

# **Towards Designing a Performance Measurement System for the CGIAR**

*Draft*

*September 2003*



CGIAR Working Group on Performance Measurement

September 9, 2003

Mr. Ian Johnson  
Chair, CGIAR ExCo

Mr. Francisco J. B. Reifschneider  
Executive Secretary, CGIAR/ExCo

---

Dear Ian and Francisco:

The CGIAR Working Group on Performance Measurement (WGPM), established at ExCo's request by the System Office, has completed its initial work and recommends that ExCo/CGIAR initiate the design and development of a CGIAR-wide performance measurement system along the lines described in the attached report of the Working Group.

We would like to stress that this is not an easy task, but that advances in this area would improve the CGIAR's overall orientation towards performance and avoid unnecessary duplication in development of performance indicators to be used system-wide. The next stage of the effort (i.e., design and development) will require oversight by a Steering Committee. Our colleagues in the Working Group and we are prepared to continue helping the CGIAR and ExCo in whatever capacity is appropriate during the next stage of this effort.

We propose a tight timetable for the design and development of an initial performance measurement system. Although it would not be perfect, we should have an implementable performance measurement system by July 2004.

At our meeting on September 5, the Working Group identified a number of issues that should be taken into account in designing and developing a performance measurement system. Most of these are highlighted in the attached report of the Working Group. However, some merit additional emphasis. These relate to the purposes and the principles that should guide the effort.

## Purposes

The WG agreed that the CGIAR's PM system should serve multiple purposes. The primary purpose should be the promotion of high Center performance, through providing incentives and stimulating learning and change. The system should be developed and implemented in full consultation with the Centers to encourage Center buy-in and use of the system. Secondly, we see the establishment of a PM system as a tool for improving transparency, demonstrating accountability and thirdly as an additional instrument that could be used by interested members in making their resource allocation decisions. However, the WG stresses that performance measurement information should be used thoughtfully in assessing and comparing Center performance. It is vital that decisions made on the basis of assessments of Center performance be made with a full understanding of the relevant circumstances. Performance information is not a substitute

for detailed evaluations and reviews. Rather it should be viewed and developed as a complement to existing processes for assessing Center performance.

Centers are currently burdened with a variety of ad hoc evaluations and data collection requirements. The PM system should be developed with these requirements in mind so that it could serve as a possible substitute for some of the information collection by individual CGIAR members and other stakeholders.

The complexities of developing a performance measurement system and the need for consultation and buy-in at all stages, suggest that a long process will be needed. However, our discussion revealed pressures from various quarters for a more rapid development process, that could run counter to these needs. These two competing pressures will have to be carefully balanced. One way of reaching this balance is to establish several “expert teams” that would develop different elements of the system in parallel (e.g., finance, science quality and relevance, HR, etc) The recommended Steering Committee could provide the overall coordination.

## Principles

The WG group proposes that the performance measurement system initially rely, to the extent possible, on center self-reporting, which would require a process for external validation of the reported information through the System Office or another mechanism. As the system evolves, the need for additional information collection, such as regular client surveys may become evident. Even where there may be existing data to draw on, Center information systems may need to be upgraded and some aspects of Center information collection harmonized to develop performance information that meets the needs of the CGIAR.

There was much discussion of the need for a balanced set of indicators and measures. Reliance on individual indicators could divert Center performance in a single direction, to the detriment of achieving other organizational goals. Also, because of the diverse nature of Center missions and mandates, variations in their size, and the different circumstances under which each operates, certain indicators and elements would not apply equally to all. Indicators should be chosen to apply to as broad a range of centers as possible. Measurement techniques, such as those used by EMBRAPA, should be developed to make appropriate adjustments where this is not possible.

There will also need to be a careful balancing of quantitative and qualitative measures. For example, some center accomplishments measured through qualitative means may be of such significance that they could outweigh others. These would not be made visible if there was exclusive reliance on quantitative measures.

The WG identified 8 elements of performance, which are summarized in the paper. The discussion at the WG meeting revealed issues with all of them that need to be addressed in the next stages of development. There was consensus that the first element “Center

accomplishments” would be the most challenging, because of the inherent difficulties of assessing the outcomes and impacts of scientific activity. There was much concern that although information on Center outputs could be collected annually, this would not be true of outcomes and impacts. Impacts were considered especially challenging to address via performance measures. Because of these challenges, the new Science Council should play an important role in the design of the PM system, particularly with regard to designing indicators of science quality, relevance and accomplishments. Other issues that arose in the discussion of the indicators included the need to address overlap among the elements of performance, to develop clearer view on science quality and relevance, and to have clearer definitions of partnerships.

In conclusion, we commend the report of the Working Group for ExCo’s consideration. The recommended performance measurement system would fill a major void that exists in the CGIAR.

