CHAPTER 2

THE ANALYTICAL FRAMEWORK

This chapter deals with the classification and selection of indicators. The logframe is used to differentiate between project inputs, outputs, outcomes and impact. Indicators are needed at each level for effective monitoring and evaluation; each have their own defining characteristics and are discussed in turn. Tools to facilitate the collection and use of such indicators are reviewed. The main focus of the chapter is, however, devoted to outcome and impact indicators, and to the measurement of results, in particular, early results. The Sourcebook suggests that, for early results, a service delivery approach can work well. For longer-term results and impact measurement, a menu of core statistics is proposed. The chapter concludes with recommendations for selecting indicators for the ARD sector as a whole and for the various subsector programmes. Nineteen priority indicators are proposed. The process may also be assisted by reference to Annex 1, which contains a menu of potentially useful indicators.

THINKING LOGICALLY ABOUT INDICATORS

A good M&E system should, in principle, be integrated into all stages of a project or programme cycle, from identification through the evaluation. At each stage, it should seek to answer the question, "Are we on track?" At the end, it should answer the question, "Did we achieve what we wanted to achieve?" Throughout the duration of the project, the M&E system should generate timely reports on project progress, sounding alarms where necessary, and providing project management with the necessary information to help keep the project running as smoothly as possible. In the end, sufficient information should have been accumulated for an evaluation to be conducted to inform the appropriate stakeholders on whether the project had achieved its expected objectives and to highlight any unexpected outcomes. This is what should happen – in principle.

A project or strategy preparation team will find the situation on the ground much more complex. Development is the result of a complex interaction of forces that cannot be easily summarized as a simple flow of causes and effects. Most development goals are achieved as the result of a number of different interacting interventions. Much of the M&E literature places a heavy emphasis on the "evaluation" aspect of M&E. It suggests that the purpose of M&E should be to measure the extent to which the development goal has been achieved and then identify the contribution made by each intervention or project. In practice, just getting an answer to the question – "Are we moving in the right direction?" – is difficult enough. Answering the question – "Are there better ways we could be moving?" – is almost impossible. In the real world, the problem is that, in most cases, the data are just not available to carry out the kind of analysis that in principle seems so logical.

When choosing indicators, the starting point should be the question, "Is this proposed indicator measurable?"

A great deal has been written on the selection of appropriate indicators, and extensive lists have been prepared suggesting suitable indicators for monitoring different types of projects. These are useful reference materials, but in many cases, impractical to apply. Not only are there hundreds of indicators, but also the data that underpin them usually cannot be secured with the necessary precision or regularity. When choosing indicators, the starting point should be the question, "Is this proposed indicator measurable?" This helps considerably in the quest to identify a minimum list that requires the lightest of M&E structures. Even so, the range of possible indicators is still sizeable, which reflects the fact that the M&E systems still have to satisfy the needs of a broad range of users, which are not identical by any means. Annex 1 is there to serve as a checklist – a menu offering a selection of indicators. The actual selection of indicators should be a reflective and participative activity involving the key stakeholders who are most intimately associated with the project design and implementation – not an imposition of demands from outside. This chapter

A systematic approach can help prioritize the selection of the most critical indicators.

outlines a systematic approach that can be adopted to help prioritize the most critical indicators that need to be selected. It provides examples of how the methodology can be applied and used for different ARD subsector programmes.
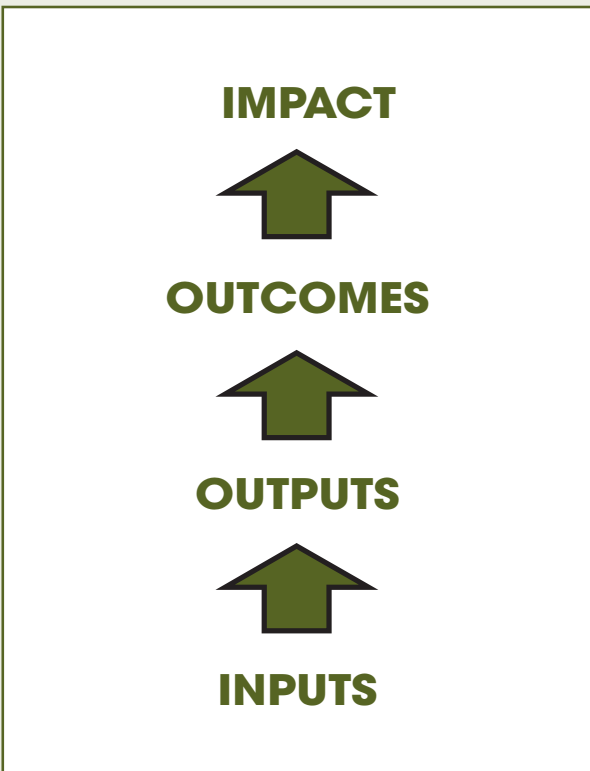
But first a word of caution. The number of indicators and the data required to compute them can grow rapidly. Even though there will always be good reasons for which the list of indicators needs to be expanded, there are also good reasons for starting small and making use of whatever data are available before collecting more. The Sourcebook strongly encourages the idea of integrating statistical

capacity building into national M&E programmes from the beginning, so as to ensure a reliable supply of core statistics from which the required indicators can be extracted.

The focus of this chapter is on indicators, but indicators are only signals. They can be helpful in highlighting whether the project or programme appears to be moving (or to have moved) in a particular direction, but they are, at best, rough instruments that can easily give wrong impressions and lead to misdiagnosis. Indicators alone are not

Indicators are still only rough instruments.

sufficient for serious evaluation. They are merely the first step in a potentially complex and time-consuming analytical exploration. Good M&E also involves blending qualitative and quantitative information that together can enhance understanding of the situation on the ground.

The methodology for selecting indicators is initially introduced in the context of a project-level M&E system, but the process is the same even if one is working on indicators for monitoring a national PRS. The starting point is to establish a framework using the widely used logical framework approach (*logframe*). In very simplified terms, this is a conceptual device that describes the project in terms of its intended goal or *impact*. In order to achieve this impact, people's behaviour is expected to have changed in a way that will help with the achievement of the project goals. These behavioural changes are known as the project *outcomes*, and it may take several years before they become apparent. In order for these outcomes to occur, the project must generate *outputs* (goods and services). These outputs in turn require that the necessary combination of *inputs* (financial, physical and human) become available at the right time, place and quantity. Thus, in reverse order, the inputs will generate outputs, which will yield outcomes and eventually an impact. Take for instance the example of a small-scale irrigation project. Inputs in the form of staff training, equipment, and capital are used to generate outputs in the form of irrigation infrastructure, establishment of extension service, farmer training courses and research on improved crop varieties. The outputs then have to be made accessible to, and used by, the farmers whose changed farming practices in turn will generate outcomes in the

IMPACT

↑

OUTCOMES

↑

OUTPUTS

↑

INPUTS

form of improved yields. Finally, these outcomes should lead to a positive impact in the form of higher revenues and greater food security.

