

Revising Statistical Standards To Keep Pace With The Web

Jacob Bournazian

Energy Information Administration, 1000 Independence Ave., SW Washington DC
20585

Abstract

In 2002, the U.S. Energy Information Administration (EIA) revised its statistical standards as a result of the 2002 Information Quality Guidelines issued by the Office of Management and Budget. As EIA increased its use of the Internet for releasing information, issues arose that related to implementing information quality guidelines. New technology also created new capabilities that were not available in the past. EIA reviews its statistical standards annually and applies several criteria to determine whether a statistical standard adequately supports the information quality guidelines. This paper discusses some changes to the agency's statistical standards over the past seven years that were needed to keep pace with changes in technology and business processes.

Keywords: Statistical Standards, Information Quality

Background

In February, 2002, the Office of Management and Budget (OMB) issued final information quality guidelines to ensure and maximize “the quality, objectivity, utility and integrity of information” that Federal agencies release.¹ As part of that final guidance, the Federal agencies were directed to issue guidelines for ensuring and maximizing the quality of information they disseminate. In October, 2002, the Office of the Chief Information Office for the U.S. Department of Energy (DOE) issued information quality guidelines that applied to all DOE components.² These guidelines described the policy and procedures to ensure and maximize the quality, utility, objectivity, and integrity of the information that DOE disseminates to members of the public.

EIA is a statistical agency of the U.S. Department of Energy (DOE) and is one of the fourteen³ statistical agencies or units in the Federal government. EIA collects, analyzes, and disseminates independent and impartial energy information to promote sound policy-making, efficient markets, and public understanding of energy and its interaction with the economy and the environment. Its information products are used by public and private analysts and policymakers to monitor the current status and trends of energy supply and consumption in the United States.

EIA's statistical activities involve developing concepts and statistical methods, planning and designing surveys and other means of collecting, processing, and editing data, analyzing information, producing estimates and projections, and disseminating information in published reports, electronic files, and other media requested by users.⁴ For many years, EIA followed a set of twenty-five (25) statistical standards, and other operational guidance and policies to ensure the quality of its activities.⁵ These standards were based on the potential sources of error that may affect the quality of counts and estimates generated from survey data as well as principals of timeliness and relevance. EIA's statistical standards followed the guidance provided in Statistical Policy Working Paper No. 15, “Measurement of Quality in Establishment Surveys.”⁶

During 2002, EIA formed an inter-office team comprised of representatives from each program office to review the agency's policies and standards to develop a comprehensive set of statistical standards that relate to its business process and that comply and support the principles of the information quality guidelines set forth by OMB and the DOE. The team reviewed agency business practices, available informational technology tools, and OMB guidance as a basis for their review. The standards were written to ensure and maximize levels of quality that were appropriate to the nature and timeliness of the disseminated energy information and available software and related information technology resources. The standards also promoted the transparency about information and methods using the latest software and information technology tools available at that time. In September, 2002, the team issued their report and the agency adopted a set of 32 statistical standards to ensure the quality (i.e., objectivity, utility, and integrity) of information disseminated to the public.⁷ Many software developments and new information technology tools that were developed since 2002 have significantly improved a user's access to metadata and documentation.

Objectivity means ensuring that the substance of the information is accurate, reliable, and unbiased, and the information is presented in an accurate, clear, complete, and unbiased manner.⁸ Most of the information that EIA releases is based on surveys sponsored by EIA. EIA-sponsored surveys are conducted using methodologies that are consistent with generally accepted professional standards for all aspects of surveys including frame development, statistical design, questionnaire design and testing; data collection, and control of sampling and nonsampling errors through nonresponse analysis, imputation of missing data, and development of weights, adjustments, and variance estimates, as appropriate.⁹ EIA surveys and information systems are documented and explanatory materials are made available for EIA information products.

Utility refers to the usefulness of the information to its intended users.¹⁰ The main efforts by EIA to ensure the utility of information products include: establishing and adhering to data release schedules to ensure high quality information is released in a timely manner and users have equal access; providing information products in ways that are accessible to a broad range of information users; providing explanatory materials to assist users in understanding and interpreting the information; analyzing customers' information requirements using specific tools, such as customer surveys; and assessing the products themselves to help ensure timeliness and relevance. Integrity refers to the security of information from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification. EIA employs a group of information technology controls and data handling procedures that provide appropriate safeguards against unauthorized access to its data systems.

