

A Concept for Performance Management for Federal Science Programs



Open-File Report 2017–1143

Cover: Photograph of the tail section of a brown trout. The individual components of the trout that integrate to make an organism whole including coloration and morphology can be a metaphor for components of science strategy and planning and performance measures that enhance the functionality of science programs. Photograph by Kevin Whalen, U.S. Geological Survey, 2017.

A Concept for Performance Management for Federal Science Programs

By Kevin G. Whalen

Open-File Report 2017-1143

**U.S. Department of the Interior
U.S. Geological Survey**

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
William H. Werkheiser, Acting Director

U.S. Geological Survey, Reston, Virginia: 2017

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <https://www.usgs.gov> or call 1-888-ASK-USGS (1-888-275-8747).

For an overview of USGS information products, including maps, imagery, and publications, visit <https://store.usgs.gov>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Whalen, K.G., 2017, A concept for performance management for Federal science programs: U.S. Geological Survey Open-File Report 2017-1143, 16 p., <https://doi.org/10.3133/ofr20171143>.

ISSN 2331-1258 (online)

Contents

Abstract.....	1
Introduction	1
Section I.—Strategic Planning and Science Program Objective Synthesis.....	3
Strategic Planning in Publicly Funded Science Programs	3
Science Program Objectives “Activity Types”	3
Defining Activity Types	4
Relating Activity Types to Performance Measurement Tools	4
Example of Relating Activity Types to an Integrated Performance Measure	4
Benefits of Typing Science Program Objectives—Relation to GPRA, Efficiency Measures, and Timelining.....	7
Annual Output Reporting—GPRA.....	7
Efficiency Performance Measures.....	7
Science Program Objective Milestones and Timelines.....	8
Section II—Integrated Performance Management	9
Linking Performance Measures to Root Causes	9
Justification and Rationale for Building an Integrated Performance Measure.....	10
Research Performance Scorecard.....	10
Setting Targets for Metrics.....	11
Weighting Factors—Relating Annual to Long-Term Performance	12
Interpreting Integrated Performance Measure Results	12
Cumulative Measurement Aspects Promote Flexibility in Achieving Long-Term Outcomes	13
Discussion.....	13
Acknowledgments.....	15
References Cited	15

Tables

Table 1. Relation of research activity types (that is, functional category of research identified in the planning objective) to performance metrics, and associated time scales of measurement	5
Table 2. Integrated performance measure for publicly funded science program, including six individual metrics and outcome measure benchmarks.....	11

This page left intentionally blank

A Concept for Performance Management for Federal Science Programs

By Kevin G. Whalen

Abstract

The demonstration of clear linkages between planning, funding, outcomes, and performance management has created unique challenges for U.S. Federal science programs. An approach is presented here that characterizes science program strategic objectives by one of five “activity types”: (1) knowledge discovery, (2) knowledge development and delivery, (3) science support, (4) inventory and monitoring, and (5) knowledge synthesis and assessment. The activity types relate to performance measurement tools for tracking outcomes of research funded under the objective. The result is a multi-time scale, integrated performance measure that tracks individual performance metrics synthetically while also measuring progress toward long-term outcomes. Tracking performance on individual metrics provides explicit linkages to root causes of potentially suboptimal performance and captures both internal and external program drivers, such as customer relations and science support for managers. Functionally connecting strategic planning objectives with performance measurement tools is a practical approach for publicly funded science agencies that links planning, outcomes, and performance management—an enterprise that has created unique challenges for public-sector research and development programs.

Introduction

U.S. Federal government research and development programs have become increasingly accountable for the public funding investment (Government Performance and Results Act of 1993 [GPRA; Public Law 103-62, 107 Stat. 285]), U.S. Office of Management and Budget [OMB] Program Assessment and Rating Tool [PART], circa 2005). A primary challenge for science programs is organizing and linking strategic science objectives at the program level to measurable outcomes based on performance metrics. Specific methods linking strategic objectives to cogent, measurable outcomes while sufficiently allowing for the uncertainty inherent to the scientific process are lacking (Loch and Tapper, 2002). In fact, the process of linking strategic science objectives and outcome-oriented performance measures remains largely unstandardized and undefined across any temporal scale.

Effective performance measures are tailored to the needs of the organization (Brown and others, 2003) with the goal of improving organizational performance by linking strategies and goals with resource allocation, such as funding/budget (Kaplan and Norton, 1996). Performance management must not be limited by organizational culture or resistance to change, but it must accommodate ways in which science is delivered and communicated to ensure buy-in from all levels in an organization (Brown and others, 2003). Few approaches are less successful than performance measures designed with disregard to organizational cultures, blindly applied, or externally driven (Alcock, 2004).

Nonfinancial performance management is wrought with many uncertainties, namely capturing the correct performance indicators at the appropriate temporal and spatial scales (Heinrich, 2002; Loch and Tapper, 2002; Ittner and Larcker, 2003; Radnor and McGuire, 2003; Alcock, 2004). Common problems include:

1. Measures used for capturing data are insensitive to true performance or lacking appropriate linkages to actual performance (Heinrich, 2002; Alcock, 2004),
2. Setting “soft” targets and gaming performance results by “picking winners” (Alcock, 2004), and
3. Improper consideration of underlying statistical properties of both targets and measured data that can lead to spurious conclusions regarding performance benchmarks (Maleyeff, 2003).

These factors may detract from the ability of a government agency to demonstrate and justify a measurable outcome for the investment of public funds (Hernandez and others, 1998).

The limitations of nonfinancial performance management potentially are quite severe for science agencies producing research outcomes and benefits that may not have clear linkages to current funding (Committee on Science, Engineering, and Public Policy, 1999, 2001; Loch and Tapper, 2002). Performance-management systems assume strong causal linkages between the input of resources (for example, effort, time, funding) and outcomes. Measurement of science outcomes is difficult because of the uncertainty of scientific endeavors (Committee on Science, Engineering, and Public Policy, 1999; Loch and Tapper, 2002). For example, deciding years in advance what long-term measures will ultimately be most relevant for measuring science program outcomes is extremely challenging. To cope with the uncertainty of science discovery, a performance-management system must respond to changing internal and external drivers and conditions over time and be sensitive to short- to long-term post-funding timescales within which the benefits of science discovery mature. Clarity of program objectives and outcomes, therefore, becomes both critical and challenging for those agencies that focus on basic rather than applied science (Committee on Science, Engineering, and Public Policy, 1999).