With best wishes,

Kevin Cleaver and Luis Arango  
Co-Chairs

## **Table of Contents**

### **1. Introduction**

- 1.1 Why Develop a Performance Measurement System for CGIAR?
- 1.2 The Worldwide Move Towards Performance Measurement
- 1.3 Key Definitions

### **2. Performance Measurement in the CGIAR**

- 2.1 Implications for CGIAR of Advances in Performance Measurement
- 2.2 The Purposes of Performance Measurement in the CGIAR
- 2.3 How does PM fit into CGIAR's Planning and Evaluation Processes?
- 2.4 Possible Elements of a Performance Management System for the CGIAR

### **3. Recommendations**

References

Annexes:

- 1. Composition of the CGIAR Working Group on Performance Measurement (WGPM)
- 2. Background on Global Trends, Concepts and Lessons Learned in Performance Measurement
- 3. Synopsis of the EMBRAPA "Evaluation and Rewarding System Based on Results (SAPRE)"
- 4. Illustrative draft of potential performance indicators and possible information sources

# 1. Introduction

## 1.1 *Why Develop a Performance Measurement System?*

Over the past year, the need for taking a fresh look at performance measurement issues in the CGIAR was stressed by various CGIAR bodies such as CGIAR Executive Council (ExCo, Sept. 2002) and the Working Group on the Establishment of the Science Council (WGSC, August 2002). Also, the CDC expressed interest in developing measures of Center performance, as part of an overall effort to reform the CGIAR's evaluation system.

The CGIAR Working Group on the Science Council urged a “much wider use of self-assessment in reviewing and enhancing the quality of science in the CG System” and suggested consideration of a performance assessment framework based on quantitative indicators. The recent iSC proposal on “Changing Monitoring and Evaluation in the CGIAR System” echoed the sentiment by pointing to the need for a continuous self-assessment that covers such matters as the relevance of research and related activities, quality of science, outputs and impacts, the efficiency and effectiveness of operations, including partnerships (iSC Secretariat, June 2003).

At its third meeting ExCo concluded that the System Office should bring recommendations to ExCo on how to approach the question of performance measurement on a system level. As a follow up, the System Office developed a concept note on establishing a Working Group on Performance Measurement in the CGIAR. A two-phase approach was recommended: in phase 1 the WG would develop options for performance measurement and in phase 2 the chosen option would be designed, developed and implemented.

The System Office established the Working Group on Performance Measurement in May 2003, under the Co-Chairmanship of two ExCo members: Kevin Cleaver (ExCo/FC Chair) and Luis Arango (ExCo/PC member). Other members represent a cross section of interests and expertise from inside and outside the CGIAR (see Annex 1 for the membership of the WG). Three sets of activities were carried out in preparation for the first meeting of the WG on September 5, 2003:

- (1) The CGIAR Secretariat prepared a *Sourcebook on Performance Measurement in Research Institutions and Programs* as background on approaches and methods of performance measurement being used in similar organizations globally.
- (2) Members of the WG shared additional information relevant to the objectives of the exercise (e.g., papers, articles);
- (3) A sub-group of the WG (made up of technical experts and resource persons<sup>1</sup>) met for a two day workshop on August 11-12, 2003 to discuss and outline

---

<sup>1</sup> Flavio Avila, Stan Divorski, Ruben Echeverria, Doug, Horton, Maria Iskandarani, Mortimer Neufville, Selçuk Özgediz, Ray Rist.

performance measurement options that could be considered by the WG at its September 5 meeting, as a means of facilitating the task of the WG. This paper reflects the outcome of this preparatory workshop.

The paper is organized as follows: the remainder of this chapter discusses the rationale behind the worldwide trends towards Performance Measurement and offers definitions of some key terms. Chapter 2 focuses on the CGIAR, describing potential purposes and uses of performance measurement, identifying possible key elements of a performance measurement system, and outlining how such a system could fit into the planning and evaluation processes of the CGIAR. The final chapter summarizes the main conclusions and recommendations.

## **1.2 *The Worldwide Move Towards Performance Measurement***

National and state governments around the world are implementing requirements for performance measurement and annual performance reporting by government agencies. This trend was fuelled by the economic downturns and increasing government deficits of the 1980's and 90's and the concomitant decrease in public confidence. The use of information on the results of government programs was seen as a rational basis for encouraging the efficient use of resources and making the hard allocation decisions that needed to be made. Use of performance information as a tool for greater transparency and reporting of government accomplishments has been seen as a response to lowered confidence.

Performance measurement is also a response to management's need for current information in a world of increasingly fast paced change. In the early 1970's Canada and the U.S. had implemented policies requiring agencies to assess how well their programs were working through detailed analytic studies, called "evaluations" that were conducted from time to time. Eventually, dissatisfaction grew with these studies because of the time they took to complete and because the costs of conducting them meant that only a limited number of programs could be studied each year. Consequently, managers began developing "performance measurement" or "performance monitoring" approaches that collected and reported performance information on a recurring and timely basis. These approaches provide an "early warning" system that allows managers to detect and solve potential problems at an early stage.

Some brief examples countries of countries that have implemented requirements for government agencies to set performance targets and report performance against them are provided in Annex 2. The requirements for national level agencies to set performance targets and report against them have cascaded down to the non-governmental agencies these support. Meeting their performance reporting requirements have led federal government agencies in Canada and the U.S. to require performance information from NGO's. Although NGO's initially developed performance measurement systems to meet the myriad reporting requirements placed on them, they have gradually seen PM as a means of improving their own management and effectiveness. For example, a 1999

Canadian report of the Panel on Accountability and Governance in the Voluntary Sector encouraged NGOs to build voluntary organizational capacity and recognition by funders of the need to focus on performance; identify outcome goals; develop ways to measure progress and achievement of goals; collect and analyze the data; and disseminate outcome assessments to stakeholders and use them in planning.

