



The attached material is posted on regulation2point0.org with permission.



J O I N T C E N T E R
AEI-BROOKINGS JOINT CENTER FOR REGULATORY STUDIES

Should Researchers Be Required to Share Data Used in Supporting Regulatory Decisions?

Linda R. Cohen and Robert W. Hahn

Regulatory Analysis 99-1

May 1999

Ms. Cohen is professor of economics at the University of California, Irvine. Mr. Hahn is director of the AEI-Brookings Joint Center for Regulatory Studies, a resident scholar at the American Enterprise Institute, and a research associate at Harvard University. The views expressed in this paper reflect those of the authors and do not necessarily reflect those of the institutions with which they are affiliated. We would like to thank Irene Chan for valuable research assistance and Robert Crandall, Christopher DeMuth, Donald Elliott, Teresa Gorman, Robert Litan, J. Gregory Sidak, and Matthew Spitzer for helpful comments.



J O I N T C E N T E R

In response to growing concerns about understanding the impact of regulation on consumers, business, and government, the American Enterprise Institute and the Brookings Institution have established the new AEI-Brookings Joint Center for Regulatory Studies. The primary purpose of the center is to hold lawmakers and regulators more accountable by providing thoughtful, objective analysis of existing regulatory programs and new regulatory proposals. The Joint Center will build on AEI's and Brookings's impressive body of work over the past three decades that has evaluated the economic impact of regulation and offered constructive suggestions for implementing reforms to enhance productivity and consumer welfare. The views in Joint Center publications are those of the authors and do not necessarily reflect the views of the staff, council of academic advisers, or fellows.

ROBERT W. HAHN
Director

ROBERT E. LITAN
Codirector

COUNCIL OF ACADEMIC ADVISERS

MAUREEN L. CROPPER
University of Maryland
and World Bank

JOHN G. GIBBONS
formerly Office of Science
and Technology Policy

PAUL L. JOSKOW
Massachusetts Institute
of Technology

JOSHUA LEDERBERG
Rockefeller University

RODNEY W. NICHOLS
New York Academy
of Sciences

ROGER G. NOLL
Stanford University

GILBERT S. OMENN
University of Michigan

PETER PASSELL
Milken Institute

RICHARD SCHMALENSEE
Massachusetts Institute
of Technology

ROBERT N. STAVINS
Harvard University

CASS R. SUNSTEIN
University of Chicago

W. KIP VISCUSI
Harvard University

Executive Summary

The scientific establishment is deeply concerned over a proposed regulation that would require data to be shared on projects that are federally funded. Specifically, the proposed amendment to OMB Circular A-110 would require data collected by researchers at universities, hospitals, and non-profit institutions to be shared with interested parties if (1) the data are produced as part of a grant or agreement funded by the federal government; (2) the data are used in a published study; and (3) the data or study is used in formulating a policy or rule. Parties could request the data under the Freedom of Information Act. The proposed rule responds to a provision by Senator Richard Shelby in the 1999 Omnibus Spending Bill that requires data generated under federal awards at universities and non-profit institutions to be available to the public.

This regulatory analysis develops an economic framework for evaluating proposals to provide greater access to research data. Our analysis also offers specific recommendations for improving OMB Circular A-110 as well as the broader regulatory process.

We argue that the economic analysis of sharing research findings can be separated into three parts: the impact of requiring public access on incentives to produce data, research, and innovation; the impact of that requirement on the quality of research; and the impact of required access on the efficiency and transparency of policy.

The economic analysis demonstrates that the standard property-rights framework used to justify time-limited property rights for the use of data is not sufficient for addressing broader problems in which research and data could be used to help inform public policy decisions. The value of sharing data for public policy must also be considered. A second conclusion is that traditional peer review done by scientific journals is not adequate for purposes of relying on research for major public policy decisions. A third conclusion is that scientists who are reluctant to share their findings are more likely to have errors in their analysis than the average researcher. A fourth conclusion is that requiring the release of data could slow the development of data and delay the publication of results.

Although substantial costs and uncertainty may be associated with greater public access to data, our analysis suggests that academic norms alone provide very limited access to scientific data. We recommend improving Circular A-110 by narrowing and clarifying the scope of the proposed regulation. The proposed regulation should apply to economically significant regulations that have an annual economic impact of at least \$100 million. In addition, we recommend that Congress create an agency that would be charged with replicating the findings of regulatory agencies before such regulations could be implemented. The recommendations concerning replication would require additional legal authority. Taken together, our recommendations would help lay the foundation for a regulatory system that is more accountable and has more scientific integrity.

Should Researchers Be Required to Share Data Used in Supporting Regulatory Decisions?

Linda R. Cohen and Robert W. Hahn

1. Introduction

The scientific establishment is deeply concerned over a proposed regulation that would require data to be shared on projects that are federally funded.¹ Specifically, the proposed amendment to OMB Circular A-110 would require data collected by researchers at universities, hospitals, and nonprofit institutions to be shared with interested parties if (1) the data are produced as part of a grant or agreement funded by the federal government; (2) the data are used in a published study; and (3) the data or study is used in formulating a policy or rule. These parties could request the data under the Freedom of Information Act. If the agency obtains the data only in response to the FOIA request, the requester would be required to pay a user fee to the agency. The proposed rule responds to a provision by Senator Richard Shelby in the 1999 Omnibus Spending Bill and states in broad terms that any data generated under federal awards at universities and nonprofit institutions should be available to the public.

