

Benchmarking the Performance of Statistical Agencies

Marilyn M. Seastrom

National Center for Education Statistics
1990 K St., NW, Washington, DC, 20006
marilyn.seastrom@ed.gov

Abstract

The Federal government has used performance measures for several decades. This paper traces the history and development of the current measures that have their roots in the 1993 Government Performance Results Act. A performance measure framework developed by a working group of the Interagency Council for Statistical Policy is described. That framework is used in an analysis of the implementation of performance measures across statistical agencies. Finally, the recent emphasis on the need for improved performance measures is discussed.

Background: Why Talk About Performance Measures Now?

Performance measures have been with us in government in one form or another going back a number of decades. Our current performance standards date to the August 1993 enactment of the Government Performance Results Act of 1993 (GPRA) (Public Law no. 103-62, 107 Stat 285). In enacting this law, the Congress found (emphasis added) that

- (1) **Waste and inefficiency in Federal programs undermine the confidence of the American people in the Government** and reduces the Federal Government's ability to address adequately vital public needs;
- (2) Federal managers are seriously disadvantaged in their efforts to improve program efficiency and effectiveness, because of insufficient articulation of program goals and inadequate information on program performance; and
- (3) **Congressional policymaking, spending decisions** and program oversight are **seriously handicapped by insufficient attention to program performance and results.**

Agencies were directed to develop strategic plans before fiscal year 1998 and annual performance plans by fiscal year 1999 such that each program activity in the budget of an agency had objective, quantifiable and measurable performance goals (unless authorized to have an alternative form). The performance plan for each program activity plan was to include a description of the processes, skills, human, capital, information, or any other resources required to meet the goals. Additional requirements included performance indicators to use in measuring or assessing the outputs, service levels, and outcomes; a basis for comparing actual results to the performance goals; and a method for verifying or validating the measures..

As agencies moved toward implementation of GPRA, both OMB and GAO released reports identifying challenges agencies were facing in implementation. For example, GAO identified problems in coordinating overlapping programs, measuring how an individual program activity influences results when there are intervening factors not under the program's control, improving performance data, moving from a process-orientation to a results-orientation, and linking performance plans to agency budgets.

Work progressed implementing GPRA over the next several years through the end of the Clinton administration and into the start of the Bush administration. Then in fiscal year 2002, after concluding that the implementation of GPRA "fell far short of its authors' hopes" (FY2004 Budget, p. 49), due to performance measures that were deemed to be "meaningless, vague, too numerous,..." (FY2004 Budget, p. 49), initial experiments with a new performance rating tool were undertaken. Thus, the fiscal year 2003 budget included assessments of 130 Federal programs, with ratings ranging from "effective" to "results not demonstrated." During the next year, another 234 Federal programs were assessed. The new rating tool, the Performance Assessment Rating Tool (PART), was formally presented in the fiscal year 2004 budget.

The tool consists of approximately 30 questions distributed across four critical areas: purpose and design, strategic planning, management, and accountability. The answers result in a numeric score, ranging from 0 to 100, in each of the four areas. The four scores are ultimately combined to produce an overall qualitative rating of Effective, Moderately Effective, Adequate, Ineffective, and Results not Demonstrated, with the last category assigned to programs that do not have acceptable performance measures, or have not collected performance data. To ensure sufficient time for each agency's evaluation, PART was introduced with a plan to review approximately one-fifth of all Federal programs each year, with a completion date for the first round of reviews to coincide with the 2008 budget submission.

It is relevant today, to note that in the 2004 roll-out of PART, it was identified as a work in progress with limitations and shortcomings. Included among these were needs for increased consistency in the scoring of similar answers, minimizing the subjectivity in the interpretation of answers to PART questions, more work defining "adequate" performance measures, measuring progress towards results (before the results are realized), institutionalizing program ratings as a routine component of the annual budget process, assessing interrelationships between programs and their results, and increasing the use of information in budget decisions.

In the summer of 2003, OMB issued a report on "Performance Measurement Challenges and Strategies" that was viewed as a complement to the PART guidance document. This report was a result of a spring 2003 workshop on performance measurement that was co-sponsored by OMB and the Council for Excellence in Government. That report first identified and discussed key definitions and concepts that are used in the PART and then went on to six common performance measurement problems—the difficulty in measuring a program's outcomes, accounting for the fact that the program is one of many contributors to the desired outcome, results will take many years to achieve, the program is intended to deter or prevent specific behaviors, the program is multipurpose by definition and includes a range of activities, and the program is intended to be administrative or process oriented.

