

Lessons Learnt From the Pilot Year 2004 Implementation of the CGIAR Performance Measurement System¹

Executive Summary

The new CGIAR Performance Measurement (PM) System was piloted in 2005 with data from 2004. Overall, the exercise has been a positive experience for all involved - on one side it helped Centers to better understand their own performance, and on the other side members expressed satisfaction with increased accountability and welcomed new information on performance which also helped in their funding allocation decisions.

The CGIAR Secretariat and the Science Council's joint assessment of the pilot experience indicate that the exercise has been useful for Members and Centers, but that several refinements are necessary. An improved PM System is expected to be in place for 2005. Suggestions for refining the performance elements and specific indicators include the following:

1. Eight Performance Elements

RESULTS DIMENSION

- **Outputs** – The output indicator “ % of output targets achieved” is appropriate, but the SC will need to work with Centers to guide the planning of output targets so they are measurable and clear, and significant enough to result in a difference towards achieving CGIAR goals;
- **Outcome** – During the initial years of the PM System implementation, Centers should be provided guidance on reporting their 5 documented outcomes that fully meet the definition of outcome (adoption, use or influence of Center outputs). The SC will then assess and rate 5 documented outcome cases. The SC rating will constitute the performance measure;
- **Impact** – Both impact indicators should be maintained with minor modifications: For the first indicator Centers should specify and document the ex-post impact assessment (epIA) activities, outputs and outcomes over the past year, particularly related to: (a) epIA studies; (b) innovation and advancement of epIA; (c) communication/dissemination and capacity enhancement; and (d) impact culture. For the second indicator Centers should be requested to submit every three years their two best completed impact studies that document the change towards achieving CGIAR goals. The PM indicator will be a SC/SPIA rating of the studies for rigor.
- **Stakeholder Perception** - The SC agrees that measuring stakeholder perceptions is important, but does not consider the tested indicator as a suitable results indicator and therefore suggests dropping it until a suitable one is identified and tested. On the other hand, shareholder and client views are considered a critical and important component of performance in most institutional performance measurement systems (or balanced score card systems)—because customer/stakeholder satisfaction is one

¹ Based on findings from the CGIAR Secretariat, the SC, the CGIAR Internal Audit Unit and considers feedback from Centers.

of the main targets for-profit and non-profit organizations aim at. This year's pilot covered only the CGIAR Member perceptions, and could be improved based on the experience with the pilot. In addition, the survey or surveys could be structured so as to capture more differentiated perceptions of a wider range of stakeholders.

POTENTIAL TO PERFORM DIMENSION

- **Quality of Research Staff:** The definition of the publication indicator should be further clarified by developing criteria for inclusion of publications in the count. Also the inclusion of the “impact factor” of journals should be considered in the final calculation of the indicator; the “no. of awards” indicator showed validity weaknesses in reflecting quality of staff. The SC recommends dropping this indicator. Alternatively, the definition of the indicator could be either designed to be further restrictive by developing criteria for the inclusion of awards in the count, or a substitute indicator yielding more valid and reliable data should be developed; the indicator “% of Center’s top choice candidates for research and managerial positions who accept the offer” should be either dropped (as suggested by the SC) or replaced by another indicator that better captures a Center’s performance or ability to attract high quality candidates. This element could be merged with Element 6 under the heading of “Quality and Relevance of Current Research”.
- **Quality and Relevance of Programs:** The publication indicator is considered suitable, but it should be explored whether it would be more appropriate to measure the publications with developing country partner authors per scientist as a percentage of total papers rather than the actual number. In addition, two other indicators could be tested next year: (i) EPMP rating of quality of research programs, and (ii) Citation indices.
- **Institutional Health**
 - **Governance:** Results from the pilot indicate that the checklist has been helpful in understanding Board performance in terms of only some aspects of organization and management of Board business and Board engagement with Centers strategic business. Therefore, complementing this checklist with other measures would strengthen this sub-element of institutional health. Options could be (i) requesting annual Board statements confirming/not confirming their fulfillment of certain Board oversight responsibilities, and (ii) conducting a Board member survey.
 - **Culture of Learning and Change:** The questionnaire should be expanded to also capture additional dimensions of learning and change, including knowledge sharing and innovation.
 - **Diversity:** The diversity indicators should be maintained and definitions clarified. Two additional indicators could be considered as complementing one of the pilot indicators (% of management positions occupied by women): (i) % of female scientists, and (ii) % of female postdocs.
- **Financial Health:** Although both indicators seem to be well established in the CGIAR, their adequacy for reflecting the financial health of the Centers should be examined, including their recommended range. Also the inclusion of additional

financial indicators should be explored, with inputs from the CGIAR finance community.