Many public-sector science programs share the difficulties of measuring science outcomes; synthesizing science program objectives; and linking budget, planning, and performance (Committee on Science, Engineering, and Public Policy, 2001; Heinrich, 2002). Section I of this report focuses on strategic planning and synthesizing science program objectives using activity type identifiers that link planned outcomes to performance measurement tools. Section II introduces an integrated performance-management system that tracks performance on individual metrics and provides explicit linkages to root performance drivers, while simultaneously assessing progress towards long-term outcomes. The approach can be adapted to specific programs by adjusting the metrics, targets for metrics, and scale of measurement that depend on the nature of science conducted by a program (for example, basic or applied science) and the performance measurement data/systems available.

Section I.—Strategic Planning and Science Program Objective Synthesis

Strategic Planning in Publicly Funded Science Programs

Strategic science planning is the process linking organizational and programmatic goals and objectives with research focus areas, as determined by scientific capacity, primary partner needs, and future opportunities. Strategic plans transparently communicate a program mission (Sneyd and Rowley, 2004), provide decision-makers with a common fact base, and ultimately outline outcomes to be achieved over specified time scales (Kaplan and Norton, 2000; Kaplan and Beinhocker, 2003). Maintaining strong links between strategic direction and performance management allows for continuous organizational learning that leads to increased operational effectiveness (Kerssens-Van Drongelen and De Weerd-Nederhof, 1999). In the U.S. Federal Government, Department-level strategic goals are stepped down to Bureau (agency) strategic objectives, and then captured in individual plans for science programs. Such program plans provide the mid-scale guidance that may be adapted on an annual basis to select and amplify specific research areas in the short term.

Outcome-oriented performance management provides the nexus and integrative link between planning and budget, where achieving outcomes is linked to the budgeting process through the funding of priority science objectives. Through strategic science planning and the prioritization of objectives, the stage is set to guide the funding of projects addressing critical knowledge gaps (Loch and Tapper, 2002). Strategic planning and the resultant prioritization of program objectives direct resources to the highest priority research. The goal, theoretically, is to shorten the generation time for the delivery of new scientific knowledge and provide a high return on investment (based on funding, time, and new knowledge) that advances the strategic goals. For organizations operating in complex and seemingly stochastic funding environments, accelerating generation time of scientific knowledge must be adaptive, particularly because Federal science agencies are continually asked to demonstrate their societal relevance.

Science Program Objectives “Activity Types”

Research objectives set the stage to achieve strategic programmatic outcomes. To assess progress toward outcomes, performance measures can be developed and used to track factors affecting the achievement of targets (see “Section II. Integrated Performance Management”). This is challenging for science programs in the business of “knowledge production,” where the connection between planning objectives and performance measures may be weak or indistinct (Alcock, 2004). Within and among science programs, inherent differences may exist in the tenor and type of research activities captured in objectives. In essence, the planning objectives of a science program represent a variety of “functional” categories of research (for example, the “tier metaphor”; Hauser and Zettlemeier, 1997; Hauser, 1998). The explicit identification of the functional categories, or the research activity types, of science planning objectives has direct implications for program management (for example, synthesizing across programmatic strategic goals and science functions). Research activity types functionally connect strategic planning objectives to appropriate performance measurement tools, which reduces incongruity between planning research objectives and measuring progress toward future outcomes that may be inherently indistinct. This activity typing approach also may be used within science programs to direct funding allocation efforts to ensure a balanced research portfolio, or to adapt allocations as progress towards program objectives is achieved.

Defining Activity Types

Five science research objective activity types are considered as follows (table 1):

1. *Knowledge discovery (KD)*—Discovery of new, fundamental knowledge.
2. *Knowledge development and delivery (KDD)*—Development and delivery of research knowledge, guides, tools, and models.
3. *Science support (SS)*—Direct technical support to managers in applying research knowledge, guides, tools, and models.
4. *Inventory and monitoring (IM)*—Design, implementation, and analysis of the results of inventory and monitoring systems to evaluate the condition of resources at relevant spatial scales and organizational levels.
5. *Knowledge synthesis and assessment (KSA)*—Synthesis and assessment of research knowledge to address specific management objectives.

The knowledge discovery, knowledge development and delivery, and the knowledge synthesis and assessment activity types are based on research activity descriptors found in existing Federal research grade evaluation guidance (U.S. Forest Service, 1995). The science support and inventorying and monitoring research activity types were added to more completely represent Federal science programs with strong basic and applied research components that emphasize customer or client needs.

Relating Activity Types to Performance Measurement Tools

Each of the five activity types relates to specific performance measurement tools or approaches, which vary by time scale (table 1). For example, science objectives focused on continual knowledge discovery (basic ongoing research) are most suitable for an annual output metric, objectives focused on knowledge synthesis and assessment (periodic summaries of knowledge) are most suitable for a milestone-oriented metric (for example, 5–7 year program reviews), and objectives focused on science support and technical assistance are most suitably assessed with customer satisfaction and (or) customer decision-making survey metrics. The knowledge development and delivery activity type may function as an efficiency measure and is discussed at length in section, “Example of Synthesizing Research Objectives by Activity Type.” The inventorying and monitoring activity type may relate most closely to an annual or milestone-oriented measure depending on the scale and timeframe of the activity.

Example of Relating Activity Types to an Integrated Performance Measure

Strategic planning typically produces a greater number of science objectives than can be reasonably addressed or funded in any given year. Activity typing enables the synthesis of objectives across science programs that contribute to a long-term outcome. Science objectives with a common research activity type show interconnectedness and provide a basis for linking outcomes among research programs. The research objective activity typing process also sets the stage for development of an integrated performance measure (table 1), a process shown in section, “Example of Synthesizing Research Objectives by Activity Type” and discussed at length in “Section II. Integrated Performance Management.”