Research organizations that are government sponsored or that obtain a significant proportion of their funding from government sources have been affected by the requirements for performance measurement, and many have moved to implement performance measurement systems. Some examples are provided in Annex 2 and are discussed more fully in the *Sourcebook on Performance Measurement in Research Institutions and Programs* (CGIAR, 2003). In particular, a model developed by EMBRAPA deserves careful consideration by CGIAR as it is one of the most advanced models for performance measurement by a scientific institution. For this reason, a lengthier description of the model is attached as Annex 3.

### **1.3 Key Definitions**

It is useful to clarify differences between three related concepts to avoid confusion and foster a better understanding of performance measurement. The three concepts are: *evaluation*, *performance management*, and *performance measurement*.

**Evaluation** refers to a systematic and objective assessment of an on-going or completed project, program, policy, or institution, covering its objectives, design, implementation and results. The aim is to determine the relevance and fulfillment of objectives, efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process.

Evaluation is an analytical attempt to determine the worth or significance of an activity, policy or program, or of the accomplishments of an institution. The focus is on the past and the present and on results (outputs, outcomes, impacts). Because it is comprehensive, it would be inefficient to conduct evaluations frequently. In the case of the CGIAR, the System Reviews and External Program and Management Reviews serve as examples of major evaluations.

**Performance management** (which is sometimes referred to as *results-based management* or *managing for results*) refers to using information on performance to guide decision making on future goals, plans, and institutional actions. It is a continuous process of translating overall institutional goals into individual actions and outputs, or aligning strategic goals with intermediate outcomes and activities at all levels within the institution (e.g., teams, individuals). A comprehensive *performance management system* would include integrated planning, performance measurement, evaluation and reward systems for different levels in the organization.

**Performance measurement** is a necessary but not sufficient element of a performance management system. It refers to the ongoing process of gathering information on dimensions of performance considered to be important by the organization for the purposes of monitoring the institution's standing with respect to these dimensions. The information gathered usually covers *results* (so that it could be used as an instrument in performance management systems) as well as data on *inputs* (e.g., resources, staff), *processes* (e.g., the science that is practiced, human resource management processes), and *efficiency* (e.g., productivity). Ideally, the set of indicators making up a performance measurement system would reflect the ability of an organization to achieve and sustain excellent results.

## **2. Performance Measurement in the CGIAR**

### **2.1 Implications for CGIAR of Advances in Performance Measurement**

The conceptual advances that have been made in the field of performance measurement in recent years, and the experience gained by other organizations in using it, suggest that it is timely for the CGIAR to take a fresh look at the way it monitors the implementation of the activities supported by its members. This is reinforced by increasing demands by donors and other stakeholders for greater transparency and accountability in the operations of the CGIAR. It should be recognized that performance measurement is new enough in most jurisdictions that not much is known about the extent to which the implementation of PM systems actually improves organizational performance. However, some examples can be found of the successful implementation of performance management. The American Society for Public Administration<sup>2</sup> has published 28 case studies outlining the successful implementation of performance management approaches in the U.S. Studies of managing for results by the Office of the Auditor General Canada<sup>3</sup> have identified some examples of successful performance management in the federal government of that country.

Common conceptual models used for monitoring organizational performance and the lessons learned by other institutions are summarized in Annex 2. These show that performance measurement should be more than a simple measurement of a set of output indicators. Generally, the CGIAR Centers use *resources and inputs* (funds, personnel, equipment, and facilities) to undertake their *research operations* in order to produce *outputs* (e.g. agricultural technologies and services) for the benefit of farmers and other users. The *outcomes and impacts* of adopting or applying these outputs are measured by

---

<sup>2</sup> <http://www.aspanet.org/cap/cases.html>

<sup>3</sup> Auditor General of Canada. 1997. *Report of the Auditor General to the House of Commons*. Ottawa.

their effects on for example production cost, yield, income, natural resources, etc. Ideally a performance measurement system would capture this whole chain of causality.

Experiences of others also show that that for performance measurement to succeed there must be agreement on the purposes to be achieved by the performance measurement system. Another lesson is that a PM system should not be “copied” from another organization, rather it needs creative development, stepwise adjustment, and, most importantly, commitment from all stakeholders. Moreover, the CGIAR has its own history, established values and other special characteristics that need to be taken into account in designing such a system. Therefore, one should not be surprised that development of such a system would take time, and start-up problems would need to be overcome before a functioning PM system could deliver robust information for use by all.

## ***2.2 The Purposes of Performance Measurement in the CGIAR***

The CGIAR has a strong and well-recognized track record when it comes to performance. The centers are known as “centers of excellence” because of the reputation of their scientific achievements. External reviews of the System and of the centers have documented the significant achievements and impact of the centers.

The CGIAR and the centers also employ a range of tools for monitoring the quality and impact of the System’s activities. However, the existing tool set needs to be adapted to the needs of a rapidly changing environment and the demands from stakeholders. The various partners and stakeholders in CGIAR have been asking a variety of questions about the continued financial and institutional health of the centers, the performance of the Boards, maintenance of the centers’ comparative advantage in quality of science, the range of outputs generated annually, etc.

A well-designed performance measurement system could provide answers to most of these questions. The information generated through such a system would have a variety of potential uses, including:

- Serving as a tool for decision-making and/or performance management by the centers,
- encouraging centers, through incentives, to perform better,
- stimulating change and learning through benchmarking performance,
- demonstrating accountability, and
- aiding in resource allocation.