The logframe is well known as a tool for project design and is a useful aid to better understand the logic that defines the development process. It has a second application, however, which is to provide the framework for developing a project M&E system that includes all stages of the project from beginning to completion and beyond. Once the logic of the project had been defined using the logframe, it should then, in principle, be a relatively simple process to monitor progress at each of the four levels. This idea has immense appeal because it helps to reduce the information needs for monitoring the project's success down to a relatively small number of key indicators, which, as already noted, is a desirable feature.

> *Logframe is useful and effective tool but has limitations.*

The logframe does have its limitations, however. First, it promotes a blueprint approach to development. Project design can become a relatively inflexible and uncreative activity. Second, it reduces the process of development to a two-dimensional cause-and-effect formula – clearly a gross simplification. The third is that the project is conceived as an isolated entity and the complex interactions between projects with complimentary or competing goals tend not to be recognized, nor is the relationship between the project goals and the country development goals.

Nevertheless, the logframe can be effective, as evidenced by the fact that it has been widely used for a number of years and has heavily influenced the design of M&E systems. These systems have been most effective at the lower end of the causal chain, in monitoring inputs and outputs. As the project progresses, however, the functions of the M&E system change. This link to the project cycle provides a very useful framework for deciding what information is needed, when and for what purposes.

At this point, it will be useful to introduce two further concepts: **performance** and **results**. These are terms that were introduced after the logframe had popularized the notion of inputs, outputs, outcomes and impacts.

Performance refers to implementation or efficiency, and measures *actual* against *expected* results; it is a proxy measure of the quality of management. In general, it covers all four levels of the logframe causal chain, but focuses mostly on the bottom-end inputs and outputs and on how efficiently the project can convert inputs into outputs. Sometimes, the concept of performance is extended to include outcomes as well.

Results are the outputs, outcomes or impact of a development intervention. Results include the effects the project goods and services have on targeted beneficiaries and others. They may also include the negative effects, such as on the environment. Results are generally, but not necessarily, longer term and more complicated to measure than performance indicators.

Initially, the focus of M&E systems was on monitoring performance (i.e. a concentration on the lower-level input/output indicators), but with the growth of interest in "results-based development", it shifted to a higher level towards the monitoring of outcomes and impacts. A complete project M&E system should include the monitoring of both performance and results.

MONITORING PERFORMANCE (INPUTS AND OUTPUTS)

Tracking inputs and outputs

The monitoring of project performance is M&E at its most basic level. It is the tracking of human, physical and financial resources and the recording of how they are converted into outputs (project goods and services). Strictly speaking, it includes financial monitoring and the analysis of financial records. In addition to generating financial reports, the data are used for cost-benefit analysis and analysis of costs per unit of output, etc. Cost data also lend themselves fairly easily to aggregation and merging with other data sets at higher levels. It is therefore relatively straightforward to integrate performance monitoring indicators into higher level (regional or global) tracking systems. Input and output indicators are generally simple to construct, and most of the information is readily available in project accounts and records. These are usually stored and disseminated through a Management Information System (MIS) that may or may not be connected to the financial management system. Information stored in the MIS includes data on unit costs (costs per hectare or per kilometre, etc.) and can also be useful for analysing the links between inputs and outputs, calculating key input/output ratios and for monitoring projects/programme performance and efficiency. The key to successful operations of the MIS is the ease with which data and monitoring indicators can be accessed and used by project management and others.

Performance monitoring is an essential part of good management.

Regular M&E reports should be generated at least annually and timed so as to serve as an input into the preparation of an Annual Work Plan and Budget. The allocation of budget resources of the following year should, in normal circumstances, be heavily influenced by the results and performance of the project during the current year – as recorded by the M&E system. Performance monitoring is now well established, particularly in projects receiving significant external funding.

A fundamental output of the M&E system at this level should be the production of regular performance monitoring reports.

17

Tools for monitoring inputs and outputs

At its most basic level, performance monitoring (inputs and outputs) is essentially a matter of "keeping the books". Proper and systematically maintained financial records are the starting point. At one time, they used to be maintained by hand, but are now handled electronically using an appropriate commercial financial management package.

Financial and management information systems

For most development projects that receive external financial assistance, it is perfectly satisfactory, indeed recommended, to use an off-the-shelf package as long as it can handle multiple currencies. In the early days, projects were given *carte blanche* to use whatever package they preferred. In an effort to improve the standardization of procedures, a number of countries now specify that public service institutions all use a single, nationally approved package. In addition to bookkeeping, the more general task of reporting on activities and outputs is required. But again, at its simplest level, this involves the establishment of simple reporting procedures and the collation of results into progress reports. As with the accounts, this could be done manually, but is now largely handled on the computer using an MIS. The choice of which system to use is a little more complicated, since it depends more on the nature of the project/programme. In general, the tools needed to operate the basic performance monitoring system at the project level need not be too complicated, and may even become easier as further technical advances are made.

> *Effective monitoring, open reporting and transparency strengthen local government and support the devolution of responsibility to local authorities.*

Integrated local government information systems

When it comes to tracking sector- and subsector-level inputs and outputs, one finds significant variations from one country to the next, but the trend is shifting from a largely uncoordinated and disparate collection of project and sector monitoring systems towards the installation of a single coordinated set of procedures. This process has been assisted by the dramatic improvements in "connectivity" technology. Coupled with improved connectivity is the need to have a well-designed MIS that is adopted universally by all government offices, both at the national and subnational levels.

The United Republic of Tanzania is a country where such a programme is being successfully implemented under its Local Government Reform Programme (LGRP). The aim of the LGRP is to strengthen delivery of public services at the local level by a process of devolving administrative responsibilities to the local government authorities (LGAs) and making them the main conduit through which nearly all government and public services are channelled to rural areas.

Box 4. The national management information system for local government reform of the United Republic of Tanzania

Tanzania's local government reform programme (LGRP) aims to strengthen local authorities and transform them into effective instruments of social and economic development at the local level. It aims to improve quality, access and equitable delivery of public services, particularly to the poor, and thereby contribute to the government's efforts of reducing the proportion of Tanzanians living in poverty.