A number of the problems and limitations identified in the PART roll-out in the FY2004 budget and in the list in the 2003 OMB report issued to complement guidance for PART submissions repeat those identified in the GAO list of measurement problems associated with GPRA.

With the 2009 beginning of the Obama administration, attention has turned again to performance measures. In the Fiscal Year 2010 Budget that was released in May 2009, the President announced plans underway to form a team within the White House to work with agency leaders and OMB "to improve results and outcomes for Federal Government programs while eliminating waste and inefficiency" (FY2010 Budget, p. 39). The head of the team is the Chief Performance Officer (CPO), who is charged with the responsibility for working with Federal agencies "to set tough performance targets and hold managers responsible for progress" (FY2010 Budget, p. 39).

The original PART implementation plan called for all Federal programs to be assessed by the time of the fiscal year 2008 budget submission; and the last of the 16 statistical programs was assessed in fiscal year 2007. At this point, assessments have been conducted for all of the ICSP member agencies and their programs that were identified for the assessment (there are three ICSP members that are programs that were assessed as a component of a larger organizational unit that are not included in this analysis). With the transition to a new administration, it is likely that there will be changes in the Federal government's approach to performance management. As a result, this is a good time for the statistical community to stop and take stock of where we are on our performance measures.

How Did Statistical Agencies Respond to PART?

The Federal statistical agencies and statistical units represented on the OMB sponsored Interagency Council for Statistical Policy (ICSP) collaborated on the development of a framework for an initial set of common performance standards. The framework includes two general areas of focus: Product Quality and Program Performance. Each area of focus is subdivided into three dimensions.

Product Quality includes measures of relevance, accuracy, and timeliness. Relevance is defined as "qualitative or quantitative descriptions of the degree to which products or services are useful to users and responsive to users' needs." Accuracy is defined as "qualitative or quantitative measures of important features of correctness, validity, and reliability of

data and information products measured as degree of closeness to target values.” Timeliness is defined as “qualitative or quantitative measures of the timing of information releases.”

Program Performance includes measures of cost, dissemination, and mission achievement. Cost is defined as a “quantitative measure of the dollar amount used to produce data products and services.” Dissemination includes “qualitative or quantitative information on the availability, accessibility, and distribution of products and services.” Mission achievement is defined as “qualitative or quantitative information about the effect of, or satisfaction with, statistical programs.”

This framework was used by participating agencies as they developed performance measures to be used in their individual assessments with the Performance Assessment Rating Tool (PART). This review provides a summary of the outcomes of these efforts. By January of 2009, there were 16 separate performance assessments among the 10 statistical agencies that were assessed using the PART¹. Eight agencies have one overall assessment that covers all programs [Bureau of Economic Analysis (BEA), Bureau of Justice Statistics (BJS), Bureau of Labor Statistics (BLS), Bureau of Transportation Statistics (BTS), Energy Information Administration (EIA), Economic Research Service (ERS), National Agricultural Statistics Service (NASS), and National Center for Health Statistics (NCHS)]. The National Center for Education Statistics (NCES) has an overall assessment for the statistics programs and a separate assessment for the National Assessment of Educational Progress (NAEP); and the Census Bureau has separate program assessments for six programs—Decennial Census, Demographic Statistics, Intercensal Demographic Estimates, Survey Sample Design, Current Economic Statistics and Census of Governments, and Economic Census. Each program was initially assessed between fall 2003 and fall 2007. At their initial assessments, 9 of the agencies or programs were rated Effective (the highest rating), 6 were rated as Moderately Effective, and 1 was rated Results Not Met (Chart 1).

Each of the initial assessments included improvement plans with action items for follow up. In subsequent annual PART evaluation cycles, the status of completed improvement plans is summarized and the status of ongoing improvement plans is updated for each agency or program previously assessed. In a number of cases the improvement plans resulted in the addition of new performance measures. For example, in 2006 NCES submitted a review of the performance measures used by other statistical agencies, with a review of their appropriateness for NCES, as directed in their performance plan, and in 2007 NCES submitted a related proposal for additional measures for use in assessing program performance. The end result of this analysis was the addition of performance measures for NCES on response rates/nonresponse bias analysis, measures of cost per case, web statistics, education associations’ secondary use and distribution of NCES data, and participation in the American Consumer Satisfaction Index Survey. Several other agencies’ assessments also note that new performance measures were added after their initial performance assessments. All 16 assessments show an update as of January 9, 2009.

