Over the past decade, donors have made a concerted effort to improve aid effectiveness. This intent was codified by the Development Assistance Committee of the Organisation for Economic Co-operation and Development in the Paris Declaration of 2005, under which development agencies agreed to manage their aid budgets on the basis of five principles, two of which included managing development funds for results and mutual accountability (Chapter 6, this volume). The focus on results-based management and accountability further emphasized evaluation methods and monitoring systems in contractual arrangements associated with the investment of international public funds in Africa south of the Sahara (SSA), including investment in such public goods as agricultural research. The focus on performance monitoring in agricultural projects was reinforced by the entry of the Bill & Melinda Gates Foundation into agricultural development in SSA and by the integration of the private sector’s “bottom line” into project development and results frameworks.

The emphasis on accountability within the timeframe of project investment shifted the focus in performance monitoring from ex post impact assessment to monitoring and evaluation (M&E) methods, and thus from economic methods to methodologies coming from the expanding disciplinary field of evaluation. This disciplinary differentiation created boundaries between M&E frameworks incorporated in donor projects, where the focus was direct benefits and project targets and impact assessment associated with measuring both first- and second-order economic benefits, such as real income gains to consumers and producers, environmental protection, and institutional learning (Chapter 11, this volume).

Because most donors, except the World Bank, were not funding national agricultural research capacity and development projects, the focus on developing M&E methods for the relatively new area of agricultural research...
shifted donor investments to Africa’s subregional organizations (SROs), the Comprehensive Africa Agriculture Development Programme (CAADP), and CGIAR. This is where M&E capacity was principally built, with a resulting shift from incorporating M&E into projects, to developing M&E frameworks for institutions. Developing M&E frameworks to improve institutional management and learning, as opposed purely to accountability, was a natural outcome of this shift. M&E approaches have always been a compromise between the needs of the funders to know “Did it work?” versus the needs of the implementers to learn “How did it work?” This is the balance between accountability and learning.1

Donor support to agricultural research in SSA essentially comes from development budgets. Over time, the overall development objectives against which investments are measured have expanded; this applies equally to agricultural research. The initial focus on efficiency gains in smallholder farming systems that resulted in the large production increases under the Green Revolution has been systematically extended over the past several decades to include reduced rural poverty; sustainable natural resource management (NRM); and, most recently, improved child and maternal nutrition. Research planning and the design of M&E systems become much more complex with this increasing array of objectives. Holding agricultural research accountable for these development objectives at the same time that the timeframe to demonstrate results has telescoped in the shift from ex post impact assessment to results-based management has significantly moved agricultural research toward the development end of the R&D spectrum. This shift has moved research off station; has expanded the scope of research into postharvest, marketing, and support services; and has put a premium on institutional partnerships (Chapter 13, this volume). In the process, this has put more emphasis on learning.

This chapter begins with a discussion of the performance monitoring challenge in African agricultural research with respect to monitoring change in smallholder agricultural systems, the time lag in agricultural research, the context for technology adoption, and causality in agricultural research impact pathways. Next, four M&E cases illustrate the need to balance the objectives

---

1 Some of this balance is achieved when internal project-level monitoring processes track moves “along the crawl space” of projects, while the program or institutional processes track higher-order moves along the impact pathway, which does not exclude rigorous evaluation, such as randomized control trials. Structured experiential learning and real-time performance data with internal feedback to decisionmaking can use within-project variations in design as their own counterfactual, and thus reduce the incremental cost of evaluation (Pritchett, Samji, and Hammer 2013).
of accountability and learning made possible by M&E: (1) monitoring inputs versus outputs, (2) monitoring research process versus outputs, (3) monitoring research outputs versus broader innovative processes, and (4) monitoring the research process versus the development process. Thereafter follow discussions of the evolution of evaluation tools, the most-used logic models and their perceived strengths in dealing with the impact pathways from research outputs to development outcomes, and evaluation methodology and the efficacy of pilot studies versus formal evaluation designs. The chapter concludes with the search for accountability in investment in agricultural research and development (R&D).

The Challenge of Performance Monitoring in African Agricultural Research

Monitoring change in African agricultural systems is both complicated and data intensive because of the heterogeneity of agriculture in Africa, which creates market inefficiencies and increases the costs of adaptive research and extension. The InterAcademy Council (IAC) (2004) recognized this heterogeneity, noting that Africa would need thousands of mini-revolutions rather than a single Green Revolution. While the IAC identified four predominant systems that, in aggregate, were the likely levers for growth, they were in fact composed of many smaller systems operating within local constraints of markets and policies, reflecting the “small-country problem” in SSA. The small scale and heterogeneity of “development domains” (areas characterized by similar agroecologies, population densities, and market access), and the inherent weakness of public services, complicate planning of agricultural research and thus the M&E processes that support it. Impact pathways were less predictable, and monitoring change in farming and market systems is commensurately more labor and data intensive than in the more homogeneous Green Revolution Asia.

Research has raised agricultural productivity in SSA, but at a rate well below that of other developing countries (Chapters 2 and 3, this volume). To achieve the 6 percent agricultural growth rates postulated by CAADP, Africa would have to achieve productivity-based growth of 3 percent, matching rates achieved by Brazil and China, which have both invested more heavily in

---

2 The International Food Policy Research Institute’s addition of population and market factors to the traditional crop “recommendation domain” highlights the importance of public services and the complexity of planning agricultural research (Johnson et al. 2011).
research and rural services than Africa has. Even with doubling investment in research, ending discriminatory policies against agriculture, and doubling irrigated areas, individual countries would still encounter the small-country problem of high costs of policy and regulatory change to serve small development domains. Cooperation with neighbors, organizing technological spillovers, and building adaptive research capacity will be critical in improving the effectiveness of agricultural R&D (Chapter 14, this volume).