2. Performance Data Collection

- Instructions for collecting the data and definitions of indicators should be made clearer in the 2006 exercise;
- PM web site should be enhanced by including a FAQ section.

3. Verification of reported data

- Centers should be requested to submit most of the supporting documentation during the data collection phase, and should be guided as much as possible in providing the needed information to allow for more consistent back up information.
- Guidance for organizing an “*audit trail*” should be given, so that data verification can be completed more timely and efficiently.

I. Introduction

In early 2005 the new pilot CGIAR Performance Measurement (PM) System was launched and Centers were requested to submit a set of institutional performance data for 2004 through an online data collection system by mid April. The results were made available to Centers and Members in August.

Experience shows that PM systems usually need calibration over time in order to enhance their validity and reliability. Therefore this document draws lessons learnt from the pilot PM exercise for refining the set of indicators and related definitions, instructions as well as the data collection process for the coming year. The paper is based on findings from the SC, the CGIAR Internal Audit Unit (IAU), the CGIAR Secretariat and feedback from Centers.² The annex includes a summary table of the SC recommendations for implementing the CGIAR PM System in 2006 and beyond.

II. General observations and lessons learnt from the pilot performance measurement exercise

In January 2005, the pilot PM system was launched with the understanding that the pilot would start with a limited set of indicators, and based on experience it would be further developed and refined. All results from the pilot, including SC comments and assessments, have been made available as an online report to Centers and Members and the response received was constructive and encouraging.

Some Centers and Boards welcomed the pilot exercise as it helped them to better understand their own performance. Clearly internal data collection had to be newly organized but generally did not place an undue burden on Centers. It can be expected that next year's data collection will be more straightforward as the initial steps for establishing a data collection system have been undertaken. Feedback received from Centers after making available the full results indicate that Centers were actually expecting some specific commentary on their individual submissions and were waiting to provide responses to these. This is something to be considered for the next PM round, as individual advice would be appreciated not only for improving performance but also for the preparation of the following year's submission.

The donor community has been increasingly looking for objective annual performance information of (global) programs as an input into their strategic funding decision. The CGIAR decision to develop and pilot a new PM system has therefore been timely. CGIAR Members expressed satisfaction with increased accountability and welcomed new information on performance. Some members used the information as an input into their funding allocation decisions and felt increased comfort about having both

² Feedback received from individual Centers and as compiled by CDDC in a commentary on Performance Indicators 2004/5.

information on past results and indication about a Center's potential to continue to perform in future.

In general, the PM system yielded useful information and insights to all involved, while it is recognized that the system is a work in progress and will need continued refinement to achieve a reliable and valid reflection of performance over time. The following chapter describes observations on weaknesses and strengths of certain indicators and draws conclusions for improvements.

III. The Eight Performance Elements: Observations and lessons learnt

The pilot system comprises a diverse set of performance indicators grouped under eight performance elements which fall under two dimensions: (i) Results (ii) Potential to perform.

The indicators used in this pilot represent an initial set of indicators generated from a larger indicator set developed by the CGIAR Working Group on Performance Measurement (WGPM) with input from expert teams from within the System. This allowed for a transition period during which

- the CGIAR can test the validity and reliability of indicators,
- Centers can establish their own performance data collection systems,
- the new M&E for CGIAR Centers can be finalized and eventually aligned with the PM system.

The following describes the key observations made about the reliability and validity of each performance indicator, and draws some conclusions for planning next year's exercise.

