

FOR PARTICIPANTS ONLY

REFERENCE DOCUMENT

DDR/1

17 May 2003

ENGLISH ONLY

ECLAC

Economic Commission for Latin America and the Caribbean

Second meeting of the Statistical Conference of the Americas of the
Economic Commission for Latin America and the Caribbean

Santiago, Chile, 18-20 June 2003

MANAGING DATA QUALITY

This document was prepared by Nancy J. Kirkendall, PhD in Mathematical Statistics, Chief Statistician in the Energy Information Administration, United States. The views expressed in this document, which has been reproduced without formal editing, are those of the author and do not necessarily reflect the views of the Organization.

03-5-288

Abstract

On September 28, 2001 the Office of Management and Budget (OMB) in the Executive Office of the President of the United States formally informed all administrative agencies within the government of new requirements for Information Quality Guidelines. During the year, the guidance was fine tuned as a result of further input and collaboration. The ultimate requirement was that by September 30 of 2002, each agency was required to inform the public concerning the availability on the Internet of new information quality guidelines that it intended to follow. There are three co-responsibilities:

1. Agencies must commit to a basic standard of quality for the information they disseminate;
2. Agencies must develop information management procedures to prevent dissemination of poor-quality data, with peer review playing an important role;
3. Agencies must have an administrative mechanism that allows affected parties to request corrections of information. The burden of proof is on the requester to demonstrate that the information fails to meet OMB or agency guidelines. If the request is denied, there must be an appeals process.¹

This paper will discuss some of the requirements of OMB's Information Quality Guidelines, describe the process of implementation used by the statistical agencies, and provide a summary of the impact of the information quality guidelines activity from the viewpoint of a statistical agency.

Background

The intent of the OMB Information Quality Guidelines is to improve the quality of information that all agencies of the United States Federal Government disseminate to the public. The stated purpose is to "ensure and maximize quality, objectivity, utility and integrity of information disseminated by Federal Agencies". The OMB guidelines give a generic definition of "influential" information, and require agencies to specifically identify which of their information products are influential. Influential information is required to satisfy more stringent requirements for transparency and reproducibility. Finally, the OMB guidelines require agencies to establish a formal process by which affected parties can request correction of information that does not adhere to applicable guidelines. Agencies are required to report annually to OMB concerning any comments received and responses given.

In September of 2001, OMB's Interagency Council on Statistical Policy (ICSP) established a team to pursue a coordinated approach. The ICSP consists of the heads of the 10 largest statistical agencies plus 4 large statistical units within other agencies. These statistical agencies/units represent 12 Departments. The Interagency Quality Guidelines Team prepared a common public announcement describing the quality principles held by the statistical agencies. This public announcement was signed by 13 of the ICSP members, and was published on June 4, 2002. During the process the Team met to agree to the wording in the public announcement, debated questions surrounding the information quality guidelines activity, and shared their agency level ideas and plans for Information Quality Guidelines.

¹ This list is from the publication "Ensuring the Quality of Data Disseminated by the Federal Government", a Workshop Report of the National Research Council, National Academies Press, Washington, D.C. 2003, (www.nap.edu).

The joint public announcement was a statement of the philosophy for the attainment of quality by statistical agencies. It included a list of statistical activities and a generic discussion of the responsibilities a statistical agency/unit assumes to maintain quality. It also provided a single source of internet addresses for each statistical agency/unit's quality guidelines.

The statistical activities defined in the Federal Register include:

- Development of concepts and methods;
- Planning and design of surveys and other means of collecting data;
- Collection of data;
- Processing and editing of data;
- Analysis of data;
- Production of estimates or projections;
- Establishment of review procedures; and
- Dissemination.

One of the Team's informal agreements was that statistical agencies were encouraged to organize their standards around these activities. However, it was agreed that agencies would not strive to agree to common standards at that time.