The lag between investment in agricultural research and actual impact on farm productivity is particularly difficult to specify. Pardey and Beddow (2013) argue that, for the United States, it is the accumulation of research results over the long run that accounts for differences in observed productivity among states: scientists build on the work of previous research going back much longer than the 10–12 years commonly estimated for the development and release of a crop variety, and up to 50 percent of the benefits can be attributed to spill-ins from outside. Crop varietal technology timelines illustrate the lengthy innovation process typical of agriculture, and the complementary roles of the public and private sectors. Moreover, the timeline between research discovery and commercial interest is longer than is often assumed. In developed countries, a large share of basic research is conducted in universities and public research institutes, whereas the private sector is predominantly investing in developmental research. In Africa, the public sector—made up of both national and international research agencies—must undertake the whole research chain, from basic to developmental research, which in more developed countries is largely in the hands of the private sector.

The heterogeneity of African agricultural systems, the large range of commodities in these systems, and the relatively small development domains involved often overwhelm limited research budgets. In Africa in the past two decades, the relative weakness of the private sector and underfunded public systems define three needs: (1) to catch up in the area of basic science that underpins key commodities, (2) to build capacity to access knowledge widely, and (3) to create linkages with the wider innovation system.

What Gets Monitored?

Project Versus Institutional Funding

A recurring preference of donors for project rather than long-term program or institutional funding has concerned analysts since the early 1980s (Ruttan 1982). The advantages of long-term program or institutional support
include lower transaction costs, the ability to plan and complete long-term path-breaking projects, and the opportunity to develop human and institutional capacity around challenges. On the other hand, contracts, short-term programs, and competitive grants create flexibility to change programs (not just at the margin) and assert individual donor interests. Such approaches require special attention to the continuity of the scientific core and to capacity building, which are not secured by other budgets.

The original CGIAR, with core support directed to addressing major global challenges, typified the type of institutional innovation that could bring results. The erosion of the core-funded model, with its periodic External Program and Management Reviews, gave way to a proliferation of project-based activities, new initiatives, and enabling structures and mechanisms in which individual donors carried out their separate reviews, often as sub-components of larger programs. This situation meant that donors ended up doing their own M&E by research project and theme, with donor-specific criteria. The resulting burden on national systems was recognized in the Paris Declaration on Aid Effectiveness (OECD 2005), which recommended greater coordination among donors and the provision of budgetary support, rather than project aid.

At the regional level, agricultural productivity programs, such as the East and West African agricultural productivity programs (EAAPP and WAAPP, respectively), were created in the World Bank under multidonor trust funds that provided loans to participating countries for a mixture of national and regional activities. SROs became involved, essentially managing competitive grant mechanisms. Given the importance of CGIAR in regional networks and the new intention to align with African agendas, M&E tools at all levels have been rapidly sought to structure a results framework around consensus on an African agricultural development agenda.

Accountability and Learning in M&E

The common belief that accountability and learning are necessarily competing objectives is disputed by Guijt (2010). However, project formats usually require accountability statements about predefined goals, and a specification of the activities and interim results that lead to their achievement (for example, in “results-based management”). Theories of change and the learning that is embedded in them can become overly rigid, as project contracts often require proof of deliverables and set milestones that negate the adaptive management that is so essential in research programs. Often there is a fundamental disconnect between the rhetoric about the need for learning and the reality
of procedures that funding agencies require. Whether it occurs because of an inability to deal with complex realities, a preference for the easily measurable over more difficult-to-measure objectives, or human and institutional capacity constraints, the tendency for accountability to prevail over learning is commonly acknowledged.

- Learning is necessary for practical improvement, strategic adjustment, and rethinking driving values. Nongovernment organizations (NGOs) are often in the lead in promoting learning through evaluation. ActionAID International (2011) is one among several NGOs that seeks “triple loop learning”: (1) Are we doing things well? (2) Are we doing the right things? (3) Were our assumptions right? There needs to be clarity about who is expected to learn, for what purpose, and at what level the learning is aimed, which is then incorporated into the design of the M&E system.

- Accountability has many forms: to managers or funders (upward accountability), to civil society (social accountability), to partners (mutual accountability), and to researchers themselves (strategic accountability, in the sense of “Did we act as effectively as possible?”).

Guijt (2010) argues that M&E must adapt expectations of accountability and learning to contextual characteristics. In ordered situations, cause and effect are clear and primarily linear, and impact pathways can be analyzed within a results-oriented system. In complex situations, such as research on smallholder systems in Africa, learning is needed to recognize emerging patterns, to respond, and to adapt. For example, the attempts to define an impact pathway for agricultural approaches to health and nutrition outcomes have identified multiple interactions, in part that derive from three separate goals, each potentially requiring its own dedicated policies, instruments, and intervention strategies. This recognizes not only that agricultural research is a “blunt instrument” for achieving policy goals, but also that in adding a goal, such as nutrition, a new instrument is needed. The skills required for compliance with donors, for strategic accountability, and for learning are convergent, so capacity building and the design of M&E systems should go beyond compliance to ensure that all three needs are met.