Proponents of the rule argue that interested parties have a right to know the basis for a regulation; that recipients of federally funded research have an obligation to share their data; and that regulatory decisionmaking is likely to be improved by making data available to all interested parties. Opponents of the rule, including many in the scientific community, believe that the rule will substantially diminish the productivity of the scientific community, will expose scientists to unfair attacks, and place unnecessary burdens on the research community. They also argue that the rule could place severe restrictions on those researchers who might obtain their data only on guaranteeing anonymity to subjects.² Some opponents also argue that current legal protections and the peer review process are adequate for the purpose of regulatory decisionmaking. Further, researchers and institutions with ties to industry fear that forced disclosure of proprietary

¹ Kaiser (1998, 1999).

² National Science Board (1999).

information will jeopardize their relationship with the private sector, which often requires a level of confidentiality as a condition for funding.³

The purpose of our regulatory analysis is to develop an economic framework for evaluating this and other proposals to share research data. We also offer some specific recommendations that could improve regulatory decisionmaking. While the focus of the Shelby provision is quite broad,⁴ we restrict our attention to regulatory decisionmaking because that was a major motivation behind the legislation and because the potential gains from improving regulatory decisionmaking are large.⁵

We argue that the economic analysis of sharing research findings can be separated into three parts: the impact of requiring public access on incentives to produce data, research, and innovation; the impact of that requirement on the quality of research; and the impact of required access on the efficiency and transparency of policy.

The economic analysis begins with a standard analysis of the economics of property rights—similar to the justification used for providing patents. In the case of patents, an inventor has an incentive to develop a new product because of the likely returns he would receive under the terms of the patent, which usually grants an exclusive right to a particular invention or product for several years. The point of patents and other related policies is to provide adequate incentives for developing new ideas and products. Such concerns also arise in research that could be used in a public policy setting. Adequate rewards must be provided to the researcher to develop new knowledge, but that is not the end of the story.

The rewards to an individual researcher must be balanced against broader concerns related to the quality of research. The replicability of published findings in refereed journals under the current peer review system is not something that can be taken for granted. Though publications have high standards, peer review almost never requires that reviewers reproduce the basic results. Providing greater access to data would provide an incentive to improve the quality of studies because researchers will be aware that their conclusions may be more easily checked for validity.

³ Ibid.

⁴ McGinley (1999).

⁵ Morrall (1986), Tenges and Graham (1996), and Hahn (1996).

Making data available before passage of a regulation offers important benefits to regulatory decisionmaking. First, making the information publicly available could improve the quality of information, thus leading to a decision that is more economically efficient.⁶ Suppose, for example, that the findings of the researchers were shown to be false before the development of a costly, environmental regulation. Then, the regulation could have been withdrawn or revised. Second, public access to data ensures greater transparency, which lends legitimacy to the regulatory process. Transparency is a valuable aspect of public decisionmaking in a democracy.

Although substantial costs and uncertainty may be associated with greater public access to data, our analysis suggests that academic norms alone provide very limited access to scientific data. We recommend improving Circular A-110 by narrowing and clarifying the scope of the proposed regulation. We also recommend that Congress create an agency that would be charged with replicating the findings of regulatory agencies before such regulations could be implemented.

The next section provides some background on the proposed amendment to OMB Circular A-110. In section 3, we present our economic framework. Section 4 summarizes our conclusions and provides recommendations for improving the proposed rule and improving the regulatory process.

2. Background

Congressional concern over access to scientific results is not new. While a general consensus exists that publicly supported research at universities is worthwhile, that consensus has broken down periodically over substantive areas of research, accounting practices, and commercial activities on campuses.

The current controversy over public access to data bears some similarity to the earlier debates over public rights in publicly supported activities, but its focus on regulatory concerns is different.

In 1993, researchers at the Harvard School of Public Health published the “Six

⁶ Opponents of access may argue that it could lead to a less efficient decision, since the political process does not necessarily promote economic efficiency.

Cities Study” in the *New England Journal of Medicine*, which was funded by the National Institute of Environmental Health Sciences and the Environmental Protection Agency (EPA).⁷ The conclusions in that study about the health effects of fine particles played an important role in the air pollution standards that the EPA developed for particulate matter in July of 1997.⁸

Several congressmen and a number of industry organizations—all unhappy with the regulatory response to the study’s results—requested that the EPA obtain the data and then release it through FOIA. The EPA initially wrote to the Harvard researchers to request the data but subsequently agreed to an alternate plan proposed by the Harvard dean for academic affairs. The researchers agreed to give the data to the Health Effects Institute, an independent research institute funded by industry and the EPA, so that the institute could convene an expert panel to reanalyze the data. The results of this study are not expected to be available until June of 1999—two years after the regulation was finalized.⁹

In response to the difficulty of obtaining such data, Representative Aderholt (R-AL) proposed legislation in 1997 that would have required any data generated from federal grants to be made public. His provision was defeated. A second attempt in 1998 to require the OMB to study whether sufficient public access to such data existed passed Congress as part of a spending bill. President Clinton vetoed that bill for other reasons. A similar provision, proposed by Senator Shelby (R-AL), finally passed in the 1998 Budget Reconciliation Act. The OMB responded to the provision with a proposed modification to the regulations that govern federal grants to universities and nonprofits that it issued February 4, 1999.

Shelby’s provision states that

the Director of OMB amends . . . Circular A-110 to require Federal awarding agencies to ensure that all data produced under an award will be made available to

⁷ Fumento (1997) and Kaiser (1997).

⁸ This discussion is drawn from Thurston (1998) and Kaiser (1997).

