

Regulatory Checkbook

August 13, 2001

Ms. Brooke Dicksen
Office of Information and Regulatory Affairs
Office of Management and Budget
Washington, DC 20503

Dear Ms. Dicksen:

In this letter I am providing my comments on OMB's "Proposed Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies" (66 Fed. Reg. 34489 *et seq.*, June 28, 2001).

Regulatory Checkbook is a nonpartisan, nonprofit, 501(c)(3) public interest corporation whose mission is to enhance public understanding and involvement in federal regulatory matters. Improving data quality is a critical element of these objectives.

My comments are intended to make constructive suggestions that would improve the final guidance. As currently drafted, OMB's data quality guidelines can be expected to have little, if any, effect on actual agency procedures and practices. Should OMB finalize the guidelines in any manner similar to this proposal, Congress may feel compelled to issue additional and stronger instructions to OMB that deny it much interpretative discretion.

SUMMARY

It is hard to disagree with OMB that agencies should not disseminate data whose quality is below some "basic level." It is also difficult to find any fault with OMB's statement that the more important the information, the higher this basic level of quality ought to be. Problems arise immediately, however, once this concept is made operational. OMB does not provide any useful guidance concerning how one is supposed to judge whether an agency has met this test in any specific case. Moreover, it is not at all clear that OMB will assert its authority under the Paperwork Reduction Act to deter or prevent the dissemination of data that fails to meet this standard of quality.

I also agree with OMB that agencies ought not develop "duplicative" or "contradictory" processes for ensuring the quality of data within its jurisdiction. However, I strongly disagree with OMB's suggestion that agencies should rely on existing processes. If existing processes were adequate, Congress would not have enacted the law OMB is now attempting to implement. Existing agency processes should first be independently audited and reformed where they are not adequate. Similarly, OMB provides no evidence that "well-established information quality standards" or ombudsman provisions created pursuant to Circular A-130 are adequate for data quality

purposes. Further, it's not obvious that agencies have consistently followed these existing procedures even where they satisfy the intent of the new guidelines. Again, Congress would not have directed OMB to issue new government-wide data quality guidelines if existing procedures and current practices were fully adequate.

I am troubled by OMB's attempt to fuzz the distinctions in meaning among the four statutory terms Congress used to characterize "data quality." While I understand OMB's desire for simplicity and the difficulties it faces, I do not believe that it has yet succeeded in this task.

In the remainder of my comments I focus on four specific issues that I believe pose significant problems in OMB's proposed guidelines. These issues are:

- 1 The proposed guidelines are merely hortatory, and therefore will not have any demonstrable effect on agency conduct.
The proposed guidelines have very little content, and therefore represent an inadequate effort to actually guide agency practice.
- 3 The proposed guidelines rely too heavily on existing agency procedures and practices without any evidence that these procedures and practices are adequate (despite an implicit congressional finding that they are not).
- 4 The proposed definitions of "quality," "utility," "objectivity," and "integrity" are incorrectly targeted, incomplete, and potentially inconsistent with law.

For each issue I provide suggestions concerning how OMB might proceed.

ISSUE #1: DRAFT OMB GUIDELINES ARE MERELY HORTATORY

Section 515 of the Treasury and General Government Appropriations Act for FY 2001 (Pub. L. 106-554) requires OMB to issue government-wide guidelines that enhance the quality of data disseminated by federal agencies. However, OMB's draft guidelines consist of hortatory admonitions but no actual requirements for agency action beyond those explicitly specified in the law. Agencies must issue their own guidelines within one year after the finalization of OMB's guidelines, but OMB's guidelines include no requirements that agencies actually follow the guidelines that they issue.

The hortatory nature of OMB's draft guidelines can be easily seen by comparing how often OMB would direct agencies to take certain actions with how frequently it would suggest but does not actually impose any specific requirement. Directive statements are typically preceded by such words as "must" and "shall," whereas suggestive statements are preceded by such words as "should" or "may."¹

In almost all cases where OMB uses the word "must" to indicate that an action is nondiscretionary, the context is one in which OMB is merely repeating statutory

¹ In some cases, "may" indicates permission to engage in an activity or take an action but does not necessarily convey the notion of a suggestion.

requirements. In only one instance does OMB use “must” to prescribe an agency action, and in one other case “must” is used in a way that is implicitly permissive.³ The word “shall” appears only once, and in a context in which OMB is merely repeating statutory requirements. In contrast, the word “should” appears 21 times in the draft guidance and its associated preamble. Virtually everything contained in OMB’s proposed guidance is hortatory and agencies may follow OMB’s recommendations or ignore them.

This seems to be a peculiar strategy for “ensuring and maximizing the quality, objectivity, utility and integrity of information.” To faithfully execute the law, OMB must decide which actions agencies are essential to meet the objectives of the statute; which other actions agencies must take except in extraordinary circumstances the boundaries of which are well-defined *ex ante*; and which additional actions agencies ought to take depending on the situation at hand. Not only does OMB fail to provide any guidance concerning where the lines are drawn among these categories, OMB does not even admit to the categories’ existence.