Although each of the agencies and programs assessed have measures fitting the described framework, the specifics of the measures vary across agencies. In addition, within each agency some of the indicators may be used internally and are discussed in the detailed assessment, while others may be performance measures that are updated annually in the Performance Assessment Rating Tool (PART). To provide a snapshot of the shared measures and their similarities and differences, this analysis will focus on the performance measures, with supporting information drawn from Questions and Answers in the detailed assessments.²

The detailed charts showing the performance measures used in each agency’s or program’s PART, are available upon request from the author.

Product Quality

While there are many aspects to quality, the three that statistical agencies identified as essential are outcome-oriented and are key to usability. Some of the measures cited in the various agencies’ and programs’ performance plans are qualitative—describing specific actions taken, while others are quantitative tied to specific measurable performance indicators.

¹ Statistical units in NSF, IRS, and SSA did not receive separate performance assessment. In contrast, Census has six separate programs assessed and NCES has two performance assessments. The results have been combined across programs to report for Census and NCES.

² Some of the detailed internal measures were discussed in Agency Strategic Plans and other Agency documents cited in the assessments.

Relevance of data products and analytic reports is monitored through ongoing contacts with data users, surveys of customer satisfaction with product content, customer use of products, and through a professional review process. Product relevance may also be demonstrated through product improvements, especially in response to customer input. Several agencies have developed composite relevancy measures that incorporate multiple aspects of product quality, such as the percent of output, timeliness, accuracy and long-term improvement targets achieved.

Each of the agencies and programs reports one or more forums for ongoing interaction with their constituent communities. In terms of quantitative measures, 9 of the 16 agencies and programs reported using customer satisfaction surveys to monitor various features, such as the usefulness, relevance, timeliness, accessibility, comprehensiveness, and quality of products and services. For example, NCES and NAEP use a measure from the American Customer Satisfaction Index survey that measures the percent of customers who would recommend the agency/program to others and who would rely on the agency/program in the future.

One agency, BTS, cites a target number of visits to the agency's data warehouse as a performance measure. In addition, BTS reports focusing on tracking and increasing the number of BTS products that contribute to the accomplishment of important, established goals, and on increasing the coverage of legislative mandates in its annual report. BTS also cites a target for the number of congressional and governmental information requests as a performance measure. Related to this, BJS includes reports of the number of citations in the Congressional Record, in congressional testimony, in federal and state courts, and in professional journals as evidence of the relevance of their programs. ERS cites the availability of issue analyses as an indication of responsiveness to customer interests. NCES and NAEP both have relatively new measures of the number of times their data are cited and/or disseminated on education associations' web sites.

Finally, some agencies cite specific product improvements as improving relevance. NCHS cites targets for the number of new or revised charts in their flagship publication, *Health U.S.* Other examples that yield quantifiable performance measures include specific targets for the expansion of coverage of populations or economic sectors being measured or improved GPS locations for counties.

ERS reports the use of a Portfolio Review Score which is based on a qualitative assessment of external experts of the relevance, quality, and performance of ERS research portfolios to enable better informed decisions on food and agricultural policy issues. EIA uses a similar measure of the percent of outside experts rating EIA's analysis and forecasting program as relevant and reliable, consistent with changing industry markets, and the quality of products as high.

Accuracy includes features such as the correctness, validity, and reliability of data and information products, typically measured as closeness to target values. In particular, statistical data accuracy is defined as the closeness to the target value as measured by the sampling error and specific aspects of nonsampling error (e.g., response rates, size of revisions, coverage, and imputation rate). Accuracy may be established and monitored through review processes. Similarly, accuracy may be assessed by comparisons of data across different surveys or through the linkage of survey data to administrative record data and it may be improved upon by sample redesign, sample expansion, or reductions in specific sources of nonsampling error. Many of these factors are quantifiable and can be turned into measurable performance indicators.

