



Contribution to the discussion on Mega Program M&E and the independent evaluation arrangement in the new CGIAR
interim Independent Science and Partnership Council

The CGIAR Joint Declaration establishes that: “The Consortium, Centers, and Fund donors are mutually accountable for Mega Program outputs financed by the Fund” and “The Consortium, Centers, all CGIAR funders and their respective partners have shared responsibility for managing toward outcomes, i.e., uptake of outputs resulting in longer-term improvements of livelihoods of end users”. The new CGIAR encourages an inclusive approach to the Mega Programs, which will involve new partnerships. There is need to agree how the CGIAR can develop an M&E system for the Mega Programs that supports and strengthens learning. The concepts of accountability (for outputs) and responsibility (for outcomes and impacts) need to be operationalized so that monitoring and evaluation of the Mega Programs stimulate innovative research and effective partnerships, and enhance the plausibility of development impact from the research. The *independent evaluation arrangement* is part of the overall M&E of the CGIAR.

The paper has three components:

- The first section discusses the important features of evaluation science and of research for development.
- The second section discusses options for the *Independent Evaluation Arrangements* regarding independence, other key considerations and cost. Further elaboration of the option of co-locating the evaluation function with the ISPC is given in Annex 1.
- The third component discusses M&E of research for development in the new CGIAR where the division of duties is between doers and funders, distribution of evaluation responsibility is among Consortium, Independent Evaluation Arrangement and SPIA, and the SRF defines the objectives for research.

1. The Research for Development Mission of the CGIAR; implications for evaluation

Evaluation of science and research for development is different from evaluation of development interventions. It needs to take into account the specific characteristics of research, which includes the unpredictability of results - serendipity and risk of failure to achieve the expected results; the often long time lags before outcomes are accumulated in full, and even longer time lags for impacts to accrue; and the highly specialised areas of research activity in terms of methodology and the state-of-the art in any particular field. Peer review process is considered the linchpin of research for development evaluation. Furthermore, the evaluation of research needs to apply the same rules of rigorous analysis as is expected from research itself in order for the findings to be credible for researchers¹.

The CGIAR is mostly involved in the generation of research outputs for the use of their first intended users (often not the ultimate users). Before the impacts on poverty, food security and environmental sustainability are visible several actors other than the CGIAR and its partners will have engaged in potentially a long chain of activities and iterative loops of further

¹ See Scriven, M. and Coryn, C.L.S. (2008). The logic of research evaluation. In C.L.S. Coryn & M. Scriven (Eds.), *Reforming the evaluation of research. New Directions for Evaluation*, 118, 89-105.

research, adaptation and development over several years. Generating impact requires the interventions, support and assessments of research partners and other stakeholders, and it requires commitments from national and regional players and extended donor support. The upscaling and outscaling of the CGIAR's research interventions for broad scale development impact are not directly under the control of the CGIAR—although the CGIAR donors, through their bilateral activities, and other agencies, may be facilitating such uptake. In research evaluation it is essential to appreciate this long-term focus where the research results themselves may be uncertain and their outcomes can not be accurately predicted. However, enhancing the likelihood of impact requires evaluation of the R&D landscape – i.e. how research partnerships are built and science is mobilized to support the CGIAR agenda. Even if impacts cannot be measured early it is possible to assess the plausibility of impact pathways and the extent to which the science design shows awareness of all the connections and pre-conditions that will determine ultimate impact.