**Causality in Impact Pathways**

The evolution of evaluation approaches in research is associated largely with CGIAR, particularly during a period when it was challenged to prove its contribution to development goals, and when donors were withdrawing their
support for research at the national level. CGIAR’s Standing Panel on Impact Assessment (SPIA) provides a useful classification of different types of assessments and evaluations on research for development (Figure 12.1).

Until CGIAR reform in 2008, evaluation of investment in agricultural research in CGIAR centers was based almost wholly on periodic external reviews in the form of ex post Impact Assessment (epIA). Such assessments were based on where research outputs were known to have been widely adopted and had impacts. These impacts were modeled through economic surplus methods that depended on translating cost reductions and production increases at the farm level into market impacts on price and the distribution of net benefits between consumers and producers. Only those parts of the research portfolio that had demonstrated impact were assessed, and epIA gave an estimate of the returns to investment on that particular line of research. Such rate-of-return studies had little to say about the market and institutional context within which the technology diffused, how the technology was disseminated, the effect of the technology on ecosystem services, or the potential for reproducing such impacts under other contexts.

In sum, there was little potential for deeper learning about causality within the impact pathway from such epIAs. Moreover, by the time of the impact study, the research program had often moved on to other research lines. Such

---

**Source:** Maredia (2013).  
**Note:** M&E = monitoring and evaluation.
assessments essentially provided an ex post measure of the returns on investment, but they provided little information on how investments in agricultural research could best achieve the development objectives that framed donor funding strategies. It was at this point that the expanding field of evaluation, which had been applied widely in the health and education sectors and was being applied internally in donor programs, was integrated into agricultural research programs.

The demands from investors in international public goods, such as agricultural research, were that aid recipients must demonstrate contributions to development objectives from their funding. Since investments in agricultural research had a longer time lag than most other investment, it would be necessary to monitor more immediate outcomes that could be causally linked to these broader development objectives—what became known within CGIAR as “intermediate development objectives.” For agricultural research in Africa, this has had a number of effects.

• First, research programs have had to be much more explicit about their impact pathways and the chain of causality between the production of research outputs, their dissemination, and the impacts on a possible range of intermediate outcomes from gender equality to enhanced ecosystem services.

• Second, in an African context, this has expanded the scope of research from a sole focus on productivity-enhancing technologies, to research on improved market efficiency and smallholder access to markets, innovations in crop or livestock insurance, more cost-effective dissemination methods using information and communications technologies, or farmers’ organizations that improve access to capital and technologies by the rural poor or women in rural households. Such research in SSA recognizes that innovations are needed in market, institutional, and farmers’ organizations to support adoption and effective use of more traditional production technologies.

Finally, enhanced accountability for impact has necessitated the development of more effective partnerships with other actors in the agricultural sector in order to achieve broad-scale development outcomes, which in turn has increased the focus on research within a larger agricultural innovation system (Chapter 13, this volume). M&E systems are developed on the basis of such increased understanding of impact pathways. Moreover, Maredia (2013) notes that in moving down the impact pathway, the assessment of results in terms of
intermediate development outcomes becomes less a case of analyzing “causal attribution” and more one of analyzing “causal contribution.”

**What Needs to Be Monitored?**

Impact pathways for agricultural research, especially in an African context, are highly context dependent, have long timeframes, and require functional organizational partnerships where many institutions have relatively weak capacity. With the expansion of agricultural research into institutions, policy, land use, and the environment, there is an expanding range of processes that research institutes develop and deploy to meet development objectives. This expansion results in significant complexity in designing monitoring systems that meet accountability needs along the impact pathway, but at the same time provide information for assessment, program adaptation, and learning. Such monitoring systems for agricultural research may be conceived in terms of different stages or levels. The discussion that follows is only suggestive of the design issues in developing such a monitoring system.

**Monitoring Inputs Versus Outputs: The Case of Plant Breeding**

Agricultural research is in many ways a design process that combines “engineering” methods with scientific creativity. This qualitative and uncertain nature of agricultural research has made it difficult to measure the performance of the agricultural research process itself. Much of the focus in such monitoring systems has been on the inputs into the process. Agricultural Science and Technology Indicators data collection is a very good and representative example of such monitoring, as the data collection focuses on primary inputs of scientific personnel, operating funds, capital investments, research infrastructure, and training.

What has been more difficult is to relate such inputs, including organizational alternatives, to the performance of research institutes and the overall efficiency of the research process. The clearest performance indicator has been publications, as measured in terms of journal impact and the number of citations. However, knowledge production does not easily translate into relevant technologies and management systems, nor does it necessarily translate into impact on intermediate development outcomes. In fact, many development agencies suggest a trade-off in publishable research and what is increasingly termed agricultural research for development. The problem thus remains how to monitor the research process itself as the first stage in an impact pathway.

Plant breeding would be a tractable example of monitoring at this stage. Numerous design choices go into a breeding program, particularly the number
of breeding projects (the number of agroecologies that require separate adaptation), the number of traits, whether populations or elite lines are to be developed, the breeding methodology employed (for example, recurrent selection versus broadening parents in the crossing program), whether to employ marker-assisted selection (as is increasingly becoming the case), and the number and location of testing sites. All of these choices have cost and resource implications and essentially determine the inputs that go into a breeding program. The question becomes whether the program could be more cost-effective and in relation to what performance criteria.

