

MERCATUS CENTER
GEORGE MASON UNIVERSITY

REGULATORY STUDIES PROGRAM

Public Interest Comment on
Draft Guidelines for Ensuring and Maximizing the
Quality, Objectivity, Utility, and Integrity of Information
Disseminated by the Environmental Protection Agency¹

The Regulatory Studies Program (RSP) of the Mercatus Center at George Mason University is dedicated to advancing knowledge of the impact of regulation on society. As part of its mission, RSP conducts careful and independent analyses employing contemporary economic scholarship to assess rulemaking proposals from the perspective of the public interest. Thus, this comment on the Draft Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information by the Environmental Protection Agency does not represent the views of any particular affected party or special interest group, but is designed to evaluate the effect of the Agency's proposals on overall consumer welfare.

I. Introduction

Data quality is of paramount importance for making sound policy decisions. Without assurances that agency decisions are based upon high quality data, the public will lack confidence in the legitimacy or efficacy of government action. Further, from a consequentialist perspective, the likelihood of success of a given policy is substantially reduced if it is constructed on the basis of questionable data. Effectively, garbage in leads to garbage out.

Even in those instances where government agencies simply act as the distributor of data, leaving the public to draw inferences and take action on the basis of that information, ensuring quality is still an important goal. Information disseminated by government agencies takes on a certain imprimatur. It carries with it a *de facto* agency seal of approval, whether the agency intends such a certification or not. Given this, there is a tremendous responsibility on the part of government agencies, in their roles as public

¹ Prepared by Jonathan Klick, Ph.D., Dorothy Donnelley Moller Research Fellow, Mercatus Center. This comment is one in a series of Public Interest Comments from Mercatus Center's Regulatory Studies Program and does not represent an official position of George Mason University.

servants, to make every effort to either warrant that the information they disseminate or act upon is of the highest quality, or to disclose fully the limitations of any suspect data.

It is with this in mind, that the Office of Management and Budget initially directed federal agencies to develop guidelines for assuring that the data and information they act on and disseminate meet some criterion of quality. It is indeed important that agencies examine and improve their data quality control procedures.

The Environmental Protection Agency asserts that it already has in place significant procedural mechanisms to guarantee the quality of the data it uses and disseminates. Repeatedly, EPA's guidelines refer the reader to extant guidelines that the agency has relied upon for years, suggesting that they meet or exceed any quality control concern suggested by OMB's directive. However, even accepting EPA's assertion that its policies are sufficient to ensure high quality data, its insistence on exempting numerous types of information from such rigorous procedures raises significant concerns and raises the question of whether or not EPA's guidelines truly fulfill the intention of OMB's directive.

It is the purpose of this comment to attempt to reconcile OMB's desire to ensure the quality of data used or disseminated by federal agencies and EPA's desire to maintain reasonable flexibility in its procedures.

II. Statutory Basis for Guidelines

In section 515(a) of the Treasury and General Government Appropriations Act for Fiscal Year 2001,² Congress directed OMB to issue guidelines providing "policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies."

The resulting OMB guidelines instruct agencies to: 1) Issue their own quality guidelines to meet the information goals; 2) Establish administrative procedures by which affected persons may seek and obtain correction of information maintained and disseminated by the agency that does not comply with OMB guidelines; 3) Report to the Director of OMB the number and nature of complaints received by the agency regarding compliance with OMB guidelines.

OMB also provided some basic standards for agencies to consider as they develop their own guidelines. These standards indicated that the guidelines needed to be flexible enough to address all of the agency's information products. Further, OMB recognized that the standard of quality would not necessarily be constant across all information. That is, some types of information might be subjected to an especially high standard of quality. Agencies, however, were directed to consider the costs and benefits of maintaining a given level of data quality. Finally, agencies should attempt to build upon

² PL 106-554; H.R. 5658.

their existing quality control procedures to minimize the administrative burden imposed by the OMB directive.