The monitoring and evaluation of on-going research can focus feasibly mainly on progress, volume and merits of research results and early outcomes. Monitoring is by nature, short term, and focused on incremental information, often annual (or shorter term), and based commonly on numeric data and records, diagnostic indicators or reports that have descriptive content. In research, monitoring needs to involve dialogue with management and analysis of the incremental information in its broader context. Monitoring information needs to be accumulated over time to be of use for evaluations. Periodic evaluation then need to take a more holistic view of the research endeavour over a longer period of time. In evaluation of research for development it is possible and necessary to assess reasons for research progress, success (i.e. outputs achieved), reasons for lack of achievement of results and, in some cases, serendipity. Thus it is essential that peer scientists are involved in the evaluation of research programs. Only they can properly appreciate the risks and challenges in research, the level of advancement in the research program relative to what is done elsewhere, and the merits and significance of results, including breakthroughs. Furthermore, only peers can distinguish research “failure”² and negative results from shortcomings in research planning, scientific methodology, implementation and interpretation. The evaluation process and findings should facilitate learning and lead to improvement and changes, if needed, in the scientific program due to analysis of the past experience, feed-back and assessment of changing conditions and expectations.

The general characteristics of research that influence the approach to M&E include:

- Need to set appropriate incentives for innovative research
- Need to accept that all research involves certain risk where the level of risk depends on the topic of research and ambition of the plan
- Need to distinguish informative failures—typical in good research—from poor research implementation
- Need to stimulate transparency in presenting and interpreting research results, including negative or unexpected.
- Need to be open to serendipity
- Need to involve peer debate for improving the methodological approaches, interpreting results and cultivating more innovative thinking
- Need to appreciate the lag time from research interventions to outcomes to development impact.

² Failure here means not achieving the intended results. Such research can nevertheless accumulate important scientific information.

In order to oversee such an evaluation by peer scientists that capture the elements described above, the *independent evaluation arrangement* will require a capacity in scientific expertise and an in-depth understanding of the research enterprise, with its inherent risks and opportunities. It will require knowledge of the scientific community from which the most appropriate peers can be selected for the task of peer review. It needs scientific knowledge and understanding of research for development to identify and assemble the best peer review panels for each specific MP evaluation.

2. Independent Evaluation Arrangement

Independence

The external review of the CGIAR recommended that the CGIAR follow global good practice in evaluation by ensuring a clear separation of responsibilities and accountabilities that are associated with “independence”. Following the guidance of the review and the Integrated CGIAR Reform Proposal, the CGIAR at AGM08 decided that “an *independent evaluation arrangement* would periodically take place at the Program and System level”. Independence was to mean that the persons/institutions conducting the evaluation would be free of the control of those responsible for the design and implementation of the program.

The Consortium is responsible for the monitoring of short- term research progress, outputs and intermediate outcomes of mega-programs for performance management purposes and for reporting to the Fund Council. It is also responsible for external evaluations of CGIAR Centers and Mega-program components. The *independent evaluation arrangement* needs to complement the internally organised M&E and, at the level of the entire Mega-Program, evaluate the longer-term outputs and outcomes and likely impacts and evaluate the quality and relevance of the science which underpins these development driven results.

Robert Picciotto³ noted that “optimum independence” is not “absolute independence”. He says, “Accurate and fair evaluations combine intellectual detachment with empathy and understanding. The ability to engage with diverse stakeholders and secure their trust while maintaining the integrity of the evaluation process is the acid test of evaluation professionalism. This is why diminishing returns set in when evaluation independence assumes extreme forms of disengagement and distance. It leads to isolation, a lack of leverage over operational decision -making and a chilling effect on learning. Thus, the basic challenge of evaluation governance design consists in sustaining full independence without incurring isolation”.

³ Former Director General of Operations Evaluation at the World Bank. Taken from a report to the Independent Advisory Committee for Development Impact (IADCI) (p5, August 2008)

Options for independent evaluation arrangements

Two options for operationalizing the independent evaluation arrangement are emerging in the recent discussions. One, raised by the TMT proposes housing an independent evaluation unit in a multilateral development agency (such as IFAD or the World Bank). The other proposed by Rob Bertram (USAID) would have “the locus of independent evaluation co-located with the locus of both science quality and impact assessment expertise” as all three functions depend on an underlying, in-depth knowledge of research for development. He proposed establishing the position of an Independent Evaluation Chair to be co-located with the ISPC in a similar manner to that for impact assessment as now done by SPIA with support staff based at the ISPC Secretariat and enforcing the IEC’s independence by (i) independent budget, (ii) independent selection and nomination of the IEC and (iii) direct reporting by the IEC to the Fund Council. In Table 2, these options are compared for the criteria that have been suggested for assessing their suitability.