Process of Revising Statistical Standards

Beginning in 2005, EIA did a comprehensive review of its statistical standards to determine if any standards needed revising. Since, 2005, EIA reviews its statistical standards on an annual basis. One criterion used in the review process for identifying standards to revise is how has the agency's business processes relating to the collecting, processing, editing, producing, and disseminating data changed over time. The second criterion is how have the information technology tools that support the agency's business

processes changed. The third criteria is are the business practices that the program offices follow consistent across the agency and what, if any, are the best business practices that support a particular activity. The fourth criterion is whether revising a current statistical standard promotes the principles of objectivity, utility, and integrity using best business practices and available informational technology tools across the agency.

The revision of a statistical standard is generally confirmatory in nature. When a standard is identified as needing revising, a review of all practices within the agency is done. The best practices within a business process are identified and ideas are shared across the agency on what is feasible, what are the obstacles, and what resources are needed for an office to adopt and follow best practices that promote information quality. Staff input is solicited from all levels, (senior management, mid-level managers and staff) of the agency. This process maximizes “buy-in” from staff and management before any proposal for revising a standard is circulated for review. The end result is the adoption of a revised standard that incorporates best practices within the agency’s business processes in the use of the latest information technology tools, and a staff that is motivated to apply the requirements of the revised statistical standard.

Statistical standards are generally revised one at a time. Changes in the technology for releasing information over the web may require revisions to components of several different standards. For example, the use of hyperlinks is a valuable tool for providing access to source data in graphs as well as other metadata for referencing model specifications, sampling methodology, and data definitions. The use of hyperlinks on the web pages improved user accessibility to metadata information which led to changes to the statistical standards on data utility, graphs, and dissemination of derived estimates. In the past, revising a standard involved a complex communication process to revise one standard and modifying several standards at the same time was difficult to coordinate. In the past, this meant that identifying changes in statistical standards that are driven by changing web technology could take several years to complete the revision process for all the standards that are affected.

EIA uses an electronic dashboard for circulating proposed revisions to statistical standards. The dashboard contains folders and all staff are encouraged to write their comments in the folder. The staff have read and write access privileges and only the agency’s standards officer has access privileges to delete or modify content in the folder. This electronic dashboard is a fast and efficient mechanism for promoting internal discussion among the staff. It allows each person to read everyone else’s comments and add to the dialogue. Using the electronic dashboard to revise standards results in collecting more feedback from staff and management and shortens the review period.

The Need to Revise Statistical Standards

During the past 10 years, EIA phased out the printing of hardcopy publications. It currently relies almost entirely on its web site available at www.eia.doe.gov for disseminating energy information. Some information is printed on an “as needed” basis depending on the information product and user needs. As the use of hardcopy reports declined, publications and reports were released in various electronic formats such as Hyper Text Markup Language (html), Text or flat files (txt), Portable Document Format (pdf), and Microsoft Excel spreadsheet files (xls). Providing information in multiple formats promotes access to information by meeting the needs of different user groups who are interested in the same information.¹¹ Each format has its pros and cons for

meeting user needs and also presents different challenges in terms of presenting information on source data, notations for revisions, documentation, and notices concerning release dates and times, and other metadata issues.

The development of online query systems has been an important change in releasing statistical information for EIA over the past ten years. Information from different fuel groups is integrated and released by classifications such as production, imports, and consumption. EIA releases approximately 70% of its survey data and all of its international data through online query systems. The development of online query systems required standardization of data codes across survey systems. The development of integrated databases that draw on multiple data sources across all fuel groups was instrumental in designing state level profiles that contain detailed data on the different energy markets within a state.