Optimally, the performance criterion would be the number of adoptable varieties; however, there is a significant time lag between investment decisions in breeding programs and a sufficient understanding of actual farmer adoption, so more intermediate performance criteria are needed. These could include the number of varieties released under national varietal release regulations, the results in the national yield performance trials, the results in the multilocalional variety trials, or trait performance in the selection process. All of these factors are used in varietal evaluation (with increased precision from the first to the last listed), but with increasing time lags in terms of feedback into breeding program design and adaptation. Developing a monitoring system involves inherent trade-offs among precision of outcomes, time lags, data costs, and the need to make adjustments in the breeding program—that is, to answer the question of whether the program is performing well.

**Monitoring Process Versus Outputs**

To deal with the lag problem in adaptive management of breeding programs, innovations have been in process. Participatory breeding and participatory varietal selection have been designed to telescope the lag time between varietal development and farmer evaluation. The potential trade-off has been in how locally adapted such varieties will be, which may be overcome by incorporation into multilocalional trials. To a significant extent, the results of the process provide the outcome indicator, and the focus is on making the process as cost-effective as possible. In this way, monitoring shifts from a focus on outputs to more of a focus on process variables; at the same time, it extends the research to a closer interface with development agents and outcomes—what is more broadly termed “client-oriented research.” Research moves primarily on farm, which in turn requires a specification of context and representativeness, particularly given the heterogeneity of African farming systems. In this regard, the research moves to another stage in the impact pathway, with significant differences in how the monitoring process is designed.
Monitoring Research Outputs versus Innovation Process

As previously discussed, there has been an evolution of systems approaches to the changing objectives of national systems and funding agencies, the need for scientists to design results to fit in wider knowledge systems, and the need for institutions to work together in a broader innovation systems framework (Elliott 2008; Chapter 13, this volume). The convergence of policy, science, and institutional perspectives on innovation systems evolved as follows:

1. **Policy and Systems Perspective**
   - National agricultural research institutes →
   - National agricultural research systems →
   - Agricultural knowledge and information systems →
   - Agricultural productivity programs →
   - National agricultural innovation systems

2. **Science Perspective**
   - Crop improvement →
   - Cropping systems →
   - Natural resource management →
   - Climate change/sustainable intensification

3. **Institutions and Change**
   - Client-oriented research →
   - Farmer-to-farmer dissemination →
   - Demand-driven advisory services →
   - Integrated agricultural research for development/innovation platforms

Evaluation systems have grown up around benchmarking systems and measuring these evolutions (Spielman and Birner 2008; Spielman and Kelemework 2009a, 2009b; Ragasa, Abdullahi, and Essegbey 2011). The target for such analysis would be institute directors and ministers of agriculture who wonder whether they are doing the right things, whether they are making the correct assumptions, and whether they are linking to the rest of the national agricultural innovation system.

Monitoring the Research Process versus the Development Process

The emphasis on development outcomes has fundamentally shifted the monitoring of agricultural research, driven by the introduction of “results frameworks.” The research process is expected to identify scenarios of progress along impact pathways leading to development outcomes. Since national agricultural
research systems (NARSs) and CGIAR are both dependent on donors (and their funding comes from development budgets, rather than budgets for scientific cooperation), it is understandable that the agenda has moved toward the development end of the research-for-development spectrum. The challenge for research is to respond to the concerns of development (and “speak their language”) without abandoning the primary role of researchers. This is discussed below in the context of designing results frameworks and defining intermediate development outcomes along the impact pathway.

The move to “results-based management” (RBM) has been led by donors who have to reassure their funders, policymakers, or constituents that public money is being spent responsibly. RBM shifts the justification of the investment in agricultural research from ex post impact evaluation to direct monitoring of the results emanating from the investment. For donors, these results should be measured in terms of impact on development outcomes. For agricultural R&D, the difficulty is the timeframe over which such results are expected to be generated. Moreover, the evaluation tools that have traditionally been used, such as logical framework (“logframe”) analysis, outcomes mapping, and participatory impact pathway analysis, have been designed, essentially, for development programs. The approaches all involve some defined investment or intervention that is expected to yield gains to some target group, while identifying the policy and institutional assumptions under which that gain can be achieved with some stated probability of success. Most critically, within RBM, if a constraint is binding and there is no budgetary or program plan within the original project to change it, the project must include within itself the means of removing the constraint.

RBM systems are being instituted in most agricultural research organizations that depend on donor funding, particularly the Alliance for a Green Revolution in Africa, the Forum for Agricultural Research in Africa (FARA), the SROs, and CGIAR. CGIAR is often a vehicle through which research breakthroughs come into practice in the NARSs (Chapter 3, this volume). CGIAR’s chief executive officer recently noted (CGIAR 2013b):

Starting today, we are presenting and discussing the Theories of Change, Impact Pathways and Intermediated Development Outcomes for the CGIAR Research Programs (CRPs), with our donors and our partners. What we are talking about here are clear, development outcomes that have been reached by consensus. The Programs, together with their partners, can be held accountable for delivering these outcomes. Program level outcomes tied to outcomes at CGIAR System
Level . . . are in turn linked to the new Sustainable Development Goals. Together with solid monitoring and reporting, these outcomes constitute the foundation, the building blocks for the CGIAR’s results based performance management system.

The Essentials of a Results-Based Management System

An RBM system must be designed for a specific user (or partnership) working within a described context if it is to serve both learning and accountability objectives. However, many elements are common to an RBM, as is illustrated by CGIAR’s template for CRPs (summarized in Appendix Table 12A.1). The template is an appropriate example, because it is designed for partnerships involving CRPs, NARS organizations, and partners in advanced research institutes, and it highlights some of the complexities and responsibilities inherent in an RBM system.