Table 1. Relation of research activity types (that is, functional category of research identified in the planning objective) to performance metrics, and associated time scales of measurement.

[Research activity types and measures represent those for a standard applied Federal science program with explicit external customers or constituencies responsible for managing resources Example measures included in accordance with the integrated performance measure are described in the text]

Research objective activity type	Metric	Reporting time scale (years)	Example measure
Knowledge Discovery (KD) —Discovery of new, fundamental knowledge.	Annual output	1 year	Expected proportion of planned publications completed.
Knowledge Development and Delivery (KDD) —Development and delivery of research knowledge, guides, tools, and models.	Efficiency	1–3 years	Expected proportion of planned products focusing on tool/model/methods development completed.
Science Support (SS) —Direct technical support to managers in applying research knowledge, guides, tools, and models.	Customer satisfaction	1–2 years	Expected customer satisfaction score to be equaled or exceeded.
	Customer decision-making	3–5 years	Proportion of delivered products that have affected species and resource management decision-making.
Inventory and Monitoring (IM) —Design, implementation and analysis of the results of inventory and monitoring systems to evaluate the condition of resources at relevant spatial scales and organizational levels.	Annual output	1 year	Expected proportion of planned publications completed.
	Science product milestones	2–5 years	Proportion of milestone products expected to be completed.
Knowledge Synthesis and Assessment (KSA) —Synthesis and assessment of research knowledge to address specific management objectives.	Program review	5–7 years	Proportion of program science focused on identified priority objectives.

Example of Synthesizing Research Objectives by Activity Type

Consider a program with a single funding line-item that is internally separated into subprograms to improve the delivery of science and to better relate to key external partners. The goals and objectives of the subprograms naturally relate to the overall mission of the formal program. Planning efforts are conducted at the subprogram level to develop 5-year strategic science plans, tiered from the strategic plans of the bureau and the agency. Goals identified in each subprogram plan typically are visionary and thematically outcome-oriented, and are stepped down to operational segments as research objectives that capture the specifics of the activities to be funded. No standard guidance exists for defining a uniform operational scale of program goals and objectives and their subprograms. Subprogram objectives are assessed an activity type according to the five categories table 1, and each is assigned a performance measurement tool.

Subprogram	Objective	Activity type ¹	Performance measurement tool
A	A-1	KD	Annual
	A-2	SS	Customer survey
	A-3	KDD	Efficiency
B	B-1	KD	Annual
	B-2	KSA	Milestone
	B-3	IM	Milestone
C	C-1	KDD	Efficiency
	C-2	KD	Annual
	C-3	KSA	Milestone

¹KD, knowledge discovery; KDD, knowledge development and delivery; IM, inventory and monitoring; KSA, knowledge synthesis and assessment; and SS, science support.

When activity types are assigned to subprogram objectives, it is possible to synthesize subprogram research objectives as follows to build an integrated performance measure. In this manner, research objective activity types and, thus, performance measurement tools serve as unifying factors for achieving and planning research outcomes.

Integrated measure metric	Subprogram	Objectives
Annual	A, B, C	A-1, B-1, C-2
Efficiency	A, C	A-3, C-1
Milestone	B, C	B-2, B-3, C-3
Customer survey	A	A-2

Benefits of Typing Science Program Objectives—Relation to GPRA, Efficiency Measures, and Timelining

Research objective activity typing provides guidance on linking outcomes envisioned in strategic science objectives to appropriate performance measurement tools. The following examples discuss practical applications of research activity typing for:

1. Formalizing connections among strategic research objectives and extant Federal performance reporting requirements (for example, GPRA);
2. Identifying selected research objectives to construct and manage efficiency performance measures; and
3. Establishing measurable, time-dependent connections between strategic objectives and the delivery of significant scientific syntheses and assessments.

Annual Output Reporting—GPRA

U.S. Federal science programs have output-oriented GPRA performance targets based on numbers of publications, technical reports, and (or) analyses, and are required to report numbers produced at regular intervals throughout the fiscal year. By typing research objective activities, science program managers can identify strategic objectives focused on ongoing knowledge discovery, or research efforts focused on producing publications or the basic unit of output counted for GPRA. GPRA products are not exclusively found within knowledge discovery research objective activity types, however. The output-oriented linkages between knowledge discovery objectives and annual GPRA products also is where elemental connections may be observed between funding and output (production). The product-oriented outcomes of knowledge discovery objectives are sensitive for immediately relating performance to changes in funding. Research activity typing to identify which strategic science objectives produce ongoing science outputs thus aids the management of science program performance.

Efficiency Performance Measures

The specific benefits of typing science program objectives are exemplified by the knowledge development and delivery activity type (table 1). Applied science programs with strong links to partner-oriented management needs often seek to focus research efforts to develop tools, models, and methods to improve efficiencies; to extend limits of detectability both in scope and magnitude; or to guide decision-making in complex information environments. Many examples exist in science where investments in new tools, models, and methods increase the resolution of scientific knowledge (for example, Zydlewski and others, 2001).

Efficiency outcomes seek to reduce the cost of producing knowledge, the amount of effort required to deliver information, and the time needed to produce and deliver research results (see section, “Efficiency Performance Measures in Research and Science—It Is Not Just about Cost”). Improved efficiencies are achieved by either increasing the amount of information delivered per a standard unit of effort or investment, or producing the same amount of information for less effort or investment (including time), or both (Tangen, 2005; see section, “Efficiency Performance Measures in Research and Science—It Is Not Just about Cost”). Research activity typing enables science program objectives focusing on tool, model, or method development to be explicitly identified and funded. Research activity typing also enables high-priority objectives focused on knowledge delivery and development efficiencies to be synthesized across programs and openly discussed, when included in the context of a transparent annual science planning and funding allocation process.