Table 2: Comparison between two different proposed *independent evaluation arrangements* for criteria listed in the M&E Framework

Criteria	Co-location of IEC with the ISPC	Hosting of a unit by a multilateral agency
To limit added bureaucracy	No added bureaucracy as IEC is established in existing framework of ISPC and its Secretariat	More bureaucracy as this would require establishment of a new unit under the control of a new host.
To enhance evaluation professionalism in the CGIAR	IEC needs to be highly competent and well linked to the evaluation community. If there is a standing panel, evaluation professionalism can be enhanced through membership.	Multilateral agencies have experience of evaluation of development programs; science evaluation professionalism would need to be guaranteed.
To retain institutional memory and promote institutional learning	Due to linkages to the ISPC and co-location in its Secretariat institutional memory is fully retained. Institutional learning depends on the linkages among <i>evaluation arrangement and Consortium</i> , ISPC and Fund Council. Co-location with ISPC that deals with science quality, foresight and impact makes institutional learning likely. Strong orientation on science evaluation enhances relevant institutional learning and willingness of research programs to learn from evaluations.	Creation of new units would most certainly lead to loss of institutional memory. If the new unit would be staffed with professional evaluators only, Institutional learning might be more emphasised, however if appreciation of science were lost, the learning might be limited or not relevant. Anchoring evaluation in a multilateral development agency may influence the perception of researcher managers of its relevance.

Criteria	Co-location of IEC with the ISPC	Hosting of a unit by a multilateral agency
To stimulate methodological advancement in the evaluation of the “Research-Development Continuum”	The continuum is somewhat artificial; all CGIAR research ought to be research explicitly for development. Good understanding of research organizations and science, and the CGIAR enhances the likelihood of appropriate methodological advancement. Close linkages to the ISPC are an advantage.	At IFAD or WB, the research aspect is lacking which could be negative for advancing appropriate methodologies for evaluating research for development.
To leverage potential synergies in agricultural research outcome evaluation, i.e., interventions leading to uptake of technologies and other research outputs by partners and stakeholders that are in common	Fulfilled.	Absence of research at host institutions could affect ability to evaluate how research leads to outcomes (vs. development interventions). It is important to appreciate both research risk and potential long lag times before full outcomes accrue.
To support harmonization efforts in light of the Paris Declaration on Aid Effectiveness and the Accra Agenda for Action	Fulfilled.	Fulfilled.

Cost

It is obvious that the new independent evaluation arrangement must be cost effective. The estimated direct costs of the CGIAR’s past EP MR (external program and management review) evaluation system shown in Table 1 offer a useful benchmark.

Table 1: Direct costs⁴ for the CGIAR EP MR evaluations

Cost item	Cost per year	What is included
Direct Panel costs	705,000	An average of three EP MRs annually. Panel (6 on average) and travel costs (initial and main phases, field trips and Chair’s travel to SC meeting; Secretary’s travel to initial and main phases)
SC Secretariat costs	170,000	Staff salary costs (1.2 FTE): Secretary; administrative staff. Printing and proportion of general administrative costs
CGIAR Secretariat	150,000	CGIAR Secretariat support ¹
SC costs	10,000	Honorarium costs of SC members for a) helping identify the peer scientists for the panel and b) providing the quality audit of the report
Total	1,035,000	

¹ Source: System Office budget data.

Thus the benchmark for how much independent external evaluation costed in the past in terms of direct costs is around 1.1 mill \$ US annually. In estimating the cost-efficiency of the external evaluations it needs to be considered that appropriate and high quality evaluations provide considerable *ad hoc* advice to the program and support the other evaluative processes.

