



## **Policy Steering**

# **The Role and Use of Performance Measurement Indicators**

Pascal DELORME and Olivier CHATELAIN  
Aid Delivery Methods Programme

February 2011

**This document has been prepared for the European Commission, EuropeAid but its content is the sole responsibility of the authors.**

## Table of contents

<b>PART 1.....</b>	<b>1</b>
<b>PERFORMANCE .....</b>	<b>1</b>
1.1 WHY MEASURE PERFORMANCE? .....	2
<i>Making the most of limited resources.....</i>	2
<i>Improving decision-making processes and how administrations operate .....</i>	2
<i>Achieving democratisation and transparency .....</i>	3
<i>Difficulties and adverse effects.....</i>	3
1.2 MEASURING PERFORMANCE IN COOPERATION POLICIES.....	3
<i>In-country demands.....</i>	4
<i>Obligations .....</i>	4
<i>Indicator matrices (Performance Assessment Frameworks, ‘PAFs’) .....</i>	5
<i>The usefulness of indicators .....</i>	5
<b>PART 2.....</b>	<b>7</b>
<b>MEASUREMENT.....</b>	<b>7</b>
2.1 HOW TO APPROACH THE PROBLEM.....	8
<i>What is an indicator? .....</i>	8
<i>The functions of indicators: classification difficulties.....</i>	8
<i>Means of obtaining indicators: the concept of sources .....</i>	10
<i>Indicator quality: numerous complex criteria.....</i>	10
<i>Choosing indicators: key stages in the process.....</i>	12
<i>‘Pressure points’ .....</i>	13
2.2 STAGE 1A – A SOUND ANALYSIS OF THE POLICIES IN QUESTION .....	14
<i>How is the policy formulated?.....</i>	14
<i>What is the purpose of measuring performance? .....</i>	15
<i>Which aspect of the policy is to be the focus? .....</i>	16
2.3 STAGE 1B – WHAT ARE THE MOST RELEVANT LOGICAL INDICATORS?.....	18
<i>Option 1: Satisfy the greatest number of users.....</i>	18
<i>Option 2: Satisfy the greatest number of uses .....</i>	19
2.4 STAGE 2A – CAN PROGRESS BE MEASURED?.....	21
<i>Is it possible to compare values?.....</i>	21
<i>The effect of the calculation formula on comparability.....</i>	21
<i>The reference value: largely implicit assumptions.....</i>	22
2.5 STAGE 2B – HOW OFTEN IS THE INDICATOR NEEDED? .....	24
<i>Whenever the functioning of the statistical system is altered .....</i>	24
<i>Various solutions.....</i>	24
2.6 STAGE 2C – CAN A TARGET BE SET? .....	25
<i>A three-pronged question .....</i>	25
<i>What is a solid baseline?.....</i>	25
<i>Growth rates and exit values.....</i>	26
2.7 STAGE 3 – TAKING STOCK OF THE APPROACH: DOCUMENTING INDICATORS .....	28
<i>Adopting or rejecting indicators.....</i>	28
<i>Documenting indicators .....</i>	28
<b>PART 3.....</b>	<b>31</b>
<b>THE CONTEXT IN WHICH INDICATORS ARE PRODUCED: STATISTICAL QUALITY.....</b>	<b>31</b>
3.1 MEASURING STATISTICAL QUALITY .....	32
<i>Aspects of the problem.....</i>	32
<i>What is statistical quality? .....</i>	32
<i>Different but ultimately overlapping conceptual frameworks .....</i>	33
<i>The five dimensions of quality .....</i>	33
3.2 SOME DIMENSIONS OF QUALITY IN GREATER DEPTH.....	34
<i>Methodological soundness .....</i>	34
<i>Accuracy and reliability .....</i>	35

<i>Serviceability</i> .....	38
<i>Periodicity and timeliness</i> .....	38
<i>Accessibility</i> .....	39
3.3 IMPROVING STATISTICAL QUALITY .....	41
<i>What is to be supported?</i> .....	41
<i>Do not forget the users</i> .....	41
<i>Improving availability: quick results</i> .....	42
<i>Improving accessibility: exerting pressure through demand</i> .....	42
<i>Accuracy and methodological soundness: a long-term effort</i> .....	42
<i>Precision: no decision without a cost-benefit analysis</i> .....	42
ANNEX 1: GLOSSARY .....	43
ANNEX 2: FURTHER READING.....	45
<i>Some essential documents on performance measurement</i> .....	45
<i>Some essential documents on indicators</i> .....	45
<i>Some essential documents on statistical quality</i> .....	46
ANNEX 3: SOME QUESTIONS TO ASK IN RELATION TO AN INDICATOR .....	49
ANNEX 4: DATA QUALITY FRAMEWORKS USED BY THE WORLD BANK/IMF AND EUROSTAT .....	51
ANNEX 5: BRIEF DESCRIPTION OF THE EUROSTAT GUIDE .....	52

**Table of examples**

Example 1: Adverse effects.....	3
Example 2: Correctly classified indicators.....	14
Example 3: An indicator has no theoretically defined function.....	16
Example 4: Matrix measurement or isolated measurement .....	16
Example 5: There is no such thing as a one-stop indicator!.....	18
Example 6. A sound ‘vertical’ logic approach: the water sector in Mali.....	19
Example 7: Are the values comparable? The collection effect .....	21
Example 8: What is growing? .....	22
Example 9: The importance of the reference value.....	22
Example 10: Comparing dynamics. Progress as percentages or indexes.....	23
Example 11: Choosing the reference year.....	25
Example 12: Target and progress. The Abuja target for health .....	27
Example 13: A good identification sheet .....	29
Example 14: The importance of defining the statistical coverage in data collection.....	34
Example 15: Data sources .....	35
Example 16: Types of error.....	37



**List of abbreviations**

<b>CFAF</b>	French Community of Africa franc
<b>AIDCO</b>	EuropeAid Cooperation Office
<b>CV</b>	Coefficient of Variation
<b>CWIQ</b>	Core Welfare Indicator Questionnaire
<b>DC</b>	Developing Country
<b>EU</b>	European Union
<b>GAVI</b>	Global Alliance for Vaccines and Immunisation
<b>GDP</b>	Gross Domestic Product
<b>GER</b>	Gross Enrolment Ratio
<b>GFATM</b>	Global Fund to Fight AIDS, Tuberculosis and Malaria
<b>GIS</b>	Geographic Information System
<b>HDI</b>	Human Development Index
<b>IMF</b>	International Monetary Fund
<b>MDG</b>	Millennium Development Goal
<b>NIS</b>	National Institute of Statistics
<b>NSS</b>	National Statistical System
<b>PAF</b>	Performance Assessment Framework
<b>PCR</b>	Primary Completion Rate
<b>PEFA</b>	Public Expenditure and Financial Accountability
<b>PFM</b>	Public Finance Management
<b>PRSP</b>	Poverty Reduction Strategy Paper
<b>RC</b>	Reception Class
<b>TPF</b>	Technical and Financial Partners
<b>UIS</b>	UNESCO Institute of Statistics
<b>UNESCO</b>	United Nations Educational, Scientific and Cultural Organisation
<b>WAEMU</b>	West African Economic and Monetary Union]

## **The purpose of this guide**

The aim of any reform is to offer the best possible policy choices to bring about a lasting change to the living conditions of the people concerned.

Outcome indicators linked to reform objectives are now indispensable when it comes to measuring the impact of a given policy, improving decision-making processes or making parliaments, civil society and the general public clearly aware of the impact of the reforms in progress.

However, the great variety of situations and objectives means that there are no 'turnkey' tools or lists of pre-defined indicators which match the specific circumstances and priorities of each partner country.

To help tackle the delicate issue of choosing sectoral indicators, this guide offers a reasoned method, divided into three stages, to aid the reader and everyone involved in the implementation of a performance-based approach and the definition or choice of the associated indicators.

- The first stage involves identifying needs: what information would shed light on or reflect the specific policy aspects and challenges on which attention is to be focussed?
- The second stage makes it possible to determine whether the indicators relevant to the identification of needs have the necessary characteristics to be used to set targets, to monitor progress and to be available as required.
- The third stage involves formally recording the chosen indicators on a fact-sheet designed to provide comprehensive information about the calculation of the indicator, document the sources used, communicate any reservations as to the quality of those sources and give a brief idea of how the indicator can be used and interpreted.

Policy dialogue relating to the achievement of outcomes must be accompanied by a discussion on the data on which the performance measurement approach is based. That issue is addressed in this guide through a series of comments and examples which illustrate the main dimensions of statistical data quality in such a way as to make the key concepts and potential impact of quality defects clear to all readers.

## Summary

Performance-based management means reconciling stakeholders' objectives and activities with the higher-level objectives. In the case of public policies, the different levels include: the government, the ministry of finance, sector-specific ministries, directorates-general, directorates and departments, with each level seeking to accomplish the goals it has identified. To do so, it must determine what has been achieved. There are various ways of determining outcomes; indicators are just one method.

It is only in the last 10 or so years, however, that performance-based management approaches have been implemented in aid recipient countries, mainly as a result of growing national calls following the democratisation of a good many political regimes and, consequently, a greater public desire for transparency. This approach stems from external calls for the introduction of new development policies, which, from Monterrey to Accra, have been increasingly specific about the fact that development partners and beneficiaries are jointly responsible for achieving tangible and measurable results.

It is the case, therefore, that, at present, not enough time has passed to be able to evaluate with clarity the good and bad points of using performance assessment frameworks and whether or not adverse effects have been observed in northern countries as a result of the application of performance-based management principles, while in the southern countries, there is – strictly speaking – no evaluation of such practices.

Choosing one or several indicators to measure the outcomes of a given policy or policies depends, tautological as it may seem, on the sound formulation of those policies and a good grasp of the thinking behind outcome-focussed management. The first step in the proposed approach is therefore to clarify firstly the relationships between indicators and the objective levels to which they are linked, and secondly those between the indicators and what they measure as far the phenomenon studied is concerned.

Whilst the first step is more to do with a good understanding of outcome-focussed management and only brief reminders are provided, the second is directly linked to the main characteristic of an indicator, its function, which can be defined as the nature of the measurement taken by indicator. To use metrology as an analogy, a thermometer measures temperature and a barometer atmospheric pressure. Similarly, a Gross Enrolment Ratio (GER) measures the degree to which a population is covered by a school system.

This concept of function, which is, after all, independent of the policies implemented, is put in perspective when a statistical (or logical) indicator is used within a performance assessment framework. Indeed, it will go from being a mere statistic to being linked to other characteristics that will make it possible to determine the following at least: (i) the programme to which it relates; (ii) the objective level to which it is linked; (iii) the department responsible for achieving the objective(s); (iv) the indicator value prior to the programme and the target (indicator value at the end of the programme); and (v) the desired progress from the start value to the target.

On the basis of the list above, a measurement indicator can be defined as an order of magnitude measurement (of either a statistical or logical type) which is generally defined by its function, the means by which it is obtained, its nature and the limits on its use. Nevertheless, it should be noted that, in many cases, an indicator is also a communication tool, in the sense that it is used to convey the outcomes of policies. To this end, the indicator must be easy to understand and interpret and be generally accepted.

This guide aims to explore, step by step, the various aspects of defining a measurement indicator and, if possible, to bring them together in a logical approach that begins with an analysis of the policies and ends in the drawing-up of terms of reference detailing each indicator chosen for the measurement framework in question.

This approach consists of three stages, in the course of which the person responsible for defining the indicators will have to address a number of issues, some of which will be easy, some of which will be complicated and, in extreme cases, there will not be a clear-cut solution.

The first stage involves: (i) analysing the formulation of the policies in question, including the logical nature of the links between objectives, activities and the expected outcomes, and examining whether or not all or part of a given policy should be monitored or looked into; and (ii) exploring which are the most relevant logical indicators that will make it possible to measure the realisation of activities ('output indicators') and of sectoral objectives ('outcome indicators'). This stage provides a theoretical matrix reflecting only the desire to provide information about a specific policy aspect.

The second stage aims to apply that theoretical matrix to the reality of the information systems available, whether they are statistical systems or administrative monitoring systems. It is a question, therefore, of establishing whether the indicators are measured or measurable with the tools available, by paying particular attention to one of the characteristics mentioned above, namely whether the measurement in question allows the expected progress to be evaluated. Returning to analogy made above, is the thermometer precise and reliable enough to be used? The end result of this stage will be a matrix which can be used in the measurement framework in question and may lead to the rejection of certain indicators which although particularly suited to measuring the phenomenon are lacking certain essential characteristics: indicators which are too complicated (difficult to explain), too costly to obtain or which provide measurements too infrequently.

The final stage is a 'summary' of the previous two, as it involves bringing together all the abovementioned characteristics on one information sheet. This stage is often overlooked but it is essential since the end-users of the indicators are often far removed from the considerations which prevailed at the identification stage. That sheet should therefore cover: (i) the programme(s) which use the indicator to measure performance and the departments responsible for achieving the outcomes; (ii) all the methods used in the calculation (information sources, formulae, units of measurement); (iii) the interpretation which should be given and the limits on use, and (iv) the period over which they are obtained and the department responsible for their provision.

This approach, in particular its latter part, draws heavily on concepts associated with statistical quality. Dialogue on the use of indicators can typically be categorised (or caricatured) as one of two kinds: (i) dialogue relating to whether or not outcomes have been achieved, 'policy dialogue'; and (ii) dialogue concerning the intrinsic quality of the indicator (clearly in the case of statistical indicators).

Dialogue on statistical quality is, in fact, often simplified in the extreme among proponents of the 'right figure' approach, generally producers of information and critics, users and competing producers. In order to move away from these sorts of debates, which can very soon become stale, the final section of the guide attempts to define and illustrate the different dimensions which characterise statistical quality and assess to what extent quality defects affect the use of a piece of data as a performance indicator. While some characteristics, such as precision, are relatively well known, there are other quality criteria that play a far more important role, including: (i) whether the quality of the datum remains constant over time; (ii) whether it is available within known time-frames, and (iii) whether users have been made aware of any methodological changes.

Finally, plans are frequently made to improve the quality of statistical systems, but the associated timescales and costs vary according to the characteristic that is to be altered. While some quality factors may be amenable to relatively quick improvements for marginal costs, in other cases, in particular where an increase in the frequency of collection and precision are targeted, the costs can soon become unacceptable when compared with the benefits obtained.

## **Part 1**

### **Performance**

---

This part provides a reminder of the basics of performance measurement in both developed countries and aid recipient countries. It also provides some examples of possible unwelcome developments.

Key words: efficiency, effectiveness, accountability, performance assessment framework, Millennium Development Goals.

## 1.1 WHY MEASURE PERFORMANCE?

Most of the large industrialised countries have radically reformed their budgetary procedures over the past few years. These reforms have generally sought to establish budgets which are no longer focussed on resources alone but also take account of the objectives to be evaluated using the relevant outcome indicators.

**Performance measurement places emphasis on the outcomes of public policies.** There are three main reasons why such measurement has become necessary for sound public management: (i) to make the most of limited resources; (ii) to improve decision-making processes and how administrations operate; and (iii) to achieve democratisation and transparency.