The joint public announcement also defined the responsibilities a statistical agency has for implementing quality programs saying, "in establishing their information programs, statistical agencies must determine sources of information, decide on the appropriate measurement methods, develop and use appropriate methods of data collection and processing; employ appropriate methods of analysis; and ensure the public availability of the data and documentation. Statistical agencies also assure the widest possible dissemination of information, and seek advice and input from customers and stakeholders."

Individual agency guidelines

Each statistical agency/unit pursued its own approach, and dealt with its parent Department/Agency. Agencies that based their guidelines on a program of Statistical Standards included the National Center for Education Statistics, the Energy Information Administration, the Census Bureau, the Science Resources Statistics Division of the National Science Foundation, the Bureau of Transportation Statistics, and the Bureau of Justice Statistics. Two of these agencies (the National Center for Education Statistics, and the Energy Information Administration), already had statistical standards in place and took the opportunity to refine, revise and implement updated standards. The remaining agencies implemented new standards programs, and one of them, the Bureau of Transportation Statistics, developed Standards for the entire Department of Transportation. Some of the new Standards are referred to as performance principles, or guidelines to good practice —rather than as Standards that *must* be followed. The Bureau of Labor statistics incorporated a partial Standards program with the implementation of Data Integrity Guidelines —focusing only on the integrity or security part of quality.

Some of the statistical units do not have separate guidelines. Instead, their activities are covered by the guidelines of their parent organization. These are the Statistics of Income Division of the Internal Revenue Service; the Office of Research, Evaluation, and Statistics of the Social Security Administration; and the Office of Information within the Environmental Protection Agency.

The Census Bureau, National Center for Education Statistics, Bureau of Transportation Statistics and Bureau of Justice Statistics organized their standards/guidelines around the list of activities in the public notice. The Energy Information Administration and the National Science Foundation used the list of activities in its Guidelines (but did not organize their standards around that list), and most other agencies used other concepts from joint public announcement (particularly the description of the responsibilities of a statistical agency for operating a high quality data program) in their individual guidelines.

Definitions of influential, transparent, reproducible

The terms “influential information”, “transparent” and “reproducible” are key concepts in OMB’s Information Quality Guidelines. OMB defines influential information to be that information that has a “clear and substantial impact on important public policies or important private sector decisions”. Examples of information that is clearly influential are the principal economic indicators that are kept to a strict release schedule, and information that is used in support of regulatory standards that are economically significant.² The Guidelines state that agencies must determine which of their data are influential. For influential information, agencies must assure that it is transparent and reproducible. Transparent simply means that the information, its sources and limitations must be clearly documented. Reproducible means that it is feasible for a replication or reanalysis to be conducted. The Guidelines recognize that information collected under a pledge of confidentiality will not be available to a member of the public for verification. Therefore the Guidelines require only that data based on information collected under a pledge of confidentiality are *capable* of being reproduced.

The mission of the Energy Information Administration, similar to the mission of other statistical agencies, is to “produce high quality, policy relevant information to support public and private decisions”. Hence, statistical agencies strive to produce influential information. Statistical agencies prepare information products for use by the public or other agencies. It is the use of or reliance on information that makes it influential. Statistical agencies strive to ensure that information disseminated is as useful as possible. Statistical agencies/units believe that good statistical practice results in assuring that all information products are transparent and reproducible. While there was great debate within the Federal Government about what information might be counted as influential, from a statistical agency/unit point of view the important thing is to assure that all information disseminated is transparent and reproducible whether it is currently influential or not.

The only two agencies that specifically stated that all their data are influential are the Census Bureau and the Bureau of Economic Analysis, both components of the Department of Commerce. Seven statistical agencies/units specifically state that they assure all information is transparent and reproducible, but do not mention influential. These are: the Bureau of Justice Statistics, the Economic Research Service, the National Agricultural Statistics Service, the National Center for Education Statistics, the National Center for Health Statistics, the Science Resources Statistics Division of the National Science Foundation, and the Social Security Administration. The Department of Energy (hence the Energy Information Administration (EIA)) and the Department of Labor (and hence the Bureau of Labor Statistics (BLS)) designate information used in economically significant rulemakings and embargoed data (including

² A rulemaking is defined to be “economically significant” according to Executive Order 12866, if it is estimated to have an annual effect on the economy of \$100 million, or more or will adversely affect in a material way the economy, a sector of the economy, productivity, competition, jobs, the environment, public health or safety, or state, local and tribal governments.

principal economic indicators) as influential. However, EIA & BLS also state that all information they produce will be transparent and reproducible.