⁴ The costs in the table include only direct costs from EP MRs (comparable to the costs of an independent arrangement), not any costs borne by the Center, such as direct internal preparation or transaction costs.

3. Integration M&E of research for development in the Mega Programs

At the core of the M&E Framework document there are four key principles for M&E that apply to all the operations and functions within the CGIAR System and to the major actors: the doers and the funders within the overall system. Three of the M&E principles are particularly relevant for the M&E of Mega Programs.⁵ Guided by the basic premise of accountability, systematic accumulation of information and self-assessment data are needed for performance management that, over time, generates the inputs for periodic external evaluations.

M&E Framework principle of accountability:

The Consortium, Centers and CGIAR Fund donors are mutually accountable for Mega Program outputs financed by the Fund. As agreed in the performance agreements, (i) the Consortium and Centers are accountable for high-quality science and technology products and services, and (ii) fund donors are accountable for an aligned provision of funds to support the development of research outputs.

An interpretation of what is meant by **accountability** is needed for defining how the accountability for outputs (as stated in the CGIAR Joint Declaration) can be enforced through M&E. Accountability should entail progressive improvement at the institution and program level. The Centers and other research providers participating in a Mega Program engage in research that entails variable degree of risk of not succeeding as planned. It is important, however, that the research effort involves competent researchers, state-of-the art research methodologies and clear scope and focus of the research commensurate with the estimated resources, and that the research management and evaluation culture supports continuous learning and improvement. From the donors' side, the effort needs to involve sufficiently large and long-term funding so that a research output does not fail due to insufficient funds. Mutual accountability therefore requires understanding of the inherent risk in research on one hand and the time and resource dimension on the other hand to produce a significant output for development impact.

M&E Framework principles for monitoring and evaluation:

The monitoring system for research under the SRF is the overall responsibility of the Consortium and is designed to provide real-time information about program outputs and outcomes to research managers in Centers and the Consortium. This information also serves as a basis for regular progress reports of the Consortium to the Fund Council, and thus for annual performance reviews by the Fund Council. A common system and set of metrics will be used for reporting program performance information to the Consortium and the Fund Council.

Evaluation of performance to achieve the Strategy and Results Framework and governance of the CGIAR will follow international best practice and will include evaluations that are independent and impartial to the policy-making process and delivery and management of programs.

⁵ The fourth principle applies to the entire CGIAR system: *The evaluation system provides periodic objective assessments of the extent to which Mega programs and other aspects of the CGIAR are likely to have achieved their stated objectives, as articulated in the SRF and the CGIAR Joint Declaration.*

Monitoring is the Consortium's responsibility and the primary purpose of monitoring is to support the development of high quality research outputs. Thus it is an activity that primarily benefits the doers (not funders). Transparency of how the monitoring information will be used increases the trust among researchers to be frank and reflect on what the progress information tells them. There are two important internal requirements for management that monitoring of research should facilitate: (i) establishing best practices in management for securing important processes; one of the most essential being the management of data and information over time that forms the basis of further research (by the CGIAR and others) and is essential for monitoring and evaluation; (ii) reporting progress in research, and related activities, towards reaching near-term output goals and advancing outcomes to the extent feasible. Monitoring should accumulate information that is necessary for any periodic evaluation and for periodically determining whether the objectives set in a performance agreement are being achieved. Thus, although the principle states a *common system and set of metrics*, monitoring should be sufficiently tailor made for each MP to benefit the particular research being conducted in the program.

To support the overview role that is retained for the Fund Council in the CGIAR Joint Declaration, there is need to agree on the purpose and content of the annual performance reviews submitted to the Fund Council and the specific measures that the Fund Council can and should take in case a performance report gives cause for serious concern. To make an annual report as useful for the Fund Council as possible, the report might present the Consortium's assessment of progress highlighting issues that are of particular interest for the Fund Council. The report's intention could be to keep the Fund Council abreast of developments at Mega programs informing the FC, for example, on completion of longer-term outputs that are in the MP performance agreements, or changes in MPs following from Consortium-commissioned evaluations. Thus, the annual reporting would contribute to information flow in a continuum towards the MP's medium-term goals rather than lead to "piece meal" reporting. It is particularly critical that the annual progress reporting mechanism is not in conflict with the longer term funding commitment that is required by the scientific process to generate major outputs.