⁹ Kaiser (1997).

the public through the procedures established under the Freedom of Information Act.¹⁰

The OMB's proposed rule considerably narrows Shelby's provision, requiring that

in response to a Freedom of Information Act (FOIA) request for data relating to published research findings produced under an award that were used by the Federal Government in developing policy or rules, the Federal awarding agency shall, within a reasonable time, obtain the requested data so that they can be made available to the public through the procedures established under the FOIA.¹¹

Neither the Shelby provision nor the OMB proposal represents a broad departure from formal policy. Federal granting agencies typically state that results of research be published and data made available within reasonable periods of time. The data provisions have been applied loosely, however; more often, scientists make such information available as part of university norms and customs rather than in reference to grant conditions.

Federal regulatory agencies, for their part, can typically obtain any data for use in developing policy or rules, without regard to whether the data collection was financed by a federal grant.¹² Moreover, once an agency has such data, the data can be accessed by the public, subject to the exemptions of FOIA. Thus, the key change in the proposal is that an agency must obtain data from researchers in response to a FOIA request to the agency; currently, an agency has discretion as to whether to request the data.¹³

The Shelby proposal and OMB's proposed rule have generated a great deal of controversy. Congressman Brown (D-CA), the ranking minority member of the House Science Committee has proposed repealing the Shelby provision entirely, a position

¹⁰ Office of Management and Budget Salaries and Expenses, Fiscal Year 1999 Omnibus Appropriations Act, Public Law 105-277. Both the Shelby provision and OMB's proposed changes allow for the requesting party to pay a fee to cover part or all of the cost of producing the data.

¹¹ Notice of Proposed Revision of Circular A-110, 64 *Fed. Reg.* 5684, issued February 4, 1999.

¹² Elliott (1999). See, for example, Sec. 114 and Sec. 307 of the Clean Air Act.

¹³ Zarcoli (1998).

endorsed by NSF Director Rita Colwell.¹⁴ Other critics have focused on a number of ambiguities in the proposal, including the specific definition of a publication, the definition of data, and whether FOIA adequately addresses legitimate privacy and secrecy concerns of scientists and universities. On the other hand, some critics argue that the provision is excessively restrictive and should apply to all data rather than merely to those the federal government uses to develop policy or rules.

Our analysis focuses only on data used by federal regulatory agencies to develop policies and rules. While the more general access provisions contained in Senator Shelby's provision could be valuable, our analytical focus was dictated by a belief that regulation is a critical area where sharing data could lead to substantial improvements in policy.

3. The Economic Framework Analyzing the Impacts of Public Access

The economic framework is designed to address the question of whether and when research data from universities and nonprofits should be made public.¹⁵ We consider the impact of providing greater public access to data on three areas: first, the incentives of academic researchers to produce data, research and innovation; second, the quality of published research results; and third, the efficiency, transparency and legitimacy of the regulatory process.

A. The Impact of Greater Access to Data on Data Production, Research and Innovation

We can frame the incentive effects of the access rules in classical property rights terms. That framework is valid and provides a useful starting point, although it is incomplete. Specifically, it usually incorporates an inadequate conception of the role of university research in the innovation system in the United States today and ignores important benefits of public access.

¹⁴ ScienceScope (1999).

¹⁵ While we focus on university researchers, our analysis also applies to other researchers at nonprofit institutions who have similar incentives and reward structures to those researchers employed by

Patent policies are based on the idea that some protection of intellectual property is needed to create incentives for innovation. Unlimited patent rights are thought to be inefficient and inappropriate because of the creation of a permanent monopoly. Conversely, no protection fails to provide an adequate incentive to develop new ideas and products.

One can similarly argue that requiring release of data will depress its production. Private data confers an advantage on a scientist by increasing the number and quality of his publications—the fundamental standard by which university scientists are judged. Early publication of data may destroy the ability of researchers to complete their papers and publish before their peers. In some cases, a researcher will not collect the data in the first place.¹⁶

Even limited requirements for public access to data create problems. The proposed OMB regulation states that data may remain private until a study based on the data is published. But many papers are typically published from a major data set. One possible outcome of the OMB proposal is that researchers will delay their first publications until subsequent papers based on the data are also completed. Another possibility is that researchers may save some time and money by collecting more limited data—only those necessary to support one or a few publications—when a more comprehensive initial study could have been done for a modest additional cost. Neither of those outcomes is desirable.

The proposed OMB regulation would restrict expanded access to scientific data collected on research projects supported by federal grants or agreements. It might be thought that such federal support would ameliorate the incentive problem discussed above, but that is not the case. The grants could solve the incentive problem if all that were needed was capital to produce good research. But research differs from standard “widget” production in that it is a highly speculative activity.¹⁷ As is well known in the employment literature, additional monitoring and incentives are often necessary to induce

universities and who are treated equivalently under the OMB-proposed regulation.

¹⁶ Other problems arise when subjects are reluctant to participate in studies if the data are made public. See Kaiser (1998) and National Science Board (1999).

¹⁷ Here, we are concerned only with the incentive or efficiency effects of the public subsidy, not with whether it is fair for the public to retain ownership in projects it pays for. The latter consideration might well argue for modifying policy as suggested by the OMB-proposed regulations, and we consider it in the subsequent section.

an appropriate level of effort from researchers.

The speculative nature of the research enterprise explains why researchers rarely work for simple wage contracts, even in the private sector.¹⁸ The efficiency of such contracts depends on the ability of management to monitor fully and easily the activities of scientists. In fact, inventive activity is particularly difficult to evaluate on a day-to-day basis. As a result, companies routinely negotiate incentive contracts with employees who conduct research—that is, contracts that give employees an incentive to work effectively without requiring extensive monitoring by managers. Incentive contracts reward employees for outcomes, rather than or in addition to, paying a fixed wage. Examples of such rewards include prizes for inventions, sharing revenues from patents, and paying part of an employee's compensation in stock options or other equity.