ISSUE #2: DRAFT OMB GUIDELINES CONTAIN VERY LITTLE CONTENT

Most of OMB’s proposed guidelines either repeat shopworn homiletic admonitions of no great moment or contain virtually no content. For example, OMB’s first guideline states:

Overall, agencies should adopt a high standard of quality ... as a performance goal and should take appropriate steps to incorporate information quality criteria into agency information collection processes. (III.1)

Because no agency would explicitly choose to adopt low quality standards, this guideline serves little purpose except to restate the obvious. That it is also hortatory (prefaced by “should” rather than “shall”), clouded with murky caveats (high quality standards matter *overall*, but not in any particular instance), and aimed a performance *goal* rather than actual *performance*, all serve to undermine the extent to which even the most obvious of criteria contains any tangible meaning. This guideline concludes with an infinitely elastic locution: “Quality is to be ensured and established at levels appropriate to the nature of the information to be disseminated.”

Similar comments can be made with respect to OMB’s second and third guidelines, which also generally restate the obvious and plow no new ground. There is, indeed, little in the proposed guidelines that changes the way in which federal agencies collect, interpret, or disseminate data.

² “Sometimes, in disseminating certain types of information to the public, other information must also be disseminated in order to ensure an accurate, clear, complete, and unbiased presentation.” 66 Fed. Reg. 34492.

³ “[A]gencies must apply these standards flexibly, consonant with existing agency information resources management and administrative practices, and appropriate to the nature of the information to be disseminated.” 66 Fed. Reg. 34490.

ISSUE #3: DRAFT OMB GUIDELINES RELY TOO HEAVILY ON EXISTING AGENCY PROCEDURES WHOSE PERFORMANCE HAS NOT BEEN DEMONSTRATED

OMB's proposed guidelines rely on established agency procedures and practices whose actual performance have not been evaluated and deserve to be considered suspect. Limiting the implementation burden to established practices is tantamount to a ringing endorsement of those practices. An OMB endorsement would directly conflict with the law it is charged with implementing.

In its third guideline, OMB suggests that agencies "establish administrative mechanisms allowing affected persons to seek and obtain correction of information maintained and disseminated by the agency that does not comply with ... OMB guidelines." Leaving aside the difficult matter of discerning when data might fail to comply -- which is not illuminated at all in these proposed guidelines -- OMB then implicitly undermines its own admonition by limiting such mechanisms to those that are "consonant with established agency practice."

If established agency practices were adequate, it seems highly unlikely that the Congress would have directed OMB to issue new government-wide guidelines for data quality. Congressional action suggests that there are significant and material problems with the quality of federal data, but one would never learn that from OMB's proposed guidelines.

OMB might be correct in its belief that some agencies' existing practices already meet data quality objectives, or that these practices could be tinkered with a bit to make them conform. The problem is that OMB provides no evidence in support of either proposition and appears to be attempting to evade its statutory responsibilities. OMB's final guideline must include objective measures that can be applied by independent, external observers to discriminate between those agency practices and procedures that satisfy OMB's data quality guidelines from those that do not.

ISSUE #4: OMB'S PROPOSED DEFINITIONS OF "UTILITY," OBJECTIVITY," AND "INTEGRITY" WARRANT GREATER CLARITY

OMB has chosen not to establish clear definitions for the statutory terms "quality," "utility," "objectivity," and "integrity." "Utility," "objectivity," and "integrity" are quality attributes that could have been independently defined, but it is not clear how the statutory term "quality" should be separately understood. Thus, OMB's preference for identifying three "aspects" that ostensibly underlie the four terms may be superficially appealing. Looking at OMB's three "aspects," however, shows that each is a wordy proxy for "utility," "objectivity," and "integrity," respectively. That is, OMB's use of "aspects" rather than straightforward definitions may be a distinction without a material difference.

Utility

The first of OMB's "aspects" appears to be intended to capture the notion of utility in a way that is much broader than the term "practical utility" as set forth in the Paperwork Reduction Act. In particular, OMB would expand the reach of practical utility to include the utility obtained by the public when federal information is disseminated. In its discussion of this aspect, OMB presumes that the objectives of public data users are as valid as those of the agency. While this may well be true in many cases, there will also be instances in which public data users seek to use federal data for objectives that contravene the statutory purposes for which an agency collected the data. For example, an agency charged with managing or informing the public about a risk may also collect information that is integral for estimating the risk but is not itself a measure of risk. There may be public data users whose political objectives require the misrepresentation of this information, and OMB presumably does not want to confer equal legitimacy on such users as on the agency that collected the data pursuant to its statutory mission.⁴

At a minimum, OMB should revise its discussion of this "aspect" of data quality to make clear that utility (or practical utility) does not encompass uses that are inconsistent with the statutory purpose for which a federal data set was collected. In addition, the utility of federal data is suspect if an agency disseminates it for a use or purpose contradicted by the justification it provided to OMB in the supporting statement for the applicable Information Collection Request (ICR).