Two types of MP **evaluations** are envisioned in the M&E framework: Consortium-commissioned external evaluations on Mega Program components; and Fund Council commissioned independent external evaluations. The content and conduct of the former type needs to be designed such that they are both (i) an important tool for internal management feed-back and decisions about directions at program content level and (ii) high quality input for the independent external MP evaluations. As will be discussed below, the external evaluations of entire Mega Programs need to cover research progress of the MPs (they are also mandated to evaluate management aspects), the achievement of outputs and the significance of the results, and further progress along the impact pathway where evidence is available of outcomes and early impacts.

As the Consortium is responsible for management oversight and funding allocations within each Mega Program, it should use the findings from evaluations (as also from monitoring) to oversee performance management. Indeed, the premise of mutual accountability requires that the Consortium has authority to conduct corrective adjustments within Mega Programs.

The likely Consortium actions following an evaluation could include:

- closing some avenues of investigation that lose promise;

- opening up new research or expanding some research due to new opportunities or progress;
- changing task/project managers;
- adjusting partner composition and engaging new necessary partners;
- shifting resources according to re-established needs, etc.

The independent external evaluations would best serve the Consortium and the Fund Council for making progress-based decisions at the MP level for the subsequent evaluation period. In the M&E continuum described above, only the independent external evaluation would be used for major actions that require the Fund Council decision, such as confirming or significantly changing the funding level of the whole Mega Program or implementing other significant changes at the Mega Program level such as winding down a program or its major component. As the strategic priorities in the new CGIAR are determined through regular iterations of the SRF, the independent external evaluation take more of a role of reviewing the MP's performance and progress for learning and decision making.

The monitoring-evaluation continuum

According to international best practice independent evaluation depends on self-evaluation provided that the latter is regular, timely and adequate.⁶ Self-assessment, including monitoring, and independent evaluation have a reciprocal relation. Self-assessment gains from intellectual guidance and demand coming from the independent evaluation which in turn should have some authority over the design of self-evaluation processes, programs and products.²

In the new CGIAR there is an opportunity to achieve this streamlined and synergistic cycle where self-assessment through monitoring and self-evaluation (Consortium-commissioned external evaluation) feeds into independent evaluation (commissioned by the Fund Council and operationalized through an independent evaluation arrangement) and receives feed-back.⁷ The internal M&E activities need to be designed so that they fully support the 4-year independent external evaluations. This means that although the monitoring is systematic and covers performance aspects of a similar nature in all Mega Programs, the information needs to be sufficiently content specific to accurately reflect performance in a particular MP (or its component). It may well be desirable to have the leadership in the independent evaluation arrangement involved in designing and fine tuning of the internal self-assessment processes.

From research to impact

The CGIAR is mostly involved in the generation of research outputs and, through partnerships and capacity building, for example, it is engaged in embedding its research for desired outcomes to accrue. The current M&E framework puts a lot of emphasis on the annual reporting of output targets, outcomes and impacts for *control* purposes. However, experiences from the CGIAR's current PMS, including lengthy discussions among SPIA and

⁶ Presentation by Robert Picciotto at the UKES: Independent Advisory Committee on Development Impact: Independence in Evaluation, An Assessment Framework.

⁷ In the past this streamlining was not fully achieved for two main reasons: (i) the Center commissioned external reviews that were expected to cover individual programs for quality, relevance and progress were designed for Centers' own purposes and were only variably useful for EPMRs (as numerous recent EPMR reports verify) due to differences in scope, approach and quality; and (ii) the performance monitoring system that was supposed to feed information into EPMRs did not serve that purpose well due to its limitations

the impact focal persons at Centres, indicate that impacts from a given research project cannot and should not be monitored annually. Even outcomes take often several years to accumulate to a sizeable scale, although initial uptake and influence of research findings and new technologies can be expected already while the research is still going on. Thus, while the ultimate measure for the CGIAR's success as an entity is on the impact on the broad CGIAR objectives, the regular M&E measures cannot establish that success.