***Decision-making and Performance Management.*** Centers could benefit from having a performance measurement system that not only monitors accomplishments, but also monitors and provides timely information on their potential to perform well in the future. Such information, appropriately adjusted, could be used by the centers as part of their internal *performance management* effort. It would help them to identify areas needing

early attention and support efforts to learn from their own performance, innovate, and make appropriate changes to improve performance. In addition, having uniform indicators of performance across the centers would enable the System to monitor its progress towards desired goals—as a System.

***Incentives for High Performance.*** Using performance measures to reward high performance would provide incentives for centers to strive for even higher performance. Using performance information solely to reduce allocations to centers that perform less well could result in efforts to undermine the system through non-cooperation, poor quality information, efforts to “game” performance measures, and other efforts to hide less than desired performance. There is thus, need for a balance between positive and negative incentives for performance.

***Stimulating Change and Learning through benchmarking.*** Performance measures that are consistently and uniformly applied across centers support institutional learning and change by allowing center performance to be benchmarked against other centers, and the collective CGIAR performance to be benchmarked against similar organizations. Benchmarks would allow a center, or CGIAR as an institution, to identify areas of high performance so that best practices can be identified and adapted, as appropriate, to the needs of individual centers.

***Accountability.*** CGIAR’s structure as an alliance of equal and autonomous partners (i.e., centers) is similar to structural models that many governments have been experimenting with, and performance measurement in such a system is important for many of the same reasons. Each must do their part if agricultural research is to contribute to reducing hunger and poverty and each is accountable to the other for the results they achieve, how they achieve the results, learning from experience and improving their contribution to the common goal, and the efficient use of resources. A performance measurement system is a valuable tool in providing the information needed to demonstrate these accountabilities to each other, to the stakeholders, and to their investors.

***Resource Allocation.*** In the public sector, one frequently cited purpose of performance measurement has been to use the information as a basis for resource allocation. Unlike the public sector, in the CGIAR there is not one, but 62 (as of this writing) actors which face a resource allocation problem, mostly on a yearly basis. A well-conceived performance measurement system could provide valuable information to each of these “investors” that they could take into account when making their resource allocation decisions. This could be but one factor that enters into each investor’s calculus in allocating resources. Because the reasons for low performance one year can be complex and varied and can include factors largely outside the control of the organization, the performance data need to be used cautiously. When the reasons for low performance one year are assessed, the appropriate response could very well be to increase, rather than decrease resources.

## **2.3 How does PM fit into CGIAR's Planning and Evaluation Processes?**

Performance information is more likely to be used and useful if it feeds into an organization's decision making processes. The CGIAR has various planning and evaluation processes in place, which are currently being reviewed. Introduction of a performance measurement system at this point in time would enable the CGIAR to make adjustments to the currently existing processes in order to maximize efficiency.

One of the existing tools of relevance to performance measurement is the Logframe approach, introduced as a program or project planning tool in 1998. The intention was to increase the efficiency and effectiveness of research and research management at all levels. The main objectives were: (1) to have an output oriented approach in research planning and research management (2) increased accountability at all levels, (3) increased transparency, (3) increased compatibility of management tools. Although it was not comprehensively implemented across the system as a management tool, some Centers used the Logframe and the centers continue to report their program and budget proposals using the output categories of the Logframe. The lessons learned from its implementation would be of great value for the design of a PM system.

On the evaluation side, the CGIAR System and Centers have developed an elaborate system of monitoring and evaluation, consisting of Center reviews, inter-Center thematic or systemwide program reviews and System reviews. These include the following:

1. **CGIAR System reviews** – undertaken periodically (three since the founding of the CGIAR in 1971) to evaluate the overall performance of the CGIAR System at large
2. **External Program and Management Reviews of Centers (EPMRs)** – organized by the Science Council and the CGIAR Secretariat every five years to evaluate respective Centers' programs and management;
3. **Centre Board Commissioned External Reviews (CCERs)** - commissioned and managed by the Centers themselves, as in-depth evaluations of the relevance and quality of science in specific research programs or as assessments of specific areas of operations and management;
4. **Inter-center thematic (stripe) reviews** - commissioned by the Science Council to evaluate specific high priority themes, e.g., Inter-Center Roots and Tubers Review; Review of Plant Breeding Methodology;
5. **External Reviews of Systemwide Programs** – SC-commissioned reviews of the established Systemwide Programs (SWPs), e.g., Systemwide Genetic Resources or Integrated Pest Management Programs;

**6. Project reviews** - commissioned and/or conducted by donors at specific intervals in the project life-cycle (typically at the end of the project) or on a periodic basis, e.g., annually (e.g., donors' special project reviews or reports);

**7. Center managed reviews** - i.e., internal mechanisms to ensure science and management quality control (audits, internal project evaluations, annual work-planning meetings, staff performance assessments, etc.).