A critical component of the programme is the adoption of information and communications technology (ICT) and the development of a management information system (MIS) to facilitate the dissemination of reliable, accurate and timely information to a number of stakeholders, both within and beyond the government system. The MIS contains a number of separate systems, two of the most important of which are the Planning and Reporting database (PLANREP) and the Local Government Monitoring Database (LGMD).

PLANREP enables all local authorities to:

- create a performance budget framework of objectives, targets and activities;
- link any target to the national strategy for growth and poverty reduction (MUKUKUTA) cluster strategy;
- calculate projected revenue from formula-based and other grants from central government, own sources, the community and development partners;
- allocate conditional projected revenue to performance budget targets;
- allocate unconditional projected revenue to local authority departments and sections;
- export budget information to the Ministry of Finance;
- enter expenditure from manual or electronic accounting system;
- enter reports on the physical implementation of development targets.

LGMD is a local government monitoring system for capturing and reporting service delivery and socio-economic profile data. These data include information on education, health, agriculture, lands and water. It is also used to capture data from villages, wards and districts. The data are used to calculate 90 indicators. Data from the local authorities are forwarded to both the region and the centre for aggregation. These tools are being introduced to all local government authorities, albeit in a phased approach depending on the issues of local capacity, ongoing support and development of the systems. The software systems, infrastructure and equipment is simple to use and robust, and has been a good support system.

A key element of the LGRP is the development of MISs and the information and communication technology (ICT) infrastructure for the LGAs. Another key feature of the MIS is the development and support of systems that allow LGAs to collect, process and use the data needed for their own purposes and other local government stakeholders (Box 4).

When complete, the LGRP will make it possible for all districts to use the MIS to develop their own plans; prepare their own budgets; review their budget allocations; track expenditures; monitor their outputs in terms of the quantity of goods and services provided; and produce regular quarterly and annual reports – all with the help of the MIS. The country vision is for effective monitoring, open reporting and transparency that will contribute to more effective implementation of national strategic plans and improved governance.

Public Expenditure Tracking Surveys (PETSs) and Quantitative Service Delivery Surveys (QSDSs)

Not all countries are as advanced in the establishment of their M&E infrastructure as the United Republic of Tanzania, however; other solutions must therefore be sought under the less-than-ideal conditions where financial accounting systems are not functioning well. In such cases, countries have been undertaking Public Expenditure Tracking Surveys (PETSs) to track the flow of public funds and determine the extent to which resources actually reach the target groups. PETSs examine the manner, quantity and timing of releases of resources to different levels of government, particularly to the units responsible for the delivery of social services such as health and education. While a PETS traces money through the organization, a Quantitative Service Delivery Survey (QSDS) works to identify organizational weaknesses that can be addressed

> *Possibly, the most basic performance monitoring activity for sector-level programmes is the tracking of public expenditure.*

through reform. QSDSs address the issue of service delivery from the perspective of the supplier. These are surveys based on a random sample of facilities or service providers that focus on quality of service, characteristics of the facilities, their management and incentives structures. One output of the survey instruments is a case-by-case diagnosis of public service delivery, helping to identify weaknesses in implementation capacity and suggesting where reform efforts should be concentrated. PETSs and QSDSs are useful for diagnosing problems in service delivery and for providing evidence on delays, "leakage" and corruption in situations where little financial information is available.

MEASURING RESULTS (OUTCOMES AND IMPACT)

This chapter now shifts from performance monitoring to results measurement, now concentrating on higher-level indicators. It is at this level that the

demand for core indicators is strongest. A results-based system attaches the highest importance to providing feedback on outcomes and goals, rather than on inputs and outputs. In fact, with the advent of results-based management, there has also been a subtle but significant change in terminology whereby the terms "outcomes" and "impact" are frequently replaced by "early results" and "long-term results". The difference is slight,

Measuring results means turning the spotlight on the intended beneficiaries.

although the more recent terms better capture the time dimension. Both are used interchangeably in this Sourcebook. Box 5 presents the chief characteristics of the

Box 5. Characteristics of different classes of indicators

	PERFORMANCE (Efficiency of the project or programme)		RESULTS (Changes resulting from the project or programme)	
LOGFRAME LEVELS	INPUTS	OUTPUTS	OUTCOMES	IMPACT
M&E ACTIVITY	Monitor resources and activities.	Track delivery of goods & services.	Assess early results (access, use and satisfaction with respect to services by users).	Evaluate long-term results.
CHARACTERISTICS OF INDICATOR	These indicators relate to physical, human and financial resources. Sources are MIS and administrative records.	Outputs are generated by the project/ programme. Outputs may include physical outputs, services, training, advice, etc. Sources include the MIS and administrative records.	Indicators should respond quickly and be easy to measure. They should measure the extent to which beneficiaries have changed behaviour due to project. Typical indicators include access, use and satisfaction with respect to project services. Sources include surveys of beneficiaries and service providers and service delivery data from surveys and administrative records.	Indicators may move slowly and be difficult to measure. They must show evidence of change and analysis must establish the extent to which change is attributable to project/programme being evaluated. They are derived from ongoing monitoring activities plus dedicated evaluation studies.
FREQUENCY OF REPORTING	Quarterly to annual.	6-18 months.	1-5 years.	5 years and over.

different classes of indicators and shows how the "results" terms fit with the more traditional logframe terms.

The shift in emphasis from performance monitoring to results monitoring has profound implications for M&E. Unlike performance monitoring, where the data are relatively easily available from internal institutional information systems, results monitoring turns to the targeted beneficiaries (clients) for information on the project and how it has affected them. A key objective of monitoring outcomes (results) is to highlight who is benefiting from the development programme or intervention, and how. At the same time, it is also important to know about the clients who are not benefiting and to understand why. This needs to be done while the programme is being implemented so that corrective action can be taken – simple in principle, but not so easy in practice. To make the task easier, it has now become good practice to separate the monitoring of short-term (or early) indicators from the monitoring of medium- to long-term indicators (which equate more closely to indicators that would be used to measure impacts). For the early indicators, rapid reporting now becomes a critical factor, which in turn affects the choice of indicator. Indicators that change slowly are not good indicators for measuring short-term outcomes, nor are those that are subject to extreme random fluctuations, that exhibit a long time lag or that take time and are expensive to measure. What are needed are indicators that respond quickly and that are easy to collect. Again, they should all be able to be disaggregated and presented for different subgroups of the population (e.g. by gender, vulnerable population groups, or the poor) and also be aggregated upwards and used to calculate indicators at the national, regional or global level.