Several agencies cite performance measures involving the percent of data items or surveys that are accurate (BLS), correct (BEA), or meet quality targets (EIA)—sometimes with specific targets for the percent meeting the goal and sometimes not. NASS cites the use of the percent of survey point estimates of key total expenditure items with standard errors reported. BLS also monitors the number of annual estimates released with standard errors. Taking this element one step further, a number of agencies cite measures that incorporate the standard error in quality indicators. For example, BJS calls for the reporting of the ratio of the design or target coefficient of variation to the actual coefficient of variation observed in the sample data; NASS reports the percent of key survey estimates meeting target coefficients of variation with a goal of 90 percent by 2009, Census reports a target coefficient of variation for the Current Population Survey in 2007. NASS also monitors the percent of county estimates with statistically defensible survey precision, and reports counts of statistical releases that include the root mean square error.

At least 8 agencies cite monitoring response rates, either generically, or with respect to specific targets for individual data collections. For example, BTS cites a target response rate of 80 percent for their Commodity Flow Survey and a target response rate of 50 percent for their Omnibus Household Survey. The Census Bureau has target response rates for the American Community Survey (92 percent), the Current Population Survey (90 percent), the Economic Census (84-86 percent), and the Current Economic Statistics (75-83 percent). NCHS reports response rates for the National Health Interview

Survey and for the Ambulatory Medical Care Survey relative to response rates from comparable studies conducted elsewhere. BJS reports monitoring both resident and agency level response rates in their surveys (with a target of 95 percent averaged over all agency data collections and a target of 90 percent averaged over all household/person level data collections). Taking a broader approach, NCES monitors and reports on the percentage of survey data collections with a response rate of 85 percent or above or a nonresponse bias analysis, with nonresponse weight adjustments for any potential bias identified in the nonresponse bias analysis.

Statistical agencies that conduct universe data collections or census collections monitor the coverage of the population being measured. For example BJS, BLS, EIA, NASS, Decennial Census, and Economic Census each include discussions of population coverage. In particular, EIA has a 2007 target of 85 percent for the percent of key survey frames with sufficient industry coverage to produce reliable supply, demand, and price statistics. NASS has a measure that uses total farm cash receipts from ERS to measure the percent of total U.S. agricultural production covered by their official statistics, with a target of 100 percent in 2007. Census is monitoring the expansion of the service sector coverage in the Economic Census and plans to improve the accuracy of the Decennial Census, with a target of reducing the overall net coverage error by 0.5 percent in 2010 relative to 2000.

BJS monitors error rates by monitoring the number of valid complaints received through the Information Quality Guidelines process and by monitoring the number of errata notices issued on an annual basis. Census has three programs in the Current Economic Statistics and Census of Governments with specific revision rate targets for initial versus final published measures. Census also has programs in Intercensal Demographic Estimates that have specific error reduction targets for 2010. BLS also monitors the number of revisions required.

Other specific examples of performance measures related to accuracy include meeting targets for number of cases interviewed (CPS), number of estimates released (BLS), reducing geographic coding errors in the Decennial Census, monitoring imputation rates by BJS and by the Economic Census, and establishing detailed project plan goals for the Economic Census, the Census Survey Sample Redesign, BEA programs, and BLS programs.

Timeliness can be measured in several ways, including success in meeting scheduled and publicized release dates, time from end of data collection to release of the information, or customer satisfaction with timeliness. In addition, agencies that have comparable international data sometimes compare the timeliness of their releases to those of their international peers.

The most common performance measure cited for timeliness is the release of data on an established schedule. A number of the releases in this group are those defined by OMB as Principal Economic Indicators that are required to have established release schedules.

Timeliness is measured as the percent of releases that occur on schedule—in some cases it is for recurring releases, in some it is for those reports identified as official reports, and in other cases, it is for reports on a specific topic. Agencies using measures of this type include BEA, BLS, BTS, EIA, ERS, NASS, NAEP, Census—Demographic Statistics, Survey Sample Redesign, Economic Census, and Current Economic Statistics and Census of Governments. In addition, BJS reports the number of reports that are issued within one month of the expected release.

NCES, NAEP, NCHS, and Intercensal Demographic Estimates report measures that include specified reductions in the time from end of collection/production to release. Related to this, BEA reports plans to accelerate the release of specific products. By way of example, NCES started in 1996 with a target of releasing 90 percent of initial release reports within 18 months from end of data collection (or demonstrating a 2 month reduction in time to release relative to the prior cycle of release for the same data); the target decreased 2 month per year to a target of 12 months in 2009. In a related vein, the Current Economic Statistics and Census of Governments are reducing data collection and processing time by facilitating reporting and reducing burden to respondents through use of electronic reporting.