Although this system of reviews is quite elaborate and a lot of information is gathered, there is concern about the effectiveness and efficiency of the current monitoring and evaluation system (iSC, 2003; CGIAR Sec, 2003):

- The external reviews of centers are the main mechanism through which the System gathers information on center outputs, outcomes and impact. But much of this evaluative information is qualitative and does not permit comparison across centers or groups of centers, or against benchmarks. In addition, there are several impact studies commissioned by SC-SPIA or the centers, but these are uneven in terms of their coverage.
- A five-year cycle of external reviews may not be frequent enough for monitoring a Center's evolution over time, nor for purposes of detecting and resolving problems at an early enough stage. It is desirable to have a more systematic and continuous process that can complement the five-yearly reviews.
- Donors tend to have different formats for the reporting they require on project activities. The numerous inconsistent reporting requirements lead to substantial time commitments by center scientists.
- CCER's are of variable quality and coverage, although some centers have made significant strides in establishing a comprehensive peer review system.

A performance measurement system for the CGIAR should be designed in such a manner that it helps to overcome the shortcomings of the current review systems, helps better serve donors' annual information needs about the system as a whole and gives both, donors/stakeholders and Centers a tool to assess potential performance. A properly designed system will complement the current evaluation system of the CGIAR, allowing the peer-review based evaluations (EPMR, CCER) to be more focused, and to benefit from the annual performance measures. The framework should be designed in such a way that it provides:

- an annual assessment of performance that identifies successes and areas where performance has been less than expected,
- an annual assessment reflecting the center's potential to perform in the future,
- a framework for promoting results-based management,
- an approach that satisfies donors and other stakeholders transparency and accountability needs, and
- an instrument for rewarding performance.

## **2.4 Possible Elements of a Performance Management System for the CGIAR**

The conceptual models of performance reviewed in Annex 2 are not specific to international agricultural research. They must be adapted to the unique nature of CGIAR center accountabilities and international relationships, and to the specific and varied nature of their research. For this reason, a performance management system for the CGIAR should be developed from the ground up, not by “copying” the model used in another organization. Having said this, it is also clear that the generic models presented do offer many useful insights to what factors contribute to performance in organizations.

The experiences of others, and the literature reviewed, suggest that two groups of factors should be thought of as candidates for a performance measurement system:

1. Elements reflecting the *results* dimension of the organization’s work, i.e., its outputs, outcomes, impacts, and efficiency ;
2. Elements reflecting the organization’s *potential to perform*;

The Working Group made an attempt to identify a set of elements that fit the special circumstances of the CGIAR that could be included in a future performance measurement system. These include the following:

### **Results**

- Center Accomplishments
- Efficiency
- Stakeholder Views

### **Potential to Perform**

- Science Quality and Relevance
- Partnerships
- Governance/Institutional Health
- Financial Health
- Culture of Learning and Change

The following is a summary discussion on these eight elements. Examples of indicators that could be developed for each of these elements are presented in Annex 4.

#### **ELEMENT 1: Center Accomplishments in Relation to Mission and Objectives**

At the core of assessing “how well a center is doing” is information on the extent to which the center provides outputs that lead to outcomes and impacts consistent with its mission and objectives. “Outputs” are the immediate products of a center activities, e.g. research publications, policy recommendations, improved plant types, etc. EMBRAPA’s

efforts at developing and reporting output measures can serve as an example for developing center output measures.

“Outcomes” refer to the immediate consequences of these products on center stakeholders and clients, e.g. changes in knowledge, attitudes, policies, research capacities and agricultural practices. Typically the outcomes of scientific research have been measured through qualitative means, particularly expert opinion provided by peer reviews such as the CCERs. There has been some development of quantitative indicators of short term outcomes, such as the volume of research stimulated elsewhere and the economic benefits stemming from research. There is often a considerable time lag between the completion of scientific research and the development of visible outcomes, making annual measurement and reporting difficult. Some techniques have been developed to mitigate these challenges. Periodic surveys of clients have been used to obtain estimates of outcomes. Some organizations have reported annually on the average of outcomes achieved over a several year interval (“moving averages”). For example, the U.S. General Accounting Office tracks the response to its recommendations over several years and reports annually on accomplishments of work completed in the previous three years. It may also be advisable for CGIAR to consider the adoption of new mechanisms, such as peer review panels that examine center performance on a more frequent schedule. However, none of these solutions are perfect, and the challenges remain.

“Impacts” refer to the longer range social, environmental and economic benefits that are consistent with the center’s mission and objectives, e.g. increased agricultural productivity, improved food distribution, etc. The impacts of science are particularly difficult to assess. They are often influenced by a number of factors, with science playing a limited contribution. As a result, it is difficult to attribute changes in these areas to any particular scientific input. Time lags for their realization can be even longer than for outcomes, perhaps taking more than a decade. The assessment of impacts often requires complex evaluation studies, which can only be conducted from time to time. The impacts of science can also differ widely in terms of their importance. Some breakthroughs may be so significant that in and of themselves they justify the resources used by a center. It is important that the performance measurement system reflect these differences and that centers be given the opportunity to highlight these “significant accomplishments”.

The problem of attributing results to a particular action is often a problem for measuring outcomes as well as for measuring impacts. In the absence of clear evidence for attributing results, it is important to establish a clear logical link. A basic principle of performance measurement is that it should provide a picture of how outputs lead to outcomes which in turn contribute to impacts consistent with the center’s mission and objectives. A common practice is to develop a framework that depicts these relationships and provides the basic direction for selecting performance indicators for accomplishments. The current CGIAR Logframe could serve as a starting point for developing indicators and demonstrating linkages among outputs, outcomes and impacts.