As the agency relied on the web as its primary dissemination tool, EIA staff were trained on how to write for the web. The text on a web site generally supports three objectives: first, draw the reader in; second, convey information to the user; and third, help the reader adjust and navigate through information on the page.¹² Short sentences replace long sentences and simple words are substituted for difficult or obscure words.¹³ Web users skim and scan text rather than read the words closely. They generally do not have the patience to read through a lengthy report. Web users seek to have their questions answered in the first few sentences.¹⁴ Reports became shorter with less text and graphs near the text became smaller to dramatize or summarize the main points. Hyperlinks were embedded in the text and graphs to direct the user to additional supporting information, other graphs, source data, documentation, and other metadata. The size of the graphs shown on a web page decreased as the number of graphs on the web page increased.

Customers have also become increasingly sophisticated in accessing information from federal agency web sites. The development of automated data retrieval programs, i.e., “robots” posed new challenges to federal agencies that pledge equal access to the public for their most popular information products. Robotic activity absorbs band width as they ping agency servers at a rate of over 50 times a minute searching for the latest and most current files containing market sensitive information.

Prior to the development of web sites, federal statistical agencies generated performance measures on data utility primarily based on the size of subscription mailing lists, number of media citations, number of telephone calls and letters. As the web became the primary tool for releasing information, new web based metrics became increasingly important for measuring information utility. The number of visits, page views, time spent viewing a page, keywords used by search engines are some examples of these web metrics. The software for generating web metrics had slow response time and provided a limited set of web analytical metrics. Over time, software has become more sophisticated in the metrics that can be generated from large volumes of data in relatively short response times. Web analytics have become one of the most important measures for determining what information products are popular and identifying categories of users who access information. Web metrics may also be used to determine what information products are no longer useful. Agencies post files to their web sites and rarely look back to remove obsolete or stale information. Over time, this can cause search engines to work less efficiently when a user is looking for current data or the latest analysis report.

Some federal agency websites allow users to rate their satisfaction with a web page by entering a score of 1 to 5 stars. In 2010, EIA expanded its communication outreach by using social media services such as “Twitter” and “Facebook” to collect feedback from users on various energy issues, generate dialogue on energy issues, and inform users on the availability of the latest energy information from EIA’s website. The application of third party websites to collect information from the public is another practice that is expanding among the federal agencies for a range of activities from hosting a conference, collecting feedback on a specific issue, or other measurement issues. The rapid pace of change in releasing information over the Internet continues to cause revisions to statistical standards for maintaining a high standard of informational quality.

Software companies continue to market updated versions of their programs by offering more features and easier navigation. The growth of Application Programming Interfaces over the past 5 years creates new capabilities for generating time line charts, graphs and maps and reduces the time for producing a new information product. (An Application Programming Interface (API) is an interface implemented by a software program which enables it to interact with other software).¹⁵ More agencies are beginning to include some type of social networking interaction on their websites such as “blogging” on their web sites. Guidance is needed from the statistical standards on the appropriate use of these new tools that supports information quality principles.

Revising EIA’s Statistical Standards

The changes in business processes and the advancement of software tools over the past 10 years prompted EIA to revise six statistical standards. Other standards were revised during this time but mainly due to new legislative requirements and other issues unrelated to the Internet.

EIA Standard No. 10 on “Revisions” and EIA Standard No. 13 “Survey Data Evaluation” were updated to provide guidance on new tools for showing when statistical estimates are revised in data releases that use multiple file formats and standardizing coding for preliminary and revised data. These changes promoted the objectivity principle by providing updated guidance on presenting information in an accurate, clear, complete, and unbiased manner. Standard No. 12 on “Policy For Releasing Information” was revised to provide guidance when the use of software was appropriate for controlling the release of popular information products and for new legislative mandates relating to controlling access to sensitive unclassified information. EIA has one product that is designated as a principal economic indicator, and several that are market-sensitive. Information is considered market-sensitive by EIA if the agency reasonably expects that a product’s public release may have a substantial effect on the pricing of energy products traded in financial markets. Other products are popular, such weekly retail motor gasoline prices, but are neither market-sensitive nor a principal economic indicator. Guidance was needed on structuring appropriate release procedures and data safeguards depending on the nature of the information product. These changes supported the quality principles of integrity by providing updated guidance on securing information from unauthorized access or revision and the principle of transparency.