Efficiency Performance Measures in Research and Science—It Is Not Just About Cost

1. Efficiency measures may be expressed as a ratio of a “unit of information” (UI) [numerator] produced for a given investment/effort, often cost [denominator], but not exclusively related to cost. Other investment/effort-related variables considered are time and full time equivalent (FTE), which ultimately are directly correlated with costs. The UI commonly is considered in numbers of products, which for research and science are easy to measure, but generally not informative for communicating and (or) managing performance.
2. Increased efficiencies are achieved by driving down costs (decreasing denominator), increasing the magnitude (scope) of the UI (increasing numerator), or both (where $\Delta \text{UI} > \Delta \text{cost}$). Undue focus on “cost” to achieve higher efficiencies is short-sighted and risk-prone—it is referred to as “managing the denominator” in the private sector (Hamel and Prahalad, 1989). This can lead to short-term decisions (cost-cutting measures) that have negative consequences for long-term strategy and market position. Excessive focus on costs can also lead to selection of only “safe” projects with reliable return-on-investment, which may quash riskier projects with higher innovation potential (Loch and Tapper, 2002).
3. For science, investing in tools/methods/models may substantially expand the scope, resolution, and scale of the UI per unit of cost. When sufficient basic information is known about a scientific issue, identifying research objectives focused on tool/model/method development and prioritizing these objectives for funding are the primary operational means by which Federal science programs will manage improved efficiencies.

Science Program Objective Milestones and Timelines

Research objectives focused on knowledge synthesis and assessment provide the template for establishing measurable, time-dependent connections between science strategy and milestone-oriented outcomes. Strategic planning in science may connect disparate approaches and research avenues in logical linkages, syntheses, or summaries. Many strategic science plans develop lists of goals and objectives, but fail to communicate timeframes within which plans for key objectives will be completed (Kaplan and Beinbocker, 2003). The knowledge synthesis and assessment objective type relates directly to the production of synthetic science products that are highly useful for summarizing bodies of science and recommending future research priorities given the state of knowledge. Because synthesis products and knowledge synthesis and assessment objectives typically are planned *a priori*, it is possible to develop milestones for these products that can be used directly in a performance-management context.

Committing to specific timeframes for the delivery of scientific products may not always be practical when funding is uncertain or potential future outcomes are highly speculative. However, research, even if not milestone oriented, can be collated at certain time points into syntheses where insights and conclusions can provide significant efficiencies into future science planning. All time-based uncertainty in meeting science product milestones must be balanced with the benefits of projecting the strategic future of a science program. Additionally, linkages among program objectives in time and space need to be communicated as individual science program plans are integrated at higher management scales.

Timelining the planned delivery of key scientific milestones and dependent intervening steps is a straightforward communication tool to show potential benefits of funding investments. Establishing timelines for strategic plan goals and objectives is the first step for turning visionary mission statements into operational strategic road maps (Ittner and Larcker, 2003). For strategic planning, identifying specific milestones for science plan objectives is an explicit approach for communicating strategic direction. Science plan objectives also may hold inherent interdependencies that need to be explicitly managed, where the completion of one objective may be predicated on the accumulation of a number of milestone steps. Research activity typing to identify milestone-oriented objectives is a potentially useful tool for managing the strategic implementation of a science program and developing associated measures of performance.

Section II—Integrated Performance Management

This section provides an overview and rationale for developing an integrated performance measure for science programs. Numerous limiting factors for measuring performance in science programs are identified and subsequently used to justify a multi-time scale measurement approach. An example integrated performance measure is constructed and used to detail important aspects of metrics, such as target setting, weighting factors, and interpretation of results.

Linking Performance Measures to Root Causes

Performance measures provide the integrative link between programmatic outcomes and strategic direction (Loch and Tapper, 2002; Sneyd and Rowley, 2004). In Federal programs, the intent of performance management is to ensure that funding and delivery of programs is occurring in the most cost-effective and efficient manner and is achieving stated outcomes (Hernandez and others, 1998). The emphasis of performance management is to determine if program budgets are aligned with the program goals in such a way that the impact of funding, policy, and legislative changes on performance are readily known.

Three factors are key drivers for Federal research and development programs related to demonstrating the interrelatedness of planning, budget, and outcomes in the measurement of performance:

Factor 1—Strong Linkages.—Adequate performance measures need to demonstrate causal linkages and, if available, to provide indicators of underlying factors limiting full performance (Ittner and Larker, 2003; Alcock, 2004). Indicators of failed or lagging performance that can point toward root causes (such as insufficient budget, poor coordination, or failed interdependencies) have potential utility for managing science outcomes. One-dimensional measures of success, or measures insensitive to program progress, may fail to identify root causes of lagging performance (Ittner and Larker, 2003).

Factor 2—Uncertainty of Science Discovery.—Performance measures of scientific research outcomes presents particular difficulties, as adequate performance management assumes strong causal linkages between the input of resources (effort, time, funding) and outcomes (Committee on Science, Engineering, and Public Policy, 1999; 2001; Loch and Tapper, 2002). Performance measures for science programs must adequately reflect the uncertainty of science discovery and contain metrics responsive to internal and external drivers that vary over time, such as funding and stakeholder inputs. Performance measure that is integrated over time scales and includes internal and external program drivers is designed to address uncertainty associated with the production of science outcomes (Loch and Tapper, 2002)—that is, no one time scale alone is sufficient for tracking progress for science programs. A performance-management system based on several metrics tracked at a range of time scales may best be able to cope with the uncertainty of science discovery.