The work of the SC Standing Panel on Impact Assessment is an important source for the development of indicators of outcomes and impacts. Developing indicators of outcomes

and impacts will in all likelihood involve finding the proper balance between qualitative and quantitative assessments of impacts. Even when qualitative judgments are reached it is important they have a sound factual basis.

Indicators will also have to take into consideration the differing mandates of individual centers. Not all centers will be able to show achievements with regard to germplasm development and conservations, for example. Others may not conduct policy research. Capacity building of NARS, for example, is an objective of all centers. Indicators will have to be developed in such a way that a common level of assessment is available while allowing sufficient flexibility so that the assessment of a given center's accomplishments reflect its mission and mandate. For example, EMBRAPA groups specific individual output indicators that may not apply to all of its research units into general categories that have a broader pertinence. In combining individual indicators into overall indices, EMBRAPA applies a multiplying factor (weight) that adjusts the importance of each indicator according to its relevance to the unit's research emphasis.

Several existing sources of information can be drawn upon in developing indicators and related measurements, including impact assessment studies, records of research outputs currently maintained by centers, external CGIAR reviews and external center reviews. However, it is likely that additional information sources will need to be developed to report systematically on certain accomplishments, especially outcomes. It is likely that centers do not currently all collect the same information on accomplishments. Some harmonization of information collection will be needed, and information systems in many centers will have to be upgraded if centers are to be compared or if their accomplishments are to be added together.

## **ELEMENT 2: Efficiency**

A center's outputs are an important indicator of its progress in moving toward its goals and objectives. The efficiency with which it produces its outputs is an important indicator of its ability to manage its performance under changing circumstances. By efficiency the WG means the human and financial resources required to produce a unit of output.

In a changing environment, a center will need to continually find new ways to maintain and improve its efficiency, therefore, indicators should not merely provide a static snapshot of efficiency, but should track changes in this area. The basic measure of efficiency is usually the ratio of outputs to inputs. However, it is important to link this measure to an indicator of continued improvement, such as year to year trends in this ratio. It is also important to have an indicator of the flexibility that a center has in improving its efficiency, for example by tracking the ratio of its fixed costs to its total expenditures. Much of the required information on the input side should be available from existing center sources and financial statements. The challenge will be in generating data on outputs. Another challenge will be to recognize the differences among centers in terms of their mandates (and outputs), perhaps by developing a weighting system such as the one used by Embrapa.

### **ELEMENT 3: Stakeholder Views**

A positive image of the centers and the CGIAR among donors, stakeholders and the public fosters continued public support for their objectives and activities. This support contributes to the establishment of external partnerships and to the take-up of CGIAR and center research products, thereby contributing to the outcomes and impacts to be derived from these products.

Satisfaction with individual centers may vary and may not represent satisfaction with the CGIAR as an entity. This should be factored into any effort to develop indicators and measures of stakeholder satisfaction.

Assessment of center/CGIAR image would require such new information collection activities as surveys of clients/stakeholders and content analysis of public statements about nature of the centers and CGIAR. An example is provided by Embrapa's development of an Image Index based on a survey of client and stakeholder opinion.

### **ELEMENT 4: Science Quality and Relevance**

The CGIAR is a "science system", central to its credibility is the quality and relevance of research undertaken by the centers. Science quality and relevance also provides an indication of the extent to which current research will continue to yield accomplishments consistent with centers' missions and objectives.

Science quality can play a useful role in helping to estimate center output accomplishments and efficiency. Many of the usual science output measures, such as the number of publications in journals, and efficiency measures based on them, can be misleading if the quality of the published research is not taken into consideration. Estimation of a center's research quality and the development of a quality based weighting can help adjust output measures to make them a more valid indicator of accomplishments. An example is provided in the weighting system for EMBRAPA's efficiency indicators, which adjusts the number of journal articles published using a multiplier that reflects the importance of the journals in which the articles were published.

Science quality is best estimated through review process involving experts. In addition, for performance measurement purposes. Consequently, for performance monitoring purposes, it may be useful to develop short term quantitative indicators such as the number of honors and awards and citation counts in the Science Citation Index (taking into proper account the disadvantages of citation analysis). It may also be useful to consider developing a system of more frequent quality ratings by panels of independent experts.

Much of the information required on science quality can be obtained from existing sources such as CCER, other external reviews and center annual reports. Relevance would need to be assessed on a case by case basis, in view of the mission of each Center.

## **ELEMENT 5: Partnerships**

Partnerships with other centers and external organizations allow centers to accomplish more than they could by themselves by complementing existing human and financial resources. Increased involvement with other CGIAR centers enhances progress toward common goals, while involvement of external organizations extends the reach of center activities. Partnerships also allow centers to contribute to a variety of development goals, such as enhanced North-South cooperation. Hence, measurement of the extent to which centers enter into partnerships and of the nature of these partnerships provides an indication of their capacity to maintain and enhance their performance under changing circumstances.

The nature of partner involvements can vary. Partners can be involved in assisting centers to prepare and conduct research, and prepare products resulting from center research and other activities. Different kinds of partners are also possible. For example, partners can be internal to CGIAR, such as other centers and the CGIAR System Office. Other partners can be external, e.g. NARS and ARIs. Indicators should reflect the mixture of partnerships involved and perhaps be weighted according to significance of the partner, e.g. give greater weight to relationships with NARS or other CGIAR centers. Inter-center partnerships can help the CGIAR System function as a “system.”