Whose Agenda Is Being Monitored and Evaluated?

The influence of donors on the M&E agenda is often determinant. In response to their own constituents’ demands for justification, donors have pulled the whole planning, monitoring, learning, accountability, and evaluation system toward accountability for final outcomes. At the same time, donors have prescribed processes, funded particular technical partners, and recommended institutional arrangements—sometimes with limited evidence for their prescriptions.

The introduction of the theory of change (TOC) as the latest refinement of logic models is an attempt to force planners to be more explicit about their assumptions and evidence for presumed links between efforts and results. The original logframe analysis outlined a development hypothesis by which inputs were transformed into outputs, outputs contributed to strategic objectives, and strategic objectives contributed to goals. Other variations of the planning tool argued that activities contributed to outcomes, outcomes contributed to purposes, and purposes contributed to goals. The approaches all required that planners specify the assumptions under which the development pathway would

3 The order in which planning, M&E, learning, and accountability are presented is significant. While the temptation to create an acronym “PMEAL” is present, most practitioners assert that planning, monitoring, and learning come before evaluation and accountability and represent the interest of the scientists and implementers of development for efficiency and outcomes.
logically lead to its goals. Where planners identified missing resources, skills, or institutions, the project would be required to include the additional resources or actions needed to ensure the result.

A contemporary approach, called “outcome mapping,” worked backward from desired goals in a participatory mode to ensure the necessary skills, processes, and institutions for success were in place. Since these logical exercises became an integral part of project documents, their targets often became rigid deliverables, where missed milestones were treated as failures, rather than as learning opportunities.

In a more adaptive mode, many agricultural research institutions use participatory impact pathway analysis (PIPA), which develops the logic model and impact pathway in participation with stakeholders. PIPA separates the performance management stages of inputs-into-outputs from the final evaluation. Monitoring and learning are continuous and separated from “evaluation,” which takes place after a project’s completion. The participatory learning process permits correction of the impact pathway and even revision of the expected goals themselves. Participants derive outcome targets and milestones, which are regularly revisited and revised as part of the project’s M&E framework. This goes beyond the traditional use of logic models by engaging stakeholders in a structured participatory process, promoting learning, and providing a framework for “action research” on processes of change (Douthwaite et al. 2008). To do this, the monitoring indicators must be SMART—that is, specific, measurable, attainable, relevant, and time-bound or trackable (Chapter 11, this volume). Thus, stakeholders know which predictions or hypotheses are being confirmed and which are being corrected.

**Understanding the Data Requirements of a Full Results-Based Management Approach**

Recent contributions to conceptualizing the need for detailed TOCs and attempts to measure them come from the World Agroforestry Centre, which reviewed attempts to scale up agroforestry to sustainably increase production and maintain environmental services (Coe, Sinclair, and Barrios 2014, 73):

Evidence suggests that this will not be achieved by wide scale promotion of a few iconic agroforestry practices. Instead, three key issues need to be addressed. First, fine-scale variation in social, economic and ecological context and how this creates a need for local adaptation. Second, the importance of developing appropriate service delivery mechanisms, markets, and institutional contexts, as well as technologies. Third,
appropriate research design, within the scaling process, that enables co-learning amongst research, development and private sector actors. This requires a new paradigm that builds on previous integrated systems approaches, but goes further, by embedding research centrally within development praxis.

Achieving development outcomes with agricultural technologies, particularly in an African context, requires fine-scale targeting and adaptation, where results critically depend on context, which in turn influences the design of the dissemination or scaling program. Understanding the adoption of new techniques and the impact on household welfare requires sex-disaggregated, household-survey data across institutional, economic, and ecological contexts within the target region or population. Developing a systematic household-survey capacity to monitor change and impact on household welfare requires efficient sampling frames, standardized survey instruments, and the ability to resurvey. Moreover, household panel survey designs usually have to be supplemented with surveys of marketing agents, service providers, and agroclimatic conditions in order to locate the household within its socio-economic context. Such surveys are costly within an African context, particularly if conducted independently for every research program. This has led to efforts to develop standardized national household surveys within a panel survey design under the Living Standards Measurement Studies–Integrated Surveys on Agriculture. However, these efforts have been instituted only in a few countries, and essentially shift the public-goods nature of the costs to national statistical agencies, often supported by development partners.

A renewed commitment to agricultural statistics and information may be an important force for improved knowledge and attention to agriculture. Regular collection of intercensus data on agriculture will be enhanced, but the flood of information on markets, trade, and prices made possible by mobile telephony will be a potential game changer for M&E. Remote sensing will be another technology harnessed for agricultural development. As discussed above, household surveys will be a third tool. The adoption of the Global Strategy to Improve Agriculture and Rural Statistics (GSARS) (WB–FAO–UN 2010) aims to produce a minimum set of baseline data, improve mainstreaming of agricultural statistics in national statistical systems, and build capacity. With support from the Bill & Melinda Gates Foundation and the UK Department for International Development, tailored strategies are being developed for individual countries. Open data, combined with information from statistical databases, may eventually provide practical solutions
for farmers. As the coordinator of GSARS cautions, however, statistics is an inferential science that requires representative samples, not just more data (Guillaume-Gentil 2015). Harnessing the data revolution for planning and performance management will require careful design.