Element 1: Outputs

In the pilot year Centers were asked to identify for each of the major program areas outlined in the center's MTP for 2004, the five most significant outputs and the output targets ("milestone" in the earlier MTP terminology) expected to be produced that year. The pilot indicator was the % achievement of the entire set. This was requested knowing that the earlier MTPs were not drafted with the intention to allow for a clear measurement of achievements as it was requested in this pilot PM. Nonetheless, the WGPM thought that there would be value added in getting a sense of how Centers would describe their actual achievements in relation to planned achievements and the practicality of this indicator.

In this context, the SC team examined the pilot outputs and output targets provided by the Centers in 2005 and gauged the usefulness of the PM output indicator and to make

suggestions for improving the indicator and instructions for providing data in the future (Science Council, 2005).³

Observations

Variability in scope of the reports

Center reports reflected variably in the MTP 2004 submissions. In many cases reporting of the precise wording from the MTP made it difficult to assess what kind of a deliverable had been produced. Some Centers described the actual achievements, which helped the assessment. With some Centers, the reported outputs and output targets/milestones did not match with the MTP information. Some Centers reported multiple milestones as a single Output target/milestone. Because Centers were asked to report only a limited number of targets, and because some Centers used large programs as the unit of reporting while others reported on MTP Projects, there was a large variability in the total number of targets reported. There was also a large difference in the proportion of original milestones/verifiable indicators reported (ranging from 4% to nearly 100%) (Science Council, 2005).

Imprecise use of terminology: “output” and “output target”

The SC noted that in some cases the output description was clearer than the output target description, reflecting different interpretation of the instructions. The specification of clear output targets across all Centers was also problematic. Based on a SC review of the reported output targets, about 40% met the definition of a deliverable, and 20% were considered not deliverables. Even among those that were deliverables, many were not quantified or specified.

The SC concluded that the output data reported in 2005 were not suitable to provide a meaningful basis for measuring Centre performance in producing outputs in this pilot year (Science Council, 2005).

Conclusions

The output indicator “% of output targets achieved” is appropriate, but the SC will need to work with Centers to guide the planning of output targets so that they are measurable and clear, and significant enough to result in a difference towards achieving CGIAR goals.

Element 2: Outcome

Centers were requested to describe the five most significant outcomes that appeared in 2004 out of outputs expected under the Center’s MTPs for 2001, 2002, and 2003. Outcomes were defined as changes resulting from uses of Center outputs by stakeholders and clients, e.g., changes in knowledge, attitudes, policies, research capacities and agricultural practices. The maximum size for each description was 250 words. In

³ Science Council (2005) paper “Comments from SC Task Force on the PM Test Indicators in the Pilot Year 2005”

addition, the provision of a reference to the documentation that supports each description was requested.

Observations

A common important problem in the outcome reports was the confused use of the terms outcome, output or activity. Also the evidence available for the outcomes was not always clear. In addition, the definition of “outcome” in the MTP guidelines and PM instructions has been inconsistent.

Conclusion

During the initial years of the PM system implementation, Centers should be provided guidance on reporting their 5 documented outcomes that fully meet the definition of outcome (adoption, use or influence of Center outputs). The SC will then assess and rate 5 documented outcome cases. The SC rating will constitute the performance measure.

Element 3: Impact

The pilot performance indicator for impacts focused on a Center’s capacity to measure the impacts of its previous research. Centers were requested to report two indicators:

3A. What steps has your center taken to establish a comprehensive set of impact assessment results across research areas, i.e., performance in terms of developing and implementing an impact assessment *process* and *culture*? [500 words maximum] Based on this input, SPIA will produce a progress rating.

3B. Submit the two best impact studies your center has done in the period 2002-2004 that document the difference that the center has made in the lives of poor people. SPIA will rate these for rigor.

Observations

The main SC task force observations were (SC, 2005):

“Indicator 3A

- the Center narratives were of variable consistency;
- some submissions were good on describing their impact assessment (IA) strategy and on the processes they have set in train to institutionalize IA;
- many (narratives) described completed IA studies but others did not;
- frequent mention was made of future activities (intentions) rather than focusing on IA activities during the last 3 years. Due to the lack of specificity in the time frame in the pilot question posed to the Centers, some described the complete history of IA while others only described the recent past.”