Early results/outcomes

What, then, are examples of good indicators of short-term results? An examination of recent World Bank ARD Project Appraisal Documents (PADs) showed that project preparation teams have serious problems in identifying suitable indicators. There is a tendency to jump straight from performance monitoring to long-term outcomes. This leaves an important gap in the logical chain, which has sometimes been referred to as the "missing middle". The problem is that there is a time lag between the provision of project outputs and the outcomes on the target population; the result will not be felt in time to take corrective action – often not until several years after the project is complete. Such indicators are therefore of little value for providing quick feedback on early results: they either move too slowly or, due to their complexity or cost, can only be collected every five years or so. In the long run, it is clearly essential to have some objective quantifiable measure of the project impact – for instance, an increase in agricultural and non-agricultural rural

> *Monitoring service delivery is the key to tracking early outcomes.*

income – but some other measure is needed in the short run, as it is impractical to think that such information can be collected and supplied on an annual basis.

So what can be done to fill the gap and catch the early signals of change? What sort of indicator can one use to measure short-term results? How can we know who have benefited from the project or programme and who have not? One solution is to ask the clients directly to evaluate how useful they feel the programme services have been. Consumer satisfaction is, after all, the standard measure used in market research to improve the quality of service delivery. So why not use a service delivery approach for monitoring development activities?

Access, use and satisfaction

A service delivery approach considers that most projects have one thing in common: they are essentially vehicles for making a product or products available to a target population. The concept of the "product" is a broad one, which may include:

- a tangible product such as a loan, a rural road, or a package of technological innovations for increasing yields;
- a service, such as an extension programme, local health care, or land registry service;
- something more abstract, such as "an enabling environment" or a "community development project".

It may even be a combination of the above – a package of products and services that the beneficiary might be expected to adopt. Even policy reform programmes can, with a little adjustment, be viewed through the service delivery lens. For instance, a decentralization policy should result in improved public services to the rural areas. These services are essentially the "product" resulting from the policy.

At its most simple level, a project comprises two elements: a **product** and a **delivery system**. For the project or programme to achieve its desired goal, not only must the product be something that the target population wants and needs, but the delivery system must ensure that they get it. An efficient delivery system may need to be capable of targeting relatively specific subgroups of the population such as women, the poor or the vulnerable. The basic questions that need answering are:

- Do the intended beneficiaries have **access** to this product? (Do they know about it? Is it physically accessible to them? Can they afford it?)
- Do they **use** this product?
- If yes, are they **satisfied** with the product?
- If not, why not?

From these questions, it is then possible to generate three basic indicators:

- *access* – **percentage of the target population having access to the project product.** The term "**access**" has to be clearly defined. It may be "time taken to reach" or "distance" or possibly "ability to pay".
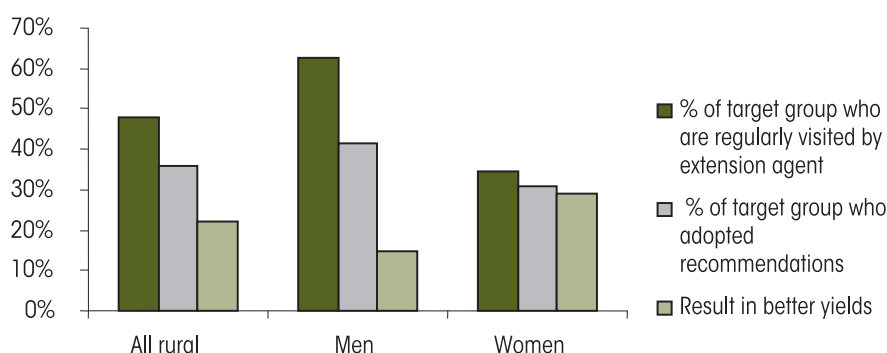
- *use* – **percentage of the target population that uses the project product.** Similarly, the term "**use**" has to be defined. It could for instance be "adoption" as in "percent of smallholders adopting a practice recommended by extension".
- *satisfaction* – **percentage of users satisfied with the product.**

Box 6 shows how these indicators can be applied and adapted to monitor agricultural extension services. Although they are simple indicators, they have a number of qualities that make them attractive as outcome indicators. They are relatively quick to process. This means that the results can be presented very soon after data collection and can consequently be used to sound an alarm in the case of unexpected results.

They can also be collected regularly in order to build up time series, with the first year serving as a baseline. This is important for making before-and-after

Box 6. Adaptation of research and extension service delivery indicators (access, use and satisfaction) to the new Technology Transfer Paradigm

The graph shows how traditional service delivery indicators collected through a household survey of smallholders may be used to monitor the effectiveness of an agricultural extension programme. Access has been defined as "persons having had contact with an extension agent in the last two weeks". Use is defined as "persons who have adopted a set of technological recommendations". Satisfaction is defined as "persons who considered that the recommendations had contributed to higher yields or had otherwise been beneficial". The indicators have additionally been disaggregated by gender.
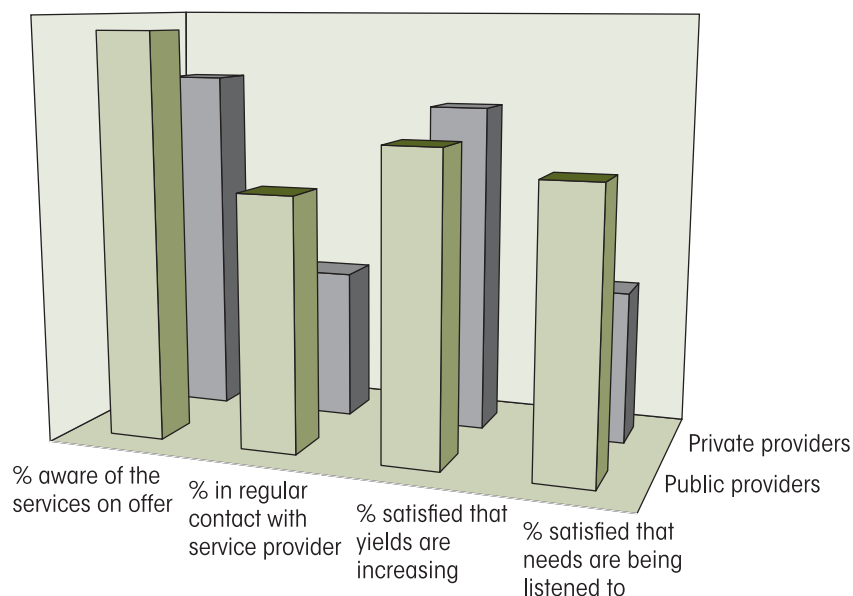

Legend:
- % of target group who are regularly visited by extension agent
- % of target group who adopted recommendations
- Result in better yields

The indicators used in the above example were developed at a time when agricultural extension programmes were based on a view of technology transfer in which farmers are passive recipients at the very end of the innovation process.