The updates to EIA Standard No. 14 “Dissemination of Information Based on Reported and Derived Data” promoted the principles of transparency and reproducibility by providing updated guidance on requiring that metadata information regarding sampling, imputation, and other adjustments be accessible directly from the information product.

This provides flexibility for the program office to either use hyperlinks to the information or the traditional footnotes or reference to explanatory notes. The requirements in the revised standard also led to standardized icons for accessing source data and documentation on web pages to promote brand recognition among users for accessing metadata information.

EIA Standard No. 17 “Information Utility” promoted the data utility principle by requiring offices to use web metrics such as web visits and page views to assess the utility of information release. Program offices wrote customized reports to generate extensive web metrics on their most popular information products. Agency software already generates a list of the 100 most popular information products from each office based on the number of visits to the web page where the information product is shown. The standard also requires offices to review, on a quarterly basis, the keywords used in search engines to find info products on EIA web pages. This activity creates a mechanism for offices to monitor user traffic on the web and review what information users are interested in on a periodic basis. If the web traffic signals a rise in the use of a keyword (or related cluster of keywords) by users accessing the web site, then this is an important signal to the agency that the utility for a specific data category is increasing.

EIA Standard No. 17 also requires offices to use web metrics to assess information that has either lost utility or has declining utility. The information products listed in the bottom quintile of the list of information products released by each program office must be reviewed each year to determine whether to retain that product in current or historical folders on the web for another year. Search engines only look at files in current and historical folders and do not search archive folders. So search engines work more efficiently by removing stale information from the primary navigation, i.e., current and historical folders, and allow the most current information to be shown at the top of the search results. Information products removed from current and historical folders are moved to archive folders, an archival section of the website, so the products are still electronically available to users upon request.¹⁶ The revised standard creates a process for archiving out-of-data information but “keeping it live” for users to access if necessary. The program office should attempt to identify who are the customers and the justification for not archiving the least popular information products as part of their review process.

Current graphic art software applications provide users with more capabilities for illustrating trends and patterns. EIA Standard No. 25 “Statistical Graphs” was revised to permit greater flexibility and creativity in generating graphs as long as the source data was directly accessible from the graph. This standard was revised to require that the information used to interpret the graph should either be visible from the web page without scrolling, accessible from the web page through a link, or on the same page in the printed product where the graph appears. This required action supports the reading performance of users. Users can read from a monitor at the same speed as reading from paper.¹⁷ A user’s reading performance substantially declines when they need to look at information on one page and remember it while reading the information on a second page.¹⁸ One rule of scrolling is that “if people expect something to be in a particular place, they will not look for that item elsewhere or scroll to find it.”¹⁹ Most agencies design their web pages for limited scrolling. However, the placement of critical elements of information on the page affects a user’s need to scroll.²⁰

Through the use of hyperlinks to source data, graphs can dramatize or summarize the factual statements in the text while providing comprehensive source and metadata

information. Links are used to provide definitions, descriptions to clarify technical concepts and reference the source data. A common page layout would position the graph so the text supporting the graph may be viewed without scrolling vertically or horizontally.²¹ Horizontal and vertical scrolling is a slow and tedious way to view an entire screen.

EIA Standard No. 25 was also revised to require that web graphs contain alternative “ALT” text tags that describe and summarize the graph for use by screen readers. This requirement promotes data utility for web users with visual impairments by allowing the visual and audio assistive technologies to work more efficiently so that a user can understand the hierarchy of the information shown on a web page.

Conclusion

Statistical agencies need to review and assess how releasing information over the web has affected their business practices. Statistical standards need to be updated to support appropriate web design, maintain high standards of information quality and satisfy user’s expectations. Statistical standards are an important tool for agencies to maintain a high level of quality for the information they release. Updating the standards is an ongoing process that requires regular review. The fast pace in which the Internet continues to develop has been a motivating force for revising statistical standards. New issues continue to arise each year such as the use of star ratings on pages and the increased use of social services media and third party web sites by federal agencies.

Relying upon the Internet to release statistical information is a common trend across all federal statistical agencies. OMB should review and revise, as appropriate, the statistical standards that implement the 2002 Information Quality principles to provide guidance on web related issues.