“Indicator 3B

- the cases submitted were highly variable in quality and rigour;
- a number of the studies were refereed publications in international journals and books;
- some of the submissions are adoption studies rather than traditional impact assessments; and some of the studies are qualitative and descriptive with little analytical work involved.”

Conclusions

Both indicators should be maintained with minor modifications:

- For the first indicator, Centers should specify and document the ex-post impact assessment (epIA) activities, outputs and outcomes over the past year, particularly related to: (a) epIA studies; (b) innovation and advancement of epIA; (c) communication/dissemination and capacity enhancement; and (d) impact culture.

- For the second indicator Centers should be requested to submit every three years their two best completed impact studies that document the change towards achieving CGIAR goals. The PM indicator will be a SC/SPIA rating of the studies for rigor.

Element 4: Stakeholder Perception Survey

For this pilot year only a short CGIAR Member perception survey was carried out by the CGIAR Secretariat (outsourced to an independent consultancy), which comprised three questions and in addition provided space for open-ended comments (see box 1).

Box 1 :

<p>The survey instrument that was used:</p> <ol style="list-style-type: none">1. I believe Center X makes significant contributions to achieving the CGIAR mission.2. I believe that Center X achieves significant results in generating global or international public goods.3. I believe that Center X is effectively engaged in establishing science partnerships for development. <p>Answer categories: strongly disagree, disagree, neither agree nor disagree, agree, strongly agree</p>
--

Members were requested to only share their perceptions about those Centers that they are familiar with.

Observations

During the pilot year only a survey of CGIAR members was conducted, by an external firm. The response rate was 44%, i.e., 28 of the 64 Members responded. In terms of funding, the 28 Members who participated reflected 68% of the CGIAR total budget in 2004. On average there were 13 responses per Center. In addition, some Members provided additional comments along with their expression of agreement with the statement. Generally, the survey provided a small but valuable insight into the perceptions of one important stakeholder about the work of the Centers.

Conclusions

The SC agrees that measuring stakeholder perceptions is important, but does not consider the tested indicator as a suitable results indicator and therefore suggests dropping it until a suitable one is identified and tested. The SC notes that the PM system is intended to inform the stakeholders, and that it should provide reliable and verifiable data about

performance as input to stakeholders for formulating their perceptions. As a source of information about performance, stakeholder perceptions are vulnerable to low response rate, variable levels of familiarity, and subsequent biases.

On the other hand, shareholder and client views are considered a critical and important component of performance in most institutional performance measurement systems (or balanced score card systems)—because customer/stakeholder satisfaction is one of the main targets for-profit and non-profit organizations aim at. This year's pilot covered only the CGIAR Member perceptions, and could be improved based on the experience with the pilot. In addition, the survey or surveys could be structured so as to capture more differentiated perceptions of a wider range of stakeholders. These surveys may well be conducted once every three years to reduce the risk of survey fatigue by the stakeholders. One possibility would be to have a rolling schedule for surveying different stakeholder groups by turns (e.g. one year CGIAR Members, second year NARS).

It should be made clear to the users that this element reports on subjective, impressionistic data and should not be confused with the more objective indicators on outputs, outcomes and impacts.

Element 5: Quality of Research and Managerial Staff

Element 5 included three pilot indicators aimed at covering different dimensions of quality of research and managerial staff:

- (i) No. of all papers published in peer-reviewed journals, books, monographs as well as book chapters per scientist in 2004;
- (ii) Number of national, regional, international, and CGIAR scientific and development awards per scientist;
- (iii) Percentage of Center's top choice candidates for research and managerial positions who accept the offer.

Observations

The following observations were made:

- The publication indicator (i) showed great variation in the results across Centers. Verification of the data showed that there is need to further clarify the definition of “scientist” and “peer-reviewed” journal as there has been slight divergence in the application of the terms. Although not defined in the guidelines, the Centers generally applied, through the efforts of the Information Management community of practice, the following definition of a refereed journal: 1. International editorial board; 2. international audience; 3. clear review procedures (feedback from technical editors; not all submitted articles are accepted - some are rejected); 4. Journal is indexed in an international database (e.g. AGRIS, ISI or any other international bibliographic database). Researchers often publish or co-publish in regional or national journals that are not included in ISI or other databases as refereed journals, though in some cases - after a process of internal consultation between information management specialists and researchers – it was determined that the other criteria were fulfilled, and so the

publications were included (IAU, 2005)⁴.