Objectivity

According to OMB's characterization of this "aspect" of data quality, data that satisfy the desired minimum quality level also must be "presented within a proper context" and be "accurate, clear, complete, and unbiased." The purpose of providing context is to enable the public to "assess for itself whether they may be some reason to question the objectivity of [data] sources." OMB views independent reproducibility (V.1.B.ii.a) and sound statistical methods with error sources disclosed (V.1.B.ii.b) as the two essential ingredients of accuracy, clarity, completeness, and unbiasedness.⁵

These attributes are useful but they are not comprehensive, nor do they highlight the most essential elements of accuracy, clarity, completeness, and unbiasedness. Context surely matters, but OMB's proposed guidance does not provide any mechanism for determining when data are being employed out of context or what an agency (or others) can or should do about it. For example, data derived from a convenience sample cannot be generalized to any known population of interest and are presumptively biased. Sources of error cannot be fully disclosed because they are not known. Nevertheless, these data might satisfy a high level of data quality for a specific, narrow purpose but be utterly

⁴ The converse also may occur. Some federal data may raise doubts about agency performance. In such cases, utility to the public should outweigh a narrow view of practical utility to the agency.

⁵ It is in this context that OMB proposes its only directive language in the proposed guidelines: "results *must* be substantially reproducible upon independent analysis of the underlying data." See 66 Fed. Reg. 34492, emphasis added.

inappropriate in almost all other situations. OMB's proposed guidance offers no illumination for the difficult task of discerning whether context has been properly addressed. OMB should revise its guideline to clearly state that use or dissemination in a manner contradicted by the agency's supporting statement for the applicable ICR is presumptively out of context and therefore below the required "basic level" of quality.

Reproducibility is generally a desired attribute, but it is a relatively primitive and undemanding one. Independent reproducibility means only that disinterested parties with appropriate skills could perform the same calculations on a data set and obtain the same answers. Federal data that fails this test surely fail to meet minimum data quality standards.

Reproducibility is a minor concern when compared with *replicability*. That is, can independent researchers repeat a data collection using the same or similar methods and obtain the same or similar results? Data that cannot be replicated might not satisfy a reasonable test of objectivity regardless of the quality of the methods used or the transparency with which sources of error are disclosed.

In similar vein, sound statistical methods and disclosed error sources surely improve transparency and the ability of outside parties to evaluate data quality, but these attributes are not sufficient to assure objectivity. For example, a data set obtained using state-of-the-art methods that is severely biased because of sample frame problems would not gain a high level of data quality merely if its fatal limitations were fully disclosed.

If there is a general theme that OMB seems to be trying to make, it is that the dissemination of federal data ought not be accompanied, whether by commission or omission, by embedded policy views and judgments. This may be closer to the Congress' intent when it set forth "objectivity" as an integral element of data quality. OMB should amend its proposed guidelines to explicitly include such a provision.

Integrity

OMB has adequately captured the notion that federal information must be "protected from unauthorized access or revision, to ensure that the information is not compromised through corruption, or falsification." See 66 Fed. Reg. 34492. Each of these sentiments deals with situations in which malevolent actors might seek to advance their objectives by harming the integrity of a federal data set.

This is fine as far as it goes, but it does not go far enough. First, OMB's language does not adequately address the different circumstances attending to primary and secondary data sets. Primary data are original collections and their accuracy, clarity, reliability and unbiasedness are the responsibility of the principal investigators who collect them. Secondary data are derivative of these sources, and provision needs to be made for correct transcription, reporting of sources, and propagation of error. Secondary data collections should not be relieved of responsibilities for accuracy, clarity, reliability

and unbiasedness, and potential contextual problems rise dramatically as distance from the original data increases.

Second, OMB should treat the quality of documentation supporting federal data to be as important as the quality of actual data. My experience using federal data and models has been that even when data can be readily obtained, the documentation for these data (and models designed to utilize them) is missing, incomplete, inaccurate, or impenetrable. Documentation must be considered an integral element of the integrity attribute of data quality. OMB should amend its proposed guidelines to clearly state that dissemination of federal data without full, transparent, and accurate documentation is per se below the “basic level” of data quality required of all federal agencies.

Finally, OMB should integrate concerns about data integrity with existing provisions of the Paperwork Reduction Act. Federal data obtained pursuant to an approved ICR have a presumptive legitimacy that may not be warranted. For example, agencies (or their grantees and contractors) may not have actually followed the research protocols that were integral parts of the applicable ICR submission that OMB approved, and OMB lacks any mechanism for ensuring that research protocols are actually followed. To ensure an appropriate degree of data quality, OMB should explicitly require agencies to document their actual compliance with these protocols as part of the dissemination of data or the use of results. Failure to document compliance with stated protocols should be interpreted as a per se violation of minimum data quality standards. If the documentation shows that compliance with applicable ICR protocols did not in fact occur, agencies should bear the burden of proof that these departures did not materially affect the quality of the data.

Thank you for the opportunity to comment on these proposed guidelines. I look forward to seeing a revised proposal that answers basic questions about OMB’s intentions.

Sincerely,

/s/

Richard B. Belzer, Ph.D.
President

✱