**Targeting and Scaling Out Interventions**

A recent study by the CRP on Climate Change, Agriculture and Food Security deals with the conceptual issue of targeting and scaling out interventions through an iterative mapping process. The authors developed guidelines for evaluating and prioritizing potential interventions (Herrero et al. 2014, Abstract):

There are real needs and opportunities for well-targeted research and development to improve the livelihoods of farmers while at the same time addressing natural resource constraints. The suitability and adoption of interventions depends on a variety of bio-physical and socio-economic factors. While their impacts—when adopted and out-scaled—are likely to be highly heterogeneous, not only spatially and temporally but also in terms of the stakeholders affected. In this document we provide generic guidelines for evaluating and prioritising potential interventions through an iterative process of mapping out recommendation domains and estimating impacts. As such, we hope to contribute to the inclusion of such important considerations when agricultural innovations are targeted and scaled out.

Targeting and scaling out are key components of an integrated ex ante assessment process, as well as fundamental to implementing downstream adaptive research and scaling programs. “Two underlying questions are addressed by this framework: which data are required for targeting and scaling out, and how can the data be integrated to assess different impacts of a range of interventions” (Herrero et al. 2014, 8). The data used in this ex ante assessment provide the basis for establishing the indicators by which progress can be monitored.

The complexity (and cost) of such an exercise is evident in the steps necessary for discerning how scalable specific practices might be for improving food security, NRM, and livelihoods, and for mitigating the impacts of climate change. The steps involve identifying

1. the characteristics of the intervention that may affect its use and adoption in agricultural systems;
2. the recommendation domain for the products of research and where these are likely to be applicable;

3. the groups of people who are likely to be affected by the output of the technology/intervention; and

4. the nature of the impacts, in terms of their type, their magnitude, and the associated trade-offs at different temporal and spatial scales, and among the different types of impacts.

The detailed knowledge of population structures and behavior, as well as the ability to link this knowledge to land quality and potential, underlines the data intensity of such an exercise. It is definitely beyond the capability of most NARSs and requires collaboration with advanced research institutes.

**Approaches to Developing an M&E Capacity within African NARSs**

The development of an M&E capacity within African NARSs cannot be separate from the development of sufficient well-trained graduates at the MSc and PhD levels with the technical skills and field experience in agriculture demanded by employers. This demand comes from the banks, private agribusiness, universities, research institutes, and donor agencies that hire away the best candidates while lamenting the inability of NARSs to retain staff. Since evaluation can be taught to agricultural scientists, as well as economists and agricultural economists, the pool begins with a mix of agricultural skills and experience. The Regional Universities Forum for Capacity Building in Agriculture (RUFORUM) is an example of a regional approach to train high-quality postgraduates who are demanded by the market, and M&E is part of their MSc program in research methods.

A second assertion is that a great deal of capacity development is already occurring in Africa. However, it is poorly coordinated among national educational programs, donor support to higher education, and private-sector demand. Moreover, donor-sponsored training tends to target bilateral projects and initiatives. Several African countries have been leaders in moving planning and evaluation forward. With the support of the Rockefeller Foundation and the former International Service for National Agricultural Research, the Kenya Agricultural Research Institute (KARI) was an early adopter and codeveloper of formal approaches to research planning adapted to local conditions (authors’ personal experience).
Third, efforts by CGIAR to establish RBM in its CRPs will have the effect of transferring skills to the SROs and NARSs; they must speak the same language if they are to develop the partnerships needed to achieve intermediate development outcomes. Many senior African leaders started their careers in planning and evaluation within NARSs, moved to SROs to work on regional M&E standards, and subsequently joined international organizations, donor agencies, or international research centers. This career trajectory should not be lamented; rather, it should be treated as part of the incentive structure for high-quality planning and evaluation in NARSs.

Finally, capacity development for M&E is a continuing investment. The focus of evaluation changes over the life cycle of an intervention, the users of the information are different, and the tools are continuously evolving. All this militates in favor of efforts to agree on common concepts and definitions, and compatible data collection.

The following sections outline two innovative experiences with M&E related to a larger agenda of policy and technical change. The first case documents the long-term use of rigorous M&E by the International Potato Center (CIP) to prove that orange-fleshed sweet potato (OFSP) can be a successful and economic means of addressing vitamin A–deficiency blindness in children. Studies have successively proved the effectiveness of biofortification, evaluated efficiency in reaching end users, and monitored efforts to reach policymakers and agents of change. The second case looks at the data-driven, evidence-based approach of the Ethiopian Agricultural Transformation Agency (ATA), and the role that rigorous M&E has played in the early stages of its implementation of a longer-term transformation agenda.

Evolving the Evaluation Focus with the Success of the Intervention

CIP runs a multistakeholder program: the Sweetpotato for Profit and Health Initiative (SPHI), with the goal of reaching 10 million households across 17 SSA countries over 10 years, through the widespread uptake of sweet potatoes to reduce malnutrition among children under the age of five years. A major program under SPHI applies the lessons of a HarvestPlus project (HarvestPlus 2012) and is now an initiative called Reaching Agents of Change (RAC). In turn, RAC’s M&E plan is a useful example of a well-articulated RBM framework. CIP and Helen Keller International will reach out to regional and national agencies who will act as agents of change for OFSP.