### Making the most of limited resources

In the public sector, whether within countries or international organisations, resources are limited and often insufficient to meet all demands. Virtually all those countries which have begun a process of national reform have as one of their aims the improvement of public finances.

The 'traditional' method of drawing up a budget, meaning that in use prior to the 1970s, involved allocating resources according to priorities. In line with the thinking of the time, improving the outcomes of a public policy meant increasing the funding allocated to it. This approach was pragmatic in a context in which the collection and dissemination of information was far less straightforward than today, and above all too slow for the outcomes achieved by the public measure to be used to prepare for the next budgetary year. The idea that outcomes improve as resources increase has remained prevalent among the public, who are especially sensitive, for instance, to budget cuts in the areas of health and education.

Nevertheless, public-sector managers' experience and the management control methods used by companies have shown that an increase in funds does not always lead to better results and that, even where it does, the improvement in results is not proportional to the increase in resources.

With limited resources, public-sector managers have to spend better, i.e. secure better outcomes for the same budget. This is the concept of **efficiency**<sup>1</sup>, (one of the three desired characteristics of public spending in the United Kingdom since the National Audit Act of 1983). In the case of aid programmes, efficiency is of direct concern to taxpayers and donors. Efficiency is similar in nature to productivity.

Another desired characteristic is **effectiveness**<sup>2</sup>: have the objectives for which resources were allocated been attained? Measuring the results is the best way of answering this question. Effectiveness concerns citizens as potential beneficiaries of public policies.

### Improving decision-making processes and how administrations operate

Just like management control measures, awareness of the outcomes achieved as a result of administrative measures can facilitate decision-making. Measurement of performance highlights strengths and weaknesses, gives an idea of the progress made over time and helps decision-makers to compare courses of action and identify the most effective mechanisms.

Awareness of the outcomes achieved can also have an impact on the members of the administration themselves. The measurement of their performance may provide an incentive for improvement. Some authors believe that the introduction of a system of performance measurement is a factor independent of modernisation and cultural change within an administration.

---

<sup>1</sup> Efficiency is the ratio of resources consumed (inputs) to results (outputs).

<sup>2</sup> Effectiveness means achieving the outcomes expected of the public policy.

## Achieving democratisation and transparency

There is a growing need for accountability to the public and, in the case of aid programmes, to donors. The ability to access information quickly and at any time raises public expectations (among citizens, public policy beneficiaries and taxpayers). The information available also allows comparisons to be made. The democratic aspirations driving the planet, particularly since the 1980s, lend this demand for **accountability** a new topicality.

## Difficulties and adverse effects

There are difficulties inherent in applying concepts drawn from industry to a public administration context. While businesses are concerned above all with achievements (production or outputs), public services have to focus on results (outcomes). However, 'goal setting in the public sector is more difficult than in the private sector. Output and outcome goals are difficult to define credibly to those who seek to achieve them, to those managing the effort, and to those benefiting from the goal achievement (Perry 2006)<sup>3</sup>.

### EXAMPLE 1: ADVERSE EFFECTS

Placing emphasis on the expected outcomes may lead services to act inadvisably to achieve objectives. One author studied the effects of the application of a performance-based approach in primary schools in Amsterdam. The city authorities had observed that their schools were scoring below the national average in the tests sat by pupils there. They therefore decided to set schools the goal of improving results until they reached the national average, with contractual penalties to be applied if that target was not met. The consequences of the approach varied from changes to lessons to organised cheating.

Results were seen far sooner than expected. To meet the test requirements, the schools altered the content of lessons in order to boost teaching of basic skills (reading, writing and arithmetic), on which the tests focused heavily. This decision reflected the objectives set; the issue of the importance given over to basic skills in primary education is a classic education policy concern, even if not all stakeholders approve of the decision that was taken. However, the success in meeting targets was also the result of less honourable practices.

According to the study, the following adverse effects were observed: (i) schools introduced computer-based learning sessions to prepare pupils for tests, which improved their results, but not necessarily their knowledge (experts were divided on this point); (ii) the weakest pupils were not entered for tests or their results were not recorded; (iii) invigilators used pre-arranged signals to point out difficulties to pupils during the tests. This example demonstrates that the choice of objectives, and the related indicators, is important but insufficient by itself to measure the quality of public policy outcomes.

## 1.2 MEASURING PERFORMANCE IN COOPERATION POLICIES

In addition to the practical reasons set out in the previous chapter, which apply both to industrialised nations (see example 1) and developing countries (DCs), there are other grounds for requiring that performance is measured in aid programmes.

Indicator matrices (Performance Assessment Frameworks, 'PAFs') are one way of responding to this need.

<sup>3</sup> Testimony of Robert M. Tobias, Director Public Sector Executive Education American University before The House Subcommittee on Federal Workforce, Postal Service and the District of Columbia. On Personnel Reform in the Federal Government February 12, 2008: <http://oversight.house.gov/images/stories/documents/20080214120014.pdf>.

## In-country demands

Emerging countries face problems very similar to those experienced by industrialised nations: the growing role of the State (e.g. in infrastructure and health), the investment burden, limited resources and the ever greater demand for transparency. For all these reasons, and because experience of the performance-based approach is now well established, emerging countries can apply the same methods, whether at national, regional or local government level.

In least-developed countries, while the observations and needs are the same, it is often more difficult to incorporate a performance-based approach into public policies. The main reason for this is the quality of the information system (availability, collection, processing and dissemination of both budgetary and social information).

In all situations where countries are recipients of development aid, a performance-based approach is not only desirable but imperative on account of international commitments.

## Obligations

International agreements signed by recipients and donors require the use of results-based management and, accordingly, the measurement of performance in aid programmes.

The **Millennium Development Goals** (MDGs) of 8 September 2000 make all aid subject to a performance-based approach. The MDG require the expected outcomes to be stated, either in the form of general objectives or as indicators to which targets are attached.

- Example of a general objective: to formulate and apply strategies which will give young people all over the world a genuine chance of finding a decent and worthwhile job.
- Example of an objective for which the target is specified by the MDG: reduce by 2015 the maternal mortality rate by three-quarters and the under-five mortality rate by two-thirds.

The **Paris Declaration** (2005) made results-based management a condition of aid delivery: 'Managing for results means managing and implementing aid in a way that focuses on the desired results and uses information to improve decision-making.'

These two texts are not the only ones on this subject (see also the *Second International Round Table on Managing for Development Results*, held in Marrakech on 4 and 5 February 2004). While their effects may vary and they occupy a low place in the hierarchy of rules in States and in the European Union (EU), there is no argument as to their moral weight, which is reasserted in the **Accra Agenda for Action** adopted in September 2008.

The **European Consensus on Development** is a unilateral commitment by the EU (Parliament, Council and Commission):

- *'The European consensus on development sets out common objectives and principles for development cooperation. It reaffirms EU commitment to poverty eradication, ownership, partnership, delivering more and better aid and promoting policy coherence for development. It will guide Community and Member State development cooperation activities in all developing countries, in a spirit of complementarity [...]*
- 95. *MDG-related performance indicators will be strengthened to better link sector and budget support to MDG progress and to ensure adequate funding for health and education. [...]*
- 115. *The Community will consistently use an approach based on results and performance indicators. Increasingly, conditionality is evolving towards the concept of a 'contract' based on negotiated mutual commitments formulated in terms of results.'*

## **Indicator matrices (Performance Assessment Frameworks, 'PAFs')**

Whether by spontaneous decision or in accordance with their international commitments, aid recipient countries should, ideally, be able to set their own objectives, choose indicators that will allow them to measure progress and adopt targets for these indicators. Indicators and the targets to be achieved over time in relation to those indicators are assigned to the objectives. The set of indicators as a whole forms the indicator matrix (PAF).

In fulfilling their own international commitments, donors should conform to recipient countries' indicator matrices. Consequently, the indicators by which donors monitor the progress of the recipient country and, in particular, the subset of conditionality indicators used in the calculation of the conditional tranches of budget support, should be drawn up by the country in question.

In reality, the difficulties faced by countries using indicators and a performance-based approach for the first time mean that the technical advice given by donors plays a large part in the selection of matrices. Furthermore, the asymmetrical nature of the reciprocal commitments and the aid recipient countries' need for resources strengthens the influence donors have, although they should use the matrices (PAFs) drawn up by the aid recipients or, failing that, matrices based on them.

Almost all aid beneficiaries have already, often at donors' encouragement, established a formal or informal platform for sector policy dialogue, meaning a consultation mechanism at national level which enables the authorities to communicate with the country's main stakeholders and interested parties in order to discuss the impacts of the policies implemented in a given sector and work jointly to identify the results to be achieved. In terms of health, for instance, professional organisations and trade unions, private health-care establishments, regional government and local authority bodies, patient representatives of some description and scientific bodies etc. are consulted.

The involvement of non-state and parliamentary stakeholders is often a sign of good governance and generally serves to strengthen the government's responsibility in relation to citizens' needs and demands.

In countries where a process of decentralisation is under way, it is also particularly important to secure local government support and to be able to boost their capacity whenever necessary.

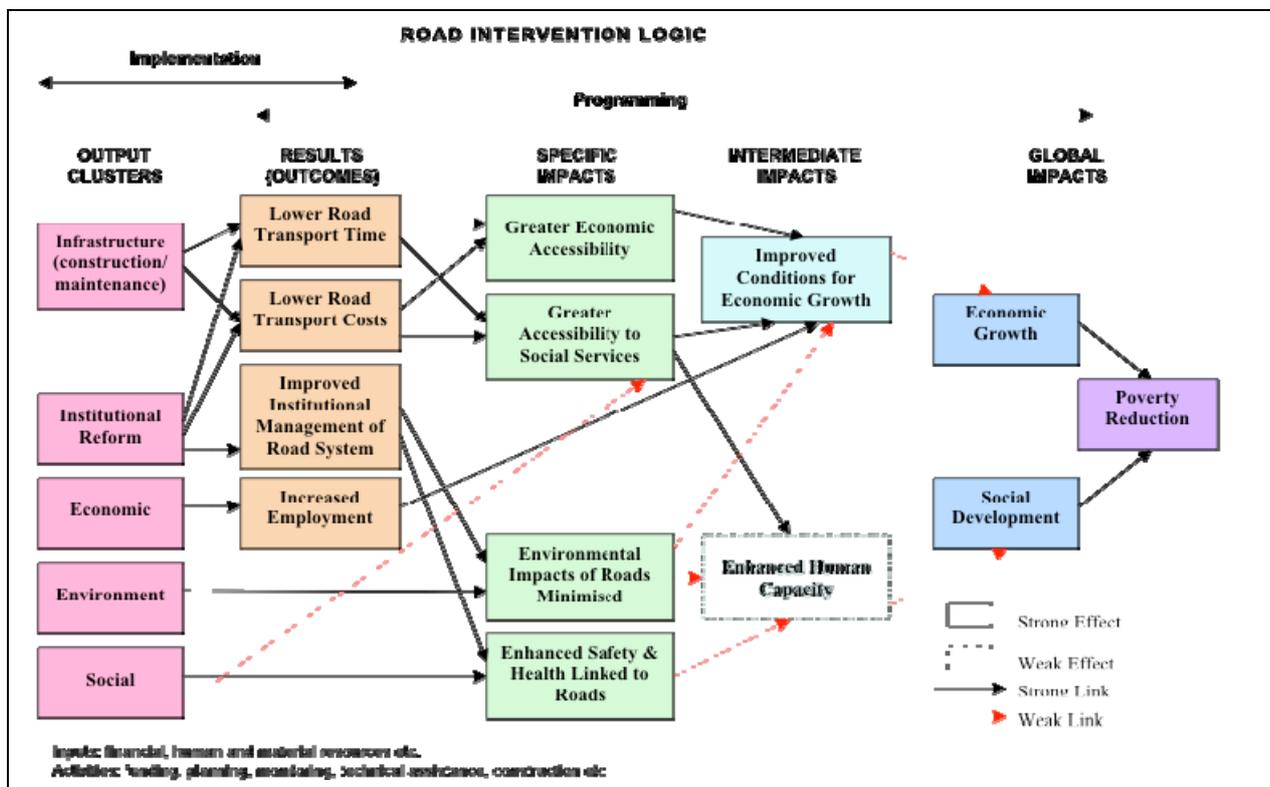
Donors also participate in sector policy dialogue when it comes to setting measurable targets so that sectoral strategies can be funded on the basis of jointly defined results. In this context, in order to reduce the additional costs arising from the need to provide specific supporting documents for each funding application submitted to individual donors, it is proposed that donors should use shared tools to assess the merits of national sectoral strategies. For instance, the World Health Organisation, working with the GFATM (Global Fund to Fight AIDS, Tuberculosis and Malaria) and GAVI (Global Alliance for Vaccines and Immunisation), is currently developing the 'Joint Assessment of National Health Strategies and Plans' tool; the EC is also involved in this work.

## **The usefulness of indicators**

The performance-based approach is not, first and foremost, an economic approach. Public services produce outputs (or 'production' in an industrial context) in order to achieve outcomes, but the aim of their policies is to have an impact on society. While it is difficult to measure impacts using indicators, this is primarily because it is often neither straightforward nor useful to quantify them. Furthermore, impacts are the result of numerous interactions which cannot be attributed to just one sector; the study of impacts is part of the assessment process.

When choosing objectives and indicators, decisions should be guided by the desired impact. There is no fixed method of converting the desired impact into objectives, indicators or targets. The arbitrary element always present in such decisions reflects the policy choices being made. Nevertheless, solid reasoning will enable the links between sectoral policies and desired impacts to become clear.

The following example (from AIDCO) demonstrates the relationship between outputs, outcomes and impacts in the road sector. The flowchart does not claim to provide an exhaustive or exact representation of reality, but rather provides a model intended to serve as a guide to the analysis and drafting of programming documents<sup>4</sup>.



<sup>4</sup> AIDCO has produced flowcharts for five sectors: education, health, water and sanitation, roads and agriculture and rural development:  
[http://www.cc.cec/dgintranet/europeaid/activities/evaluation/impact-indicators/index\\_en.htm](http://www.cc.cec/dgintranet/europeaid/activities/evaluation/impact-indicators/index_en.htm)

## **Part 2**

### **Measurement**

---

This section sets out the key points which, based on the formulation of policies, will allow for the identification of which will form the measurement matrix. It also describes how to overcome problems relating to the poor quality of statistical data in situations where a country has only a very basic statistical system.

Key words: indicator type, function of indicators, statistical significance, frequency of measurement, indicator quality, indicator relevance, reference values, target, progress.

## 2.1 HOW TO APPROACH THE PROBLEM

### What is an indicator?

An indicator is a (generally statistical, but also potentially logical) **order of magnitude** linked naturally or arbitrarily to the measurement of policy activities (in the broadest sense of governance). Indicators are characterised primarily by the fact that they provide information in summary form, are communicable and are subject to relative consensus.

An indicator is generally defined by its **function** (what it measures), **the means of obtaining it** (formula and necessary data), **its quality** (the extent to which it can be interpreted and monitored over time) and the **limits on its use** (what it does not measure or measures poorly).