Factor 3—Adequacy of Measurement Tools.—Although several standard approaches for measuring performance toward outcomes exist, a substantial limiting factor for performance measures of science is the availability of appropriate tools. Few tools are adequate for tracking performance of programs, particularly in the uncertain and complex external environment within which science programs typically operate (Committee on Science, Engineering, and Public Policy, 2001). By nature, short-term performance measures generally are more output- or production-oriented, whereas long-term measures require a more integrated view. Measurement tools for long-term assessment of progress include the achievement of milestones related to synthesis products, surveys of partners or customers affected by or using science and research produced, reviews of science programs, and potentially other metrics related to the quality or impact of science products. It is necessary for measurement tools that integrate performance management over a range of time scales to include external views of the progress of science programs (Committee on Science, Engineering, and Public Policy, 2001). This is important for communicating outcome-oriented societal benefits of investments in science (Loch and Tapper, 2002). Tracking measurement interdependencies over integrated time scales is particularly important for science that advances through a strong reliance on accumulated knowledge.

Justification and Rationale for Building an Integrated Performance Measure

The integrated performance measure proposed here is based on the approach (tested in the private sector) of integrating performance analysis across financial and nonfinancial domains (Kaplan and Norton, 1996; Nilsson and Kald, 2002). The integrated approach to nonfinancial performance management has been shown to be value-added for fully capturing (1) interdependencies in measuring program progress (Werner and Souder, 1997; Brown and others, 2003), (2) internal and external program or process drivers (Alcock, 2004), and (3) short- as well as long-term outcome achievements (Heinrich, 2002). Integrated performance measures also have been effectively used in various domains—organizational (Griffin and Page, 1996; Loch and Tapper, 2002), process (Verweire and Van den Berghe, 2003), and management (Leem and Kim, 2004).

Collectively, previous nonfinancial performance analyses showed clear potential for addressing the three factors identified for performance management of science program outcomes. A standard publicly supported science program, with both a basic and applied research focus and strong linkages to external customers and constituencies, was considered for this exercise. I assume that the primary long-term outcome of the example science program here is to increase the use of targeted analysis/investigations delivered to customers that are used in management decision-making.

Research Performance Scorecard

The integrated performance measure proposed represents a “scorecard,” as reviewed by Coe (2003), with metrics providing strong linkages to internal and external drivers (when linked to budget in a performance-based annual planning cycle), integrated over measurement scale ranging from 1 to 7 years (table 2). Specifically, the metrics include:

- An annual output-oriented metric,
- An efficiency measure; both short-term and long-term customer-oriented metrics,
- Milestones for science products, and
- A 5–7 year cycle program review (table 2).

Selection of the six metrics identified represents a parsimonious balance between annual, efficiency, and long-term measures that provide a basic resolution ability to adequately assess scientific progress, and having too many measures that simply inflate complexity (Feltham and Xie, 2005). Absent *a priori* information regarding metrics, which is common, the selected metrics represent a reasonable starting point to build an integrated measure. All selected metrics have linkages to research activity types typically found in strategic science plan objectives of publicly funded programs that have both basic and applied research components (table 1).

The integration of the six metrics supports the cumulative, long-term progress toward the outcome of providing knowledge and information to support and improve management decision-making. The integrated measure provides the flexibility to evaluate progress within each of the metrics individually (Werner and Souder, 1997), while simultaneously tracking progress across all metrics as they relate to the long-term outcome (Coe, 2003).

Table 2. Integrated performance measure for publicly funded science program, including six individual metrics and outcome measure benchmarks.

[Planned and completed data are for a specific planning year *i*. The long-term measure (cumulative) is as follows: by year *i* + 5 achieve greater than (>) 80 percent of integrated measure to support outcome of increasing the use of program targeted analyses/investigations delivered to customers that are used in management decision-making. Milestone measures planned, but not accomplished, are carried over to the following year. Achieved = "yes" if "score" > "expected." Adjusted values are the product of the "weighting factor" and the "score" or actual value achieved; adjusted values are summed and represented as a percentage for comparison with the long-term outcome measure benchmark. Measurable (positive) progress toward long-term outcome is "yes" (achieved) if summed adjusted value > long-term benchmark]

Metric	Outcome measure benchmark (target)	Weighting factor	Year <i>i</i>				Long-term measure
			Number planned	Number completed	Score (percent)	Expected (percent)	Adjusted value (percent)
Annual output	≥80 percent of planned publications completed.	0.20	19	10	53	80	10.6
Efficiency	≥80 percent of planned products focusing on tool/model/method development completed.	0.20	8	7	88	80	17.6
Customer satisfaction	≥80 percent annual customer satisfaction maintained.	0.10	(¹)	(¹)	85	80	8.5
Science product milestone	≥80 percent of planned key program milestones accomplished.	0.10	7	6	86	80	8.6
Customer decision-making survey	≥80 percent of delivered products used in management decisions by partners.	0.15	(¹)	(¹)	80	80	12.0
Program review	≥80 percent of program science focused on identified priority objectives.	0.25	(¹)	(¹)	72	80	18.0
						80	75.3

¹Benchmarks are not based on production units; see expected values.

Setting Targets for Metrics

The process for setting targets for metrics depends on many factors. In some cases, performance data may exist that can be mined to derive credible baseline values for metric targets (Loch and Tapper, 2002). Collaborative processes have been described to establish performance measures and associated baseline targets (Hernandez and others, 1998). However, as may be typical for publicly funded science programs that have not had a long history of performance management, extensive baseline target data may not be available at the outset of instituting an outcome-oriented performance measurement system.

Thus, the selection of metric baselines may use some existing data, but remain mostly an educated guess. Some basic, logical approaches are that metric targets should be reasonably ambitious. If the existing baselines are easily achieved (that is, perceptibly low), then targets may logically be increased to higher benchmarks. If data exist to show that success of a particular metric is already adequate, the program managers may emphasize maintaining high levels of success rather than unrealistically increasing benchmarks to unattainable values.

In the example here, each individual metric in the integrated measure is given a specific target score or benchmark expected value, resulting in six individual year-end targets (table 2). It is important to note that metric targets, if not based on sampling, more truly indicate a level or degree rather than a pure value. By definition, singular (*N*=1) metric targets lack statistical properties, such as variance (confidence limits), and so are incomparable to actual measures data in a traditional statistical sense (see Maleyeff, 2003). As such, inferences regarding the progress of a program toward annual or long-term outcomes may be largely qualitative even if based on numeric targets and measures data.