The independent external evaluations should be tasked to look at the early stages of the impact pathway i.e. uptake, adoption and intermediate outcomes from research to assess relevance and likely effectiveness. They should also look at issues of research process, partnerships and capacity building that are necessary for enhancing the likelihood of outcomes and impacts. Mega Programs need to conduct studies to evaluate the uptake of research results while they are emerging for feed-back for MP adjustments. Such studies will provide essential inputs for external MP evaluations and *ex post* impact assessments.

In the new CGIAR *ex post* impact assessment is primarily the responsibility of SPIA with its specific independent role. Documentation of ultimate impacts intended by the CGIAR, including effects of research on poverty, requires rigorous and credible *ex post* assessment done sufficiently long after the completion of the research. It also requires that appropriate benchmarks are set and appropriate data are accumulated along the impact pathway. There is need to establish what data are essential for the System's *ex post* impact assessment and consider to what extent MPs should be responsible for collecting such data (about adoption and subsequent changes in factors that contribute to impact) also after the active research phase. If the MPs are explicitly tasked to accumulate such data they need to be resourced appropriately to do it. Due to the time lag it takes for long-term impacts on CGIAR's goals to accrue and be documented, *ex post* impact assessments are not suitable for informing individual MP funding decisions. They can, however, help direct research to areas where it is likely to be most effective.

What to monitor and evaluate?

The fact that the CGIAR deals primarily with research frames the approach to M&E, what should be monitored and how evaluations should be conducted. Considering that the MPs are very different in size, complexity and research content, much of the content of monitoring and evaluation should be tailor made for each program.⁸ Although every Center and program monitors its activities through a number of means, there is no common best internationally agreed practice on what data, records and information ought to be systematically accumulated for efficient and effective performance management and evaluation of research. Annual performance indicators have been introduced to research organizations, but they tend to cover only such details that can be quantified. Furthermore, there is limited experience and understanding on how monitoring of only short term quantifiable performance indicators and subsequent rewarding will influence longer term research directions and success. Monitoring and evaluation in research commonly relies on bibliometric methods and while publishing high quality articles in internationally recognised, peer reviewed venues is the universal yard stick for good science, the CGIAR's mission oriented research requires many other research dimensions to be included in its M&E. CGIAR research that is expected to deliver solutions to agricultural production and development constraints, enhance their uptake by intended

⁸ The common areas for monitoring could include research quality monitored through bibliometric analyses, management or records and data, communication, capacity building and inclusion of appropriate gender analysis.

users, and deliver capacity services requires M&E that is designed for research but that is different than M&E used in academia.

Regarding results, each MP will have its own objectives, and short and medium-term targets. The MPs are directly accountable for producing outputs. The expected output results should be significant solutions to major problems that the research is directly designed to address within a reasonable, 3-5 year, term. The indicators (of progress and success), need to be developed when the programs are being designed and the clarity of planning and planned monitoring need to feature in the program selection criteria.

In the new CGIAR monitoring and evaluating the main results and progress towards the main objectives is likely to be more feasible than in the old CGIAR.⁹ In the MPs the planning and funding come together for supporting major outputs. Indeed, the approval of the MPs depends on the program components clearly addressing particular problems, and monitoring is focused on tapping progress towards achieving the expected output. It will be feasible to determine the critical sub-products by content and volume that are needed for the problem's solution. When the progress and success criteria are incorporated at the design stage, it makes it compelling to collect data, records, and evidence to form the basis of self-assessment and evaluation. The indicators for progress should be specific to the problem being addressed, but the time allowed to accomplish the output result should be flexible to accommodate risk taking and failure. If the risks are sufficiently analysed at the design stage, the level of success can be assessed in the context of well understood risks. This way of monitoring would provide early warning signs when a line of research is proving to be non-productive. It would set a time frame for the output, but not dictate the way and sequence of steps for getting there.