### The functions of indicators: classification difficulties

The concept of indicators is closely linked to objective-based planning. The logical approach to objective-based planning is the same whether it concerns national policies (Poverty Reduction Strategy Paper (PRSP) or others), sectoral policies, a programme or a project.

The function of an indicator is defined by what it measures in connection with the achievement of an objective. There are therefore various types of indicator, which are summarised in the table below<sup>5</sup>:

<i>Objective level</i>	<i>Indicator type</i>	<i>Definition</i>	<i>Examples</i>
	<b>Input indicator</b>	Provides information about the financial, human, material, organisation and regulatory resources needed to implement the policies in question.	Annual budget deployed; number of people involved in the programme etc.
<i>Operational objective</i>	<b>Output indicator</b>	Refers to what has been achieved, i.e. the products or services generated.	Kilometres of roads built; the number of people who have successfully completed a training course etc.
<i>Specific objective</i>	<b>Outcome indicator</b>	Refers to the direct, short-term effects on beneficiaries or recipients	Time gained by road users; new practices implemented by the trained staff; level of satisfaction among companies that have benefited from a consultancy service etc.
<i>General objective</i>	<b>Impact indicator</b>	Points to the consequences beyond the scope of the (public or programme) intervention itself and its interaction with beneficiaries. This includes any negative or unplanned consequences.	Trainee employment rate after 12 months; survival rate of companies established as a result of the programme; number of jobs lost after the ban on the sale of a product etc.

<sup>5</sup> The definitions given here for the various indicator types are in line with those set out in the 'Glossary of Key Terms in Evaluation and Results-Based Management' edited by the DAC (Development Assistance Committee); however, there are differences between the DAC, the EC and this guide as regards the translation of the terms 'output' and 'outcome': [http://www.oecd.org/document/21/0,3343,en\\_2649\\_34435\\_40381973\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/21/0,3343,en_2649_34435_40381973_1_1_1_1,00.html)

	<b>Context indicator</b>	Applies to the territory or population category in its entirety, with no distinction between those who have been affected by the programme and those who have not.	Number of jobs in the tourism sector; number of Internet connections in region X; unemployment rate in region Y etc.
--	--------------------------	--	--

This classification is based, however, on an approach which links the objective levels to their integration within a two-dimensional matrix. The fact that indicators are often set out in a row in tables explicitly leads to the following horizontal reasoning:

Input ⇒ output ⇒ outcome ⇒ impact

This supposes, of course, that there is a model that logically links all the arguments:

Investment budget ⇒ building of classroom ⇒ increase in GER

⇒ increase in literacy rate

While, in some cases, simplified models like the one above can be used without too many problems, in general terms, the outcome and impact cannot be reduced merely to inputs and outputs, as they are rather the result of a whole range of causal links, which are often poorly or insufficiently defined in policy or strategy documents. In many cases, the models are far more complex (see the causal flowchart on the road sector in Chapter 1).

Many measurement frameworks also contain a set of indicators which do not correspond to the same measurement levels (outputs, outcomes and impacts) and which, consequently, are not measured with the same periodicity. The exercise becomes all the trickier when the indicators used are not of equivalent quality (when they are measured) or refer to a different time period (progress from the original value). Moreover, the selection of an indicator by sector-level strategic axis leads (even where the indicator is particularly effective) to distorted analyses and even to misguided actions when the indicator is used for decision-making purposes (budget allocations or disbursement in the case of aid programmes). It is an easy transition from using figures to benefit a policy to making policies fit the figures<sup>6</sup>. By trying to fit everything within a 'simple' matrix, that matrix becomes a mirror which distorts the complex reality of a situation.

There is also the difficulty, in evaluation terms, of linking impacts to input factors. The joint RELEX/DEV/AIDCO evaluation unit has therefore drawn up some guidelines for producing objectives diagrams and impact diagrams as an aid to clarifying the strategic vision proposed in official documents and establishing the coherence and relevance of the various contributions or activities to the system of objectives and impacts.<sup>7</sup>

This type of approach is necessary as evaluators often have to contend with the following obstacles:

- Strategy or policy documents listing a range of objectives without providing sufficient details to be able to rank them;
- The absence of any reference to an overall objective;
- Low relevance of the causal links between objectives;
- Objectives at a given level which make no reference to any higher-level objective or are unrelated to lower-level objectives.



*It is often unhelpful to attempt, at all costs, to obtain a definitive classification of indicators according to their type with regard to a given framework; it is, however, imperative to define an*

<sup>6</sup> In Europe, too, we are quickly moving from measurement of outcome to measurement of the indicator alone. For example, the reduction in the number of road deaths, through speed restrictions, is now seen only in terms of the indicator used: the number of speeding offences recorded. It is a question of 'meeting quotas'.

<sup>7</sup> See [http://ec.europa.eu/europeaid/evaluation/methodology/tools/too\\_obj\\_en.htm](http://ec.europa.eu/europeaid/evaluation/methodology/tools/too_obj_en.htm)

*indicator's function in respect of the sector to which it relates: an enrolment ratio is an indicator measuring **access** to the education system, and nothing more. Whether it is an output or an outcome is ultimately less important and depends largely on the policies applied to increase such access and, above all, on the programming level to which it relates.*



*In a given matrix, the juxtaposition of various indicator types, especially where they are not classified by function, introduces a number of analytical distortions. It is better to construct a matrix by objective level, as it will be far easier to read and can be easily linked back to the logical approach.*

## **Means of obtaining indicators: the concept of sources**

When it comes to defining an indicator, three elements need to be considered: (i) the statistical variables needed for its calculation; (ii) the degree of statistical significance; and (iii) the frequency of calculation.

Many indicators are now defined in a very precise manner, including by the United Nations, in particular for sectors such as health, education and agriculture. These are, however, theoretical definitions. The difficulty, as we shall see, lies in identifying the source of each variable and, above all, verifying whether the variable provided by this source actually corresponds to that given in the definition.

With regard to the degree of statistical significance, a substantial difference remains between what is possible and what is feasible. It is, of course, possible in many cases to break down data thematically (sex, gender, socio-occupational categories) or geographically (province, region etc.). However, the specific nature of sources means that a breakdown of this sort can sometimes be termed a 'statistical blunder' (see example 16) or intellectual dishonesty.

The frequency of calculation can be defined in a number of ways and may refer to:

- the optimum period of time between two calculations for any difference to be statistically significant;
- the period of time between two collections of data used to form the indicator;
- the frequency with which the policies the indicator is meant to monitor are revised.

The second definition is more or less the default one, except where there is a model allowing forecasts and projections to be made, which is rare in many DCs. All things being equal, moreover, the other two definitions are related. In accordance with the very concept of public service statistics, it is the frequency of revisions (or decisions) which should guide methodology to ensure that the statistical difference is significant.

In practice, frequencies, where they are mentioned, are the result of awkward compromises between these three definitions (politicians want annual figures, but studies are carried out only every 10 years, so, by opting for five years, there is a chance of having some data to inform the indicator). At present, this is one of the main problems during the annual periodic reviews, be they sectoral or general.



*The means of obtaining indicators is generally defined as the source, but in many matrices the latter is described in relatively vague terms: a ministry with responsibility for a particular sector is not a source of a piece of data; the source is in fact the survey or collection carried out by the relevant department within that ministry.*

## **Indicator quality: numerous complex criteria**

Several sets of criteria have been compiled for the qualification of indicators:

- Specific, Measurable, Achievable, Realistic and Time-limited (referred to by the acronym SMART);

- Clear, Relevant, Economic, Adequate and Monitorable (referred to by the acronym CREAM);
- EUROSTAT<sup>8</sup> (logic, relevance, possibility of setting a target, frequency of data collection, appropriateness and possibility of estimating precision).

As with any theoretical definition, the practical application of these considerations is often subject to problems relating to the criteria which can be used to assess quality:

- the criterion of independence of events, for instance, is extremely difficult to analyse, as it supposes, in many cases, that it is possible to isolate the indicator in question from its context (whether physical or economic). That would amount to saying that there are micro- or macro-economic simulation models enabling any variation to be broken down into internal or external factors (which can be assigned a relative weighting). However, in many areas no such models exist;
- the criterion of relevance brings into play two different types of criteria, those relating to the indicator's ability to provide the desired measurement, but also cost-benefit criteria. The most suitable indicator is sometimes unattainable from a cost point of view or the quality gain with regard to the most relevant indicator is minimal in comparison with the use made of it. It is generally for these reasons that proxy indicators are chosen<sup>9</sup>.

In practice, it seems that the main criterion used, including with regard to numerous PRSPs, has ultimately been the feasibility criterion (the data exist or have already been collected) and, to a lesser extent, measurability (there are methods of accessing the data). This undeniably pragmatic approach raises the following question: when monitoring a policy framework, how should the theoretical matter of which indicators best meet the quality criteria listed above be reconciled with practical concerns, i.e. which indicators are actually available when the monitoring exercise is being carried out?

It is also necessary to bear in mind that:

- The selection of largely theoretical indicators does not allow a target to be attached to policies, but it does enable the ultimate aim to be borne in mind.
- The use of only feasible indicators (or a large number of proxy indicators) enables targets to be set, but will not always precisely reflect the intended objective.

---

<sup>8</sup> Statistical Office of the European Commission

<sup>9</sup> A 'proxy' indicator is one which measures an order of magnitude which is similar or closely correlated to that ultimately targeted.

This, in fact, is a large part of the problem posed by the systematic adoption of monitoring indicators, statistical or otherwise: they encourage in-depth reforms to national statistical systems (but also to public finances and even the workings of government bodies), which can only be done at revision level. Discarding the 'right' indicator ultimately means missing out on the opportunity to ensure the transparency of the results. On the other hand, retaining it means settling for 'leaving the box empty' until its production.



*It is important to remember that an indicator should be:*

- *logical, meaning that it can be linked to a policy easily and unambiguously;*
- *relevant, meaning that it corresponds to a real policy-monitoring need;*
- *monitorable, meaning that it is measured regularly using a stable methodology.*

*In addition to these qualities, it should also be possible to attach targets to the indicator, i.e. to make future projections<sup>10</sup>.*

### **Choosing indicators: key stages in the process**

The process set out below attempts to define a number of stages in the identification of indicators. It is presented in sequential form, but can also be viewed as a series of questions to which answers will have to be found. The important thing is to remember that there are three stages:

- one which is purely analytical and sets out to ask questions in relation to the policies implemented (without reference to statistics);
- a second more pragmatic and practical stage aimed at selecting indicators in connection with the means of obtaining them;
- a third involving the standardisation of the presentation of the indicators selected using a system of information sheets.

The first stage results in the identification of the challenges presented by the policy in question which merit special attention and should be monitored; these are the key issues around which dialogue with the authorities should be centred.

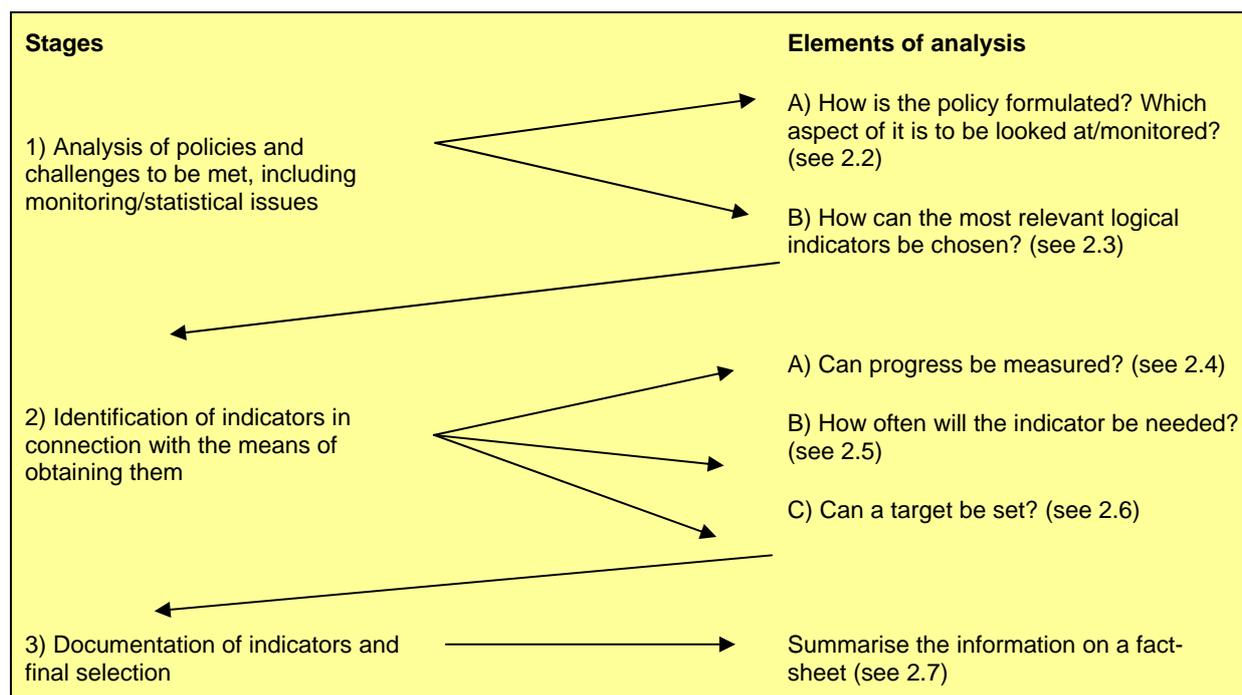
In the course of the second stage, the indicator(s) is/are identified and linked to the challenges defined during the previous stage, to see whether it is possible to compare each indicator to the desired frequency and whether realistic goals can be set. This stage culminates in an initial selection of indicators, which may be instantly available or may require further refinements.

The third stage culminates in the drawing-up of an information sheet summarising the analyses carried out in the previous stages: the question asked, mathematical definition of the indicator, sources, limits on use, instructions as to interpretation and definitive selection of the indicators to be used.

---

<sup>10</sup> In this sense, the Eurostat definition is the most operational.

The approach can be summarised as follows:



### ‘Pressure points’

Clarification is needed with regard to the general approach taken to using statistics in many DCs. Unlike in the northern countries, statistics are not usually available prior to a political decision or, at the very least, there are too many gaps in the data available when the decision is taken for them to be of any real use. The introduction of indicators (MDG, PRSP and other matrices) has proved no exception to this rule and it is only now, with the creation (or restoration) of measurement tools, that it is becoming apparent that some of the objectives set were unrealistic or that their measurement cannot be achieved by the statistical systems.

There is also some confusion as to the number of indicators that should be used, with the idea that, by restricting the overall number, it is possible to keep the cost of obtaining indicators low, which is, to some extent, untrue. The cost essentially depends on the source: a single study may provide all the indicators for a given sector; restricting the revision to just one indicator will have little impact on the final cost. Increasing the frequency of calculation, however, will have a much greater financial impact.

In theory, the number of indicators should be sufficient to shed light on the policies. Restricting this number may, on the one hand, make it far easier to reach a consensus (an ex ante decision on the number of indicators goes some way to preventing the need for the ex post elimination of ‘awkward’ indicators) and, on the other, facilitate the sharing of monitoring information. Nevertheless, the total number of indicators for a given area should be considered in relation to the level of detail of the monitoring exercise and, therefore, the political importance of the interventions in question (see example 3).