Partnerships can involve different levels of partner support, ranging from the contribution of significant financial resources to simply providing public confirmation of the need for the research. If indicators are to support benchmarking or resource allocation decisions, there must be some process for ensuring that contributions of other parties are counted as partnerships only if their participation is meaningful. Indicators should provide a clear standard as to what constitutes a meaningful partnership. For example, standards could be set as to the level of funding share or real contribution needed to constitute a “partnership”. Another possibility would be to establish a review process for determining whether a party’s involvement rises to the level of a “partnership”. Partnerships with NARS are very important because they also help develop capacity in NARS—another CGIAR objective.

Some information needed to measure partnerships will be available from existing self-reporting by centers, although this reporting may need to be modified for the purposes of performance measurement. Additional information could be made available through analyses of information from the Sciences Citation Index. Certain partnership indicators, such as shared funding with stakeholders and co-authorships, may require additional analyses of information available from center records.

## **ELEMENT 6: Governance/Institutional Health**

Continuing success and improvement depend upon a center’s being effectively governed and on the health of its internal institutional climate. In a performance measurement system, it is especially important to have indicators of the extent to which the organization’s culture fosters performance. It is also important to monitor the extent to

which the organization has in place key processes that support governance and contribute to institutional health. These include processes for:

- turnover in top management;
- human capital management that ensure the organization has and will continue to have the appropriate level and blend of capacities for accomplishing mission and objectives;
- gender balance and diversity in staffing;
- effective stewardship of resources;
- internal audit;
- following up on recommendations from reviews, audits and evaluations
- self-evaluation, including self-evaluation by boards of trustees
- leadership development;
- organizational climate/culture (e.g., staff satisfaction, processes for handling grievance)

Some of the required information may be available from existing external and internal reviews, although the current multi-year cycle for these limits the extent to which they can be used in annual performance reporting. Information on such matters as staff satisfaction would require new information collection activities for most centers. Some information, such as on processes for follow-up of recommendations and on the results of follow-up, will require centers to develop additional procedures and information reporting.

## **ELEMENT 7: Financial Health**

Financial health – the effective acquisition and management of financial resources - is fundamental to a center’s performance. For an organization to make reasonable progress towards its goals and objectives, it must have the required resources available when needed. Sufficient flexibility in its financial situation helps an organization be innovative, take advantage of new opportunities and adapt to changing circumstances. Reporting on financial health is also a fundamental aspect of any accountability relationship and helps funding organizations to make resource allocation decisions.

Financial Health indicators currently used by the CGIAR provide an example of indicators that could be used, given further refinement:

- Overall financial outcome for the year
- Liquidity (working capital)
- Fixed cost ratio
- Personnel expenditures as a % of total expenditures

A center’s financial flexibility is also reflected by indicators such as the ratio of projected to actual expenditures and the proportion of its funding that is unrestricted.

Science is unique in that its progression is difficult to predict, making financial projections difficult. In addition, in the case of the CGIAR, the instabilities in the donor funding environment introduce yet another uncertainty. For these reasons, there may be large, naturally occurring yearly fluctuations in financial health indicators. This may require the use of specific techniques such as moving averages to account for year to year fluctuations outside the control of center management. Adjustments may also have to be made to indicators to account for the fact that not all centers entail the large fixed capital costs associated with laboratories and farms.

Much of the required information should be available from existing center financial statements and annual reports.

### **ELEMENT 8: Culture of Learning and Change**

Performance measurement is not only a tool of external accountability, but also a tool for self learning that an organization can use to help maintain and improve its own performance. For an organization to use performance measures, it must have the needed capacity for change and a willingness to change. As science organizations, centers have shown a capacity to learn from research results and make appropriate changes to their research activities in order to advance science. Strengthening their orientation to learning and change and extending it to such new areas as greater responsiveness to stakeholder needs will help centers ensure that past levels of success will continue into the future and perhaps be exceeded. Indicators of a center's capacity to learn from its performance and environment and make appropriate changes to enhance performance will help donors and stakeholders assess whether past successes can be expected in the future. They will also help centers assess and improve their own orientation to change.

Current models for organizational learning and change were developed for private sector organizations. Although there has been effort to adapt these models to public sector organizations, developing the needed indicators will require identification of a model specific to international agricultural research. However, some basic principles apply that can serve as a basis for developing performance measures. For example, a center's orientation to learning and change will be strengthened to the extent that in has in place:

- adequate mechanisms to support and manage change;
- mechanisms for regularly updating strategies;
- a system for rewarding innovation and change; and,
- a system for internal evaluations that are used to improve performance.

Learning and change also require the right staff mix and training, reflected in such indicators as proportion of budget spent on training, age profile, influx of "new blood", healthy turnover rates, etc.

Much of the required information can, or could be collected from existing information sources such as external reviews, center self reports and mid-term plans. Assessment of willingness to change would require additional information collection, such as stakeholder and staff surveys.

### 3. Recommendations

The foregoing discussion indicates that a variety of approaches could be used in developing a performance measurement system. The specific approach to choose would depend on the objectives sought of such a system. In identifying possible options for the CGIAR, the WG assumes that the CGIAR's PM system should serve the purposes outlined in section 2.2. The primary emphasis should be on using performance measurement to promote high performing organizations by:

- using it as a tool for decision-making and/or performance management by the centers;
- encouraging centers, through incentives, to perform better; and,
- stimulating change and learning through benchmarking performance.