Universities have adopted some of those reward structures for research that leads to patents or other commercial products.¹⁹ In general, researchers are rewarded for peer-reviewed publications, a somewhat attenuated form of performance-based pay.²⁰ The government could find it difficult and expensive to compensate researchers for lost or less important publications by paying for the preliminary research. But even if that were possible, the monitoring problem remains. University research is at least as difficult to monitor and assess as industrial research and probably more so. If scientists are no longer judged by the quality and number of their publications, the federal government, like businesses who sign ordinary wage contracts with their scientist-employees, will be hard pressed to evaluate grantees and allocate subsequent research grants.

A more important problem with the view that publicly supported research can be easily placed in the public sector is that it ignores the actual structure of research in the United States. Intellectual property rights, when there is public and private cooperation, are inevitably the subject of difficult, intense negotiations. That is so because virtually all important innovations are joint products, depending in a fundamentally nonseparable manner on the activities of university scientists, the federal government, and a host of

¹⁸ Such arrangements hold more commonly for key research employees.

¹⁹ In part, universities are concerned that without some kind of award for patents, university researchers will fail to disclose inventions. Usual arrangements include features such as sharing gross licensing revenues between the university and the inventor.

²⁰ It is performance-based pay only if peer review provides an accurate measure of good science.

commercial firms, banks, venture capitalists, upstream suppliers, and downstream customers. Research joint ventures and longer-term research collaborations among those different actors have become increasingly common and are actively encouraged by the National Science Foundation (NSF) and other federal agencies.

Those agencies support the collaborations for several reasons. First, they want universities to provide matching funds for federal grants. In addition, they subscribe to the principle that interactive, collaborative research is an efficient way to encourage innovation. Such research focuses university scientists on important problems and facilitates technology transfer from universities to industry. Furthermore, that research provides businesses with access to fundamental science at a time when industry, in the face of increasingly competitive, open markets, has significantly reduced intramural investments in basic science. Universities cannot be characterized as stand-alone entities that produce ideas that private companies subsequently commercialize.²¹ As this brief description suggests, identifying the federal component of a project can be somewhat arbitrary. Moreover, requiring early public access to the arbitrarily defined federal part of the project may be problematic for the entire enterprise.

Not surprisingly, one of the more difficult problems that industry, universities and the federal government have confronted in organizing collaborative research concerns the assignment and sharing of intellectual property. Universities, industry, and some federal agencies have been involved in developing contracts and procedures about when and how much data will be released from the projects, how proprietary data will be protected, and how best to use opportunities to file patent applications. Since 1980, the federal government has largely pursued a very flexible policy regarding patent rights to activities that involve federal funds, and the federal courts have liberally interpreted the ability of universities to patent the fruits of scientific activities. Those policies have been important to the research collaborations as they allow for relatively clear definitions of ownership, and enable protection of intellectual property in the related commercial products.

Universities are now very active in patenting and licensing activities, although only

²¹ That view of U.S. universities was probably never correct. But the caricature is less true today than any time in the past thirty-five years and is rapidly becoming less so. See, e.g., Mowery and Nelson (1998), Stokes (1997), and Rosenberg (1994).

a few make significant amounts of money in royalty payments.²² Industrial support for academic research—increasingly provided with some kind of condition that the sponsor have differential access to inventions arising from associated projects, such as the right of first refusal for an exclusive license—constituted 7.5 percent of all academic research in 1995.²³

The commercial activities at universities have been subject to considerable controversy, in part because of claims that traditional university norms of scientific openness have been compromised.²⁴ The provisions for public access to data generated under federal grants may pose serious problems for that fairly substantial and increasingly important research structure. The proposed regulation for data access introduces potential complications and uncertainty, which may reduce the productivity of the university-industry-government collaborations.

This subsection emphasizes the costs of requiring public access to research data. Requirements that data be made public are likely to have an adverse impact on the production of the data. That conclusion is independent of who pays for the research—the federal government or some other entity. Moreover, the efficiency loss remains even if the requirement is narrowed to requiring that data be produced after publication. Finally, a broad requirement for public access to data is problematic and of potentially large consequence for the collaborative activities of universities.

Although we have emphasized costs, public access has important benefits as well. We discussed those that relate strictly to the academic enterprise—that is, production of more research or science—above. In the next subsection we turn to other categories of benefits.

B. *The Impact of Public Access to Data on the Quality of Scientific Studies*

²² According to Zilberman (1999), the top five universities for royalty revenues in 1995 were the University of California system (\$57 million), Stanford (\$39 million), Columbia (\$33 million), Michigan State (\$15 million), and the University of Wisconsin (\$12 million). The University of Virginia ranked tenth in royalty payments and earned less than \$5 million in 1995. See Zilberman (1999).

²³ National Science Board (1998).

²⁴ Cohen, Florida, Randazzese, and Walsh (1997).

The property rights framework starts with the assumption that data contain information more valuable to a scientist if he or she has exclusive access to it, and that policy requires balancing that value against the inefficiencies generated by the exclusivity. If benefits from diffusion are very large, the framework would recommend that data be made public. The more common outcome, however, would be for a balancing of the factors to lead to limited exclusivity, much like patent rights are limited in scope and time. Public access to scientific data, however, is a long-standing norm among scientists, and is based on a different logic. Specifically, a scientific conclusion may be inaccurate, and hence, making it available for scrutiny by peers (or, more generally, the public) will allow the legitimacy of the scientist's conclusions to be tested or evaluated. The scientific norm of "openness" is institutionalized in the policies of many academic journals, which require, in theory, that data used in producing articles be made available to peers wishing to replicate the results.