*The total number of indicators has sometimes played a particularly unhelpful role in the construction of matrices, enabling awkward indicators ultimately to be eliminated or replacing them with indicators which do not perform the same functions or do not measure what they are supposed to.*

## 2.2 STAGE 1A – A SOUND ANALYSIS OF THE POLICIES IN QUESTION

### How is the policy formulated?

This stage does not directly concern statistics; it is about determining monitoring needs with regard to the policy framework. The general questions posed are as follows:

*What objectives are being pursued and what are the main issues at stake in the policy in question?  
What information is required at the various levels in order to ensure it is effectively monitored?*

The answers given at the end of this stage should not refer to any indicators (in the statistical sense of the term), but to a series of questions which will be asked of the monitoring system (regardless of whether it concerns public service statistics).

The formulation of objectives and outcomes, as they appear in policy documents, should be examined carefully, as they have a bearing on the indicators that will subsequently be used:

- reducing by half the average number of people living in poverty, reducing by half the poorest section of the population and reducing the number of people living on less than USD 1 a day are related but not identical concepts which will not necessarily be represented by the same indicator;
- similarly, ‘increasing the accessibility and quality of primary education’ points far more clearly to those factors on which efforts will need to focus than ‘ensuring better primary education’.

At this stage, the measurability of the issue in question should not come into play. What can be done at this point, however, is to spell out the monitoring need: ‘the provision of indicators every two years on the accessibility and quality of education by education authority’ is, at this stage, clear to all.

### EXAMPLE 2: CORRECTLY CLASSIFIED INDICATORS

The framework for the education sector in Tanzania links education policies and the indicators selected as shown below.

Policy	Name of indicator
Increasing access to and the quality and equity of the education system Standard VII*	Percentage of cohort completing
	Percentage of cohort completing Ordinary-Level Secondary Education (Form IV)*
	Percentage of cohort completing Advanced-Level Secondary Education (Form VI)*
	Transition Rate from Standard VII to Form I Gender-Disaggregated*
	Gross enrolment (Total Enrolment) in Higher Education*
	Qualified Teacher/Pupil Ratio in Primary Schools*
	Qualified Teacher/Pupil Ratio in Secondary Schools*

However, it does not clearly relate indicators to policy objectives. By separating them, we can see what is really being measured.

Policy	Name of indicator
Increasing access to the education system	Gross enrolment (Total Enrolment) in Higher Education*
Increasing quality and coverage	Percentage of cohort completing Standard VII*
	Percentage of cohort completing Ordinary-Level Secondary Education (Form IV)
	Percentage of cohort completing Advanced-Level Secondary Education (Form VI)
	Transition Rate from Standard VII to Form I Gender-Disaggregated*
Increasing effectiveness, quality of service, equity and funding	Qualified Teacher/Pupil Ratio in Primary Schools

## Qualified Teacher/Pupil Ratio in Secondary Schools\*



*Identification of an indicator or a group of indicators depends, above all, on how clearly and precisely policies are formulated. If it is not possible to influence policy formulation, then the monitoring need must at least be expressed in terms of the broad functions of the indicators.*

*It follows that, if the proposed policies do not allow relevant questions to be identified as regards monitoring, it is not necessary to come up with a set of indicators that relate to them: an indicator addresses a given problem and a policy drafted in response to this problem in a particular context and in relation to other 'concurrent' priorities.*



*Complex matrices, i.e. those that introduce a whole series of indicators, from inputs to impacts, do not always prove the most effective in monitoring terms: on the one hand because the analysis carried out on the basis of these different indicators will not always be carried out by the same people and, on the other, because they suppose logical links that sometimes do not exist or differ too greatly from one another.*



*When a 'ready-made' indicator matrix is to be used, it is worth placing the objectives/sectoral challenges to be monitored and the indicator selected for this purpose side by side. For example, in the case of complex multi-sectoral matrices, it may be a good idea to break down and reclassify the indicators (see example 2).*

## What is the purpose of measuring performance?

There are as many sets of indicators as there are measurement matrices, and the selection of the former rests largely on a sound grasp of the indicator matrix will play in steering policies.

Even without establishing a comprehensive overview, it is still possible to identify a few major categories of matrix:

- 'incentive' matrices, which are derived, for example, from international agreements (such as the MDGs or Kyoto agreements). The key characteristic of such matrices is that they in no way draw on the policies that are to be monitored in order to arrive at the expected outcomes, and each country is free to choose the economic levers which will help to achieve those outcomes.
- 'contractual' frameworks, which reflect a commitment arising from a social contract (such as an electoral mandate) or a simple contract (such as matrices linked to aid from technical and financial partners)<sup>11</sup>. These frameworks are characterised by the existence of a form of agreement on the strategies to be implemented in order to achieve the outcome.
- matrices which can be described as analytical or observation matrices, which are technical tools (such as Public Expenditure and Financial Accountability (PEFA) exercises<sup>12</sup>). They are used in particular for policy-steering purposes. The immediate results of such matrices are used to guide decision-making.
- descriptive frameworks, used to provide a summary of a process or a context at a given point in time.

The type of measurement matrix has significant consequences for the indicators used: the same indicator can be used in various forms of matrix, but will not be viewed in the same way: the level of uncertainty acceptable in one case will not be the same in another (see the example below).

<sup>11</sup> The latter, after all, being merely a consequence of the former.

<sup>12</sup> These are indicator frameworks for assessing how public finances are working, via an overlaid scoring system.

**EXAMPLE 3: AN INDICATOR HAS NO THEORETICALLY DEFINED FUNCTION**

The indicators linked to the MDGs are in fact basic indicators, intended to represent a number of dimensions of poverty: in some cases, the choice of indicators is relatively arbitrary and they are more symbolic than anything else.

Consequently, the decision to focus some health outcome indicators on vaccination rates should not obscure the fact that these are output indicators which will fluctuate from one year to the next, depending specifically on the performance of the health system in question.

It would be perfectly legitimate to announce as an objective ‘universal vaccination for all’ within a specified timeframe; here, the vaccination rate becomes an outcome indicator for that hypothetical programme. However, for a national programme to improve health, it is merely an output indicator as, firstly, health policy cannot be reduced to the provision of vaccinations and, secondly, the means of realising this objective are many and varied. This is a trap into which many PRSP-related frameworks have fallen, considerably diminishing their analytical scope.

**Which aspect of the policy is to be the focus?**

A policy cannot be reduced to the observation of a whole series of indicators and less still to that of just a few. When this is done it is usually for purely political ends, for example reducing immigration policy to the single statistic of ‘deportations’ or purchasing power to the ‘consumer price index’.

However, if the intention is to draw attention to specific aspects of a policy, this will have major implications for both the choice of indicators and of the different calculation methods possible. If reducing geographical disparities is an objective, then the national average is an unsuitable indicator: either the number of regions below the target level or the dispersion indexes must be calculated (see example 4 below).

**EXAMPLE 4: MATRIX MEASUREMENT OR SPOT MEASUREMENT***Regional disparities*

In a good many countries there are marked disparities between regions. These disparities are reflected in a very large number of indicators and in numerous sectors. Accordingly, it may be desirable to focus attention on overall performance as well on the reduction of these inequalities.

There are several possible solutions. The one set out below is taken from an MDG contract and relates to the Gross Enrolment Ratio (GER).

*(i) Overall rate (based on national averages)*

Reference value		Targets					
2006	2007	2008	2009	2010	2011	2012	2013
75 %	77 %	80 %	84 %	87 %	88 %	88 %	88 %

*(ii) Regional disparities (the spread between the arithmetic average of the GER of the three least-favoured regions and the national GER average)*

Reference value		Targets					
2006	2007	2008	2009	2010	2011	2012	2013

19.83 points 20.23 points Reduction in spread

*Method of assessing the indicator*

The changes in the overall indicator (i) and the regional disparities (ii) are assessed *separately*. The indicator score is then calculated as the simple average of the two scores.

There is therefore always the option of focusing on the entire field covered or on a particular aspect, or on both at once by weighting the importance attached to two measures (as in the above example).

In many frameworks, it is very rare for use to be made of the possibility of calculating numerous variables derived from the same dataset, and data processing is often restricted to totals, i.e. to averages



*Always remember that, at this stage of the analysis, the indicator, with its single value, cannot reflect the complexity of certain phenomena. It is merely a tool for assessing a situation.*



*The choice of question asked also has consequences for both the number of indicators to be used and the calculation formula which will ultimately be chosen.*

## 2.3 STAGE 1B – WHAT ARE THE MOST RELEVANT LOGICAL INDICATORS?

Relevance is a multifaceted concept. In this context, the relevance of an indicator is tied up with the formulation of the policy requiring support, along with its challenges and priorities. It is then necessary to choose between possible indicators, which will involve different sorts of calculations: an absolute value, a ratio or a target index (national average, deviation from the average). This can sometimes lead to an increase in the number of more or less similar indicators aimed at meeting the same need and requiring additional work on the part of producers. This is why it is necessary to verify, as far as possible, in which analytical frameworks the indicator can or cannot be used.

### Option 1: Satisfy the greatest number of users

It was explained above that relevance is linked to the satisfaction of measurement needs in the various policy frameworks in existence. However, an indicator may be relevant to sectoral monitoring but far less relevant within a general context. In this regard, vaccination rates are a fairly good example. It is easy to see why, for the Health Ministry's policy, such indicators are chosen as output indicators. On the other hand, they are completely 'off topic' in an overall framework intended to measure the performance of a country's health system in the medium term (especially if they have reached the optimum levels of > 95 %).

In a bid to harmonise monitoring mechanisms, there is sometimes a tendency to look systematically for indicators common to the different frameworks and to select those that are most frequently used or cited. This approach soon reveals its weaknesses, notably that of detracting from any consideration of the function of indicators: the relevant indicator becomes 'the indicator measured'.

Thus, many studies on the identification of indicators set the indicators across a row and the policy frameworks in a column, resulting in a sort of 'shopping list' with no internal logic. The indicator(s) selected will no doubt tick the boxes of the multitude of frameworks in use in the country, but are they the product of a prior analysis of the context?

### EXAMPLE 5: THERE IS NO SUCH THING AS A ONE-STOP INDICATOR!

#### *Summary indicators*

Summary indicators such as the Human Development Index (HDI) deserve a mention at this point, as their purpose is to be more or less stand alone or at least to replace a group of indicators with one value. Their primary role is to allow comparisons at international level, although as soon as the position of one country is examined in relation to another it is virtually always necessary to break down the various elements. The HDI employs two structural dimensions – life expectancy and literacy – and one economic dimension – growth expressed as gross domestic product (GDP). This type of indicator is more or less unusable in the context of measuring performance, since the performance measured cannot be assigned to one policy.

## Option 2: Satisfy the greatest number of uses

In order to be able to analyse the performance of a sector or strategy at a given moment, it is necessary to have a set of indicators at your disposal. Having only one (or two) to monitor a particular sector cannot 'disguise' or simplify the reality. Whatever the framework, the relevant indicator is the indicator which can be analysed, meaning that – in addition to that indicator – there are other indicators or information which may be used to interpret its variations.

The aim of satisfying various frameworks should be replaced with a 'vertical' logic approach of satisfying the different policy analysis levels, from local to sectoral to global. The question that must be asked is the following: can the two or three indicators ultimately chosen for the policy be explained by others? Do they allow for explanations of what is happening at different levels? To which level are they particularly relevant?

Accordingly, in the example below, you will see that the rate of access to drinking water is used (with its different levels of disaggregation) in the three analytical frameworks which correspond to the decision-making levels. It is, therefore, a particularly relevant indicator.

### EXAMPLE 6. A SOUND 'VERTICAL' LOGIC APPROACH: THE WATER SECTOR IN MALI

A recent study conducted by the PRSP Monitoring Unit has identified the following sets of indicators for the drinking water sector:

To inform the decision-making level (government and partners), 3 indicators:

- Progress in projects included in the government programme;
- Number of water points created or rehabilitated/objectives;
- Change in the overall rate of access to drinking water.

To feed information into the PRSP in relation to the monitoring of the MDGs (annually), 6 indicators:

- Percentage of villages with at least one modern water point;
- Number of modern water points for every 400 inhabitants in rural areas;
- Equipment rate to meet water needs in rural areas;
- Proportion of the population in urban and semi-urban centres with sustainable access to an improved water source;
- Proportion of the population in rural areas with sustainable access to an improved water source;
- Overall rate of access to drinking water in Mali.

For the purposes of detailed monitoring by the Ministry responsible for water (annually), 19 indicators:

- Specific indicators relating to the water company concession: number of customers, rate of access to tap water, rate of access to drinking fountains etc.;
- Specific indicators for urban and semi-urban centres: number of drinking water access points, rate of access to drinking water, operation rates etc.;
- Specific indicators for rural areas: number of modern water points, rate of access to drinking water, operation rates etc.;
- Global indicators: rate of access to drinking water at national level, total number of modern water points, percentage of villages with at least one modern water point, annual investment in the water sector etc.



*In a given measurement framework, the relevance of an indicator can be measured on the basis of the following characteristics:*

- *whether it can be used in various analytical frameworks (national, sectoral or other matrices) for the purposes of monitoring key aspects of national/sectoral policies;*
- *its capacity for interpretation, i.e. the existence in the area in question of statistical or factual information which will enable, at a given time, explanations or clarification to be given of the values it will show;*
- *its compatibility, either with other existing frameworks or with internationally accepted standards.*



*Retaining one or two indicators for the measurement of global policies (at government or organisation level) must not conceal the fact that that selection was made as a result of a pyramidal process which starts with sector programmes, moves on to sector policies before arriving at overall policies (see the example of the water sector in Mali). By no means are these the only indicators available for the sector in question. If that were the case, the indicator could be described as a 'policy position' indicator which has been included in the matrix not in order to measure performance but to show that the sector exists.*



*It is worth underlining in this respect the misuse of what are called proxy indicators when the desired indicator is not measured. A proxy indicator is one that measures an approximate value. For example, the rate of HIV prevalence measured among sex workers is used as a proxy for prevalence among the population as a whole (a clearly questionable hypothesis). The pupil/teacher ratio is a proxy for the quality of education, with quality measured in terms of learning outcomes (but exam results are not always available, nor are the tests standardised, which makes the selection of this indicator problematic). By contrast, using the livestock growth rate as a measurement of the intensification of livestock does not amount to using a proxy, as it does not measure intensification.*

## 2.4 STAGE 2A – CAN PROGRESS BE MEASURED?

### Is it possible to compare values?

The formulation of objectives and targets has clear political dimensions: it is always possible to make declarations of intent without the experts being able to do anything about these unfeasible political ambitions. However, the setting of targets and intermediate values depends on two purely statistical issues: (i) the length and comparability of the series available for a given indicator; and (ii) the precision of this indicator and, consequently, its capacity to draw a genuine distinction between two consecutive positions.

The question to be asked is as follows: *Are the sources sufficiently precise for the difference between two consecutive values for the same indicator (i.e. between two reviews) to be meaningful?*

Answering that question will make it possible to determine or estimate:

- the general growth or decline trend, which enables the frequency of reviews to be determined, i.e. the frequency of calculation;
- the analytical scope, meaning the interpretation that will be made on the basis of the performance observed.