This approach is being progressively superseded by the new vision of innovation systems in which farmers, farmers' organizations and communities play a more active part in defining the content of the technology development programme and in which the concept of publicly funded and state-owned extension services is substituted by the approach of pluralistic, public/private, advisory services where farmers choose the service provider and pay for it.

Under such circumstances, the indicators have to be adapted, but the overall service delivery framework can still be maintained. This can be done first by restructuring the questions to the farmers so that a separation is made between the different service providers (public and private) and so that indicators can be separately calculated for each type of provider and second, by recognizing that the "service" is no more just the technological recommendations, but also includes the provision of opportunities for farmers to express their needs. Thus, the satisfaction questions may be expanded to include questions on the extent to which farmers feel that their needs are being listened and catered to.

% aware of the services on offer

% in regular contact with service provider

% satisfied that yields are increasing

% satisfied that needs are being listened to

Private providers

Public providers

comparisons. They can also be disaggregated so that comparisons can be made between the answers given by different subgroups of the population (such as by gender, socio-economic group or regional location). They can equally well be aggregated upwards – as long as care has been taken to ensure that consistent definitions are used – so that responses from different countries can be compared at regional and global levels. Nevertheless, a key question needs to be asked: "How easy are they to collect?" There are basically three options: institution-based surveys; community surveys, or household surveys.

Institution-based surveys aim to collect the information directly from or through the institutions that are delivering the product or service, e.g. a fertilizer distribution centre or a rural bank. Reference has already been made in this chapter to QSDSs.

Focus groups or community surveys work at the community level using a community survey with focus group discussions. Using well-trained enumerators to guide the discussions can be very effective in getting people to talk about the project or programme, and at delving below the surface to understand why a service is or is not meeting the needs of a particular user group.

Household surveys will be reviewed in greater depth below, but it can be pointed out here that these surveys are well suited to the collection of service delivery indicators.

A doubt may be raised about the validity of using "satisfaction" as a measure of success. Can one really trust the respondent to give an honest answer? How can one quantify such a subjective notion? There is no reason why a subjective assessment such as satisfaction is not a valid indicator to include among the early measures of outcomes. In fact, who is better suited to evaluate a product than the user him or herself? Monitoring and evaluation are not exact sciences but involve a process of picking up information from various sources and of combining and comparing them to arrive at the most probable assessment. The respondent's opinion is as valid as any other source of information, and although subjective, it can still be quantified. It is generally recommended that independent agencies – not the service providers – should gather the data from the intended beneficiaries so as to reduce possible bias. It can also be useful to collect information both from the service provider and the service user, and to carry out an analysis of the perception gap.

> *"Satisfaction" is a qualitative concept that can be measured in a quantitative way.*

Thus, by employing the service delivery approach, it is possible to set up a system using just a few basic indicators that can serve as a means both to track results and to signal early warnings where results stray significantly from expectations. The service delivery approach works for a large number of projects, including safe water, health care, immunization, electricity, schooling, employment, credit/financial services, roads, public transport, telephone services,

postal services, agricultural inputs and police services. But it does not work in all cases. For instance, it might be difficult to apply it to a component where the main objective was "institutional reform", or to assess the effects of a policy change. Yet even there, questions such as "How has the economic situation of your household changed over the last 12 months?" can provide very useful early indicators of changing circumstances and overall satisfaction with government performance.

In promoting the use of service delivery indicators, there is no suggestion that other measures of project outcomes should be dropped. Production and yield indicators are clearly necessary, but are problematic and long-term. Further, as shown in the next section, it may take a number of years before lessons can be drawn from them. Annex 1 contains a list of suggested indicators relevant to the ARD sector programmes. Some of these may already be available in the country but not collected on a regular basis; others may require collection mechanisms to be established. It is important that systems be put in place to start capturing them early on so that baseline measures can be taken and time series started. These indicators should be taken as a minimum set to which other indicators can be added.

Sector- and national-level outcomes

Up to this point, the discussion has focused largely on M&E of the project level. When it comes to monitoring at the sector level, the principles remain the same. However, the range of products increases and the interaction between programmes takes on increased significance since ultimately, the M&E findings will affect how resources are allocated to each of them. This could lead to the installation of very heavy M&E programmes and to difficulties in coordination.

> *Sector-level M&E must aim to compare the relative contribution of the different programmes towards the achievement of shared goals.*

Fortunately, as one moves up the results chain, one finds that the various projects/programmes are all contributing to the same common goals – the country development goals. The task of monitoring progress towards these goals is no longer a project-specific activity, but a shared one. This calls for a pooling of information and data, and for the standardization of methodology, concepts and definitions. At these higher levels of the results chain, data come partly from the accumulated body of information disseminated through the individual project M&E reports and partly from additional data that will need to be collected. Working at the top end of the results chain is less a question of monitoring indicators than of systematic analysis. It can be a very data-demanding exercise, especially since such higher-level indicators become increasingly costly to collect and complex to analyse. A weak statistical and analytical infrastructure imposes severe limitations on what can be achieved.
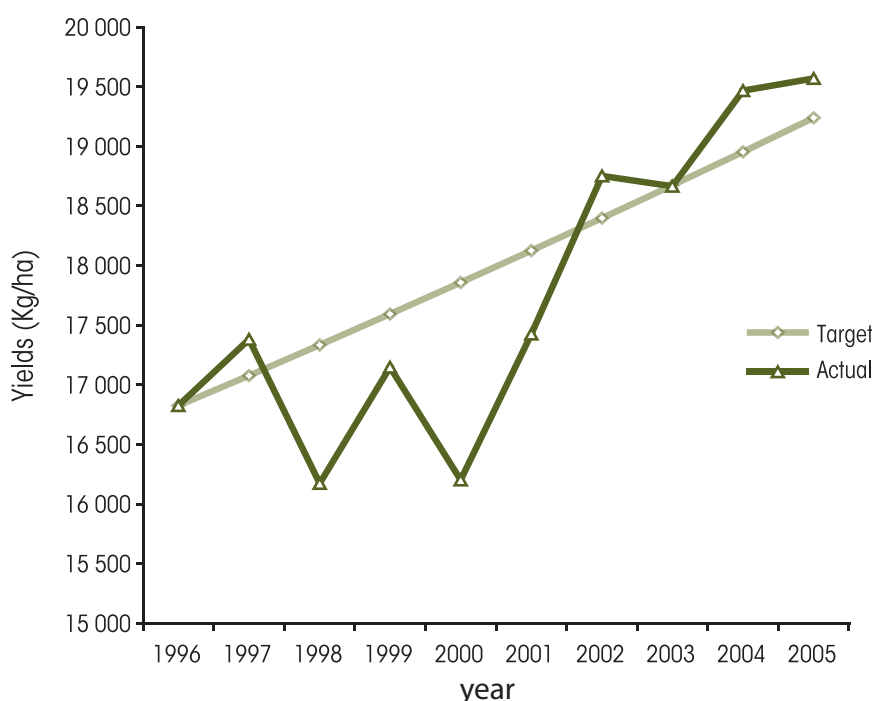
It is not so much that the number of indicators increases, but rather, that complexity increases. Many indicators at this level are quoted as ratios, and separate estimates are needed for both the numerator and the denominator, both of which are potential sources of error and bias. Indicators need to be chosen with care.