The university community does not accept the notion that peer-review before publication constitutes a sufficient check on the validity of scientific conclusions. Journals do not take responsibility for the accuracy of papers, beyond attempting to provide reasonable reviewers.²⁵

Indeed, errors in published papers are probably widespread. In the early 1980s, a now-famous study requested the data used in every published paper (with statistical analyses) published in *The Journal of Money, Credit and Banking*, a leading economics journal. The study authors found errors in nearly every paper that were sufficiently serious that the results could not easily be replicated. The authors also found that, notwithstanding both the general norm that data be available and the requirement of the National Science Foundation that data be produced on NSF-funded projects, their requests for data were ignored, denied, or otherwise frustrated in a substantial number of cases.²⁶ Another study in the *British Medical Journal* gave a paper with eight deliberate errors to 420 people to review. For the 221 reviewers that responded, the maximum number of errors detected was five, the median was two, and 16% of the respondents did not find any.²⁷

²⁵ For an interesting analysis of potential biases in the review process see Rennie (1998).

²⁶ Dewald, Thursby, and Anderson (1986).

²⁷ Smith (1997).

As discussed above, university scientists, especially in the biomedical fields, have become involved in activities with potential commercial applications over the past two decades. A growing concern exists over conflicts between the incentives produced by commercial possibilities and the openness norms of science. The recent controversy and discussion within the community suggest that the procedures for ensuring scientific credibility are strained by the possibility of commercializing results.²⁸

Academic enforcement of the openness norm is largely a matter of self-regulation. The seriousness of peer review and the extent to which journals promote access to data appear to be correlated with a discipline's view of the quality of journals. Generally, publication in the best journals is rewarded more highly than publication in the second-tier. But the system, as is common with much self-regulation, relies on the cooperation of scientists. Unfortunately, the structure of the system includes an adverse-selection bias.

The bias works as follows: if a researcher's conclusions cannot be replicated, and if, in fact, the conclusions can be shown to be seriously in error, the researcher suffers embarrassment and loss of reputation. Suppose that researchers have private information about how reliable and robust their own conclusions are. We expect that data are least available from those individuals whose assessment of their work is that it is unreliable. Such studies might be published in peer-reviewed journals, although perhaps in less prestigious ones. Eventually, the data might need to be produced.

But enforcement at many so-called leading journals is clearly lax, so that very substantial delays may occur in the production of data. Thus, it is probably the case that the *Journal of Money, Credit and Banking* project reports error rates that are biased downward, since the subset of authors who did not make their data available to the study probably had higher error rates. That raises the unsettling prospect that problems are most likely to exist among the studies of those least willing to cooperate in the self-enforcement of the openness norm.

An increasing number of the most prestigious journals, such as *Nature*, *Science*, and *American Economic Review*, now require data availability as a condition of publication. Requirements vary by field and by journal. Some require that data be posted

²⁸ Feinberg, Martin, and Straf (1985).

at the journal. In other journals the conditions appear weaker.²⁹

The discussion of the problems with access to data and the need for improved accountability³⁰ lead us to conclude that the problems quantified in the economics discipline broadly characterize access to data in academia. Providing access to data once a study is published is widely recognized as appropriate behavior. It is, however, a standard from which academicians regularly fall short.

C. *The Impact of Access to Data on Regulation*

Our discussion so far has considered the extent to which academia encourages public access to and the sharing of data. The question arises as to whether such access is sufficient for regulatory purposes, as specified in the proposed OMB rule. Considering the benefits of access in a regulatory context changes its desirability in several important ways.

First, one benefit of providing greater access is to increase the likelihood that a mistake is found in an analysis. If the analysis is the basis of a regulatory strategy, then an error may be extremely costly. The particulate matter standard provides a good example. Projected to cost from \$9 billion to \$37 billion annually, that regulation will give the EPA vast new powers to regulate a variety of sources ranging from power plants to barbecues.³¹ In many cases, it will be impossible to meet the standard with known technology. An analysis that casts doubt on the science could have changed the outcome of the regulatory process, saving billions of dollars.³² Moreover, if a proposed regulation entails irreversible costs, the importance of a solid scientific basis and analysis is higher still. Frequently, once a regulation is passed, it becomes more difficult to modify because constituencies grow in support of the regulation, both inside and outside government. Thus, if the strategies involve large expenditures by consumers or businesses, large

²⁹ For example, *Cell* requires that data be made available when there are disputes.

³⁰ George Thurston, a vocal opponent of the Shelby provision and the federally enforced data requirements in general, warns that without greater responsibility for accuracy by journal editors, outside regulation might occur. See Thurston (1998).

³¹ Environmental Protection Agency (1997).

³² Analysis does not always play a critical role in changing regulations because political concerns frequently override objectives related to economic efficiency or scientific merits. Nonetheless, in specific situations, analysis can be helpful, particularly when a regulatory policy is being developed.

bureaucracies to regulate or enforce the regulations, or other significant investments or expenditures, then the value of greater public access could be substantial.