#### EXAMPLE 7: ARE THE VALUES COMPARABLE? THE COLLECTION EFFECT

Crime statistics are, in fact, proxy indicators, derived from cases tried in courts. Between 2000 and 2006, criminal cases in one country increased by 84 %, rising from 599 to 1 120 new cases a year; however, during the same period, the number of courts almost doubled, from 10 to 19. Should we conclude that crime has risen or that the justice system has improved?

Nevertheless, between 2005 and 2006 specifically, the number of new criminal cases shot up by 8 %, while the number of courts rose by just one.

In the first part of the series, it is impossible to conclude that there has been a rise in crime; it would be necessary to take the 10 courts in existence in 2000 as the basis of calculation and compare them with the courts in operation in 2006. However, it is possible to conclude that crime increased in 2005 and 2006.

### The effect of the calculation formula on comparability

Indicators may be absolute values, ratios or percentages. However, these different types are not interpreted in the same way and the systematic use of percentages to measure progress often obscures more than it reveals (an increase of 100 %, i.e. a doubling, can be used just as easily for absolute values of both 3 and 3 million). For instance, the French *département* which holds the 2008 record for reducing the number of road deaths is Manche, with a 50 % drop, the absolute value having fallen from 10 to 5<sup>13</sup>.

Generally, measurement frameworks use very few statistical tools in order to standardise indicators and make them easier to interpret. Other than deflators<sup>14</sup> applied to a population or sections of a population in the fields of health and education, it is strange that the other indicators used are rarely adapted to population; however, growth (whether in production, infrastructure or equipment rates) is greatly restricted by the population growth rate (between 2 % and 3.5 % in the countries of the Sahel, for example). Progress which is not deflated by population growth often presents a distorted picture of reality (see example 8 below). Similarly, in countries with high inflation, budget indicators are useful only

<sup>13</sup> By way of reminder, percentages should not be calculated for figures lower than 100.

<sup>14</sup> The term 'deflator' is used at the start of the price analysis; 'deflate' means to cancel out inflation in analyses of price changes. By extension, it means making the numerator independent of changes to the denominator.

on a constant currency basis (i.e. in real terms/at real prices), i.e. deflated in the strict sense of the word.

---

**EXAMPLE 8: WHAT IS GROWING?***The effect of population on growth*

The total area of land given over to agriculture has increased markedly, from 2 139 901 ha in 1984 to 4 112 217 ha in 2006. It has more than doubled in the space of 20 years, with farmland increasing by 3 % on average each year. Within the same period, the population has grown by 2.3 % each year, from 7 million to over 13 million. At the same time, the area farmed per member of the working population remained virtually stable, fluctuating between 0.57 ha and 0.73 ha, with an average of 0.67 ha.

Cereal production has gone from 2 450 000 tonnes in 1991 to 3 600 000 tonnes in 2006, an increase of 50 % within 15 years for a population that has risen from 8 to 13 million, translating to a decline in the amount available per inhabitant from 0.3 to 0.27 tonnes.

Similarly, the agricultural value-added per capita rose from CFAF 91 620 in 1993 to CFAF 124 297 in 2005, an increase of 36 % in current francs (in nominal terms). However, if the devaluation of the CFAF in 1994 and inflation are taken into account, the rate can be said to have remained stagnant.

Analysis of increases in terms of absolute values is extremely flattering in terms of results. However, as soon as they are deflated by population in order to determine the profitability and intensification trends, the situation is quite different.

---

**The reference value: largely implicit assumptions**

In practically all frameworks, the reference value taken is the variable value for year 1, which is the implementation variable. The unspoken assumption is one of positive linear growth, which in fact addresses the idea underlying the model linking inputs to outcomes, which can work for classrooms or vaccination campaigns, but far less successfully if the indicator is not solely dependent on policies. Agricultural yields, the annual values for which partly hinge on climate conditions, cannot be compared from one year to the next. If there is a real desire to measure a trend, the average for the last 5, or even 10, years would have to be used as a comparison, with that value being taken as the reference value as opposed to the year 1 value.

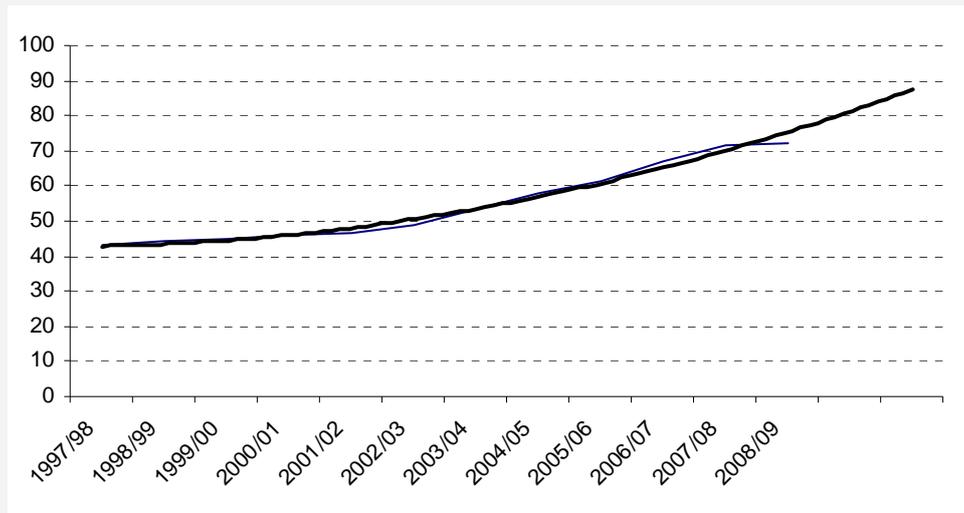
---

**EXAMPLE 9: THE IMPORTANCE OF THE REFERENCE VALUE**

The example below illustrates how the value for a given year can influence a trend (and, consequently, the setting of targets). The upset to the statistical series is purely administrative, since in the first series of enrolment ratios the legal age related to 7 to 12-year-olds, whereas the second concerned 6 to 11-year-olds.

On the basis of the trend curve (shown in bold), the forecasts point to sustained growth. On the basis of the second curve (not in bold), the effect of the administrative measure is to inflect the output curve and cancel out the predictable nature of the trend. As a result, the change made requires that assumptions are based on the reference value.

---



This is a fundamental problem in the matrices used in the context of budget support, since the ratings which will trigger payment of (variable) performance tranches chiefly evaluate growth. Therefore, the choice is either to restrict indicators to those displaying only linear growth (generally the preferred option, which explains the sometimes unusual sparseness of matrices) or to alter the basis to allow for more effective assessment of the structural and non-economic trend (an option which is rarely taken up).

Simple indicators, with a reference value (100) adopted at the beginning of the programming exercise, are used very infrequently (see example 10). However, their use would allow for a better comparison of trends and, most importantly, make it possible to compare indicators that are different in nature (budgetary efforts and outcomes).

**EXAMPLE 10: COMPARING DYNAMICS. PROGRESS AS PERCENTAGES OR INDEXES**

This example is taken from an education scorecard. The first table shows changes in the PCR (primary completion rate) since figures were first recorded.

	1997/98	1998/99	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07
Boys	28.4	27.7	30.0	31.8	32.2	33.6	35.1	36.6	37.6	40.3
Girls	19.4	19.8	21.5	22.4	23.6	25.7	27.3	28.7	30.4	32.4

In terms of the dynamic, it is clear that the indicator has increased in a linear fashion; however, if that dynamic is to be compared, it is much more effective to use an index.

The table below reproduces the same data using the year 2000/01 as the baseline index of 100.

	1997/98	1998/99	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07
Boys	89.1	86.9	94.2	100.0	101.1	105.4	110.4	114.9	118.1	126.6
Girls	86.6	88.4	96.3	100.0	105.6	114.8	122.1	128.4	135.8	144.9

It is immediately evident from the second series that the dynamic is much stronger for girls than for boys: double the number of index points.

## 2.5 STAGE 2B – HOW OFTEN IS THE INDICATOR NEEDED?

### Whenever the functioning of the statistical system is altered

The question to be asked is as follows:

*Are there regularly-financed collection mechanisms within the information systems (sources) which allow indicator values to be obtained?*

- If so, the frequency of calculation can be determined and there are two possible scenarios:
  - ✓ either the frequency of calculation is compatible with the frequency of revision, in which case there is no problem;
  - ✓ or the frequency of calculation is incompatible with the frequency of revision, in which case consideration has to be given to whether, and under what conditions, the frequency of calculation can be increased.
- If not, efforts have to be made to define the frequency of revision, which will remain a theoretical frequency, until a collection mechanism has been set up and the indicator in question can be calculated.

This issue of the frequency of revision and that of calculation ‘contaminates’ much of the policy dialogue surrounding indicators and the ‘completion of frameworks’ ends up becoming an activity that occupies entire units (many units responsible for PRSP monitoring more or less rely on secondary data collection from sector-specific services in order to carry out the annual revision report).

### Various solutions

The frequency issue actually spans several problems, making it difficult sometimes to tackle it head on:

- the problem of the indicator’s function, in particular in the case of some outcome and impact indicators which are usually structural indicators and therefore relatively unresponsive in annual and infra-annual terms;
- the type of matrix (the indicator matrix’s role in steering policy);
- the erratic nature of in which collection systems are financed;
- the unpredictable way in which the data needed to calculate indicators are supplied.



*The frequency with which indicators are calculated must be considered in connection with the nature of matrices, i.e. whether they are analytical or descriptive (see the definitions in chapter 2.2)*



*For a given framework, it is most important to verify:*

- ☛ *the nature and frequency of the associated collection system;*
- ☛ *the exhaustiveness of the associated publications (if all the variables are collected, all the indicators must be available);*
- ☛ *the ‘timeliness’<sup>15</sup> and punctuality of the publications in the light of the revision system.*

*There can be a set of analytical indicators and a set of descriptive indicators for one and the same matrix. The sets of indicators will not necessarily be observed with the same frequency. This in no affects the nature of the indicators or their function (many indicators can be used for various functions – it all depends on the framework).*

<sup>15</sup> The period of time between the collection of data and the date on which the data are made available.

## 2.6 STAGE 2C – CAN A TARGET BE SET?

### A three-pronged question

The question can be phrased as follows:

*Is there a reference value, i.e. a starting value which is sufficiently sound for it not to be called into question once the relevant actions have scarcely got underway?*

This involves: (i) choosing a reference value (or 'baseline'); (ii) selecting a growth rate; and (iii) determining an exit value which is the product of those first two values.

The trickiest issue is comparing the progress proposed with developments in reality. There is no denying that, in this respect, the indicators used in the second generation of PRSPs or budget aid are far better thought-out than in the first, for which the targets were sometimes set in a purely mechanical fashion, often by inverse calculation. This is particularly true of the MDGs, where, using the exit value, only the annual growth rate needed to meet the target was calculated.

### What is a solid baseline?

The baseline selected is often the year in which a measure is introduced. However, in cases other than of linear progress, greater attention should be given to this type of decision. The concept of 'soundness' addresses three concerns:

- the capacity for discrimination, having regard to the phenomena studied;
- knowledge of past progress in order to be able to compare the 'political' objective with the reality of the situation (realism);
- the reliability and regularity of the information system producing the necessary variable(s) (see next chapter).

#### EXAMPLE 11: CHOOSING THE REFERENCE YEAR

In the example below, the pass rate for the certificate of basic education has – since measurement of it began – been relatively stable, except for the year 2005/2006, which shows something of a deviation.

	1999/00	2000/01	2001/02	2002/03	2003/04	2004/05	2005/06	2006/07
<b>Rate</b>	<b>46.8</b>	<b>50.4</b>	<b>45.3</b>	<b>50.4</b>	<b>45.5</b>	<b>47.3</b>	<b>69.4</b>	<b>55</b>

The target the Education Ministry set itself in 2006 was 75 % by 2010.

A mere glance at the curve should have led people to ask two questions: one relating to the reference year 2005/2006 (as there would appear to be a calculation problem connected with this year) and one concerning the rate of progress, since an indicator of this kind is far less sensitive to proactive measures than a gross enrolment ratio. Unless the achievement level is lowered accordingly, this raises the problem of the potentially perverse effects of some indicators.

With regard to the more or less regular nature of the information system, there are two possible scenarios **in which there is – strictly speaking – no real system for collecting information** and making it available:

- either the planning framework lasts long enough to allow adjustments to the monitoring system and it will therefore be possible to make corrections to the mechanisms in place that will, at least, enable intermediate or exit values to be obtained;

- or the planning framework is too short for the variable(s) required to be introduced into the system. In this case, an alternative indicator must be identified.

The concept of the amount of time necessary for the implementation of an information system can be difficult to define; however, as an initial estimate, any period of less than three years can be considered short. This three-year period may apply as much to a National Statistical System (NSS) as to a project or programme. It is derived simply from the fact that many variables are annual, so studying and testing them and setting up a collection system is possible only over a period of several years.



*When selecting an indicator from those potentially available for a particular question, account must be taken of the following quality factors:*

- *the length of the statistical series available;*
- *the (at least theoretical) precision of the indicator and, accordingly, the ability to discriminate between two successive positions;*
- *its standardisation<sup>16</sup> within the analysis of progress, in particular its insensitivity to factors of progress other than the parameter under examination.*

**NB:** It is always possible to improve, through calculation, the comparability of the two values derived from the same indicator. However, this presupposes that it is possible to recalculate the series using a new baseline.

These characteristics of comparability and discrimination largely determine the extent to which it is possible to set a credible target for the indicator (see example 11 above).

## **Growth rates and exit values**

In theory, the growth rate should be defined either by statistical projection (see example 8) or using a modelling system. In most cases, particularly in the DCs, neither of these options exists, except with regard to population (population projections) or the economy (forecasts of general or sector growth).

The growth rate is therefore frequently calculated retrospectively with regard to the target, which is generally set on purely political grounds or as the result of international agreements.

Apart from these two cases which fall under the heading of 'policy dialogue', the techniques which can be used to adopt a growth rate are often reduced merely to empirical examination of the quantity of inputs injected (financial resources) and the effects observed (unfortunately, over short periods).

At this stage, the sensitivity of the indicator, i.e. its ability to respond immediately to financial efforts, is crucial: a vaccination rate reacts immediately to large-scale campaigns, whilst a literacy rate responds only in the context of a long-term effort.

---

<sup>16</sup> Standardisation here means the calculation (application of population and price deflators etc.) or variable transformation (smoothing) processes as a whole which allow for the enhanced interpretation of trends.

**EXAMPLE 12: TARGET AND PROGRESS. THE ABUJA TARGET FOR HEALTH**

The States signatories to this agreement committed in 2001 to ensuring spending on health would reach 15 % of total public expenditure by 2010. The following table provides a picture of the situation in 2008.