The system should also support transparency and accountability; and could serve as an input in resource allocation.

Given these purposes, the most critical choices for the CGIAR involve:

- the “content” of performance, i.e., choice of aspects (or dimensions) of performance to measure; and
- the “process” of measurement, i.e., the level of detail and intensity involved in generation of performance data.

***The “content” of performance:*** As the foregoing discussion indicates, elements of center performance could be grouped into two clusters:

1. Elements reflecting the *results* dimension of the centers' work, i.e., outputs, outcomes, and impacts and efficiency;
2. Elements reflecting the center's *potential to perform*;

The *results* cluster is essential for any performance measurement system. While the indicators reflecting center results would need to be developed, the existing Logframe provides a good starting point.

The second cluster, *potential to perform*, includes several dimensions of the centers' current institutional health. If the PM system is to serve the five purposes outlined above, it would be essential to include this cluster as well in the CGIAR's PM system. While the eventual content of the cluster (the elements and specific indicators that make it up) would be up for discussion, the five elements identified earlier present a starting point. These elements reflect both the factors commonly noted to be important determinants of performance in the literature and factors that are important for the CGIAR (e.g., science quality and relevance).

***The “process” of measurement.*** Once the purposes and content of performance measurement have been identified, the choices to be made relate to the operational

aspects of how the centers and the CGIAR could develop indicators that reflect performance in the chosen areas. Methodologically this is a most challenging task because the success or failure of the performance measurement system would rest largely on the *reliability* and *validity* of the indicators chosen to reflect performance.

Building the indicator set in such a way as to rely mainly on self-assessment and reporting by the centers, with little or no new data collection, would impose fewer burdens on the centers and would cost less. However, it is particularly important that self-assessments are based on credible data and are not purely narrative descriptions of results. Even relying on self-assessment will require investments to upgrade centers' information systems and provide greater harmonization of data collection across centers. It is not clear how reliability and validity of the performance indicators would be affected by reliance on self-assessment. It would not be possible to resolve these issues until design and development begin.

## **Recommendations**

***Recommendation 1.*** The WG favors an evolutionary, learning-by-doing approach to designing and developing a performance measurement system for the CGIAR, that includes both “results” and “potential to perform” elements, based largely on self-assessment by Centers. Development of a fuller set of indicators would follow once it was clear the extent to which readily available information could satisfy the minimum reliability and validity requirements. This approach would enable the CGIAR and the centers to start with what they have (in the form of information), without going into major new data collection expeditions. New approaches would be explored only once it was clear that they were necessary. The eventual aim would be to have a fully-developed PM system in place in the medium term.

***Recommendation 2.*** Any data generation system that is based on self-reporting needs a parallel verification system, as is practiced by organizations using self-assessment and reporting. The WG recommends that an audit system should be developed simultaneously with the design of the performance measurement system to ensure the users that the reported information is accurate.

***Recommendation 3.*** The design and development of the PM system should not be a top-down exercise. Instead, it should be a collective exercise involving all major interested groups. The Science Council should take the lead in the effort to design indicators of science outputs and quality. The System Office should coordinate the remaining aspects. The effort may require having several teams working simultaneously, and all teams should have significant involvement by centers.

***Recommendation 4.*** The effort should be started as early as possible (following discussion by the CGIAR), with ExCo monitoring the development of the PM system. Experience elsewhere has shown that performance measurement systems do not necessarily work initially. It is important to start, monitor implementation closely and

fine tune as necessary. The system put in place should be evaluated three years after implementation starts.

***Recommendation 5.*** All members of the CGIAR should be encouraged to consider using the performance data generated by the CGIAR PM system, and constitute to its further improvement, instead of developing their own parallel systems.

## References

CGIAR Secretariat, Performance Measurement of Research Institutions and Research Programs - A Sourcebook, Draft, Washington, June 2003.

Cozzens, S. 2002. Research Assessment: What's Next? Final Report on a Workshop. *Research Evaluation*. 11(2): 67-79.

European Union. 2002. *The Common Assessment Framework - Improving an organisation through self-assessment*, Copenhagen 2002, <http://www.2qconference.org/filer/404/CAF-english.pdf>

Heinrich, C. 2003. Measuring Public Sector Performance and Effectiveness. In Peters. B, and Perre, J. (editors). *Handbook of Public Administration*. London: Sage Publications.

iSC Secretariat, Changing Monitoring and Evaluation in the CGIAR System, Draft, Rome, June 2003.

Kaplan, Robert and David P. Norton. 1996. Using the Balanced Scorecard as a Strategic Management System, *Harvard Business Review*, January-February 1996, pp. 75-85.

Kaplan, R. S. and D. P. Norton. 2000. Having trouble with your strategy? Then map it. *Harvard Business Review* (September-October): 167-176.

Perrin, B. 1998. Effective Use and Misuse of Performance Measurement. *American Journal of Evaluation*. 19(3): 367-376.

Perrin, B. How to – and Not to – Evaluate Innovation. *Evaluation*. 8(1):13-28.

Portugal, Alberto Duque; Elisio Contini; Antonio F. D. Avila and Geraldo da Silva e Souza: Evaluation and Rewarding System based on Results, Paper presented at the Directive Committee of PROCISUR, Dec 11-12 1996 in Vina del Mar, Chile.