Two agencies with economic indicators, BEA and BLS, cite the fact that they release their key statistics faster than their international counterparts as an indication of timeliness.

While most agencies conduct customer satisfaction surveys, the customer survey measures cited for NCES and NASS specifically call out timeliness as a dimension. In the case of NASS timeliness is in a composite measure, along with other dimensions such as ease of use, frequency, and accuracy. In contrast, NCES has three measures that focus solely on timeliness, one for data files, one for services, and one for publications, the target for each NCES measure is 90 percent.

Other performance measures that are related to timeliness include reports from ERS of monitoring timeliness through correlation of its activities and coverage in the media, and of monitoring timeliness through correlation of its activities and public interest as measured by website visits.

Program Performance

Statistical agencies must meet the information needs of the Nation, while at the same time operating efficiently and effectively to balance costs, disseminate information, and accomplish the mission of the agency. Thus, costs of products or programs may be used to develop efficiency measures. Measures of products disseminated, and of mechanisms used to release the information provide performance indicators for dissemination. Finally, mission achievement means that the information provided makes a difference.

Cost involves the dollar amount used to produce specific data products or services. The intent is to determine the costs of various programs or alternative modes of operation. Examples of cost data include full costs of products or programs, measures of returns on investments, dollar value of efficiencies, and ratios of costs to products distributed.

The most common metric adopted by statistical agencies to date is some variation on a measure of unit cost. For example, BTS has one performance measure that targets a reduction in the cost to produce three annual reports, and a performance measure that targets a reduction in the data collection costs for international freight data. BLS monitors the cost per housing unit used in the development of the CPI, and the cost per transaction in their Internet Data Collection Facility. EIA reports plans to decrease the cost of data collections, after adjusting for inflation. Decennial Census, Demographic Statistics, Intercensal Demographic Estimates, Survey Sample Redesign, Current Economic Statistics and Census of Governments, and Economic Census have established a survey specific set of unit costs per case for individual surveys; in each case the actual costs are monitored relative to the scheduled costs. In addition, the Decennial Census has further refined these measures for the American Community Survey by disaggregating unit costs by data collection mode. NASS, NCES, and NAEP have related measures that propose to maintain level or better costs per sample unit, after adjusting for inflation. In the case of the 2010 Decennial Census, the target is to keep the cost at or below the \$12.2 billion spent on the 2000 Census.

Despite the widespread use of unit cost measures, there is considerable variation in these measures. Census Demographic Statistics, EIA, NASS, NCES, and NAEP all cite adjustments for inflation in their unit cost measures, while 10 other agencies and programs with unit cost measures do not mention adjustments for inflation, presumably they are incorporated, but not mentioned in the PART measure description. Other differences include differences in the specificity as to the activities that are included in the unit cost measure and differences in whether the cost is defined per case, or per response on individual questions.

Beyond data collection costs, ERS and the Intercensal Demographic Estimates program have performance measures that discuss an increase in the number of products produced with stable or decreased staffing levels. Census also discusses the increased use of electronic data collections as a means of cost savings in a number of their programs, monitoring the percent of reimbursable projects completed on time, and increasing the number of data products with level funding. Finally, BJS has an aggregate index of efficiency that includes goals of reducing the number of staff required per unit of data collection and of reducing the costs of providing data services per user.

Dissemination includes information on the availability, accessibility and distribution of products and services. Typical measures include product downloads, number of Internet user sessions, degree of accessibility (including 508 compliance for accessibility to people with disabilities), customer satisfaction with ease of use, citations of agency data in the media, use of newsletters to describe products, exhibits to inform the public about information products, amount of technical support to data users, number of participants at user conferences.

BJS, BLS, and Census—Demographic Statistics, Intercensal Population Estimates, monitor the number of releases, in some cases relative to specific expected targets. Census also monitors the percent of data from the Economic Census and the Census of Governments that are disseminated to the public.

BTS, BJS, NCHS, NCES, and NAEP have performance measures that focus on web visits and web downloads. For example, BTS has a measure for increasing the number of database downloads from their TranStats website, while the other agencies and programs report web site visits. BJS, BTS, and NAEP have measures of the number of reports and/or data products distributed; and NAEP and NCHS have measures of the number of users who access on-line data tools and of improvements

to on-line data tools, respectively. BTS also has specific targets for increasing the distribution of the CD-ROM for the National Transportation Atlas database.

