, J. M., Austin, E. W., Geda, C. L., & Traugott, M. W. *Sharing research data*. Paper presented at the meeting of the American Association for the Advancement of Science, Washington, D.C. January 1982.
2. Cecil, J., & Griffin, E. *Obtaining access to data: Legal issues*. Report on data sharing to the Committee on National Statistics, National Research Council, Washington, D.C. 1983.
3. Hedrick, T. E. *Justification for and obstacles to data sharing*. Report on data sharing to the Committee on National Statistics, National Research Council, Washington, D.C. 1983.
4. Cordray, D. S., & Orwin, R. G. *Technical evidence for quantitative integration of research and evaluation*. Paper presented at the IASSIST-IFDO Conference on the Impact of Computerization on Social Science Research, Grenoble, France, September 1981.
5. Wortman, P. M. *Meta-analysis of quasi-experiments: School desegregation*. Unpublished manuscript, University of Michigan, 1981.
6. Feinberg, S. E. *Sharing statistical data: An overview*. Paper presented at the meeting of the American Association for the Advancement of Science, Washington, D.C. January 1982.

REFERENCES

- American Sociological Association, Committee on Professional Ethics. Revised code of ethics for sociologists. *Footnotes*, 1982, 10, 9-10.
- American Statistical Association, Committee on Privacy and Confidentiality. Business directories: Findings and recommendations. *American Statistician*, 1980, 34, 8-10.
- Barnes, J. A. *Who should know what?: Social science, privacy and ethics*. New York: Penguin, 1979.
- Bevan, W. The sound of the wind that's blowing? *American Psychologist*, 1976, 31, 481-489.
- Bevan, W. On getting in bed with a lion. *American Psychologist*, 1980, 35, 779-789.
- Bonnen, J. T., et al. Improving the federal statistics system. *American Statistician*, 1981, 35, 184-211.
- Boruch, R. F., & Cecil, J. S. *Assuring the confidentiality of social research data*. Philadelphia: University of Pennsylvania Press, 1979.
- Boruch, R. F., & Cordray, D. S. (Eds.). *An appraisal of educational program evaluations*. Washington, D.C.: U.S. Department of Education, 1980.
- Clark, H. H. Language as a fixed effect fallacy: A critique of language statistics. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 335-359.
- Darlington, R. B., Royce, J. M., Snipper, A. S., Murray, H. W., & Lazar, I. Preschool programs and later school competence of children from low-income families. *Science*, 1980, 208, 202-204.
- David, M. Access to data: The frustration and utopia of the researcher. *Review of Public Data Use*, 1980, 8, 307-326.
- Duncan, J. W., Beresford, J. C., Chartrand, R. L., Day, M. S., & Kasputys, J. Private versus public sector responsibility for the collection, distribution, and analysis of statistical data. *Review of Public Data Use*, 1980, 8, 307-326.
- Elder, G. *Children of the Great Depression*. Chicago: University of Chicago Press, 1974.
- Fishman, D. B., & Neigher, W. D. American psychology in the eighties: Who will buy? *American Psychologist*, 1982, 37, 533-546.
- Greguras, F. M. Information practices in criminal justice research. *Review of Public Data Use*, 1979, 7, 6-12.
- Greguras, F. M., Broder, P. K., & Zimmerman, J. The impact of legal controls on human subjects protection: A preliminary case study. *Review of Public Data Use*, 1979, 7, 87-95.
- Hofferbert, R. I. Social science archives and confidentiality. *American Behavioral Scientist*, 1976, 19, 467-488.
- Holt, V. E. Federal funding trends for research in psychology and the federal budget process: An overview. In R. P. Lowman, V. E. Holt, & C. O'Bryant (Eds.), *APA's guide to research support*. Washington, D.C.: American Psychological Association, 1981.
- Institute of Human Development. *Principles governing access to the archives for longitudinal studies*. Berkeley: University of California Press, 1981.
- Kamin, L. J. *The science and politics of I.Q.* Hillsdale, N.J.: Erlbaum, 1974.
- Keppler-Seid, H., Windle, C., & Woy, J. R. Performance measure for mental health programs: Something better, something worse, or more of the same? *Community Mental Health Journal*, 1980, 16, 217-234.
- Lowman, R. P. Seeking federal funding: Finding the pot of gold. In R. P. Lowman, V. E. Holt, & C. O'Bryant (Eds.), *APA's guide to research support*. Washington, D.C.: American Psychological Association, 1981.
- Marshall, E. Briefing. *Science*, 1982, 216, 1391.
- Mason, W. M., Taeuber, K. E., & Winsboro, H. H. *Old data for new research*. Madison: University of Wisconsin Center for Demography & Ecology, 1977.
- McFarlane, J., Jones, J., & Honzik, M. *The course of growth*. Chicago: University of Chicago Press, 1976.
- National Institute of Education. *Five-year outlook: Problems, opportunities, and constraints in science and technology*. Washington, D.C.: U.S. Government Printing Office, 1980.
- National Institute of Education. *Request for proposal number NIE-R-81-0008*. Washington, D.C.: U.S. Department of Education, March 1981. (a)
- National Institute of Education. *Teaching and learning research grants announcement*. Washington, D.C.: U.S. Department of Education, 1981. (b)
- National Institutes of Health. *Extramural programs announcement (No. 80-33)*. Washington, D.C.: U.S. Department of Health and Human Services, July 1980.
- National Research Council, Committee on a Study of National Needs for Biomedical and Behavioral Research Personnel. *Personnel needs and training for biomedical and behavioral research: 1981 report*. Washington, D.C.: National Academy Press, 1981.
- National Science Foundation. *Announcement grants for scientific research (NSF 78-41)*. Washington, D.C.: Author, August 1978.
- Nelkin, D. Intellectual property: The control of scientific information. *Science*, 1982, 216, 704-708.
- Parloff, M. B. Psychotherapy and research: An analytic depression. *Psychiatry*, 1980, 43, 279-293.
- Peters, D., & Ceci, S. J. A naturalistic study of the peer review process in psychology: The fate of published articles, resubmitted. *Behavioral and Brain Sciences*, 1982, 5, 187-195.
- Smith, M. A., Glass, G. V., & Miller, T. I. *The benefits of psychotherapy*. Baltimore, Md.: Johns Hopkins University Press, 1980.
- Wolins, L. Responsibility for raw data. *American Psychologist*, 1962, 17, 657-658.
- Wolins, L. Reanalyzing studies of race differences in intelligence: Scale dependent mistakes. In R. F. Boruch, P. M. Wortman, & D. S. Cordray (Eds.), *Reanalyzing program evaluations*. San Francisco: Jossey-Bass, 1981.