A second benefit of making data widely available is to enhance the transparency of the regulatory process. The process is improved through enhancing its legitimacy, and enforcement is improved if the process by which regulations are formulated is generally perceived as fair. We cannot quantify the value of transparency, nor do we have a good idea of how it trades off against the disadvantages of access discussed above. We do, however, consider transparency to be important and valuable. At a minimum, it suggests that when data are used to formulate important regulations, uncertainties about the appropriate policy should be resolved in favor of access.

Two implications follow immediately from our discussion of the regulatory context. First, for at least some major regulations, public access may be more valuable in a regulatory setting than within academia. The tradeoff between the desirability that data remain in the hands of the researchers and the social interests in allowing access shifts toward access. Second, our analysis of the value of public access to data used in regulatory proceedings applies to any scientific data, not just data derived from federal grants or data collected by researchers at universities and nonprofit institutions.

Currently, regulatory agencies can obtain the underlying data for studies used in setting regulations. In addition, if an agency has obtained the data, it can be compelled to provide them to interested parties under the provisions of the Freedom of Information Act, which exempts certain categories of information. The excluded categories appear to include information that critics of the proposed OMB rule have been primarily concerned about, including identifying medical records and proprietary information. But it does not appear that the existing provisions routinely provide more access than that available within the norms of academia. In particular, agencies appear to have considerable latitude to determine the need to obtain underlying data.

In *Endangered Species Committee of the Building Industry Association of Southern California et al. v. Bruce Babbitt*,³³ the decision of the Department of the Interior to list the California gnatcatcher as a threatened species under the Endangered

³³ Civ. No. 92-6210 (SS), U.S. District Court, District of Columbia, May 2, 1994.

Species Act was set aside by the District Court of the District of Columbia. Raw data underlying a study used by the Department of Interior were not made available to the plaintiffs. The Court states that

where an agency relies upon data to come to a rulemaking decision, it generally has an obligation under the APA to provide such data for public inspection. (852 F. Supp. 32 at 36).

That case appears unusual, however, because the author of the data had published two conflicting studies based on the same data set. Thus, the analysis in the study used by the Department of the Interior was subject to particular skepticism. In general, the courts appear reluctant to abandon deference to agencies as to whether some data were critical to a regulation and had to be produced. For example, three years later the same court, ruling on a different endangered species dispute between similar parties, ruled that

[Fish and Wildlife Service]’s erroneous failure to make available for notice and comment study relied upon in final rule did not rise to level of arbitrary and capricious action. (979 F. Supp. 893 at 893)³⁴

We cannot predict whether the judiciary will choose to interpret the Administrative Procedure Act narrowly or broadly with respect to the production of data. One possibility is that the requirements will be similar to those of the biomedical journals; that is, that production would be required in the event of a specific dispute or contradiction. Thus, it is plausible that the scientific peer review standards will be applied to the regulatory context. Our analysis suggests that this may be inadequate. We have no way of determining whether the current peer review standards are in fact adequate, excessively stringent, or overly lax for academic purposes, although we join others in the scientific community in stressing that peer-review is not a guarantee of accuracy. We can conclude, however, that if the academic standard is appropriate for academic purposes, then it is excessively low

³⁴ *Building Industry Association of Superior California et al. v. Bruce Babitt et al.*, Civ. No. 95-0726 (PLF), U. S. District Court, Washington D.C., 1997.

for at least some regulatory purposes.

4. Conclusions and Recommendations

This regulatory analysis argues that the issue of requiring data to be shared to aid in the development of public policy can usefully be analyzed by considering an economic framework that consists of two parts: the standard property rights analysis used for providing appropriate incentives for producing new research and an analysis of the impact of greater access for both academic and public policy purposes.

The economic analysis demonstrates that the property rights framework is not sufficient for addressing broader problems in which research and data could be used to help inform public policy decisions. The need for sharing data to increase their value for public policy must also be considered when developing appropriate incentives to produce research.

A second conclusion is that traditional peer review done by scientific journals is not adequate for purposes of relying on research for major public policy decisions. That is not to say that peer review is inappropriate for the purposes of journals or scientists but simply that it does not guarantee the validity of the results to the extent that may be appropriate for at least some regulatory settings.

A third conclusion is that the problem of adverse selection could mean that scientists who are reluctant to share their findings are more likely to have errors in their analysis than the average researcher. We do not wish to push this result too far because a researcher may have many reasons for being reluctant to share data. But researchers who are confident in their findings would be more likely to share data than those who are less confident when other things are equal.

A fourth conclusion is that releasing data may have significant efficiency costs. Those costs include possible disincentives to produce data and to delay publication of results. In addition, the effect the requirements might have on efforts of universities, industry, and federal agencies to engage in collaborative research is unclear. Other critics of the proposed rule and the Shelby provision have emphasized more direct costs: the cost of preparing and cleaning the data and the possibility that the rule would be used to harass

scientists instead of for legitimate participatory reasons. We agree that those concerns are not trivial.

In short, the Shelby provision and the proposed amendment to OMB Circular A-110 address a legitimate public policy issue that we think is a problem. The question is what to do about it. Our first six recommendations offer modest suggestions directly related to the Shelby provision and the proposed revisions to Circular A-110. The remaining two recommendations directly address the problem of replicating regulatory analyses, which we think is fundamental to enhancing regulatory accountability.

Recommendation 1: The proposed regulation should be restricted to economically significant regulations developed by executive and independent regulatory agencies.³⁵

Discussion: Targeting economically significant regulations is likely to yield those with the greatest potential efficiency gains. We include independent regulatory agencies because we believe there are also potentially great efficiency gains from allowing public access to data used in promulgating those rules.