Finally turning the focus on data users and dissemination, ERS, NASS, NCES, and NAEP have customer service measures of user satisfaction with various aspects of information dissemination. NCHS monitors the number of participants in the NCHS data users conference. EIA includes a specific mention of meetings with stakeholders and data users to understand their information needs.

Mission Achievement is perhaps the most difficult performance dimension to quantify. This dimension speaks to the extent to which each agency has achieved its objectives and met the expectations of its stakeholders. In a number of cases the contributions of individual agencies are indirect, with the work of multiple agencies contributing to aggregate measures, such as the GDP and the CPI. In other cases the outputs from statistical agencies affect the programmatic outcomes of other programs; for example, the analysis conducted by NCES on alternative high school graduation rates and NAEP data on student achievement are being used by elementary and secondary education program in the Department of Education to monitor progress. Similarly, data collected by the Health and Nutrition Examination Survey at NCHS are used by health professionals to set treatment guidelines for various diseases, such as hypertension and diabetes. The narratives included in Section I of the PART on program purpose and design describe the effect of statistical information on policy or change of policy, supporting research focused on policy issues, informing debate on policy issues, or providing in-house consulting support.

Finally, the quantitative measures of customer satisfaction may indicate whether an agency has met the expectations of its stakeholders. Similarly, quantitative measures of the number of times an agency's information is used in congressional deliberations, court cases, or scholarly journals may shed light on the interaction between the outputs from statistical agencies and outcomes that result from the use of those outputs.

Ten of the 16 statistical agencies and programs use customer surveys to help monitor their performance—BEA, BJS, BLS, Current Economic Statistics and Censuses of Government, EIA, ERS, NASS, NCES, NAEP, and NCHS. Four of the 10 have participated in the American Customer Satisfaction Index Survey. BTS monitors Congressional and other government contacts; BJS monitors the use of their data by Congress, the courts, and in scholarly publications; and NCES and NAEP monitor the use of their data on education association web sites.

Where Are We Now?

This review of the types of measures and the specific indicators used suggests that statistical agencies are currently monitoring more aspects of their program performance than was the case a decade ago. Ten of the 16 agencies are now using customer surveys to measure various aspects of dissemination and in some cases overall relevance of the program. Monitoring the accuracy of their data is not new to statistical programs; however having specific measures on data quality reported in a performance management system may serve to focus increased attention on some of these measures. Attention is focused on the timeliness of statistical products, perhaps more than ever before. (NCES redesigned initial release products and streamlined the review process to meet the target of 12 months.) The emphasis on cost resulted in agencies focusing on details beyond the total budget for individual data collections, by developing measures of unit cost per case or per survey question completed. While related to relevance, focusing attention on mission achievement has resulted in measures of customer satisfaction with various aspects of the products and services delivered by statistical agencies, and more broadly has resulted in agencies and their programs turning more attention to defining their missions, the uses of their products, and the impact they may have.

What is Ahead?

In a 9/24/09 statement before the Senate Committee on Homeland Security and Governmental Affairs Subcommittee on Federal Financial Management, Government Information, Federal Services, and International Security, the Chief Performance Officer stated that “it is not enough for Federal agencies to produce performance information. The ultimate test of our performance management efforts is whether or not the information is used—not just by government agencies, but also by Congress, the public, our service delivery partners, and others” (9/24/2009 Statement, p.1). He acknowledged the role of GPRA in pushing agencies to state their goals and to measure progress. Similarly, he acknowledged the fact that the PART

questions helped to move the expectations for performance goals to the program level. However, he further noted that “too much emphasis was placed on producing performance information for the purpose of complying with reporting requirements, and too little attention paid to analyzing and acting on this information” (9/24/2009 Statement, p. 1).

The CPO announced that the new goal is to use Federal performance information as a performance improving tool to identify successful programs that should be continued, and to identify less successful programs that are in need of attention.