Working back from its goals of reducing vitamin A deficiency and food insecurity, particularly of women and young children, RAC sets up specific
objectives related to (1) generating new investments to scale up the adoption of OFSP in five countries, and (2) building the capacity of implementation agencies to design and implement technically strong and cost-effective interventions focused on (or including) OFSP. A third specific objective deals with the management of RAC itself. Its TOC is based on knowledge of what works and how change is expected to occur. Working back from a vision of success, it poses three questions:

1. What advocacy activities are necessary to generate investments by governments, donors, and NGOs to scale up OFSP?

2. What materials and capacities do implementing agencies need to deliver strong interventions?

3. What resources, capacities, and activities will the project need to manage for results?

The M&E plan is structured with biweekly activity reports, quarterly monitoring of achievements toward milestones, and a special narrative on systematic, corrective actions taken based on lessons learned. Comparing experiences across Burkina Faso, Ghana, Mozambique, Nigeria, and Tanzania, the RAC project will generate lessons on what works across different contexts and what corrective measures are needed in individual countries (Mbabu et al. 2015).

The example of OFSP illustrates several important considerations in designing an M&E system: (1) clarity in specifying goals, (2) a decent TOC that allows for adaptive management, (3) data collection instruments that identify needs for corrective action in real time, and (4) the involvement of local implementers in managing for results. The RAC program promotes policy advocacy at one end and technical skills development at the other. RAC’s M&E plan identifies the levels of responsibility for implementation and arms the actors with the skills and tools to manage within each level. Information on activities carried out is recorded biweekly, whereas quarterly reports are filed with the M&E officer.

The OFSP case highlights the way that M&E and impact evaluation methodology have adapted to the task: first, randomized controlled trials to establish proof of concept; second, less costly evaluation of the phase of reaching end users; and currently, detailed M&E of the TOC associated with reaching agents of change. The RAC project is being piloted and monitored in two countries (CIP 2012).
Monitoring and Evaluation in a Results-Based Management System

The ATA has a much broader agenda than agricultural research—notably, the transformation of Ethiopia’s agricultural economy. The goal is an agricultural sector that is more technologically intensive, creates more value up and down the value chain, contributes more to social goals, and ultimately represents a smaller share in gross domestic product through its contribution to growth elsewhere. ATA’s strategy is to focus on transformation clusters that are regionally and geographically identified for their industrial and export potential, and to address bottlenecks in support systems and value chains. ATA was able to base its planning on baseline studies carried out by the International Food Policy Research Institute and the Ethiopian Development Research Institute (EDRI) under the Ethiopian Strategy Support Program (ESSP). The studies and ongoing support enable ATA to identify priority points of entry along value chains, as well as key systemic bottlenecks, while building toward a broader transformation of policies and institutions.

ATA is unique among similar entities in Africa because of (1) its political support and organizational status; (2) the analytical base on which it was founded; and (3) its structure as a data-driven, results-based learning organization. ATA serves as secretariat to an Agricultural Transformation Council chaired by Ethiopia’s prime minister and comprising key ministries, has a reporting role to the Ministry of Agriculture, and plays an important role in building capacity within the ministry and regional bureaus of agriculture. ATA has developed internal monitoring, learning, and evaluation systems for reporting on its own deliverables in advancing more than 100 proposed solutions to overcome prioritized constraints in value chains, systemic bottlenecks, and cross-cutting issues. The deliverables are time denominated and, as such, impose a strong internal discipline. ATA reports quarterly to the Transformation Council regarding key issues and the status of each deliverable (not started, significant issues, slight issues, on track, or completed), all of which is supported by analysis and suggested actions and is included in ATA’s annual report.

ATA has been judicious in managing its dual relationship with the Transformation Council and Ministry of Agriculture. Being mandated to identify and address bottlenecks, it is not an implementing agency per se.

---

4 A “deliverable” is an activity that has not yet become an outcome or an impact, such as a strategy for a particular commodity or goal.
Nevertheless, it does work with other agencies and departments in building capacity and providing analytical and reporting strength. ATA is accumulating a track record in overcoming near-term technical and logistic constraints, while moving forward in addressing the more difficult policy and institutional constraints needed for the long-term transformation of Ethiopia’s agricultural economy. These constraints lie in trade policy, financial systems, cooperatives, and extension. ATA’s success has originally been in addressing priorities identified by the prime minister—recognizing that cooperatives and extension play an important national role in the political structure—while at the same time being part of the transformation agenda.

Even though ATA’s experience is too recent to evaluate, some lessons are emerging. First, building capacity in the Ministry of Agriculture is proving more difficult than originally expected and will require additional attention. Second, while it is still too soon to evaluate the impact of ATA’s strict monitoring of its deliverables on the broader transformation agenda, its discipline in tracking implementation does afford some credibility, which carries over to its longer-term strategic analyses. ATA’s ability to call on ESSP for research—for example, on policies relating to the importation of wheat versus enhanced domestic procurement efforts, while maintaining market stability—puts ATA at the heart of policy and institutional transformation. Third, the Transformation Council’s link to the prime minister’s office is a valuable asset. It is recognized, however, that ATA’s data-driven, evidence-based approach, led by the Bill & Melinda Gates Foundation and supported by other donors, will be challenging to replicate in other countries.

Creating the Necessary “Architecture” for Monitoring and Evaluation

With the 2013 memorandum of understanding (MOU) between the African Union (through CAADP) and CGIAR, a new “architecture” is creating the prospect of bringing order to the complexity of a global system. The elements of this new architecture, if implemented, include

1. aligned strategies at the regional, subregional, and national levels through an MOU between CAADP and CGIAR;
2. coordinated CGIAR center activities through multiactor and multi-stakeholder CRPs;
3. a fund council representing the collective will of the donors (constituent CGIAR centers receive some overhead through CRPs, although
inadequate support to the core remains a potential weak spot in the system;

4. funding from Africa’s national governments, which have committed to meet the CAADP target of investing 10 percent of their national budgets in agriculture;

5. compatible tools for planning, monitoring, and evaluation, eventually following similar results frameworks supported by FARA and the SROs at the national level;

6. regional programs, such as EAAPP and WAAPP, that have pioneered the way for the World Bank to fund regional activities through loans to participating countries and to engage the SROs in program management; and

7. independent advice from the International Science and Partnership Council (ISPC), which has been monitoring the experience with impact pathways.