The reason for restricting attention to economically significant rules is that the proposed rule will have costs, and we have concerns about how the rule is likely to be used in practice. Thus, we believe that it is important to begin by focusing on those rules where data sharing could have the highest payoff. One potential problem of concern is that the government may have market power in dealing with some researchers who have few other options for obtaining support.

The definition of *economically significant* is critical. We believe that a reasonable cutoff point is to include rules that have an annual economic impact of at least \$100 million.³⁶ Initially, however, a higher cutoff, such as \$500 million could be used to determine whether the rule is likely to have significant adverse consequences.

Recommendation 2: The proposed regulation should be limited to new federally funded grants and agreements.

³⁵ Current regulatory oversight by the OMB does not include independent agencies.

³⁶ For a similar provision, see Executive Order 12866.

Discussion: The government signed grants and agreements with researchers and should abide by the terms of those agreements or at least not impose major additional costs on researchers without compensation.

Recommendation 3: The terms of new federally funded grants and agreements that fall under the new regulation should be restricted to data used in published research in refereed journals that are directly related to the grant.³⁷

Discussion: The government should interpret the scope of the regulation narrowly. Academic researchers will tend to get more benefit if the publication is refereed than if it is not.

Recommendation 4: The researcher should be required to provide as full a rendering of the data set as possible.

Discussion: There is a natural tendency in some research fields, such as economics, to report results that are statistically significant or that will increase the chances for publication, even if they tell only part of the story. Sometimes, potentially important variables are left out of the analysis. A researcher should provide as full a rendering of the data set as possible so interested parties can check the validity of the results.

Recommendation 5: A researcher should get compensated for reasonable incremental administrative costs of producing a clean data set.

Discussion: Under Circular A-110, a user fee would be paid to the agency that fulfills the FOIA request. At a minimum, a researcher should get compensated for reasonable administrative costs.

Recommendation 6: The new rule, if implemented, should be evaluated five years later by an expert panel selected by the National Academy of Sciences.³⁸

³⁷ The definition of *published research* is a delicate issue. That is an important problem but beyond the scope of this analysis. Because we recommend an incremental approach, we would start with refereed journals, but that may be too narrow and should be revisited soon. The rules of publication are changing, particularly with the Internet.

³⁸ Measurement of regulatory impacts may be difficult, but an obvious place to start would be with the

Discussion: The panel should include academics who can evaluate the economic, social, and scientific impacts of the regulation. The panel should provide recommendations for improvement.

Recommendation 7: Congress should create an independent agency or one that reports to Congress directly to replicate findings for economically significant regulations that have an annual economic impact of at least \$100 million.³⁹

Discussion: Agencies with particular missions, such as promoting the environment or protecting public health, may frame analyses in such a way as to further their particular agendas.⁴⁰ To help avoid such bias, a separate agency should be charged with replicating the results of the analysis developed by the agency promoting the regulation.

A law would be needed to create such an agency. Currently, there is a bill in Congress to set up a Congressional Office of Regulatory Analysis.⁴¹ That office could serve that function. Alternatively, an independent agency could be created.⁴² We feel strongly that the responsibility for replication and quality control should not be placed within the executive branch because there is a greater likelihood that the “independent” analysis would be biased.

If Congress objected to setting up a separate agency to perform that function, it should consider asking an existing agency, such as the Congressional Budget Office or the General Accounting Office, to perform the replication function. Staffing would have to be modified accordingly.

FOIA requests themselves. If research suggested that the negative consequences were thought to be significant, then a more modest proposal, such as having an independent agency or group analyze the data and check the validity of the findings, may be a reasonable solution.

³⁹ That solution borrows from Breyer (1993), but it differs from his in that our focus is not on the executive branch. Moreover, his proposal to create a kind of technocratic elite within the executive branch is more ambitious.

⁴⁰ Breyer (1993) refers to that problem as one of “tunnel vision.”

⁴¹ For an analysis of that proposal, see Hahn and Litan (1999).

⁴² Replication could also be done by independent third parties, but we believe that it is best to have the agency appointed by Congress be primarily responsible for the replication.

Recommendation 8: Government should be allowed to use research findings for economically significant regulations only after the findings have been replicated by the agency created by Congress.

Discussion: Replication is a key to ensuring the quality of results. Replication should require a finding by the newly created agency that the basic conclusions drawn from the data are supported by the data.⁴³ The data would be given exclusively for the use of the agency charged with replication.

The requirement that replication be done before promulgation of the regulation is critical. While replication could result in delay, the delay could be minimized if the agency doing the regulatory impact analysis efficiently managed the transfer of the data and models. Exemptions could be provided for situations in which regulations are necessary to respond to emergencies; but such exemptions should be used sparingly. Because of the difficulty attached to changing a poor regulation once it is already in place, the benefits of such replication for improving regulation are likely to be large.

The creation of an agency would impose some additional modest costs on taxpayers. The government would need to pay for the data and replication efforts to ensure the quality of its findings. We believe that such expenditures are well worthwhile for economically significant regulations.

Compensation for researchers would be similar to that described in recommendation 5, except that it would apply to all research that was necessary for replication. Incremental costs should be interpreted broadly to include the costs of a researcher's time in helping to clean the data set.

The recommendations would provide greater access to certain kinds of data used for publications that receive federal funding. That raises an important issue. Are there situations in which the government should provide other information on regulations to the public? And if so, what kind of compensation should be provided to researchers or individuals who produce that information?

⁴³ The replication exercise could be defined narrowly in terms of reproducing the results of the initial research or policy analysis. We would prefer to define it a little more broadly, though that may make it harder to define the conditions under which the data actually support the results.