In August of 2009, OMB issued a revised version of OMB Circular A-11 Preparation, Submission and Execution of the Budget. This document may provide additional insights into new directions for Federal performance management. The Circular reiterates the fact that “Strategic plans, annual performance plans, and annual performance reports comprise the elements of GPRA...” (p. 2, Section 200, OMB Circular A-11), and directs agencies to prepare performance budgets, in lieu of annual performance plans. Emphasis is placed on the need for agencies to describe the relationship between expected results and budget resources requested, and for the annual performance report to contain information on program evaluations relevant to the Agency’s goals and objectives.

In this revision of Circular A-11, references to PART were dropped (p. 1, Section 200, OMB Circular A-11), and requirements were added for strategic plans to focus on the agencies’ most important performance goals, for personnel accountability for performance goals (p. 1, Section 210, OMB Circular A-11), and for the inclusion of trend data in annual performance reports (i.e., a period of 5 years—the current year and 4 previous years). The inclusion of trend data is intended to provide the information base needed by Congress, the public and other stakeholders to assess how a program is progressing compared to past experience (p. 5, Section 230, OMB Circular A-11). The FY 2009 Performance and Accountability Reports are to be submitted to the President and Congress by 11/16/09.

In discussing the strategic plan, Section 210 of OMB Circular A-11 notes that each agency’s strategic plan is required under GPRA to be updated or revised at least every three years, with a suggestion that the each agency’s first strategic plan issued during a new administration “represents a significant opportunity to revisit its general goals and objectives” (p. 1, Section 210, OMB Circular A-11). The strategic goals and objectives are described as statements of what the agency “wants to achieve relevant to national problems, needs or challenges, and how it expects to achieve them” (p. 3, Section 210, OMB Circular A-11). The related performance goals should reflect a limited number of long-term (3 to 5 years) goals that are the intended outcomes. Finally, there is an expectation that the Agency’s strategic plan and especially the performance goals should be aligned with the current administration’s stated goals.

For example, the Department of Veterans Affairs (VA) website entries on the fiscal year 2010 budget states that the President’s vision is to transform VA “... into a 21st Century organization that is Veteran-centric, results-driven, and forward-looking. This transformation is demanded by new times, new technologies, new demographic realities, and new commitments to today’s Veterans.” The VA web site goes on to state “VA’s budget request for 2010 provides the resources critical to achieving the President’s vision and will help ensure that Veterans—our clients—receive timely access to the highest quality benefits and services we can provide and which they earned through their sacrifice and service to our Nation.”

GPRA requires annual performance reports to include information on program evaluations related to the goals and objectives specified in an agency’s strategic plan or agency performance measures and goals. The revised OMB Circular A-11 states that an agency’s Performance Improvement Officer will determine whether these program evaluations have sufficient scope, quality, and independence. The importance placed on program evaluation by the Obama administration was further clarified in the 10/7/09 OMB Memorandum-10-01 (Increased Emphasis on Program Evaluation). More specifically, rigorous, independent program evaluations of how well programs are doing in reaching their intended outcomes, while keeping costs as low as possible, “...can help policymakers and agency managers strengthen the design and operation of programs.”

Under this initiative, as part of the Fiscal Year 2011 Budget process, agencies will make information available on-line for ongoing and newly planned Federal evaluations focused on program impacts. An inter-agency Evaluation Working Group will be formed to promote stronger program evaluations. Finally, limited funding will be available for agencies to request additional funding for high-priority evaluation activities through the Voluntary Fiscal Year 2011 Evaluation Initiative.

On October 27, 2009, the Senate budget Committee announced the formation of a new task force on government performance. The Task force was charged with responsibility to consider new ways for agencies to work more efficiently. Then on October 29, 2009 the OMB Deputy Director for Management testified before Congress that there is a need for

outcome based metrics that drive change. Stating that the federal government needs to learn from practices that work and those that do not, he emphasized the need to use performance goals and measures to set priorities, monitor progress, diagnose problems, motivate the workforce to achieve greater results, and to allocate scarce resources wisely. For a new approach to performance measurement to work, the Deputy Director stressed the need for senior leadership to take ownership of the performance management process, for the use of cascading goals and measurements, for the use of outcome-oriented, cross agency goals and measurements, and for thorough review and accountability to take place through a transparent process.

Finally, initiatives that are currently under way include the development of 3 to 8 ambitious, outcome-oriented high priority goals to be accomplished in 12 to 24 months, the use of management dashboards for IT, procurement, financial management, and personnel management, an increase in customer access to on-line information on the processing status of services, and the implementation of rigorous evaluations of social, economic and educational programs.