One recommendation is to require all metadata information be directly accessible from the same web page where the information product appears. This is a simple requirement to satisfy. Fifteen years ago this may have been too burdensome to comply. Today, with all federal agencies using the web to release information and available resources, a user will find a higher information quality experience if all federal agencies applied the same requirement.

A second recommendation is that OMB should require statistical agencies to use some web metrics to assess data utility. The number of visits, keyword searches, and page views are common analytical metrics that all statistical agencies currently use. However, there is no guidance on using web metrics in the 2002 standards. Another consideration is whether some responsive action should be required if the web metrics show a sudden increase or decrease. The standard concerning response rates requires certain actions be taken when survey responses are less than 80%. How far should we go in requiring agencies to take some action when assessing data utility?

The federal statistical agencies reliance on releasing information over the Internet has created a strong need for statistical agencies to review and revise their statistical standards on specific web related issues to maintain quality and promote the principles of objectivity, utility, integrity, and transparency. By updating statistical standards concerning the web related issues, the business processes of the agency work more

efficiently, agency employees understand the relevance of the statistical standards to their work, and users experience a high level of quality when accessing information.

References

¹ “Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies,” Federal Register Vol. 67 No. 8452, February 22, 2002. These guidelines were in response to section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554, 114 Stat. 2763).

² “Final Report Implementing Office of Management and Budget Information Dissemination Quality Guidelines,” Office of Chief Information Office, Department of Energy, Federal Register Vol. 67, No. 194, p. 62446. October 7, 2002.

³ “Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA),” Federal Register, Vol. 72, No. 115, p. 33362 at 33368. June 15, 2007.

⁴ “Federal Statistical Organizations’ Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Disseminated Information,” Federal Register Vol. 67, No. 107, p. 38467. June 4, 2002.

⁵ Office of Management and Budget, Federal Committee on Statistical Methodology Statistical Working Paper No. 26, “Seminar On Statistical Methodology in the Public Service,” 1997 -- Parts 1 and 2 (NTIS PB97-162580), Session No. 6, “Data Quality at the Energy Information Administration: The Quest for a Summary Measure,” by Renee Miller, p. 145 at 150.

⁶ Office of Management and Budget, Federal Committee on Statistical Methodology Statistical Working Paper No. 15, “Measurement of Quality in Establishment Surveys,”

⁷ EIA Information Quality Guidelines are available at http://tonto.eia.doe.gov/abouteia/information_quality_guidelines.cfm. EIA’s statistical standards are available at http://tonto.eia.doe.gov/abouteia/eia_standards.cfm

⁸ “Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies,” Federal Register Vol. 67 No. 8452 at 8455, February 22, 2002.

⁹ EIA Information Quality Guidelines, http://tonto.eia.doe.gov/abouteia/information_quality_guidelines.cfm p. 1

¹⁰ “Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies,” Federal Register Vol. 67 No. 8452 at 8456, February 22, 2002.

¹¹ Koyani, Sanjay J., Bailey, Robert W., Nall, Najice R. “Research-Based Web Design and Usability Guidelines.” Government Services Administration (2004). p. 161.

¹² Brinck, Tom., Gergle, Darren., Wood, Scott D., “Usability For The Web,” Morgan Kaufmann Publishers, (2002), p. 247.

¹³ Id. p. 247.

¹⁴ Redish, Janice. “Letting Go of the Words,” Morgan Kaufmann Publishers (2007), p. 3.

¹⁵ Application Programming Interface, (n.d.) Retrieved June 2, 2010 from Wikipedia http://en.wikipedia.org/wiki/Application_programming_interface

¹⁶ Brinck, Tom., Gergle, Darren., Wood, Scott D., “Usability For The Web,” Morgan Kaufmann Publishers, (2002), p. 143.

¹⁷ Koyani, Sanjay J., Bailey, Robert W., Nall, Janice R. “Research-Based Web Design and Usability Guidelines.” Government Services Administration (2004). p. 20.

¹⁸ Id. p. 20.

¹⁹ Nielsen, Jakob., Loranger, Hoa. “Prioritizing Web Usability.” New Riders. (2006) p. 324.

²⁰ Id at p. 324.

²¹ Id. p. 67.