III. The Need for Principles to Govern the Degree of Flexibility

As indicated in OMB's original directive, despite the importance of ensuring data quality, agencies must be allowed some flexibility in applying quality standards. Implicit in OMB's suggestion that some cost-benefit analysis be applied is the notion that optimal data quality is situation specific. If the cost of securing the highest quality data available far exceeds any differential benefit to be gained, relative to using some second-best data, clearly the public would be harmed by any effort to use the best data. Similarly, if the potential loss of using lower quality data is large, substantial efforts should be made to use and disseminate the best data.

However, suggesting that the required level of data quality is not absolute is not tantamount to allowing complete discretion to an agency. Instead, it places a burden on an agency to justify its use and dissemination of any particular data. In EPA's guidelines, the agency embraces the need for flexibility, without explicitly accepting the attendant burden of justification. In describing when its guidelines apply, EPA asserts, "Factors such as imminent threats to public health or homeland security, statutory or court-ordered deadlines, or other time constraints, may limit or preclude the applicability of these guidelines."

Such broad language eviscerates the purpose of laying out guidelines at all. Clearly, there will always be "time constraints" on an agency action, and it is unclear what constitutes an "imminent threat" in this context. This vacuous language effectively allows EPA to apply its guidelines to information at its discretion, allowing it to vacate the standards without any defined or principled justification.

In order to temper this discretion, OMB has suggested a valuable tool in cost-benefit analysis. By binding itself to a cost-benefit standard, EPA could provide a principled standard by which it could determine whether its quality guidelines apply in a given situation. That is, if EPA does encounter a situation where the costs of applying the guidelines are prohibitively high, then it should face a burden of justifying its decision. If EPA officials have reason to believe that the cost of delay imposed by applying its quality standards is large, they should articulate that, and they should be committed to comparing those costs to the foregone benefits of applying the standards. Data quality is not an absolute good; however there should be a presumption in favor of applying the guidelines that can only be overcome in extreme cases.

The EPA guidelines provide no rigorous standard for determining when the data quality criteria can be ignored. This effectively makes the guidelines worthless as a tool for improving agency behavior. Presumably, before OMB's directive, EPA was not in the habit of seeking out bad quality data for use and dissemination, and there is no reason to believe it will do this in the future. The importance of developing data quality guidelines is to commit agencies to some standard for ensuring a process leading to optimal data

quality. If EPA stipulates that its guidelines essentially apply at its discretion, no such process is achieved.

Operationally, EPA should commit to warranting that every time it disseminates information, the information meets the standards laid out in its guidelines. This should be done explicitly in the form of some kind of written notice accompanying the information. If, for some reason, EPA has decided that the costs of applying the guidelines to specific information are prohibitive, this too should be explicitly disclaimed. The disclaimer should include EPA's analysis justifying its rejection of the guidelines for the given information. By explicitly warranting or disclaiming that the information has been subjected to the quality standards, EPA would be committing itself to a principled procedure upon which the public could rely. Explicit notices of this type will inform the public and allow for important criticism in situations where the agency might have abused its discretion.

IV. The Imprimatur of Quality

One of the perverse effects of a generalized quality standard such as these data quality guidelines is the presumption they create among the public. That is, if there is a presumption of quality for a given set of information, that presumption is likely to flow to information not covered by the guidelines as well. In section 1.3 of its guidelines, EPA lays out numerous categories of information not covered by its guidelines. However, except in a few instances, EPA makes no provision for informing users of information in these categories that the information was not subjected to the data quality guidelines.

EPA is justified in exempting some kinds of information from its guidelines. For example, the cost entailed in examining and certifying all of the information contained within "Internet hyperlinks and other references to information disseminated by others" would, no doubt, be prohibitive. Further, EPA reasonably exempts other information being offered as "someone's opinion rather than fact or EPA's views."