The Interagency Quality Guidelines Team started collaborating on the requirements of OMB's Information Quality Guidelines in September 2001 —well before any other part of government. This early start and the fact that statistical agencies have always been concerned about how to assure the quality of information meant that the Statistical Agencies were poised to make a significant contribution. Recognizing that Departments and nonstatistical agencies had not started considering its Information Quality Guidelines, OMB reminded all Departments and Agencies about the requirements associated with the Information Quality Guidelines in early 2002. By that time the Interagency Quality Guidelines Team had a draft public announcement, one representative had already drafted Information Quality Guidelines for his agency, there had been substantial sharing of ideas as all worked to develop guidelines, and the team had addressed several difficult issues such as characteristics of the comment process, and a taxonomy of information products. This information was widely shared by OMB throughout the Federal Government. Individual members of the statistical agency team also became active participants on Departmental teams (in two cases, chairing the Departmental activity). In addition, members of the Statistical agency team were invited by OMB to participate in OMB collaborations and the Workshops on Information Quality Guidelines sponsored by the National Academy of Sciences.

The member of the Information Quality Guidelines Team representing the Economic Research Service of the Department of Agriculture contributed two particularly useful ideas, and also chaired the Information Quality Guidelines activity of the U.S. Department of Agriculture. He provided significant proposals concerning the mechanism whereby the public would be able to comment if they thought an agency was not following its information quality guidelines. He also proposed a taxonomy, identifying different types of information products that needed to be viewed differently from an information quality point of view. The taxonomy included financial, administrative, statistical, research, and regulatory information.

One of the most significant contributions to the OMB Information Quality Guidelines effort by the Interagency Quality Guidelines Team was the first model for agency Guidelines by the Social Security Administration (SSA). The SSA team was chaired by one of the members of the Interagency Quality Guidelines Team. SSA was the first Agency to prepare draft guidelines, and the fact that OMB shared this document early in the process resulted in a common look and feel to many Department and Agency information quality guidelines, as agencies took the SSA draft to serve as a model.

There were several key benefits to statistical agencies of the Information Quality Guidelines activity. First, it moved the statistical agencies toward commonality, while acknowledging that even the statistical agencies are all different. Second, it encouraged the statistical agencies to push quality principles from within. There have always been efforts to encourage the adoption of quality principles within statistical agencies. However, under pressures of getting the information out, sometimes progress is slow. OMB's Information Quality Guidelines have highlighted quality initiatives and made them easier to sell within an agency. Finally, both the statistical agencies and their parent Departments benefited from the experience the Interagency Quality Guidelines Team members brought to Departmental efforts.

Attachment

Background information

1. All OMB public announcements (Federal Register Notices) may be found at the following link. <http://www.whitehouse.gov/omb/fedreg/>
2. The joint public announcement by the 13 statistical agencies may be found at the following link. http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=2002_register&docid=02-13892-filed.pdf
3. The Office of Information and Regulatory Affairs within the United States Office of Management and Budget provides the following information. http://www.whitehouse.gov/omb/inforeg/agency_info_quality_li
4. The National Academy of Sciences sponsored several workshops concerning the Information Quality Guidelines. Information about the workshops is found at the website below, by clicking on “activities”. <http://www7.nationalacademies.org/stl>

The workshops were also summarized and disseminated in the following Workshop Report, “Ensuring the Quality of Data Disseminated by the Federal Government”, Workshop Report, National Academies Press, Washington, D.C., 2003. (www.nap.edu).