Generally, Center's internal performance data tracking systems have been developing over the past year, and one result of the exercise has been that Centers now have better ways to track their scientists' publication output.

- The award indicator (ii) also showed great variation in the results across Centers. Data verification showed mixed interpretation by Centers as to what should be counted. One issue is that "national" has been interpreted in some cases as being "from within the nation" rather than "at national level", or "regional" meaning "sub-national". Thus there are some awards in the Center listings that are from sub-national governmental authorities. Based on discussions with Center staff, Centers also varied in the degree of conservatism applied as to what should be considered substantial enough to count. Another issue was that all awards were treated the same and no weighting was given according to the level of international prestige of the award (IAU, 2005).
- Indicator (iii) did not yield sufficient variation to be meaningful. Also there is currently a lack of standardized approach to recruitment across the Centers, making comparison problematic. It is as well very dependent on Centre specific recruitment patterns in any particular year (10 recruitments in one year, 1 in another year).

Conclusions

The following conclusions can be drawn:

- The definition of the publication indicator should be further clarified by developing criteria for inclusion of publications in the count. Also the inclusion of the "impact factor" of journals should be considered in the final calculation of the indicator;
- The "no. of awards" indicator showed validity weaknesses in reflecting quality of staff. The SC recommends dropping this indicator. Alternatively, the definition of the indicator could be either designed to be further restrictive by developing criteria for the inclusion of awards in the count, or a substitute indicator yielding more valid and reliable data should be developed;
- The indicator "% of Center's top choice candidates for research and managerial positions who accept the offer" should be either dropped (as suggested by the SC) or replaced by another indicator that better captures a Center's performance or ability to attract high quality candidates;
- This element could be merged with Element 6, discussed below, under a new heading "Quality and Relevance of Current Research".

Element 6: Quality and Relevance of Programs

For the pilot, two indicators were identified:

⁴ CGIAR Internal Audit Unit (2005); Draft summary findings and observations from 2004 Center performance data verification.

1. Science Council assessment of program quality based on 2003 and 2004 EPMRs and CCERs;
2. Scientific papers published with developing country partners in 2004 in refereed journals, conference and workshop proceedings.

Observations

After closely looking at the 2003 and 2004 EPMRs and CCERs, the SC determined that the current CCERs can not give information useful for the PM purpose as they are often program specific and are variable in purpose and quality. The SC also looked at the most recent EPMRs to consider if that process can also be used for the PM purpose. Again, they determined that there is a great diversity in the approach and type of information provided by the EPMR in terms of the quality and relevance of the programs. The SC concluded that it should not rate Centers on the basis of the recent past EPMRs (SC, 2005).

For the second indicator data verification suggests that instructions and definitions need to be clearer. Also the question arose how to treat publications authored exclusively by NARS scientists but working in Center projects, as their consideration in some way would be in the spirit of measuring partnerships.

Conclusions

As noted above, there is merit in merging Elements 5 and 6 to have a composite element reflecting the quality and relevance of current research by the Center. The composite element could have sub-elements (such as on research and on programs), in a manner similar to the Institutional Health element.

More specifically, the publication indicator is considered suitable. But it should be explored whether it would be more appropriate to measure the publications with developing country partner authors per scientist as a percentage of total papers rather than the actual number.

In addition, two other indicators could be tested next year:

- *EPMR rating of quality of research programs* - The SC is exploring an approach where future EPMR panels are asked to give a simple rating (e.g. scale of 1-3) of both quality and relevance of programs, which the SC may endorse if it finds the report assessment credible on the basis of clear justification and evidence. The SC will develop a set of criteria for use by EPMRs for this assessment, so that all EPMRs would apply a similar approach when considering the rating;
- *Citation indices* - The mechanism and process for collecting the citation data need to be decided, and perhaps this indicator can be tested before the PM data is collected for next year's performance assessment.