However, beyond some language about "mak[ing] it clear" that these data originate with some source other than EPA, the agency does not commit to any general policy of explicitly informing the public which information has not been subjected to the guidelines. Presumably, EPA expects the public to have a complete understanding of its various distinctions between what is "information" and what is not. In fact, EPA exempts its primary form of communication with the public, press releases,³ from the guidelines. If, because of the guidelines, the public is convinced of the quality and legitimacy of EPA information, it is likely to be more willing to accept EPA pronouncements. However, if

³ "Distribution of information in press releases and similar announcements: These guidelines do not apply to press releases, fact sheets, press conferences, or similar communications in any medium that announce, support the announcement, or give public notice of information EPA has disseminated elsewhere.

there is nothing to guarantee the quality of those pronouncements themselves, EPA allows itself the potential to exploit the public's trust.⁴

The difficulty then is reconciling the fact that EPA cannot be expected to certify all of the information it is somehow related to, directly or indirectly, with the notion that the public will often assume it has. Luckily, the solution is a relatively simple one. The EPA should adopt a policy of explicitly warranting or disclaiming all information, even that falling into one of the categories outlined in section 1.3 (What is not covered by these guidelines) of the guidelines. This warranty or disclaimer should be conspicuous, simple, and standardized. In this way, the public will be made immediately aware of whether or not a given piece of information has EPA certification. Further, there will be no ambiguity involving whether or not EPA is responsible for defending the quality of the information in the face of subsequent criticism.

V. All Data Are 'Bad' Data

All information and data have flaws in the sense that, by undertaking some additional expense, an individual could improve the data in some way. However, the improvement might be infinitesimal in relation to the incremental cost of improvement. Clearly then, in such a situation, improving the data is not in the public's interest.

However, information about the quality of the information itself is often of tremendous value to a consumer. If individuals are aware of the limitations of the information or data presented to them, they are in a better position to make quality choices, and policy makers are in a better position to make informed policies.

EPA should commit itself to disclosing all known data limitations in any information it uses or disseminates. This disclosure should take place at all stages of the data quality process. Specifically, data limitation disclosures should be provided to referees during the peer review procedure, and they should be provided to the public prior to any public comment period. Such disclosures would ensure the highest likelihood that data limitations would be remedied if possible, as well as allowing for the highest level of informed criticism.

Most important, data limitation disclosures should accompany final dissemination of information. With such disclosures, the public and policy makers can accordingly temper their interpretation of agency information.

⁴ A similar point is made by Lawrence Tribe "Trial by Mathematics: Precision and Ritual in the Legal Process," 84 Harv. L. Rev. 1329 (April 1971). In that seminal article, Tribe argued that the imprimatur of objectivity and importance of statistical evidence unduly influenced jurors in the courtroom. The notion was that jurors would be willing to accept statistical evidence without questioning its significance or legitimacy simply because it seemed scientific. That is, jurors will impute the rigor of scientific statistical arguments to legal statistical arguments. In the data quality arena, if individuals recognize that EPA rigorously examines some of its information, they might be likely to impute that rigor to all of EPA's information.

These disclosures should be standard operating procedure for all EPA dissemination. They should include specific descriptions of the limitations, as well as a justification for not remedying the limitation. In cases where alternate data could have been used or disseminated, EPA should explain the analysis leading it not to use the alternate data. In cases where it is cost justified, EPA should even consider discussing any differences in conclusions the use of the alternate data might have generated.

VI. Conclusion

In many ways, EPA's data quality guidelines read more like an outline or proposal than actual guidelines themselves. Effectively, EPA exploits OMB's suggestion that the guidelines be "flexible" to maintain a huge amount of discretion over what information will actually be rigorously examined. While flexibility is important, as optimal data quality is not equivalent to the highest data quality available at any cost, if flexibility is not tempered by rigorously justified analysis or principles, the guidelines might just as well not exist. This comment presents suggestions regarding what form these principles should take.