Unit	Country (Currency)	Benin (CFAF)	Burkina Faso (CFAF)	Cape Verde (ECV)	Côte d'Ivoire (CFAF)	Gambia (Dalasi)	Ghana (former CEDI)	Guinea (GNF)	Guinea-Bissau (CFAF)
Million	Health budget	41 143	44 442	2 306	98 807	301	1 965	65 685	5 611
Million	Health budget expenditure	475 739	451 768	29 055	1 758 758	2 989	28 745	2 223 866	59 712
-	Health/budget ratio	8.6 %	9.8 %	7.9 %	5.6 %	10.1 %	6.8 %	3.0 %	9.4 %

The growth rates needed to reach the target range from between 50 % and 100 %, or even higher (400 % in the case of Guinea). Aside from the fact that the target will not be achieved by 2010, the question to be asked is by what percentage a national budget can reasonably be expected to grow each year.

This type of indicator also poses the problem of whether it should be expressed as a relative share of the budget or of spending, as it is obviously possible to meet this objective by drastically cutting other budget items. At a push, an indicator of where health spending ranks within the budget would perhaps be more meaningful.

## 2.7 STAGE 3 – TAKING STOCK OF THE APPROACH: DOCUMENTING INDICATORS

### Adopting or rejecting indicators

By this stage of the analysis two documents have been produced:

- the first, resulting from stage 1 of the analysis, reflects the monitoring needs;
- the second, arising from stage 2, is an indicator matrix or set of indicators together with the means of obtaining them.

For a given area, a table can be produced to enable the selection of:

- indicators which satisfy the various criteria listed above. They are usually classed as adopted;
- indicators which require changes to existing information systems, to frequency or to precision which are compatible with the timescale of the policy framework. These are feasible indicators;
- indicators which would require implementation of collection mechanisms incompatible with the policy framework timescale and are usually rejected.

At this point, it is necessary to return to the policy framework (stage 1) because it is at this level that the decision as to whether or not to retain certain indicators will be made.

### Documenting indicators

Documenting an indicator means drawing up an information sheet which will later enable any changes to be analysed properly. The details to be included can be relatively standard.

#### Reminder of the measurement framework

The indicator should be linked to the programmes and detailed information should be provided regarding what it measures. Information about the measurement must be worded in very clear terms. For instance, an enrolment ratio is an indicator which measures access to the education system, whereas a cereals balance-sheet is an indicator of the availability of food.

#### Description of the indicator

All information concerning the calculation of the indicator (formula, measurement unit) should be provided.

#### Indicator quality

The sources used should be documented and any reservations in connection with the quality of the sources expressed. This aspect is of the utmost importance because it will determine, to some extent, the way in which the indicator is interpreted subsequently. It should further be borne in mind that it will be other people – and not necessarily specialist – who will be interpreting the indicators.

#### Interpretation of the indicator

A few lines should be provided describing how the indicator can be used: (i) its direction, positive or negative and (ii) the means of interpretation. For example, under means of interpretation the value beyond which an increase or decrease is meaningful could be stated, or refer may be made to the observation of other statistics for the purpose of gaining a better understanding. See also **Annex 3** which sets out a few simple questions which may be asked in relation to an indicator or its value.

**EXAMPLE 13: A GOOD IDENTIFICATION SHEET**

<b>DOCUMENTATION SHEET</b>		
Indicator	<i>Title</i>	
Programme	<i>Reference to the government programme</i>	
Objective	<i>Reference to the relevant objective within the programme</i>	
Action	<i>Reference to the relevant action within the objective</i>	
Department responsible	<i>Name of the department making use of the indicator within the programme</i>	
<b>Description of the indicator</b>		
Measurement unit		
Periodicity of measurement		
Last known result	<i>Year</i>	<i>Value</i>
<b>Development and quality of the indicator</b>		
Nature of basic data	<i>Numerator, denominator, field covered.</i>	
Method of data collection	<i>Survey, administrative collection....</i>	
Departments and bodies responsible for collection	<i>State the body/bodies responsible for collection</i>	
Method of calculation	<i>State the formula</i>	
<b>Means of interpretation</b>		
Known limits and bias	<i>State the limits of use (significance) or, in the case of a proxy, the bias with regard to the desired indicator</i>	
Means of interpretation	<i>Specify its significance in relation to the area – What, precisely, is the indicator measuring?</i>	
Direction of interpretation	<i>Decrease or increase</i>	
<b>Documentation schedule</b>		
Delivery date	<i>Date on which the document containing the value appeared</i>	
Improvement in progress	<i>State whether the department producing the indicator intends to alter the calculation or collection method.</i>	
<b>Comments</b>		



## **Part 3**

### **The context in which indicators are produced: statistical quality**

---

This section provides a succinct explanation of the approach to be taken to analysing statistical quality and demonstrates, above all, that quality has many dimensions and cannot be reduced to the quality of basic data.

Key words: accuracy, reliability, availability, methodological soundness, serviceability, accessibility, survey information, administrative data, precision, error.

## 3.1 MEASURING STATISTICAL QUALITY

### Aspects of the problem

Dialogue on statistical quality is often biased and extremely narrow in scope:

- ☛ some complain of the ‘deceptive nature’ of the figures;
- ☛ others argue about the official<sup>17</sup> or scientific nature of the figure.

The end result is a ‘dialogue of the deaf’, especially since – as far indicators isolated from their context within a summary table are concerned – the sources are often manifold and, above all, ill defined. Even without developing full analytical frameworks, it is possible to use the five main dimensions of quality (see below) as a form of summary check, even if it means carrying out a more complete and exhaustive examination at a later point.

### What is statistical quality?

Statistical quality is, like all quality-based approaches relating to products or services, a system designed to **qualify a production process** in order to improve it and, most importantly, to ensure the continuity of the products made available to the public.

In general, the first aspect that comes to mind is the quality of the product, which in the case of statistics refers to the quality of a piece of data, which in turn is often reflected solely in terms of precision or, more generally, of the **validity of basic data**. Although this is the main factor, it is misleading in the sense that it is extremely reductive (even valid data can be ‘spoilt’ by inappropriate processing or simply not processed at all) and can contradict other production needs, such as the timely provision of results.

The **accuracy** of a statistic results from the production process, which extends from the design of the collection method through its implementation on the ground to the processing system employed (with all the verifications carried out in the course of these activities). Its **reliability** is the consequence of ‘scrupulous’ adherence to the entire process and its replication over time<sup>18</sup>.

Nevertheless, having accurate figures is not, in itself, the result of the service which the user expects. For a piece of data to become a piece of information, it must be communicated, i.e. made **accessible** to the audience for whom it is intended (as a reminder, the target audience of public service statistics is the citizens who pay for this service through their taxes). This audience varies greatly in terms of their expectations and needs and the ways in which they can access the information. Information must therefore be provided to the public about the **availability** of data, with that information being adapted in so far as possible to their expectations. Citizens (whether ministers or the members of the general public) are not normally statisticians; consequently, it must be explained to them how the figures are used and what they really mean. In short, this involves assisting their ‘reading’ of the figures.

However, to ensure that a piece of data is credible, it is also necessary to establish that it has been produced in accordance with **sound** practices, that it has not been subject to any manipulation (in the negative sense of the term) and that **independence** and **transparency** are the rule. To do this, the production processes must be known, or it must at least be possible to gain access to them, and the product must be delivered with complete **impartiality**, i.e. with the exception of the producing body (which inevitably has access to the information before everyone else) there must be no ‘preferred customers’.

<sup>17</sup> A figure is often considered sound when it has been ‘endorsed’ by a minister.

<sup>18</sup> The concepts of accuracy and reliability should not be confused with that of precision, which refers to the margin of error surrounding a piece of data. It is therefore generally accepted that population figures for Europe are known to a precision of approximately 1 %.

Finally, producers of public-sector statistics do not produce them on their own behalf, nor do they decide what should or should not be produced. Their relationship with their main customer, the State, is based on the **serviceability** of what they produce having regard to the policy guidance needs in the country for which the State is responsible. That relationship takes the form of a constant trade-off between resources and outcomes: in a period of full employment, it is not **relevant** to expend considerable resources on employment statistics. By contrast, when 10, 20 or 30 % of a population are unemployed; it becomes imperative for the State to have accurate and detailed figures. In order to be **serviceable**, a piece of statistical information must be **timely**, so that decisions can be taken when they are needed (and not a few months or years later). This short description summarises the various aspects of the quality of public service statistics:

**Serviceable statistics, produced soundly and with integrity, according to reliable methods and accessible to the largest possible audience.**



*What should be taken from the quality assessment framework is essentially the need to maintain dialogue with producers of data and, consequently, with the main customer, the State. The most important thing is to allow any doubts or questions about the figures provided to be voiced as effectively as possible and to move away from simplistic dialogue about the 'right' or 'wrong' figures.*

*A statistical quality framework should, above all, make use of a common language to identify the strengths and weaknesses, qualities and flaws, and, most importantly, to prioritise problems with a view to finding solutions (if the greatest problem with statistical production is that the data arrive two years late, then this defect must be remedied first of all, even if it is felt for example that the data need to be more precise).*

## Different but ultimately overlapping conceptual frameworks

The need for international comparisons have led institutions such as the World Bank, the International Monetary Fund (IMF) and the European Commission to develop similar approaches, in particular to be able to 'judge' the data provided by Member States according to identical criteria and, potentially, to call on those same States to improve their production processes.

Nevertheless, there are differences between the approaches, for example between the European approach and that taken by the World Bank (inspired by the IMF's approach)<sup>19</sup>.

## The five dimensions of quality

The quality assessment framework is an adaptation by the World Bank of the IMF's statistical quality assessment framework. It therefore incorporates the IMF's **five dimensions** of statistical quality, namely:

- assurances of integrity (objectivity in collection, processing and dissemination);
- methodological soundness (a framework and practices complying with international recommendations);
- accuracy and reliability (sound data and techniques, products which reflect reality);
- serviceability (adequate periodicity, reasonable timelines, consistent data and scheduled revisions);
- accessibility (availability of data and metadata, adequate support for users).

The assessment framework also examines the 'prerequisites of quality' (a favourable legal and institutional environment, adequate resources and an awareness of quality in statistical work).

---

<sup>19</sup> Comparison between the World Bank/IMF and Eurostat quality frameworks (Annex 4).

## 3.2 SOME DIMENSIONS OF QUALITY IN GREATER DEPTH

Detailed consideration will now be given to some of those dimensions of quality, in particular from the perspective of the possible impact quality defects may have on the use of indicators.

### Methodological soundness

This dimension measures adherence to statistical practices, i.e. whether or not the methods employed comply with the science of statistics, on the one hand, and the international environment, on the other, which is defined most frequently by the rules laid down and adopted by international or regional forums (such as adherence to terminology or calculation methods).

Simply put, as far as indicators are concerned, account should be taken of three elements:

- the reality of the statistical field;
- the existence for a given field of an exhaustive list of variables and indicators with their method of calculation (what is commonly known as a 'dictionary of variables');
- the availability of manuals or methodological guidelines.

**The statistical field** determines the scale of the collection as compared with the theoretical definition of the population. If the entire population is not questioned (not counted or surveyed), the field is not fully covered. Many of the indicators used in measurement frameworks are in fact proxies.

#### EXAMPLE 14: THE IMPORTANCE OF DEFINING THE STATISTICAL COVERAGE OF DATA COLLECTION

##### *Registration problems in the administrative collection of data*

It is often mistakenly thought that such problems relating to the scope of collection apply only to sample surveys and not to administrative collection. Nothing could be further from the truth: for example, in order to produce education statistics, an up-to-date list of educational establishments is required; similarly, for health statistics, there must be a list of all medical centres, and so on.

In 2003/2004, in a particular country, the Ministry of Basic Education carried out a dynamic census of private schools. Although private schools must theoretically be registered in order to receive the Ministry's approval, few in fact are registered. At the end of the census, it became apparent that 800 of the country's 6 200 schools had never been registered and that the percentage of private schools in the basic education sector had risen from 13 % to 23 %. In practice, 80 000 pupils had therefore not been factored into the various indicators

**Dictionaries of variables or indicators** should be updated regularly, in particular in order to take account of changes in terminology, concepts and collection methods. In terms of performance measurement indicators, having calculations for all the variables is crucial: even if, ultimately, only two indicators are selected, it is the existence of all the others which will make it possible, when the time comes, to judge how the sector has performed.

There is a plentiful supply of **manuals** (collection, processing and calculation manuals). Failing that, it is essential always to establish whether a collection manual exists, namely instructions for individuals (people responsible for conducting surveys or completing administrative questionnaires) on the correct working methods. Aside from this aspect, which will have a significant impact on the accuracy and reliability of data, manuals are in practice an easy way of helping people to understand the end figures: in order to understand what has been calculated, it is necessary to know what question has been asked. Moreover, these manuals often contain references to the concepts and terminology used.

## Accuracy and reliability

This is the best-known area of statistical quality, since it relates directly to the quality of the data produced and is therefore largely based on the **statistical methods used** and their suitability to the issue studied. However, accuracy and reliability refer to two sometimes contradictory ideas:

- **accuracy** depends entirely on the method used and scrupulous adherence to that method throughout the production chain running from collection to publication;
- **reliability** refers, first, to the maintenance over time of a production standard and, second, to the efforts which may be made to ensure that data are regularly assessed and that users are informed of the relevant findings.

---

### EXAMPLE 15: DATA SOURCES

#### *Administrative data and survey data*

There are essentially two sources of data: 'administrative data' which originates from administrative activity (the registration of educational establishments provides education statistics, police log-books provide information about crime etc.), and 'survey data', obtained when a natural or legal person is asked questions (for example, where a child is asked about his or her level of study or an individual his or her experience of fraud).

However, it should be borne in mind that these two sources of data have very different bases: administrative data are the result of an administrative process. For example, since the Ministry of Education has no authority to supervise private-sector education, the statistics derived from these data will be restricted to the public sector. Survey data usually measure a phenomenon at population level and the intrinsic value of the indicators means that they measure an outcome (in population terms) far better than administrative data.

A piece of survey data can therefore provide a good system for classifying administrative collections, provided, however, that the concepts employed are identical: the same variables and reference population. There are several collections where the value of an indicator can be assessed according to two different sources:

However, we shall see later that the comparison and, particularly, the trend analysis of two surveys require certain conditions to be in place in order to prove meaningful. In a comparison between administrative data and survey data, it is always necessary to ask whether these two sources really measure the same thing (even if the indicator has the same name).

---

Figures have a mathematical dimension which creates the illusion that they are scientific facts. There is a natural tendency to confuse *precise* and *accurate*; a vague concept can never be wholly accurate. If the results of a survey indicate that the population is 1 235 455, this is the accurate figure obtained following calculation of the variables. If this survey has a 10 % margin of error (i.e. a probable error that may account for up to 12 000 people), it is just as possible to state that there are 1 235 000 inhabitants. It is a less precise figure (in the common sense of the word) but just as accurate.

In statistics:

- the **accurate** figure is the one which corresponds to the result obtained if the methodology has been applied strictly, from collection to processing. It may be partly false, **in the common sense of the word**, either because the variable was wrongly defined at the outset or because the question asked did not allow access to the relevant variable or the respondent selected was unable to provide a correct answer;
- the **precise** figure is the one which is known with a margin of error compatible with its use;
- the **reliable** figure is the one which is produced to a consistent quality and known to be precise.