Element 7: Institutional Health

a. Governance

The 22 item governance checklist aimed at verifying whether adequate mechanisms for good governance of Centers are in place and are used. It is composed of questions related to (1) organization and management of Board business and (2) Board engagement with Centers strategic business.

Observations

Piloting the checklist (which is based on Center self-assessment) revealed a number of strengths and weaknesses related to the advantages/limitations of checklists in general, and to the questionnaire as such. It became apparent that a checklist can capture whether specific processes or mechanisms are in place, but can not adequately reflect the quality and the effective use of the same. The pilot exercise showed that yielding reliable data, for instance, in assessing the degree of Board engagement in areas of strategic planning, fiscal oversight or policy development through a checklist tool is rather difficult.

Feedback and queries received from Centers and findings by the IAU suggest that a few of the questions need to be posed more clearly as they are ambiguous or lack clear definitions (e.g. how is “professional expertise in financial management or corporate governance” defined?).

Conclusions

Results from the pilot indicate that the checklist has been helpful in understanding Board performance in terms of only some aspects of organization and management of Board business (e.g. composition, orientation) and Board engagement with Centers strategic business (e.g. whether certain policies are in place). Therefore, complementing this checklist by other measures would strengthen this sub-element of institutional health.

Options could be

(1) *Annual Board statements* where the full Board is requested to confirm/not confirm their fulfillment of certain Board oversight responsibilities, for instance,

- the receipt of adequate financial statements on a regular basis (e.g. monthly, quarterly) and adequate follow-up, or
- adequate involvement in developing and reviewing Center policies (related to science and program, finance and human resources) – criteria could be developed.

(2) *Survey of Board members* with a structured questionnaire where Board members give their judgments and assessments about Board practices and their own and the Board’s engagement in Center oversight.

b. Culture of Learning and Change

The 9 item checklist on culture of learning and change covered a range of questions, including the conduct of a staff satisfaction survey, leadership development, staff training (budget and days), internal program planning and review, and number of CCERs completed.

Observations

The questionnaire yielded useful information on some aspects of learning and change at Centers. The questions were mostly geared towards human resources development and some questions on program planning and review. Other aspects of culture of learning and change, such as “knowledge sharing” and “innovation” were not included. Data verification efforts indicate that the term “leadership development plan” needs clarification.

Conclusions

The questionnaire should be expanded to also capture additional dimensions of learning and change, including knowledge sharing and innovation. This could involve expanding the inquiry into the following areas:

- how and at what level a Center deals with learning and knowledge sharing and whether there is staff assigned to take over this responsibility;
- whether there are procedures in place that capture and preserve knowledge when staff leave or retire;
- whether staff performance review procedures and criteria consider/give incentives or reward knowledge sharing and organizational learning;
- staff perception about knowledge sharing and organizational learning.⁵

c. Diversity

This sub-element comprised 4 indicators capturing diversity in terms of gender, nationality and seniority.

Observations

Gaining insight into a balanced set of diversity dimensions seemed to have been sensible. The measurement of the indicators have been straightforward, although data verification suggests that the definition of “IRS” and “management” needs further clarification to yield more reliable results.

One concern that was raised during the indicator development period was that staffing diversity, particularly in terms of gender, needs to be seen in context of the Center location and to some extent is determined by enabling conditions at the host country of the Center. As a result there would be different benchmarks across Centers. Diversity indicator 1, attempted to cater to this concern and was made practicable due to the newly

⁵ Drawn from work by ILAC (Inter-center Initiative on Institutional Learning and Change)

initiated effort of the G&D Program to help Centers to set their own gender and diversity goals.

Conclusions

The indicators should be maintained and definitions clarified. Two additional indicators could be considered as complementing one of the pilot indicators (“% of management positions occupied by women”):⁶

- % of female scientists
- % of female postdocs.

Element 8: Financial Health

Two financial health indicators were used:

1. **Short term solvency (in days)** - current assets plus long term investment minus current liabilities divided by per day operating expenses excluding depreciation. 90-120 days is the recommended acceptable range
2. **Long-term financial stability (in days)** - unrestricted net assets less net fixed assets divided by per day operating expenses. 75-90 days is the recommended acceptable range.