The CGIAR could benefit from experiences from monitoring of large research programs emerging elsewhere. Some different approaches to monitoring have been applied in the Bill and Melinda Gates' Foundation funded programs. For example, BMGF program monitoring may include monitoring of achievement of shorter and longer term project-specific milestones combined with planning and review meetings and frequent communication between BMGF program officer (with research background) and project leader, and mutual agreement on adjustment implementation.¹⁰ CGIAR Centers are involved in these programs, and some of the programs contain also high-risk research components. Thus these programs that involve testing optimal M&E for research could provide appropriate lessons for designing the M&E for the CGIAR's MPs.

Can the Strategy and Results Framework guide M&E of MPs?

The Strategy and Results Framework is expected to provide the results framework against which the MPs are monitored for progress and evaluated for results and effectiveness. The most recent version of the SRF (18th November), however, describes the expected results mostly in terms of impact that it is impossible for any monitoring system to capture.¹¹ These

⁹ The MTP output targets were only loosely describable under the overall output heading and their design was governed by the specific requirements, activities and milestones of numerous small projects largely funded by restricted grants that reflected the reality of Centers' research. Thus the output targets were highly variable by size, original research content, resource requirement and potential contribution to the overall problem solution. It was nearly impossible to determine how and when the sum of the output targets led to the expected output.

¹⁰ A fundamental difference here is that the program officer in the BMGF case is working for the funder while in the new CGIAR the MPs would report to the Consortium office and ultimately to the Board.

¹¹ Examples of expected results that have been used for justifying MP selection include following: productivity increases of at least 10% over 10 years and help lift 60 million out of poverty (MP1); impact on 400 million

kinds of result targets in terms of quantifiable changes in poverty, productivity and environmental sustainability, for example, present *ex ante* estimates of anticipated effects and they are important for directing research to where it is estimated to be most effective for advancing the CGIAR's goals but not for monitoring and evaluating research progress. Only *ex post* impact assessments can attempt to establish the extent of actual impact of the kind indicated in the SRF but even then, there are challenges in impact assessment related to, for example, methodologies, cost, data availability (which accounts for some of the cost) and attribution.

It may not be realistic to expect that the SRF can describe in detail the specific short and medium-term results in terms of outputs and early outcomes that could be used for MP monitoring. However, when finalized, the SRF can be expected to provide more detail of the theory behind the estimates of final impacts and thereby illustrate what intermediate results should be expected from each MP in the final portfolio.¹² The SRF could therefore provide indications of the kinds of results that each MP is expected to define in its program plan that it will be feasible to monitor in an on-going MPs. The comprehensive results framework which would serve monitoring and on which the MP performance agreements would be based should be elaborated in the MP proposals.

small farm households through reduced transaction costs and risk (MP2). genetic gains that will account for 60% of the overall productivity gains for food crops (0.4% gain per year) (MP3); measurable improvement in population health (MP4); improved crop/water productivity by 20-50% over 30 years with direct benefits on up to 100 million people (MP5); 10% reduction in deforestation by 2030 and 10-fold increase in incomes of local communities benefiting from forest and timber products (MP6).

¹² For some MPs the draft SRF has listed expected results at the output level that serve as examples of medium-term result targets.

Annex. Elaboration of an *independent evaluation arrangement* co-located with the ISPC.

This arrangement would establish an independently appointed Evaluation Chair (IEC) co-located with the ISPC. There could be a small standing panel (as with SPIA) working with the IEC. The co-location would involve staff supporting the IEC in technical and administrative matters located at the ISPC Secretariat. The IEC would have a dedicated independent budget set by the Fund Council, which would include funding for program reviews, IEC and panel time and meetings, and administration. The IEC would report to the Fund Council on all matters related to evaluation. The IEC (and staff) would have the responsibility to identify and support independent expert teams to conduct program reviews the full direct costs of which would be covered by the dedicated budget.