Weighting Factors—Relating Annual to Long-Term Performance

To combine the individual metrics into an integrated score, a weighting factor is applied to each individual metric (Coe, 2003). In the example provided in table 2, the weighting factors range from 0.10 for two measures to 0.25 for the program review (which is reflective of its significance for managing progress toward long-term outcomes). Weighting factors can be adjusted as managers seek to emphasize program areas, and as understanding changes over time as to the significance and relevance of each metric to the overall long-term outcome goal (Hauser and Zettermeyer, 1997; Hauser, 2001; Heinrich, 2002).

The weighting factor is used quantitatively to adjust the individual metric score for summation for the overall integrated score. For example, if the program completes 53 percent of the planned publications, the weighting factor of 0.20 for the annual output metric is applied to the actual score to give an adjusted value of 10.6 percent. Similarly, if a customer satisfaction survey provides a score of 85 percent, it is adjusted by the 0.10 weighting factor for an adjusted value of 8.5 percent. When the individual benchmarks for each metric are not achieved, a failure of the measure is indicated, but the adjusted score is still included in the overall summation for the long-term measure.

Interpreting Integrated Performance Measure Results

In the example provided, both the annual output and the program review metrics did not achieve their individual benchmark (table 2). Clearly, performance related to the annual output was lacking, suggesting perhaps an inappropriate benchmark, insufficient budget or personnel, or poor management. Otherwise, the measure results suggest that the program is engaged in good relations with partners and that program science is having a significant effect on partner decision-making. Furthermore, the achievement of long-term products or milestones is not lacking, as was observed for the annual production of publications. When the root causes of the poor annual performance are identified, it is necessary to ensure that factors related to the lagging annual performance do not cascade or inflate (or act as a catalyst) to cause effects on more substantive synthesis products and (or) affect partner relations and decision-making (Brown and others, 2003). The slightly missed target for the program review metric suggests that minor rather than wholesale modifications are needed for achieving priority program goals.

When individual metric scores were summed over the integrated measure, the long-term benchmark of 80 percent is not met or exceeded (table 2); however, the adjusted total score of 75.3 percent is very close to the benchmark. Does failure to meet or exceed the long-term benchmark imply that the program is not making reasonable progress toward the long-term outcome? Absent the multi-scale, integrated performance metric system, such a conclusion could be erroneous. The suite of measures that are integrated over time, and that include internal and external program drivers, act as a check-and-balance system to assess progress toward the long-term outcome (Werner and Souder, 1997). Trends in the individual metric results suggest adequate progress toward the long-term outcome, with the measure value compared to the benchmark indicating that improvements are needed to continue progress.

This integrated approach is quantitative, adaptable, and, over time, will provide detailed information for managing root causes of both failed and exceptional performance. Over time, linkages may be established among these metrics to better define factors influencing the attainment of long-term outcomes (Hauser, 2001; Heinrich, 2002; Alcock, 2004). Additionally, other metrics may be added to the integrated measure as new tools become available and standardized for performance-management purposes. Changing the performance-management scorecard metrics to adapt to program drivers must be balanced against the benefits of maintaining consistency in the integrated metric through time (Coe, 2003) and the effects of adding metrics that inflate measurement variation (Hauser, 2001; Feltham and Xie, 2005).

Cumulative Measurement Aspects Promote Flexibility in Achieving Long-Term Outcomes

A necessary feature of a performance measure is flexibility, particularly for Federal science programs that, at times, need to adapt to unforeseen Administration priorities. The long-term integrated measure incorporates flexibility by being cumulative over time. If funding or changing priorities affect the completion of a science objective milestone planned for any given year, the planned milestone is simply shifted forward in time to the next year. Although such shifting may influence the integrated measure score in any given year, assuming the milestone remains a priority, the eventual success of the milestone has the potential to be realized and accumulated, thereby supporting the long-term program outcome. The example in table 2 represents a single year hence one snapshot in time. Importantly, as measures data are accumulated, it will be possible to assess relationships among individual metrics, as well as to assess cumulative progress toward the long-term outcome (Hauser, 2001).

Discussion

Described herein is an approach for linking strategic science plan objectives to an integrated performance-management system to demonstrate progress toward long-term program outcome goals, while simultaneously identifying root causes of potential factors limiting program success (Ittner and Larker, 2003). The integrated measure approach uses a multi-scale system to address key uncertainties associated with the task of assessing the performance of science programs, whose accomplishments rarely can be adequately predicted (Committee on Science, Engineering, and Public Policy, 2001; Loch and Tapper, 2002). Furthermore, the inclusion of measures reflecting internal and external drivers, focused on partner satisfaction and decision-making, provide a robust and complimentary suite of metrics for linking performance to internal and external forces (Brown and others, 2003). Overall, the integrated measure, which is linked directly to science objective activities, has the potential to provide managers with a clear picture of progress toward long-term program goals. This approach is adaptable should underlying conditions affecting the program change in the future (Coe, 2003).

The typing of science program objectives, coupled with the integrated performance measure, may have broad appeal, as it would allow comparability and benchmarking across Federal science programs with similar functions (Coe, 2003). In some cases, multiple Federal agencies have research and development programs or bureaus that support similar mission functions to provide information and knowledge for management decision-making. Conceptually, the activity typing of science program objectives across Federal science programs could be linked to an integrated performance measure. In this manner, Federal science programs with similar missions and approaches could be evaluated to compare duplicative functions in contrast to overlapping functions where “market competition” among programs may enhance innovation. In some cases, similar across-bureau scientific capacity also may be needed to address challenging high-order questions. By using the integrated performance measure concept proposed in this report, within-program management can feed and inform progress toward long-term outcomes at higher organizational levels.