Both indicators have been developed, tested and debated by the CGIAR finance community before the PM system was discussed. There has been broad agreement over the adequacy of both indicators and their subsequent inclusion in the PM system.

Observations

The reporting of the finance indicators was aligned with the submission of the Center financial reports for the CGIAR financial report. The CGIAR has recently established a peer-review mechanism, involving a team of Center Chief Financial Officers in helping to determine whether the Center’s audited financial statements are in compliance with the CGIAR Accounting Policies and Reporting Practices Manual. Any inconsistencies found by the peer-review team in the review process have been adjusted accordingly by the Centers and in the context of the PM exercise.

While they are intended to measure two different phenomena, statistically the measures used for the short term and the long term are highly correlated ($r^2=0.93$).

Conclusions

Although both indicators seem to be well established in the CGIAR, their adequacy for reflecting the financial health of the Centers should be further examined, including their recommended range. Also the inclusion of additional financial indicators should be explored, with inputs from the CGIAR finance community.

⁶ As suggested by the G&D Program Leader

IV. Performance Data Collection

a. On-line Performance Data Collection System

This pilot exercise included the establishment of an on-line Performance Data Collection System for Center data and Science Council data. For ensuring confidentiality and security, access was limited to a small group of people, including Directors General, PM focal points, SC, SC Secretariat and CGIAR Secretariat. The site is administered by staff of the CGIAR Secretariat and an IT Specialist advising the CGIAR Secretariat.

Observations

The feedback received from the Center focal points shows strong consensus that the online- data collection tool very much helped in facilitating the data collection effort. The online submission forms worked well and the site was straightforward to use. Some repetitive inquiries suggest that a “Frequently Asked Questions (FAQ)” would have been helpful to users of the site and the individuals responsible for compiling the data.

There were only very few incidences that needed technical attention and troubleshooting. The response to queries from Centers was considered to be timely and clear.

Conclusions

For next years’ online data collection exercise the PM site should be enhanced by including a FAQ section.

b. Instructions and Definitions

Observations

Queries received during the data collection phase and verification of some data afterwards suggests that instructions for collecting the data and definitions of indicators gave some room for interpretation and need to be clarified (e.g. Full-time-equivalent (FTE), scientists, IRS and the inclusion of Postdocs and Secondments).

Conclusion

Instructions for collecting the data and definitions of indicators should be made clearer in next year’s exercise. Also clear criteria for what is considered a “peer-review publication” should be developed.

V. Verification of reported data

The PM system is based on self-assessment by Centers, but has a number of mechanisms in place that allow data verification. These are the following:

- SC assessment of submissions for selected indicators;
- Submission of supporting documents and reference material;
- Random check of selected reported figures by CGIAR Internal Audit unit.

Observations

For some indicators the submission of supporting material was requested (e.g. indicator 2a-outcome), and for other indicators (e.g. 5c) Centers were requested to hold documentation if data verification is to be undertaken. The pilot showed that the supporting material provided was of varied comprehensiveness and therefore was not always fully usable.

The CGIAR Internal Audit unit was requested to verify selected information reported by Centers. IAU therefore contacted all PM focal points, who were very cooperative and helped to provide all information as demanded (e.g. publication lists and lists of awards). Some Centers retained extensive documentation to support their submissions, in other cases the information was scattered among different staff and (more importantly) not complete or easily audited. Experience from this pilot proved that requesting documentation after the data collection exercise was completed is very time consuming and tends to lead to delays.

Conclusions

Centers should be requested to submit most of the supporting documentation during the data collection phase, and should be guided as much as possible in providing the needed information to allow for more consistent back up information. For instance, this should include the submission of a complete publication list along with the reported measure, or in case of the outcome description clear evidence for the reported outcomes.

Guidance for organizing an “*audit trail*” should also be given, so that data verification can be completed more timely and efficiently.

ANNEX 1: SC Recommendations for Implementing the CGIAR PM System in 2006 and Beyond

The Science Council has participated in the development of the CGIAR PM System since the initial phases of its development and is fully committed to its implementation and success. The SC has carefully reviewed the results of the pilot implementation of the system and recognizes that the system is an evolving one in which we will have the opportunity to continue to improve the indicators. Specific issues that the SC is addressing for 2006 include the need for more clearly specified output targets, a mechanism for ensuring that Centers identify significant output targets instead of focusing on low-risk targets, and a way for poor EPMPR ratings to be upgraded in between EPMPRs if appropriate.