Improving accuracy or precision (aside from the cost it entails) can be to the detriment of reliability and it is a hundred times better to have data suffering from known errors but produced on time and with regularity than less erroneous data produced in an erratic fashion.

The types of error that may be committed during a data production process are too numerous to list in full. Only the main errors and their potential impact on the values taken by indicators are examined here. It is worth remembering a few general rules with regard to precision and errors:

- acceptable error (regardless of the source) must always be viewed in relation to the variable measurement objective: in theory, if you wish to measure developments in intervals of 1 %, the error committed must be below this value! In northern countries, the population is estimated to the nearest 1 %; in the southern countries, such an estimate has never been made, but we can assume that there is sometimes a deviation in excess of 10 %. Therefore, an indicator using population as a denominator only becomes truly meaningful if its values are greater than the 10 % error committed in the population estimate. In the same way, if cotton production is known with a 10 % or 15 % error, this is not necessarily a problem (for macro-economists) as far as the needs of the national economy are concerned, but it becomes one for the cotton mill boss who makes his margin on 10 % to 15 %;
- the error of observation increases with the number of measurements carried out. A measurement error is one committed when a questionnaire is filled in. The more data are collected, the more errors are committed and, as a consequence, the less accurate the final data. This is what we can be called 'census syndrome', which was welcomed by statisticians in socialist economies, but which continues to underlie numerous statistical operations: swelling of the sample and questions asked (up to 1000 variables in some large surveys) to the detriment of the quality of the questionnaires completed.

**EXAMPLE16: TYPES OF ERROR***Coverage error in administrative collections*

Since administrative data collections are considered to be censuses, reference is made to a 'coverage error' where not all the questionnaires addressed to 'establishments' on the register<sup>20</sup> are returned. However, strictly speaking, reference should be made to 'total non-response' rate or completeness rate, with the 'coverage error' being used only for register errors (see above).

If we take the following (very basic) data: in one country there are 6 200 primary schools with a reception class in each establishment and an average of 45 newly enrolled pupils for a reference population of 300 000 (age range). The table below shows how the indicator of the rate of admission to the reception classes changes at different rates of completeness, i.e. at different questionnaire response levels.

	98 % rate	99 % rate	100 % rate
Number of questionnaires sent	6200	6200	6200
Number of questionnaires received	6076	6138	6200
Numbers of newly enrolled pupils	273 420	276 210	279 000
Rate	91.14	92.07	93.00

If the progress measured between two years is less than 2 %, and the completeness varies from 98 % to 100 % (as is generally observed), we cannot really assess what changes have occurred. Inter-annual variation should therefore only be evaluated on the basis of equivalent completeness rates.

**NB:** Caution must always be exercised in the case of small percentages: 1 % is nothing in absolute terms. It all depends on what is being measured. In the example above, it represents 2 790 pupils or 62 school classes. Similarly, 5 % is not a very high percentage, but it would account for 310 classes.

*Sampling error and geographic disaggregation*

The majority of surveys carried out on households in many DCs generally use samples that are intended to be representative at the level of the basic administrative division in that country. The table below shows the coefficients of variation (CV) for some indicators in a Core Welfare Indicator Questionnaire (CWIQ) survey. The CV indicates, for a given probability, the relative variation in the value: so a CV of 10 % means that the value varies within an interval accounting for 10 % of the measurement taken.

	National	Rural	Urban	Region 1	Region 2
Literacy rate	4.2	4.1	2.9	17.4	12.9
Access to primary education	2.7	3.2	1.8	16.7	14.8
Access to secondary education	7.7	13.9	4.3	<b>45.9</b>	<b>42.3</b>
Access to health services	4.9	6.8	3.2	25.8	26.1
Low birth weight	2.4	2.4	7.5	5.2	9.2

This (real-life) example shows that, at national level, comparisons are possible once inter-annual variation is greater than the coefficient of variation, i.e. outside the margin of error:

- In order to be able to conclude that the rate of access to primary education is increasing, only values higher than 2.7 % will be meaningful.
- In order to be able to conclude that the rate of access to secondary education has increased,

<sup>20</sup> The register is the list of individuals identified as respondents – for instance, it could be a list of schools, health centres or courts. The word 'register' is used in administrative collections. The basic notion of a poll, which refers to the same concept, is used only for surveys.

only values above 7.7 % will be meaningful.

As regards the absolute value of estimates, the question is whether it is right to publish a figure known to have a margin of error of 50 %. This is, of course, due to all the problems of disaggregation on the basis of survey data, whether geographically or by subpopulation. The greater the level of disaggregation, the greater the margin of error, making many comparisons misleading: the regional data are not precise enough to measure progress of a few per cent.

## Serviceability

This dimension refers to two aspects which lie at the heart of the issues surrounding the use of performance indicators, namely the **relevance** and **suitability** of the data produced, with regard to the decision-making imperatives (i.e. has the data been produced in a timely fashion and in accordance with a planned timetable).

Relevance refers to the satisfaction of users' needs. It can also apply to:

- a statistical domain: are statistics on governance relevant in a dictatorship?
- coverage of this domain: is it relevant to have statistics on private medicine if it accounts for a low percentage of the treatment administered?
- the frequency of collection: is it relevant to carry out an annual measurement of structural variables (i.e. variables relating to the very basis of society and the economy, and which therefore change very slowly)?

The problem of relevance boils down, in many cases, to questions about the cost-benefit ratio of obtaining a statistic and therefore to user satisfaction. It is at this level that political dialogue between the majority of aid recipient countries and their technical and financial partners is unbalanced. While, in northern countries, statistics play a part in all decision-making processes, it should always be borne in mind that this is not true of many southern countries. The need is therefore the one voiced by partners, who are often the main users of statistics. Beyond the speeches and declarations, and aside from some very specific cases, the main driving force is the 'donor': **statistics are therefore established outside the needs of the society in question, which means that the government will have a limited interest in putting collection tools in place.**

Symptomatic of this is the fact that the availability of statistics in many countries mirrors very closely the cooperation policies pursued, which, after all, is quite normal (the principle of serviceability requiring statistics to be produced for decision-making purposes), but means that the possibilities for analysing countries' situations are far from clear.

The development of performance monitoring, in particular poverty reduction frameworks, has significantly increased statistical demand. However, this demand is growing faster than quality-related information is becoming available (the previous example of governance statistics is not merely innocent or provocative), meaning that it is necessary to consider two possible alternatives:

- quality standards are increasing as analysis tools improve, so the less reliable information should be discarded;
- quality standards are declining to fit in with the analysis tools; the information is certainly complete, but less reliable.

## Periodicity and timeliness

This quality refers to the fact that, in order to be used in the decision-making process, data must be produced in accordance with an established content and timetable. For example, it should be known that the yearbook of education statistics is published in April, or that the health statistics yearbook comes out in May.

This aspect of quality measurement is crucial and is, moreover, the subject of several international commitments (such as the IMF's General Data Dissemination System). In practice, however, there are no well-established data delivery procedures. As a consequence, it is usually a question of supply on demand (particularly with indicators relating to frameworks), the exact status of which can be relatively vague. This goes a long way to explaining the delivery of inconsistent figures or figures which alter depending on the degree to which the results have been processed.

This aspect touches on the fundamental problem of the source and how it is identified when frameworks are constructed: it is not enough to indicate that the Health Ministry is the source. By contrast, identifying the source as the health statistics yearbook should lead all other information to be discarded until this publication has been disseminated in one way or another.

The corollary is that the delivery dates should be consistent with the timetable for data use and therefore with the choice of source support: if, on date  $t$ , the reference document is not ready, there must be a document containing provisional data, which are recognised as such, which can be used as the 'official' source.



*Whether during the development of matrices or subsequent reviews, the delivery timetable for the sources identified, possibly indicating the different stages of publication (provisional or final), should always be included in the documentation. Failing this, it is necessary to:*

- *draw up a list of the publications needed to complete the matrix;*
- *establish the delivery deadlines;*
- *make an analysis of the 'status' of the data (has the information been published or 'forwarded'?) part of the review process;*
- *apply the quality requirement to publication and dissemination (in other words, be much firmer about the fact that a piece of data should be published and disseminated).*

## Accessibility

Accessibility refers to the efforts undertaken by producers to make the figures available to the public as well as to assist them to use the results.

This aspect of quality is often overlooked for probably the same reasons as those stated in regard to relevance. As the figures are not considered to be a public commodity, i.e. something which can be used just as easily by the average person as by another statistician, users are left to their own devices. It is true that, in the DCs, the community of users is relatively small compared with that in the northern countries. However, most significantly, there appears, in the majority of cases, to be a very limited statistical culture (including among many decision-makers). As a result, the supply-demand relationship is very one-sided; it is supply that drives statistics and not user demand.

This supply should therefore be easy to digest and, for the scant time and resources devoted to it, the audience is growing (in Burkina Faso, the regular revision and updating of the National Institute of Statistics website has seen it become the country's most visited government site in the space of a few months).



*With regard to indicators and, consequently, to the publications which support them, the following should always be verified:*

- *the existence of series elements and, at the very least, the corresponding curves;*
- *the inclusion of the definitions and the concepts used, including indicator function, in each publication;*

- *the existence of signs of possible changes to methodology or the collection process;*
- *precise references to the external sources used (population projections are usually included in publications);*
- *information about the department responsible for the publication (so that users can request further information).*

*In more general terms and in order to assess the accessibility of a statistical system or an area, it is possible to:*

- *conduct a quick search on the Internet using a few key words;*
- *conduct a quick search at the appropriate level to determine whether the system has given rise to publications.*

### 3.3 IMPROVING STATISTICAL QUALITY

#### What is to be supported?

It is symptomatic that the result of drawing up numerous performance assessment frameworks is ultimately the need to strengthen the system<sup>21</sup>. Without wishing to make statistics a national preoccupation, some assessment of how the system is performing is required. This ‘evaluation’ usually leads to proposals to improve its quality, but ones which are very rarely subject to prioritisation and are sometimes contradictory, such as ‘greater precision within a shorter time-frame, ideally for the same cost’. It is often impossible to succeed in all respects and, in keeping with developments in Public Finance Management (PFM), production should not be forced when the foundations are not in place: the ‘get the basics right first’ maxim applies. It is not unusual to see unwieldy surveys (costing a few million euro) designed on the basis of surveys over 10-years-old or much older still<sup>22</sup>, or the creation of a GIS (geographic information system) in production units which struggle to maintain a simple database.

The support which may be provided for a statistical system or one of its components can typically be divided into two categories:

- support for **data production** (methodology, processing, analysis, dissemination), in the form of either technical or financial support (for collection operations, for instance);
- support for the **strengthening of statistical capacity**, i.e. for system operating conditions, (training, infrastructure, equipment and master development plans, but also the management of NIS and functioning of national coordination).

It should be borne in mind that financing a survey only rarely serves to increase statistical capacity. It is clear that people may gain in expertise or that the pool of motor vehicles or IT equipment will be boosted, but this is a short-lived effect<sup>23</sup> for many reasons<sup>24</sup>, the main one being the fact that, during an *ad hoc* operation, all efforts are focused on the delivery of the final product: the data. That leaves little time (especially in systems in which a shortage of resources is a constant feature) for real development in the medium term.



*It is, therefore, particularly important when it comes to setting specific targets to remember that:*

- *the financing of production helps generate a greater supply, but not necessarily increased capacity;*
- *financing the strengthening of capacity does not necessarily lead to increased supply in terms of the quantity of data produced;*
- *it is possible, for one and the same project, to finance both types of support, but they are not suited to the same temporal and institutional terms.*

#### Do not forget the users

As stated above, in many contexts, there are statistical ‘cultural’ problems which create a marked asymmetry between supply and demand. This makes it difficult to voice demands, define requirements

<sup>21</sup> There is a fundamental difference here with the northern countries. In the course of successive rounds of EU enlargement, the upgrading of statistical systems (to meet European standards) has been an essential pre-condition for membership. The same strategy is in place with regard to the Balkans and the countries in the European Neighbourhood.

<sup>22</sup> In this case, it is advisable to consider this cost in relation to a General Population and Housing Census (GPHC) and attempt instead to direct aid towards obtaining a new survey framework.

<sup>23</sup> Project evaluations are more or less unanimous on this point.

<sup>24</sup> E.g. the staffs involved in a project are largely contract staff or the material becomes obsolete or out-of-date at the end of the project.

with regard to producers and, more simply, to use the data, which results in those data being considered of little use. Awareness-raising measures would be an option provided that a relatively sizeable critical mass can be reached (there is no point in training two politicians or two trade union members) and, first and foremost, subjects of direct concern to society could be tackled (and not subjects about which there is a desire that the public be better informed, which can be prove to be quite a different matter).

### **Improving availability: quick results**

It is thought (particularly in the DCs, but also in many emerging countries) that the NSSs produce a third of the data necessary for guiding social and economic policy and that one-third of this third is disseminated (i.e. one-ninth overall). There is, therefore, more data available than one might think. It is just a question of publishing them. The problem is often that the reason for not processing data is the insufficiency of human resources or rather the fact that those resources are largely occupied with collection operations. This type of support is often provided, therefore, by outsourcing (technical assistance, sub-contracting to research bodies etc.).

### **Improving accessibility: exerting pressure through demand**

The less developed a country is, the more one-sided the supply-demand relationship is in favour of supply. By making statistics more accessible, efforts are being made to rebalance this asymmetry. This involves providing support for users, which is a long-term activity. However, low-cost and 'quick-win' measures can be applied that will have immediate effects on performance assessment: publication of metadata, improvements to the content of publications and diversification of dissemination channels.

### **Accuracy and methodological soundness: a long-term effort**

These aspects are at the heart of statistical work and, in terms of improvements, require a whole range of tools, from methodology to IT tools and involving the recycling of staff, efforts to ensure the stability of human resources and investment in research and development and equipment, to be adopted and made more reliable. It must be remembered that it takes several years to make data production more reliable.

### **Precision: no decision without a cost-benefit analysis**

Apart from in a few specific cases, increasing precision will inevitably have a significant financial impact and result in recurring costs because it is not enough to increase it for one data collection; it must be increased for all subsequent surveys of a similar nature.



*Eurostat's 'Guide to Statistics in EC Development Co-operation'<sup>25</sup> provides a typology of support for statistical systems, as well as support for the drafting of terms of reference for the identification and formulation of such support (see the brief description in Annex 5).*

---

<sup>25</sup> See a description of the guide on the Eurostat site:

[http://epp.eurostat.ec.europa.eu/portal/page/portal/international\\_statistical\\_cooperation/thematic\\_activities/communication\\_and\\_coordination/guide\\_to\\_statistics](http://epp.eurostat.ec.europa.eu/portal/page/portal/international_statistical_cooperation/thematic_activities/communication_and_coordination/guide_to_statistics)

**ANNEX 1: GLOSSARY**

---

---

**A**

---

**Accessibility:** dimension of statistical quality which measures the conditions subject to which users have access to data.

**Accountability:** the duty to provide explanations, to answer for one's actions.

**Accuracy:** the accuracy of a figure is based on painstaking adherence to the method employed, from collection to processing.

**Administrative data:** the statistical variables resulting from administrative data arise from administrative activities (registry of births, marriages and deaths, customs, court registries), or collection from administrative departments (school censuses).