The IEC would have *ex officio* status at ISPC meetings. Thus the IEC (and the panel) would be well informed in most aspects of the CGIAR research for development regarding science quality, trends and impacts such that the IEC could ensure that the peer evaluations mesh with the rest of the CGIAR system activities and that learning from evaluations both in the research programs and at the administrative side takes place.

For logic of independence this arrangement would mean that the Chair ISPC, Chair Evaluation (IEC) and Chair Impact assessment (SPIA Chair) report directly to the Fund Council and that their activities are supported by independent budgets. It is worth noting that SPIA was originally set up as an independent unit under UNDP but with the staff member providing technical support co-located at the Secretariat of the CGIAR's Technical Advisory Committee (TAC) at FAO. Even after merger with TAC, SPIA reported directly to the CGIAR at AGM.

All three Chairs (ISPC, SPIA, Evaluation) should be appointed by Fund Council Chair. The IEC would be selected applying normal no-conflict-of-interest rules. As the ISPC is the primary scientific advisory body to the Fund Council, ISPC Chair would be involved in selection of IEC and IEC unit professional staff at the Secretariat.

Staff supporting the above three functions should be co-located at the ISPC Secretariat together with staff supporting the ISPC to maximize synergy and minimize duplication of highly technical expertise and knowledge of the system and research that are required to successfully support the respective functions.

Issues of independence with the co-location option

1. Organization independence

Co-location with the ISPC provides organizational independence from the Consortium, the ISPC and the individual donors. Independence is established through the IEC managing an independent budget and reporting directly to the Fund Council. The selection procedure of the IEC (and a panel) also needs to enforce the independence of this arrangement.

As was noted in the e-discussion, in the new CGIAR the Consortium is responsible for developing the Strategy and Results Framework and for designing, managing and overseeing the Mega-programs. The ISPC is an independent science advisory body and is itself free of the control of those responsible for the design, implementation and funding of the Mega-program.

2. Behavioural independence and conflict of interest

This refers to the independence and integrity of those who are conducting the evaluation and that their behavioural independence is protected and promoted by those who engage them. The proposal for the co-located arrangement entails that the IEC sets up the teams undertake each individual review. The peers are chosen for their expert standing and vetted for not having any conflict of interest (according to criteria that will be agreed). Not having IEC or evaluation staff participating in the evaluation of program secures independence better than id an evaluation unit staff would evaluate the programs.

3. Protection from external influence

The independent status of the IEC, independent funding and reporting prevent any parties from trying to influence the IEC or the evaluation teams (for example by withholding results/report). The selection process of the IEC needs the re-enforce the impartiality and independence of the position holder.

What is the co-location option likely to cost?

The past process of undertaking an EPMR (presented in Table 1) is the best guide for estimating the likely costs for the future evaluations of the Mega-programs. The basic evaluation arrangement is assumed to include:

- Independent Evaluation Chair and 1-3 Panel members (if included) operate as a Standing Panel (serving exclusively the evaluation function) with time requirement for Chair ~2 months per year.
- Evaluation arrangement (and Panel) is supported by 1-2 staff co-located at the ISPC Secretariat
- External evaluations are conducted by independent, peer panels as is currently the practice for EPMRs

The following factors would affect the cost compared to the benchmark (Table 1):

- Increase:
 - MPs are larger than Centers and potentially more complex
 - IEC (and Panel members) time
 - initially resources are needed for designing the independent evaluations
- Decrease:
 - MPs are fewer than Centers; with 7 MPs 1.75 evaluations per year
 - MPs have Consortium operated systematic and harmonized M&E designed to feed into the independent evaluations making the latter more cost-effective than EPMRs
 - Evaluation support more cost-effective than with EPMRs where two Secretariats were involved