Difficulties with the measurement of agricultural output

For monitoring the results of ARD programmes, the most obvious outcome indicators are those that relate to the measurement of changes in production levels (crop, livestock or fish) and yields. While these measures are central to most M&E programmes for the ARD sector, they bring their own particular problems.

Since most agricultural projects share the goal of raising agricultural output, one would think that the simplest indicator would be to measure "yields" – calculated as the ratio of production to area cultivated – and see how they change over time. Unfortunately, it is not that easy, for two reasons. The first reason is essentially a statistical one and centres around the issue of time series analysis. The problem is that agricultural production fluctuates and can vary significantly from one year to the next, primarily but not exclusively due to the

Box 7. Detecting a trend in maize yields

strong effects of rainfall, or the lack of it. This phenomenon is particularly acute in non-irrigated conditions. As a result, it is frequently not possible to detect any change in the trend until a number of years have passed – as many as seven or eight. It is common to see project appraisal documents with projected yield increases similar to those shown in Box 7 (light line). The target is a steady two percent increase in yields per year. This looks reasonable and not too difficult to monitor. But when *actual* yields (dark line) are measured and superimposed over the anticipated trend line, it becomes clear that sharp year-to-year fluctuations in yields make the drawing of any conclusion almost impossible, particularly for the first six years when it would appear that there is no upward trend at all. In this particular case, when the final four years are plotted, the trend line does in fact show an increase of almost exactly two percent a year, as anticipated. But it is statistically impossible to determine this until well past year 6. Random and erratic year-to-year fluctuations of the kind that rainfed crops are prone to experience will severely complicate attempts to carry out time series analysis within too short a period.

But that is not the only difficulty. There is also the problem of measurement errors – errors associated with the measurement of smallholder crop areas and crop production. The classic methodology is to use randomly harvested crop cuts to estimate production and yield. Although this methodology is being successfully applied in many countries, it is known that crop cutting can lead to overestimates of as much as 30 percent in specific situations. Overestimates are due to a number of reasons, including the "boundary effect"; where there is doubt whether a plant is inside or outside the crop frame, it is usually included inside. Overestimates are particularly high in Africa, where traditional plots frequently include multiple crops, irregular planting density and ill-defined, even non-existent, plot boundaries. This makes the application of the crop-cut technique difficult, particularly in less-than-ideal conditions. However, there are other ways of tackling the problem. Methodological experiments to test the viability of alternative ways of measuring production have come up with some interesting and challenging results that suggest that, at least under rainfed conditions, farmers' own estimates may provide substantially cheaper and faster measures of crop production than "objective measures". Indeed, the estimates may even be better.

Farmer estimates may, in some cases, provide cheaper and quicker estimates of production than estimates derived from objective measurements – and with fewer errors.

Methods using GPS for area measurement have the potential of increasing the efficiency of yield estimates in situations where correct estimates of area harvested may not be available. However, in some areas (hilly areas, very small plots, forest areas, etc.) or where plots are irregularly shaped, measurement errors may still be an acceptably high.

These caveats notwithstanding, the measurement of agricultural production will continue to be a central component of any ARD programme, but one should be aware of the potential for error and be on the lookout for alternative ways of assessing results. On the positive side, the introduction of modern farming practices, combined with the arrival of new measurement methods including the use of satellite imagery, are beginning to make the life of the agricultural statistician a little easier. Also, as time series start to lengthen, it becomes easier to identify and discard the obvious outlier years and to reduce the risk of misinterpretation.

The challenge of measuring poverty under less-than-ideal conditions

The ultimate goal of nearly all ARD projects and of the PRS as a whole is to reduce the level of poverty, i.e. to increase rural incomes as a whole and at the same time to reduce income disparities between the rich and the poor. If the measurement of agricultural production was deemed difficult, the measurement of living standards is even more challenging. In order to track the first MDG poverty indicator – "percentage of the population living on less than one dollar a day" – a detailed household survey is required. This may involve multiple visits to households, and the collection and processing of 200 or more items of data from every sample household to compute an estimate of household consumption. Further information has to be provided on all household members, including their age and gender, in order to estimate per capita consumption. More data is then needed on comparative prices before the complex analytical task can begin establishing who is and who is not below the poverty line. In most countries, this is not the kind of indicator that can be realistically measured more frequently than once every five years or so. At the same time, given the close correlation in most countries between household incomes and agricultural production, all the problems associated with the estimation of trend from a time series analysis discussed in the previous section apply equally to the measures of poverty and to the measures of agricultural production. This leads one once again to be on the lookout for alternative measures or methods that could be applied in countries where conditions are less than ideal. Thus, in certain countries, where the goal of regularly monitoring changes in poverty levels may be unrealistic, it may be more productive if instead of focusing on the question "What proportion of the population are below the poverty line?", the analysis focuses on the question, "Are the anti-poverty programmes and services actually reaching the poor and vulnerable as well as the non-poor?" This then becomes an easier question to answer. It focuses attention on the provision of services rather than on the measurement of poverty, but it still requires the classification of households into two classes – the poor and the non-poor. The standard way of doing this would be to establish a national poverty line based on minimum food and non-food requirements, and then establish who is above and who is below this fixed line. This is an absolute measure of poverty, but again, the establishment of such a poverty line can be difficult. An alternative and to some extent simpler solution

is to use a relative concept of poverty. For instance, instead of having a fixed poverty line, one could simply decide to classify, say, the bottom 10 percent as being "the poor". All at once, all the complexities of establishing the poverty line are removed, and the analytical task is simply to compare the services reaching the bottom 10 percent compared with those reaching the rest of the population.