**Availability:** dimension of statistical quality which measures what is made available to the user and in what form.

---

**B**

---

**Baseline:** value taken by the indicator when it is included in the performance assessment framework.  
*NB: it is not always the value of the present year.*

---

**E**

---

**Effectiveness:** the production of the results expected of public policy.

**Efficiency:** the relationship between the resources used (inputs) and the results (outputs).

**Error:** the uncertainty value surrounding a variable. The weaker the error, the more precise the figure.

---

**F**

---

**Frequency of measurement:** in the case of statistical data, the period of time between two collections.

**Function of an indicator:** definition of what an indicator measures in respect of a given phenomenon (e.g. the enrolment ratio measures the degree to which the education system covers the population).

---

**M**

---

**Methodological soundness:** dimension of statistical quality which measures whether the statistics produced comply with statistical best practices.

**Millennium Development Goals:** the eight objectives which all members of the United Nations have agreed to strive to achieve by 2015 in order to tackle poverty. Each of the goals has specific targets and indicators attached to it.

---

**N**

---

**National Statistical System:** group of public service statistics producers governed by the same legislation.

---

**P**

---

**Paris21 initiative:** joint initiative by a group of development partners, based at the OECD, which aims to promote public service statistics and the use of best practices.

**Performance assessment framework (PAF):** a framework composed of a set of statistical or logical variables, the value of which is known and for which a quantifiable objective is established in relation to a measure.

**Precision:** a precise figure is one known with a margin of error compatible with its use. *NB: a figure may be sufficiently precise for one use and not for another.*

**Progress: estimated** value taken by the indicator year-on-year to achieve the ultimate target.

---

Q

---

**Quality of an indicator:** the indicator's capacity to fulfil its measurement role in terms of time and space. It is a summary concept described according to different sets of criteria which are more or less identical (SMART, CREAM etc.).

---

R

---

**Relevance of an indicator:** indicator relevance is one of the dimensions of quality.

**Reliability:** a reliable figure is one produced to consistent quality and known in terms of precision, availability date etc.

---

S

---

**Serviceability:** dimension of statistical quality which measures how well production complies with demand (e.g. a piece of data produced very late is not very serviceable).

**Statistical capacity:** a set of (tangible and intangible) elements required to produce information continuously and within the time allowed.

**Statistical legislation:** body of law and regulations defining the role and code of ethics for public-sector statistics.

**Statistical significance:** defines the gap between the population surveyed and the reference population. For example, a sample is said to be statistically significant when it reflects the distribution of several characteristics of this population.

**Survey data:** data obtained from a survey of a sample of the population. *Be careful, however, as this term is misused: censuses are surveys, but ones which measure a variable in relation to the entire population.*

---

T

---

**Target:** the final value the indicator must reach at the end of a given time-frame (generally that of a corresponding programme). *NB: the target can be a numerical value or (upward or downward) valuation.*

**Type of indicator:** classification made according to an indicator's position in the logical chain of a process (e.g. input indicators, output indicators, impact indicators).

## ANNEX 2: FURTHER READING

### Some essential documents on performance measurement

#### ***Mesure de la performance et incitations dans la gestion publique***

Dominique Bureau, Michel Mougeot

© La Documentation française, Paris, 2007-ISBN: 978-2-11-006636-7

<http://www.cae.gouv.fr/IMG/pdf/066.pdf> - extract:

*If you have time:*

BRAUN, G., ***Étude comparative portant sur la réforme de l'État à l'étranger***, Information Report No 348 (2000-2001) on behalf of the Finance Committee, tabled on 31 May 2001 by Gérard BRAUN  
<http://www.senat.fr/rap/r00-348/r00-3481.pdf>

#### **UK National Audit Office Value for Money work:**

[http://www.nao.org.uk/what\\_we\\_do/value\\_for\\_money\\_audit.aspx](http://www.nao.org.uk/what_we_do/value_for_money_audit.aspx)

#### **Paris Declaration on Aid Effectiveness:**

<http://www.oecd.org/dataoecd/53/38/34579826.pdf>

#### **The European Consensus on Development:**

[http://www.delsen.ec.europa.eu/fr/telechargements/edp\\_statement\\_oj\\_24\\_02\\_2006\\_fr.pdf](http://www.delsen.ec.europa.eu/fr/telechargements/edp_statement_oj_24_02_2006_fr.pdf)

#### **Accra Agenda for Action:**

<http://siteresources.worldbank.org/ACCRAEXT/Resources/4700790-1217425866038/FINAL-AAA-in-French.pdf>

#### **Policy dialogue:**

[http://www.internationalhealthpartnership.net/CMS\\_files/documents/joint\\_assessment\\_tool\\_and\\_guidel\\_EN.pdf](http://www.internationalhealthpartnership.net/CMS_files/documents/joint_assessment_tool_and_guidel_EN.pdf)

### Some essential documents on indicators

DESROSIERES, A, ***Gouverner par les nombres: l'argument statistique***, T2, Presse de l'école des Mines; 2008. This book offers specific case studies, studies relating to household budgets, planning commissions, local statistics and national accounting, looking at the production of public-sector statistics and their use by public authorities. Although it does not specifically address DCs' problems, it is an excellent critical work.

For general information about indicators, the main reference websites are:

- OECD: <http://www.oecd.org>
- World Bank: <http://web.worldbank.org/WBSITE/EXTERNAL>
- IMF: <http://www.imf.org/external/np/>
- Eurostat: <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>

Though a little iconoclastic, the Pénombre website [www.penombre.org](http://www.penombre.org) regularly publishes short articles on the use of figures.

*If you have time:*

*Agriculture*

A large number of reference documents concerning agriculture in the broader sense can be found in the documents section of the FAO site:

<http://www.fao.org/documents/>

#### *Education*

The most straightforward document is one drawn up by Claude SAUVAGEOT for UNESCO as part of the COMED programme, Dakar, June 2001:

[www.dakar.unesco.org/sised/](http://www.dakar.unesco.org/sised/)

From a general point of view, the website of UNESCO's statistical body provides access to the most important documents relating to indicators and statistics in this field:

<http://www.uis.unesco.org/>

#### *Telecommunications*

The International Telecommunications Union website regularly provides quantified scorecards about sectoral indicators. The latest publication to appear on the site is: *African Telecommunication/ICT Indicators 2008: At a crossroads*:

<http://www.itu.int/publ/D-IND>,

### **Some essential documents on statistical quality**

- 1) ***Handbook on Data Quality Assessment Methods and Tools***, Eurostat, 2007. This document lays down the main quality dimensions applicable within the European statistical system.
- 2) ***Quality Assurance Framework***, Statistics Canada, 2002. This is a policy document setting out Statistics Canada's approach to ensuring quality management in the NSS.
- 3) ***Quality concepts for official statistics***, Encyclopaedia of statistical sciences, John Wiley and son. This article provides the most useful definitions of statistical quality.

*If you have time:*

#### IMF

<http://dsbb.imf.org/Applications/web/dqrs/dqrswork/>: on the IMF website, you will find numerous country reports concerning the evaluation of statistical quality, particularly in those areas dealt with by the organisation.

#### Statistics Canada

<http://www.statcan.ca/>: the Statistics Canada website is one of the best equipped in terms of online documentation and the quality-based approach employed by this organisation is probably one of the best developed.

#### Eurostat

<http://epp.eurostat.ec.europa.eu/portal/page>: on the Eurostat website you will find all the documents and reports concerning the quality-based approach within the European statistical system.

#### Paris 21

<http://www.paris21.org/pages/designing-nsds/NSDS-documents-knowledge-base/index.asp?tab=KnowledgeBase&option=doc>: on the Paris21 site, under the 'Knowledge Base' tab, you will find the majority of reference documents on the subject

International Organisation for Standardisation

<http://www.iso.org/iso>

Some experiences with the southern countries

<http://www.who.int/healthmetrics/>: on this site, you can find the results of quality assessments of health information systems.

<http://www.uis.unesco.org/>: the website of the UNESCO Institute of Statistics (UIS), which carries out quality assessments of education information systems. On this site, you will find the approach derived from the IMF/World Bank framework which is used by the UIS.



### ANNEX 3: SOME QUESTIONS TO ASK IN RELATION TO AN INDICATOR

Questions	Answers
Who is responsible for its calculation?	A key question, as if there are doubts about its quality, I need to know where I can find an explanation.
What is the collection mechanism used to obtain the data?	The data are administrative data, survey data or both.  Does this or do these mechanisms provide users with information about the methods and results? If yes, where can I get hold of it?
At what stage of processing is the figure provided?	Are the data provisional or definitive? In the latter case, in which publication was this indicator disseminated?  An official figure is one that has been published, whether it has been included in a publication or press release or made available online.
Survey data: what is the degree of precision (even theoretical) of the value?	Unless methodological publications can be stated which define the very measurements in question, common sense will have to be applied: <ul style="list-style-type: none"> <li>▪ a degree of precision of around 1 % is exceptional, even in the northern countries.</li> <li>▪ in the DCs, the majority of household surveys have desired degrees of precision of less than 5 %. In practice, at national level, they are between 5 % and 10 %, as precision soon deteriorates following geographic disaggregation (at the second level of disaggregation, e.g. regional or provincial, the value is meaningless).</li> </ul> <p>The only case where precision is not a concern in the comparison is when the sample (i.e. the individual questioned) is the same in both surveys.</p>
Administrative collection: on what basis is the calculation made?	Check whether the publication indicates the completeness of the collection, i.e. the ratio between the number of units on the administrative register identifying the individual questioned and the number of answers obtained.
Is the collection made under the same conditions?	You will often need to check whether the collection has taken place during the same periods, whether in a survey or certain administrative collections not carried out on an ongoing basis (e.g.: school surveys).  Are the budgets in place the same from one collection to the next? It is materially impossible to guarantee results of a consistent quality by spending three times less money (unless the method is changed).
Is the calculation method the same for two deliveries of the	When the data are published, always refer to the publication and look for elements indicating a possible change.

indicator value?	You should be wary of 'vague' sources. The source should refer to the collection mechanism and not to an administrative body (a ministry or otherwise).
Are the variations between two values plausible?	<p>Again, unless the producer has provided an explanation for the unusual nature of the variation, you will need to apply some common sense.</p> <p>For all variables touching on fundamental aspects of an economy or society, marked variations from one year to the next are very often down to a change in methods.</p>
What is the value of the unit?	<p>For all indicators expressed as a ratio, percentage or index, you should always keep in mind the value of the unit: in the case of 71 %, what is the absolute value corresponding to 1 %?</p> <p>This simple calculation can help to avoid erroneous conclusions about the achievement of targets: boosting an enrolment ratio by 1 % is hugely significant in terms of the number of new pupils and, consequently, the accompanying budgetary effort. At 0.9 % progress, the effort has been made, even if the target has not officially been reached.</p>
Is the variation trend known?	<p>Some basic considerations:</p> <ul style="list-style-type: none"> <li>▪ in the case of two values obtained a few years apart, it is impossible to draw a straight line linking them or to conclude that there has been an upward or downward trend;</li> <li>▪ in order to be able to confirm a trend, observations over a period of at least 10 years are required (and longer if possible).</li> </ul>

## ANNEX 4: DATA QUALITY FRAMEWORKS USED BY THE WORLD BANK/IMF AND EUROSTAT

The World Bank/IMF and European approaches differ in some respects:

- the European approach is, above all, a normative approach (in the sense that it has the force of law) aimed at qualifying the result. As a result, the dimensions of quality are predominantly methodological and technical;
- the approach taken by the World Bank and IMF is more descriptive and is closer to a holistic quality-based approach; it is concerned, certainly, with the quality of figures, but also with the entire process of quality management.

IMF system	EUROSTAT system
<p><b>0 Prerequisites of quality</b></p> <p><i>0.1 Favourable legal and institutional environment</i></p> <p><i>0.2 Resources commensurate with needs</i></p> <p><i>0.3 Relevance of the statistical information</i></p> <p><i>0.4 Quality management</i></p> <p><b>1 Assurances of integrity</b></p> <p><i>1.1 Professionalism as a basic principle</i></p> <p><i>1.2 Transparency in statistical policies and practices</i></p> <p><i>1.3 Ethical standards</i></p>	<p><b>Institutional and organisational arrangements</b></p> <p>1 Professional independence</p> <p>2 Mandate for data collection</p> <p>3 Adequacy of resources</p> <p>11 Relevance</p> <p>4 Commitment to quality</p> <p>5 Confidentiality</p> <p>6 Impartiality and objectivity</p> <p>9 Limited respondent burden</p> <p>10 Effectiveness and efficiency</p>
<p><b>2 Methodological soundness</b></p> <p><i>2.1 Concepts and definitions in line with international standards</i></p> <p><i>2.2 Scope in line with international standards</i></p> <p><i>2.3 Classification/sectorisation in line with international standards</i></p> <p><i>2.4 Base for recording</i></p> <p><b>3 Accuracy and reliability</b></p> <p><i>3.1 Appropriate source data</i></p> <p><i>3.2 Verification of source data</i></p> <p><i>3.3 Sound statistical techniques</i></p> <p><i>3.4 Assessment and validation of intermediate data and statistical outputs</i></p> <p><i>3.5 Revision studies</i></p>	<p><b>Core statistical processes</b></p> <p>7 Sound methodology</p> <p>8 Appropriate statistical procedures</p> <p>12 Precision and reliability</p>
<p><b>4 Serviceability</b></p> <p><i>4.1 Periodicity and timeliness</i></p> <p><i>4.2 Consistency of data</i></p> <p><i>4.3 Revision policy and practice</i></p> <p><b>5 Accessibility</b></p> <p><i>5.1 Accessibility of data</i></p> <p><i>5.2 Accessibility of metadata</i></p> <p><i>5.3 Assistance to users</i></p>	<p><b>Statistical products</b></p> <p>13 Timeliness and punctuality</p> <p>14 Coherence and comparability</p> <p>15 Accessibility and clarity</p>

## ANNEX 5: BRIEF DESCRIPTION OF THE EUROSTAT GUIDE

The '*Guide to Statistics in EC Development Cooperation*' was drawn up as a pilot version in 2009. While the initial version is available only in English at present, French and Spanish versions should be published shortly, along with an electronic version that can be consulted on the Eurostat website.

The guide is split into four main parts:

- The first is devoted entirely to the role statistics play in development policies and contains reminders of the MDGs and the main declarations and principles mentioned in part 1 of this guide: the Marrakech and Paris Declarations. This part also provides a detailed description of a national strategy for the development of statistics, such as that implemented by the Paris 21 initiative.
- The second part, entitled 'Statistics in Development' sets out the main principles underlying public service statistics and the characteristics of large-scale statistical operations. It also explains how a national statistical system works and defines the main types of statistical capacity.
- The third part provides some advice on the support that can be given to statistical systems in aid recipient countries. It describes the main concepts of statistical quality and the means available to improve this quality. It also proposes a set of guidelines for providing support/standard terms of reference, adapted to various types of support.
- The final part deals, more specifically, with statistics and problems of regional integration, in particular as regards the balance of payments and external trade.

The guide is backed by a wealth of supporting documents and contains references to all available information sources about statistics, both online and 'offline'.