We think that there will be cases in which the government should provide greater public access to other key information not covered by the recommendations. They include situations where data are central to an analysis supporting an economically significant regulation, but the data are not required to be shared. In such cases, we think that the government should provide compensation to the producers of that information on the basis of its market value, which may be difficult to determine. There are two principles here: first, the public should have greater access to critical data underlying important regulatory decisions to help increase the transparency and legitimacy of decisionmaking; and, second, the government should not be able to take property from individuals without providing reasonable compensation.

The question of how best to address the issue of data sharing for public policy is complicated. We have attempted to provide a useful economic framework for analyzing the issue. Even if one disagrees with our specific recommendations, we hope that the framework is useful.

Although substantial costs and uncertainty may be associated with greater public access to data, our analysis suggests that academic norms alone provide very limited access to scientific data. We recommend improving Circular A-110 by narrowing and clarifying the scope of the proposed regulation. In addition, we recommend that Congress create an agency that would be charged with replicating the findings of regulatory agencies before such regulations could be implemented. Taken together, our recommendations could help lay the foundation for a regulatory system that is more accountable and has more scientific integrity.

References

- Breyer, Stephen G. 1993. *Breaking the Vicious Circle*. Cambridge, Mass.: Harvard University Press.
- Cohen, Wesley M., Richard Florida, Lucien Randazzese, and John Walsh. 1997. "Industry and the Academy: Uneasy Partners in Technological Progress." In *Challenges to the Research University*, ed. Roger Noll. Washington D.C.: Brookings Institution.
- Dewald, William G., Jerry G. Thursby, and Richard G. Anderson. 1986. "Replication in Empirical Economics: *The Journal of Money, Credit and Banking Project*." *American Economic Review* 76(4): 587-603.
- Elliott, E. Donald. 1999. Yale University. Email correspondence, April 30.
- Environmental Protection Agency, Innovative Strategies and Economics Group Office of Air Quality Planning and Standards. 1997. *Regulatory Impact Analysis for the Particulate Matter and Ozone National Ambient Air Quality Standards and Proposed Regional Haze Rule*. Research Triangle Park, N.C.: Environmental Protection Agency.
- Feinberg, Stephen E., Margaret E. Martin, Miron L. Straf, eds. 1985. *Sharing Research Data*, Washington, D.C.: National Academy Press.
- Fumento, Michael. 1997. "Polluted Science." *Reason Magazine* 29:28.
- Hahn, Robert W. 1996. "Regulatory Reform: What Do the Government's Numbers Tell Us?" In *Risks, Costs, and Lives Saved: Getting Better Results From Regulation*, ed. Robert W. Hahn. New York and Washington, D.C.: Oxford University Press and AEI Press.
- Hahn, Robert W. and Robert E. Litan. 1999. *The Regulatory Right-to-Know Act and the Congressional Office of Regulatory Analysis Act*, Joint Testimony before the Committee on Governmental Affairs, U.S. Senate. Testimony 99-1. Washington, D.C.: AEI-Brookings Joint Center for Regulatory Studies.
- Kaiser, Jocelyn. 1997. "Environmental Policy: Researchers and Lawmakers Clash over Access to Data." *Science* 277: 467.
- _____. 1998. "New Law Could Open Up Lab Books." *Science* 282: 1023.
- _____. 1999. "Plan for Divulging Raw Data Eases Fears." *Science* 283: 914-15.
- McGinley, Laurie. 1999 "Scientists Challenge Provision Opening Access to Data." *Wall*

Street Journal, A24. March 1.

Morrall, John F., III. 1986. "A Review of the Record." *Regulation* 10: 25-34.

Mowery, David and Richard Nelson. 1998. *Paths of Innovation*. New York: Cambridge University Press.

National Science Board. 1998. *Science and Engineering Indicators*, NSB 98-1. Arlington, VA: National Science Foundation.

_____. 1999. Press Statement, February 23, 1999. Arlington, VA: National Science Foundation.

Rennie, Drummond. 1998. "Peer Review in Prague." *Journal of the American Medical Association* 280: 214-5.

Rosenberg, Nathan. 1994. *Exploring the Black Box: Technology, Economics, and History*. New York: Cambridge University Press.

ScienceScope, *Science* 283: 1427.

Smith, Richard. 1997. "Peer Review: Reform or Revolution? Time to Open Up the Black Box of Peer Review." *British Medical Journal* 315:759-760.

Stokes, Donald E. 1997. *Pascal's Quadrant: Basic Science and Technological Innovation*. Washington, D.C.: Brookings Institution.

Tengs, Tammy and John Graham. 1996. "The Opportunity Costs of Haphazard Social Investments in Life-Saving." In *Risks, Costs, and Lives Saved: Getting Better Results From Regulation*, ed. Robert W. Hahn. New York and Washington, D.C.: Oxford University Press and AEI Press.

Thurston, George. 1998. "Mandating the Release of Health Research Data: Issues and Implications." *Tulane Environmental Law Journal* 11: 331-54.

U.S. Department of Justice, Office of Information and Privacy. 1998. *Freedom of Information Act (FOIA) Annual Report Fiscal Year 1998*. Washington, D.C.: Government Printing Office.

Zarcoli, Alec. 1998. "Provision in Spending Laws Opens Access to Data Underlying Federally Funded Studies." *Daily Environment Report* 210: AA-2.

Zilberman, David. 1999. *The Performance of Offices of Technology Transfer*, paper presented at the All-UC Group in Economic History Conference, University of California, Berkeley. Berkeley, CA. March 26-28.