Performance management in general assumes strong causal linkages between effort and outcomes (Sneyd and Rowley, 2004). For both basic and applied scientific endeavors, the connection between effort and outcome may be weak, random, or indistinct, and thus exceptionally difficult to define (Hauser, 1998; Committee on Science, Engineering, and Public Policy, 2001). The generation of new knowledge, the basic underpinning of scientific endeavors, leads to often unanticipated societal benefits. In fact, many examples of disconnected, yet significant, linkages between science and societal benefit exist (Hauser, 1998). The integrated planning and performance-management approach described here incorporates both risk and negative findings, because “success” is not based on any one measure, and the approach recognizes that progress toward the long-term outcome is cumulative. The method enables innovative, riskier cutting-edge science objectives where established methods are unavailable to be openly managed as part of the larger science portfolio.

Limitations exist for private sector nonfinancial performance management systems being applied to the public sector (Brown and others, 2003). However, the integration of metrics, which has become state-of-the art in private-sector performance measurement systems (Werner and Souder, 1997; Loch and Tapper, 2002; Leem and Kim, 2004), is a readily adaptable approach for assessing progress of publicly funded science programs. In this report, the concept of performance metric integration was extended to include multi-time scale metrics that directly support the uncertainty of science discovery and the manner in which public science is delivered. In the private research and development sector (among others), innovation-to-cash processes, product commercialization, and fear of commoditization drive and feed rapid ongoing assessments of research portfolios (Hauser, 2001; The Boston Consulting Group, Inc., 2005). Absent market-driven performance forces, the multi-time scale integrated performance measure provides short-term, mid-scale, and long-term feedback for publicly funded science programs, based on strategically derived priority research objectives.

A limiting factor for a performance measure and science is the availability of appropriate tools. By nature, short-term performance measures generally are more output- or production-oriented, whereas long-term measures require a more integrated view (Hauser, 1998; Committee on Science, Engineering, and Public Policy, 2001). The use of production-oriented measures as a short term performance indicator does not preclude their potential use as a long-term measure. Over time, the accumulated achievement of short-term production goals may be viewed as a long-term index of sustained success. Measurement tools for long-term assessment of progress include the achievement of milestones related to synthesis products, surveys of partners or customers affected by or using science and research produced, reviews of science programs, and other metrics related to science quality (for example, citation indices for papers produced and [or] more detailed science citation impact analysis). The key feature of long-term measurement tools is that they include external views of the progress of science programs, which is necessary for communicating outcome-oriented societal benefits of investments in science.

The integrated performance measure proposed here is designed to provide interpretive feedback to managers and to enable assessments that can identify and initiate remedial actions to achieve long-term outcomes, when needed. Key to this analysis is the relation between targets (which are unimodal, potentially arbitrary benchmarks) compared to actual measures that may be based on assessment data. Maleyeff (2003) details a number of underlying statistical pitfalls associated with comparing benchmark to measures (assessment) data. The large potential number of underlying statistical considerations involved with analyses comparing benchmark (unimodal, potentially arbitrary) to measures data suggests that the emphasis of this analysis not be solely quantitative. For public programs, and for science programs in particular that have not generally had a long history of being “measured”, suitably robust assessment data are unlikely to be available to rigorously quantify differences between benchmarks measures, and to link these differences to specific causal agents (Hauser, 2001; Heinrich, 2002; Mayeleff, 2003; Feltham and Xie, 2005). Taking a holistic view is particularly appropriate for tracking progress toward a long-term science outcome using the integrated measure approach.

Efficiency outcomes seek to reduce the cost of producing knowledge, the amount of effort required to deliver information, and (or) the time needed to produce and deliver research results. In this report, I identify a process for “typing” science objectives focused on tool, model, and (or) method development. These types of research improve efficiencies by extending limits of detectability both in scope and magnitude and guiding decision-making in complex information environments. The objective activity typing approach, coupled with the integrated performance measure, provides the foundation to track the interrelatedness of metrics and to quantify trade-offs of unequal funding of the science portfolio. In this manner, the cost-benefits of improving knowledge delivery efficiencies can be balanced among competing needs for advancing general science discovery, as well as goals for meeting external partner expectations.

It is a difficult challenge to select specific measures, or to determine years in advance what measures will be most relevant in the future (Goddard and Mannion, 2004). Remaining loyal to a performance measure that is not adaptive could drain program resources and lead programs away from priority science research areas as external program drivers change (Alcock, 2004). Therefore, measures of progress must be flexible enough to accommodate significant changes in the operating environment over time and to accommodate inherent risks of scientific endeavors (for example, experiments often fail, but learning and progress continues). The benefits of the proposed integrated measure approach are that success or progress of a program is based on a diversity of metrics, thereby reducing concerns related to the reliance of potentially invalid measures. The integrated measure also should reduce the potential for “gaming” performance measures (Alcock, 2004), where soft targets may be selected and measures data may be generated to meet the benchmark without concern for linkage to causal factors (Feltham and Xie, 2005).

Several authors describe in-depth processes for communicating with internal and external stakeholders regarding the development of performance measures and an associated feedback system (for example, Buckheim, 2000; Wisniewski and Stewart, 2003; Sneyd and Rowley, 2004). In the private sector, post hoc statistical studies of the adequacy of performance measures for new product development and innovation have been shown to be highly data intensive (Hart, 1993; Hauser, 2001). As baseline information on the individual metrics is developed, science programs can continue to adapt and hone targets and measures using more in-depth communication procedures within the program and with external partners. Statistical treatments also can be considered when sufficient measurement data become available. Getting performance measurement systems “Up and Running” and implemented, while working to clarify metric targets, enables organizations to reap benefits from improved communication that follow performance-management discussions (Buckheim, 2000). Heightened communication is a direct benefit that follows an enhanced focus on strategy and performance management within organizations or businesses (Buckheim, 2000).

Acknowledgments

Special thanks to the following U.S. Geological Survey colleagues for providing reviews of this report: Kay Briggs; Vivian Nolan; John Thompson; Don Dennerline; and Lianne Ball.