Based on the lessons learned in the 2005 implementation of the PM System, the SC recommends using a limited set of indicators in 2006 and beyond. The indicators in the following table are judged to have the potential to accurately and credibly reflect real performance. In addition, they are a parsimonious and appropriately balanced set that do not place an undue burden on Centers or overemphasize potential to perform over actual performance. The SC recommends not adding additional indicators to the set unless the analysis over 2-3 years reveals shortcomings in the system, at which stage the shortcoming should be carefully assessed.

Pilot elements and indicators	PM Elements in 2006
A. RESULTS	
OUTPUT	
Tested: % MTP output targets achieved in 2004	Suitable. The SC will work with Centers to guide the planning of output targets so that they are measurable and clear, and significant enough to result in a difference towards achieving CGIAR goals. The Centers will report on achievement of targets.
OUTCOME	
Tested: Centers report the five most significant outcomes in 2004 in relation to the Center's MTPs submitted in 2001, 2002, and 2003	Suitable. The SC will assess and rate 5 documented outcome cases which represent adoption, use or influence of Center outputs.
IMPACT	
Tested: Steps taken to establish a comprehensive set of impact assessment results across research areas, i.e., performance in terms of developing and implementing an impact assessment <i>process</i> and <i>culture</i> ?	Suitable. Centers specify and document the ex-post impact (epIA) assessment activities, outputs and outcomes over the past year, particularly related to: (a) epIA studies; (b) innovation and advancement of epIA; (c) communication/dissemination and capacity enhancement; and (d) impact culture.
Tested: Two best impact studies done in the period 2002-2004 that document the difference that the Center has made in the lives of poor people	Suitable (every three years). Centers submit two best impact studies completed in the three year period that document the change towards achieving CGIAR goals. SC/SPIA will rate the studies for rigor.

Pilot elements and indicators	PM Elements in 2006
STAKEHOLDER PERCEPTIONS	
Tested: Donor ratings of Center performance	Not suitable. The PM system is intended to inform the stakeholders. The PM System should provide reliable and verifiable data about performance as input to stakeholders for formulating their perceptions. As source of information about performance, stakeholder perceptions are vulnerable to low response rate (observed in the test year and likely to get worse if donors are not convinced that this indicator is meaningful), variable levels of familiarity, and subsequent biases. The SC recommends not using this element in the PM system, while recognizing the value of stakeholder perceptions in priority setting processes and in Center specific external reviews.
B. POTENTIAL TO PERFORM	
QUALITY OF RESEARCH STAFF*	QUALITY AND RELEVANCE OF RESEARCH
Tested: % of top choices who accept offer for research positions	Not suitable. Dependent on Center specific recruitment patterns (10 one year, 1 in another year), and may be difficult to verify due to confidentiality of recruitment information. In test year did not allow differentiation. The SC recommends not using this indicator.
Tested: Peer reviewed publications per researcher	Suitable. Peer reviewed journal articles per researcher multiplied by impact factor.
Tested: # of national, regional, international, and CGIAR awards received per researcher	Not suitable. The indicator is too influenced by different “awards culture” in different countries, and differences between scientific fields. The SC recommends not using this indicator.
QUALITY AND RELEVANCE OF RESEARCH PROGRAMS*	QUALITY AND RELEVANCE OF RESEARCH
Tested: Scientific papers published with developing country partners	Suitable. Scientific papers published with developing country partners Potential (not yet tested). EPMP rating of quality of research programs Potential (not yet tested). Citation indices
INSTITUTIONAL HEALTH	Suitable. The SC has no recommendations for this element.
FINANCIAL HEALTH	Suitable. The SC has no recommendations for this element.

* The SC recommends that these elements are merged into a new element, QUALITY AND RELEVANCE OF RESEARCH, which includes indicators on input quality (staff quality), output quality (EPMP rating) and relevance (co-publishing and citations).