But the problem remains that households must still be ranked using some wealth-correlated variable, such as household income or consumption, which would still require a periodically updated household expenditure and consumption survey. For many countries, this is simply not practicable. However, a number of countries are now experimenting with much lighter household surveys that do not involve the collection of consumption data, but collect specific, easy-to-measure indicators of household well-being. Such indicators may include, *inter alia*, asset ownership, number of literate adults, number of children malnourished, housing quality, mean number of persons per room, and adults unemployed. These are used to create a composite poverty index. Households are then ranked using this composite indicator, and then grouped into deciles. Once this point has been reached, comparisons can be made between deciles. The point is that, even if it is not possible to measure the absolute number of households living in poverty, these short-cut methods allow to identify and isolate those households that are at the bottom end of the distribution, whatever the welfare indicators, and to observe whether they are getting any direct benefit from the various ARD programmes under review.

Evaluation

Finally, one must not forget the "E" in M&E. Monitoring and evaluation are parallel and complementary activities. It is important to be rid of the notion that monitoring is an activity that takes place at the beginning of the project, and evaluation, at the end. Wherever and whenever there is a monitoring activity, there needs to be a regular process of review – of questioning what the data mean and thinking through what the implications are for policy and for the future. Hence, both monitoring and evaluation are continuous activities throughout the life of the project. It is generally thought that evaluation is complex and data-demanding. It need not be so.

> *Without evaluation, there is no learning; without learning, there is no progress.*

There are a range of available types and methods of evaluation – programme reviews, interviews with key stakeholders, focus group meetings, performance audits, etc. – that do not require much in the way of additional data, and that can and indeed should be built into the M&E work programme.

What is true, however, is that as one progresses up the results chain, the tasks of evaluation can become increasingly more challenging, and in consequence, require more data. In the early phases of implementation, evaluation may be

no more than the annual review of inputs and outputs to guide the allocation of further resources during the next year. Further up the chain is where the problems lie.

The first task is simply to take the selected outcome indicator and to establish whether it is possible, over a predetermined period of time, to establish a trend. We have already seen how difficult a task this is, particularly where the expected outcome is an increase in agricultural yields. Just establishing a positive trend may require eight or more annual observations. But if this was difficult, then even more so is the task of determining the extent to which the change can be attributed to specific project interventions. The domain of **impact evaluation** and **social policy and impact analysis** will now be discussed. These are analytical tasks that extend way beyond the analysis of simple indicators. Impact evaluation may be undertaken at any level: project, sector or country. Ideally, it requires information on key indicators before (baseline data), during and after the specific intervention or reform. It may involve the setting up of a quasi-experimental design that controls for sample characteristics and permits testing against counterfactual hypotheses so as to compare both the before/after situation and the with/without situation. The complete evaluation should also identify any unexpected or unanticipated outcomes. A full review of impact analysis techniques is beyond the scope of this Sourcebook, but interested readers are referred to Ravallion (2008a and b) for a more complete description of the main methods for counterfactual analysis.

It is important that, where it is assumed that an impact evaluation will be carried out, the expected path that the analysis will take is mapped out as early as possible so that the data requirements can be assessed and addressed accordingly. The process that has just been described for the selection of outcome indicators is in itself a preparation for an impact analysis down the road. It sets out a specific conceptual framework and identifies channels through which the programme/project services are to be transmitted. It is also important that, when selecting the indicators, thought is given in advance to the need to select indicators in such a way that the impact on gender and on the environment can be extracted and evaluated.

What emerges from this is that if careful thought is given at the very start of the project to the selection of indicators to be monitored, and if they are selected so that they catch the most critical stages of the expected transmission mechanisms, then the additional data demands of the evaluation can be minimized.

The burden of evaluation can be minimized in countries with limited resources.

Several lessons emerge for those operating in less than ideal conditions. Not all projects/programmes need full-scale impact evaluations. These should only be conducted where it is thought that there are lessons to be learned. Second, evaluation does not always mean that much additional data is required beyond what has been routinely collected

for monitoring purposes. Third, the additional data needs can be reduced by thinking ahead at the beginning of the programme. Fourth, given the fact that most projects converge towards a single common goal, there are enormous synergies to be gained by looking at certain aspects of the evaluation of impacts at the sector or country level, rather than at the project level. Fifth, if quantitative data are scarce, good use can be made of qualitative studies that can yield valuable and important insights. Finally, where there is clearly a need of serious evaluation, it needs to be planned well in advance, include both qualitative and quantitative studies, and to take into account both expected and unexpected outcomes. It will almost certainly involve combining data from various different sources, and coming to a considered view about the impact of a particular intervention. The benefits of good evaluation are, however, frequently under-appreciated. Evaluative research also has some of the properties of a public good, in that the benefits spill over to other projects. Development is a learning process, in which future practitioners benefit from current research (Ravallion, 2008a and b). The implications of such a research agenda, with respect to the data needs, are considerable.

A CORE SET OF PRIORITY INDICATORS FOR ARD PROGRAMMES

We now complete the work on identifying and prioritizing suitable indicators by bringing together all the indicators that have been discussed so far, and linking them in with the indicators for monitoring national development objectives as specified in the PRS documents.

We started by noting that there is a difference between monitoring performance and monitoring results. We noted that, for the most part, performance indicators could be monitored using information derived from internal MISs and we looked at some of the tools now available to help improve the monitoring process. Next, we grouped our results indicators into indicators for monitoring early results and indicators for monitoring medium- to long-term results. The early results indicators consisted primarily of service delivery indicators for each of the main ARD products. These service delivery indicators should be supported where possible by quantifiable outcomes, such as yield increases, resulting from target populations adopting or using programme and subprogramme outputs. However, these may need to be tracked several years before any reliable conclusions may be drawn.

In order to establish a minimum set of core indicators, a country must comply to international standards.

There is another set of outcome indicators that is equally important. It covers those that are not directly project-linked – or more correctly, those linked to multiple projects. These include macro- and national-level indicators and indices – the indicators that move as a result of broad policy changes or of the combined effects of several programmes or interventions. They include price indices, food production, agricultural exports, fertilizer use and imports. They also include

some of the more common multi-sectoral indicators that may be used to compare the rural and urban areas, and to measure the results of the combined package of policies and interventions specified in national development strategies. Examples of these include: the proportion of population living in poverty, GDP per capita; urban/rural comparisons of multi-sector indicators such as prevalence of underweight children under five years of age; ratio of girls to boys in primary and secondary education; and the proportion of the population with sustainable access to improved water sources.