References Cited

- Alcock, P., 2004, Targets, indicators and milestones: *Public Management Review*, v. 6, no. 2, p. 211–227.
- The Boston Consulting Group, Inc., 2005, *Innovation 2005—Senior management survey*: Boston, The Boston Consulting Group, Inc., 23 p.
- Brown, K., Waterhouse, J., and Flynn, C., 2003, Change management practices—Is a hybrid model a better alternative for public sector agencies?: *International Journal of Public Sector Management*, v. 16, no. 3, p. 230–241.
- Buckheim, R.K., 2000, Developing performance metrics for a design engineering department: *IEEE Transactions on Engineering Management*, v. 47, no. 3, p. 309–320.
- Coe, C.K., 2003, A report card on report cards: *Public Performance and Management Review*, v. 27, no. 2, p. 53–76.
- Committee on Science, Engineering, and Public Policy, 1999, *Evaluating Federal research programs—Research and the Government Performance and Results Act*: Washington, D.C., National Academy of Sciences. 96 p.
- Committee on Science, Engineering, and Public Policy, 2001, *Implementing the Government Performance and Results Act for research—A status report*: Washington, D.C., National Academy of Sciences, 208 p.
- Feltham, G.A., and Xie, J., 2005, Performance measures congruity and diversity in multi-task principal/agent relations: *The Accounting Review*, v. 69, no. 3, p. 429–453.
- Goddard, M., and Mannion, R., 2004, The role of horizontal and vertical approaches to performance measurement and improvement in the UK public sector: *Public Performance and Management Review*, v. 28, no. 1, p. 73–95.
- Griffin, A., and Page, A.L., 1996, PDMA success measurement project: recommended measures for product development success and failure: *Journal of Production and Innovation Management*, v.13, p. 478–496.

- Hamel, G., and Prahalad, C.K., 1989, Strategic intent: *Harvard Business Review*, May–June 1989, p. 63–75.
- Hart, S., 1993, Dimensions of success in new product development—An exploratory investigation: *Journal of Marketing Management*, v. 9, p. 23–41.
- Hauser, J.R., 1998, Research, development, and engineering metrics: *Management Science*, v. 44, no. 12, p. 1,670–1,689.
- Hauser, J.R., 2001, Metrics thermostat: *The Journal of Production Innovation Management*, v. 18, p. 134–153.
- Hauser, J.R., and Zettelmeyer, F., 1997, Metrics to evaluate R, D, and E: *Research Technology Management*, p. 32–38.
- Heinrich, C.J., 2002, Outcomes-based performance management in the public sector—Implications for government accountability and effectiveness: *Public Administration and Review*, v. 62, no. 6, p. 712–725.
- Hernandez, M., Hodges, S., and Cascardi, M., 1998, The ecology of outcomes—Systems accountability in children’s mental health: *The Journal of Behavioral Health Services and Research*, v. 25, no. 2, p. 136–150.
- Ittner, C.D., and Larcker, D.F., 2003, Coming up short on nonfinancial performance management: *Harvard Business Review*, November 2003, p. 88–95.
- Kaplan, R.S., and Norton, D.P., 1996, *The balanced scorecard—Translating strategy into action*: Boston,, Harvard Business School Press, 336 p.
- Kaplan, R.S., and Norton, D.P., 2000, Having trouble with your strategy?—Then map it: *Harvard Business Review*, September–October 2000, p. 167–176.
- Kaplan, S., and Beinhocker, E.D., 2003, The real value of strategic planning: *Harvard Business Review*, v. 44, no. 2. p. 71–76.
- Kerssens-Van Drongelen, I.C., and De Weerd-Nederhof, P.C., 1999, The use of performance measurement tools for balancing short- and long-term NPD performances: *International Journal of Innovation Management*, v. 3, no. 4, p. 397–426.
- Leem, C.S., and Kim, I., 2004, An integrated evaluation system based on continuous improvement of IS performance: *Industrial Management and Data Systems*, v. 104, no. 2, p. 115–128.
- Loch, C.H., and Tapper, S., 2002, Implementing a strategy-driven performance measurement system for an applied research group: *The Journal of Product Innovation Management*, v. 19, no. 3, p. 185–198.
- Maleyeff, J., 2003, Benchmarking performance indices: pitfalls and solutions: *Benchmarking—An International Journal*, v. 10, no. 1, p. 9–28.
- Nilsson, F., and Kald, M., 2002, Recent advances in performance management—The Nordic case: *European Management Journal*, v. 20, no. 3, p. 235–245.
- Radnor, Z., and McGuire, M., 2003, Performance management in the public sector: fact or fiction?: *International Journal of Productivity and Performance Management*, v. 53, no. 3, p. 245–260.
- Sneyd, K.P., and Rowley, J., 2004, Linking strategic objectives and operational performance—An action research-based exploration: *Measuring Business Excellence*, v. 8, no. 3, p. 42–51.
- Tangen, S., 2005, Demystifying productivity and performance: *International Journal of Productivity and Performance Management*, v. 54, no. 1, p. 34–46.
- Verweire, K., and Van den Berghe, L., 2003, Integrated performance management—Adding a new dimension: *Management Decision*, v. 41, no. 8, p. 782–790.
- Werner, B.M., and Souder, W.E., 1997, Measuring R&D performance—State of the art: *Research Technology Management*, v. 40, no. 2, p. 34–42.
- Wisniewski, M., and Stewart, D., 2003, Performance measurement for stakeholders—The case of Scottish local authorities: *International Journal of Public Sector Management*, v. 17, no. 3, p. 222–233.
- Zydlewski, G.B., Haro, A., Whalen, K.G., and McCormick, S.D., 2001, Performance of stationary and portable passive transponder (PIT tag) detection systems for monitoring fish movements in streams and rivers: *Journal of Fish Biology*, v. 58, p. 1,471–1,475.

Publishing support provided by the U.S. Geological Survey
Science Publishing Network, Tacoma Publishing Service Center

For more information concerning the research in this report, contact the

Western Region Unit Supervisor
Cooperative Fish and Wildlife Research Unit Program
U.S. Geological Survey
12201 Sunrise Valley Drive, MS 303
Reston, Virginia 20192
<https://www.coopunits.org/>