Making the Architecture Work for Decisionmaking, Measurement, Learning, and Evaluation

This chapter has shown how M&E has evolved over the past three decades, along with planning tools that address the changing concerns of policymakers, funding agencies, and implementers of agricultural research. Their needs, respectively, are for tools that (1) reduce the risk of decisionmaking; (2) provide accountability to government and funding agencies that money was well spent; and (3) enable implementers to identify problems, flexibly adjust activities, and readjust goals in real time.

The institutional architecture is largely in place: the CGIAR centers are well located throughout the African continent through regional and subregional offices, while the CRPs target agreed-upon problems with strong participation from NARSs, advanced research institutes and development-oriented partners, NGOs, and agencies. Through the African Union, New Partnership for Africa’s Development, and CAADP, there is a policymaking forum linked to the heads of state and a technical structure through FARA and the SROs for collaboration among NARSs. Many countries, however, lack strong articulation of a national agricultural innovation system to bring research, higher education, extension, and the private sector together. CAADP compacts seek to close these gaps while committing governments to a meaningful funding target to support
### APPENDIX TABLE 12A.1  CGIAR's essential requirements of a results-based management system

<table>
<thead>
<tr>
<th>Requirements</th>
<th>Notes (abbreviated from source document)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Statement of strategic goals (1 page)</td>
<td>Strategic goals of the CGIAR Research Program (CRP) should be reachable within 12–15 years and show contribution to system-level objectives. Note that increasing productivity or income cannot be assumed to automatically lead to poverty alleviation or food security, and that productivity or income growth cannot be maintained in the long term without effective management of natural resources and effective policies.</td>
</tr>
</tbody>
</table>
| 2. Impact pathway, theory of change (TOC), key hypotheses (4 pages) | • Highlight the TOC and key hypotheses upon which the TOC is based, and which the CRP will test in its own M&E system.  
• Synthesize the major elements of the CRP’s high-level impact pathways and their supporting assumptions and TOCs. |
| 3. Justification of international comparative advantage (1.5 pages) | Explain why the CRP has a clear comparative advantage. This is equivalent to demonstrating that no institutions are better placed to produce the necessary results, given the impact pathway of the CRP. |
| 4. Intermediate development outcomes (IDOs), targets, and Indicators (3 pages) | For each intermediate IDO, specify the target population and geographical area concerned and quantify the targets. Indicate the level of uncertainty associated with those targets and, if necessary, plans to refine them. Robust indicators of progress toward these targets must be specified. |
| 5. Flagship projects and clusters of activities (15 pages) | The guidelines are specific to CGIAR and call for organization of flagship programs and clusters of activities underneath the flagships. This organizes work in the most effective manner to produce the outcomes, outputs, and IDOs and monitor progress. The flagships must show the linkages the CRP has built with other CRPs, with a description of the type of research undertaken and its sustainability. |
| 6. External partnership strategy and intellectual property issues (6 pages) | • Explain the CRP’s strategy concerning external partnerships (non-CGIAR partners) from discovery to scaling-up phases.  
• Include a typology of your partners and their functional role in the CRP, including governance.  
• Explain how intellectual property issues are managed through these different partnerships. |
| 7. M&E and risk management strategy (5 pages) | • Describe the CRP’s approach to risk management and mitigation.  
• Describe how the CRP tracks its own progress within clusters of activities and subcluster levels, how adjustments/alignments are implemented when progress is not as expected, and how this approach fits with the overall results-based management implemented at the CRP level. |
| 8. CRP governance (1.5 pages) | • Describe how the CRP management structure draws upon resources and talents across CGIAR centers and beyond the CGIAR system.  
• Indicate how the recommendations of the external review of management and governance in CRPs will be implemented by this CRP. |
| 9. Budget request (3 pages) | • Explain the budget needed to produce the outputs, research outcomes, and IDOs.  
• Indicate the budget allocated to partners outside of CGIAR over time. |

**Source:** CGIAR (2013a).  
**Note:** M&E = monitoring and evaluation.
the system. The growing divide in human resource research capacity among the region’s countries must be reversed (Chapter 8, this volume).

African countries have not been slow to adopt new practices in planning, monitoring, and evaluation when they serve national needs. KARI was an early adopter of formal priority setting before many countries in other regions of the world. Ethiopia’s ATA is pioneering harmonized data for coordination of the system. The migration of African expertise among leading countries and institutions at the regional and subregional levels has helped spread the concepts and the development of tools. Nevertheless, the uniform development of the data, tools, and practices among countries, which would make M&E a powerful tool, is lacking. Donors continue to have their preferred partners under various alliances or bilateral programs for historical reasons. The recent Science Agenda for Agriculture in Africa argued that the African Union should take up the challenge to ensure that no country is left behind in having sufficient scientific capacity (FARA 2013). Its adoption by the heads of state in Malabo is encouraging. It is hoped that they will address the inadequacy of data for planning and monitoring Africa’s agricultural growth and development, which includes linkages with R&D and national-level data sharing to draw institutions of agricultural knowledge together in a system.

References