The process of selecting a comprehensive set of indicators that meets everyone's requirements is not easy, since different users at different levels have varying information needs. Ideally, the process of selection should be participatory and take into account the needs of all stakeholders, and the principle should be retained that countries select their own indicators according to the content and goals of their PRSPs.

> *The priority indicators need to be underpinned by a database of core ARD statistics.*

The process can be facilitated, however, by drawing on the experience of what other countries have done. Annex 1 provides a menu of indicators that countries can use to help them prioritize and select the most useful indicators for their particular needs. The list is not exhaustive nor is it expected that all countries should adopt and use all of them. Some may not be relevant and others may lack the country capacity to collect them, but the list offers a choice and includes examples of good practices taken from different countries around the world. The indicators include measures of early results as well as medium- to long-term results. They are provided for all the main ARD subsectors and related themes, and countries can choose which ones to use.

For monitoring ARD goals at the international level, however, there has to be standardization. A subset of 19 essential indicators have been identified from among the full list and labelled as priority indicators. Some of these indicators already appear in the FAO statistics database (FAOSTAT), but for many countries, the series are either non-existent or incomplete, with significant gaps or with the values that have been filled by imputation. The international series are in need of urgent upgrading, but the quality of the series can only be improved if all countries commit to maintaining the same indicators at national level, and agree to adhere to common standards. These priority indicators represent a minimum core set that all countries need to maintain and update on a regular basis. Without this minimal commitment at the country level, it is not possible to improve the quality of M&E at the international level. But this should not be too onerous a burden, since the same indicators serve not only to monitor at the international level, but also at a national level. The priority indicators on their own are not enough to meet all M&E data needs, but they should be seen as an essential subset, and as far as possible, they should be included in all national M&E programmes. The priority indicators are shown in Box 8 and the expanded list of indicators are found in Annex 1.

Box 8. List of priority indicators

A Sector-Wide Indicators for Agriculture and Rural Development

Early outcome

P1 Public spending on agriculture as a percentage of GDP from the agriculture sector.

P2 Public spending on agricultural input subsidies as a percentage of total public spending on agriculture.

P3 Prevalence (percentage) of underweight children under five years of age in rural areas.

Medium-term outcome

P4 Food Production Index.

P5 Annual growth (percentage) in agricultural value added.

Long-term outcome

P6 Rural poor as a proportion of the total poor population.

B Specific Indicators for Subsectors of Agriculture and Rural Development

1. Crops (inputs and services related to annual and perennial crop production)

Medium-term outcome

P7 Change (percentage) in yields of major crops of the country.

2. Livestock

Medium-term outcome

P8 Annual growth (percentage) in value added in the livestock sector.

3. Fisheries and aquaculture

Long-term outcome

P9 Capture fish production as a percentage of fish stock (or a rating of the state of major capture fish stocks relevant to exports and local food).

4. Forestry (developing, caring for or cultivating forests; management of timber production)

Long-term outcome

P10 Proportion (percentage) of land area covered by forest.

5. Rural Micro and SME Finance

Early outcome

P11 Percentage of the rural population using financial services of formal banking institutions.

6. Agricultural Research and Extension

Early outcome

P12 Public investment in agricultural research as a percentage of GDP from the agriculture sector.

7. Irrigation and Drainage (services related to water use in agriculture)

Early outcome

P13 Irrigated land as percentage of crop land.

8. Agri-business (agricultural marketing, trade and agro-industry)

Medium-term outcome

P14 Change (percentage) in sales/ turnovers of agro-enterprises.

C Indicators for Thematic Areas related to Agriculture and Rural Development

1. Community-based Rural Development

Early outcome

P15 Percentage of farmers who are members of community/producer organizations.

2. Natural Resource Management

Medium-term outcome

P16 Withdrawal of water for agricultural as a percentage of total freshwater withdrawal.

P17 Proportion (percentage) of land area formally established as protected area.

P18 Change (percentage) in soil loss from watersheds.

3. Land Policy and Administration

Early outcome

P19 Percentage of land area for which there is a legally recognized form of land tenure.

The exercise of validating identified indicators at the country level was aimed at testing the "relevance" of the indicators to the current development activities and the feasibility of their compilation in less-than-ideal conditions.

In recommending the 19 priority indicators, greater attention has been given to the criteria of "comparability" across countries and "availability" of data for their compilation, in addition to "relevance".

Box 9. Cambodia's two-tiered system

The development of the national M&E system in Cambodia is anchored on the country's National Strategic Development Plan (NSDP). The plan is a single, overarching document containing the priority goals and strategies of the Royal Government of Cambodia to accelerate the reduction of poverty and to achieve other Cambodian Millennium Development Goals (CMDGs) and socio-economic development goals for the benefit of all Cambodians.

The M&E system adopts the "two-tiered structure" as its operational framework. It consists of a set of performance indicators, derived from the framework and the priorities of the NSDP, together with effective mechanisms for tracking progress. It aims to ensure regular and periodic M&E of the provision of inputs, achievement of outputs and outcomes of various strategies and actions under the NSDP.

At the national level (first tier), a limited and manageable number of 43 core indicators have been selected. These are aligned with macro-development goals and targets to achieve CMDGs. These are also used to monitor key dimensions of NSDP progress, and provide the basic framework on which annual progress reports are prepared.

The second tier is used at the line ministry/agency level. Each line ministry/agency is required to develop its own set of performance indicators using CMDG indicators (referring to the 43 NSDP-based core indicators) under its jurisdiction, and other indicators relevant for sector-level monitoring purposes. The aim is to create a more in-depth and disaggregated picture of the ministry/agency-level support to detailed policy/programme monitoring and analysis, and reorientation. Guided by the NSDP, the development and selection of indicators at the line ministries/agencies should:

• facilitate informed decision-making and help re-set priorities and policies;
• enhance transparency and accountability through improved information sharing;
• promote a better understanding of the linkages between NSDP implementation and resulting outcomes.

Box 9 describes how a process very similar to the one described here was used in Cambodia in the selection of indicators for monitoring their PRS.

It is not enough, however, to simply develop a list of desirable indicators without at the same time identifying the data that will be needed to calculate them. Thus, linked to the concept of priority indicators is the idea of maintaining a set of core statistics data series needed to underpin the indicators. Once these statistics are added together, the modest list of data requirements starts to grow very quickly, with significant implications for the NSS. This "shopping list" of data needs provides the basis for a dialogue with the suppliers. For most of the outcome indicators, the supplier will be the NSO. It may also include other agencies that make up part of the NSS. The objective of the dialogue is to negotiate arrangements for a programme of survey activities that will ensure the delivery of the appropriate data according to the timeline specified. This is the subject of the next chapter.