About WEDC

The Water, Engineering and Development Centre is one of the world's leading education and research institutes for developing knowledge and capacity in water and sanitation for sustainable development and emergency relief.

We are committed to the provision of effective, evidence-based and appropriate solutions for the improvement of basic infrastructure and essential services for people living in low- and middle-income countries. With over 45 years of experience, we offer expert advice and quality learning opportunities for sector professionals.

Founded in 1971, WEDC is based in the School of Civil and Building Engineering at Loughborough University, one of the top UK universities. Being a part of a leading university gives us a recognised platform of independence and quality.

What makes us stand out from the crowd is our outreach to practitioners. We use our knowledge base and our applied research work to develop the capacity of individuals and organizations throughout the world, promoting the integration of social, technical, economic, institutional and environmental activities as foundations for sustainable development.

Visit our website to find out more about our postgraduate and professional development programmes (MSc, Diplomas and postgraduate certificates available at the University or by distance learning); our research; our advisory services; our international conferences; and our extensive range of information resources which are free to download from our knowledge base.

http://wedc.lboro.ac.uk

WEDC
Developing knowledge and capacity in water and sanitation

GUIDE 11

Selecting water, sanitation and hygiene indicators

For the effective management or investigation of a water, sanitation or hygiene project, the manager or researcher has to be aware of the current state of the project at any given point to be able to review its direction and measure progress towards its goal. Many indicators of progress can be measured but collecting and analysing information is expensive, so choosing which indicators to use and deciding when, where and how to measure them is important. This guide helps with this decision-making process.

Contents

This guide explores the nature of a good indicator, whether the indicator is for the day-to-day monitoring of water utility performance, emergency assessment of water resources or an in-depth assessment of attitudes about hand-washing. It does not set out to prescribe what should be measured, but describes the process of selecting what to measure and when and where to measure it.

wedc.lboro.ac.uk/knowledge

Loughborough University

This guide is one of a series of published learning resources which are available for purchase in print or available to download free of charge from the WEDC Knowledge Base. Any part of this publication, including the illustrations (except items taken from other publications where WEDC does not hold copyright) may be copied, reproduced or adapted to meet local needs, without permission from the author/s or publisher, provided the parts reproduced are distributed free, or at cost and not for commercial ends and the source is fully acknowledged. Please send copies of any materials in which text or illustrations have been used to WEDC at the address given below.

Please note:

There are numerous publications on indicators produced for specific sectors. As WASH is multi-disciplinary, material for this booklet has been sourced from the health, environment, education, utility manage-ment, humanitarian and general development fields. The sources for this booklet are listed in the bibliography.

Learn with WEDC

Learn with one of the top, award-winning universities of excellence in the UK and partake in a quality learning experience.

Gain a recognised, respected, independent and validated qualification.

WEDC offers you a wide variety of learning opportunities in all aspects of water and environmental management, water and waste engineering and infrastructure in emergencies.

You can learn in different ways and at different levels. Come to WEDC or study at home or at your place of work.

Choose from one of our postgraduate programmes and study towards a Postgraduate Certificate, Diploma or Master of Science (MSc). Study by research towards a PhD.

Alternatively, tailor-make a programme from our wide range of stand-alone modules to suit your particular professional development requirements.

Or you may prefer to invite us to collaborate with you to devise a programme especially for your staff.

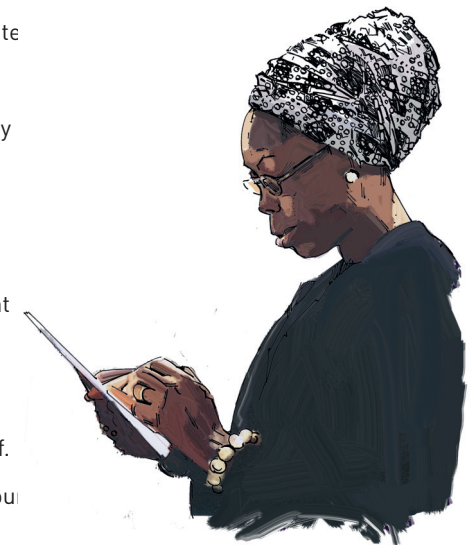Go straight to details about one of our learning opportunities.

Postgraduate programmes

- Infrastructure in Emergencies
- Water and Environmental Management
- Water and Waste Engineering

Other courses and programmes

- Special courses for groups
- Professional development
- PhDs
- Online courses

Notes

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Introduction

Many factors have to be taken into account for development projects, programmes and services to be effective and sustainable. Projects (such as the construction of a wastewater treatment works), programmes (such as a training course for hygiene promoters across a whole organization), and services (such as the provision of a water supply in an urban area) involve the management of social, human, economic, environmental and physical aspects. These can be monitored and measured using specific parameters, which, if they are not clearly defined, could result in too many measurements that are difficult to manage, or too few which could mean that necessary details are lost. Choosing the wrong indicator can distort decisions by giving undue attention to certain aspects. However, measuring what is in theory an ideal indicator may not be practical. So what should be done?

Why measure?

"When you measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be."

Lord Kelvin (1824 – 1907)

Measurements are made so decisions can be taken. The phrase 'if you can't measure it, you can't manage it', indicates the importance of basing decisions on facts rather than guesswork and supposition.

Measurements however are only the start of the process. Raw data needs to be analysed and presented to be informative. Information has to be assessed and thought about critically to create knowledge. Experience and judgement of this knowledge provide the basis for wise decisions. Facts and figures are explicit and can be written down. Knowledge is implicit and exists in the mind.
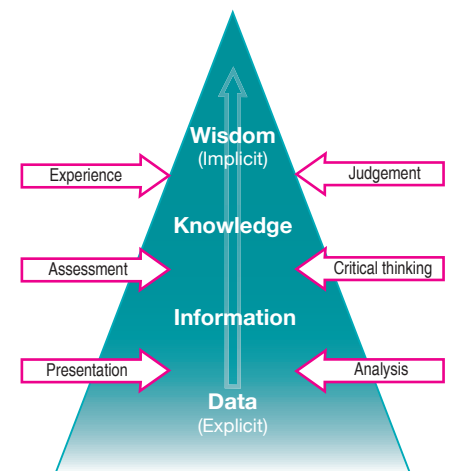
Figure 1. The interpretation of data

What are indicators for?

"You can expect what you inspect."

W. Edwards Deming (1900 – 1993)

Indicators should be chosen well and should emphasize and relate to issues that need action or attention as the general rule of thumb with indicators is that 'what gets measured gets done'. Chosen indicators should be easy to understand and relevant for decision-making, evaluation and communication. Indicators on their own will not provide sufficient explanation, interpretation and assessment to replace a full report.

Data: recordable facts

Information: meaningful combinations of data

Knowledge: the sum of what is known by an individual, or about a subject. Knowledge is created through the accumulation of selected items of information. Knowledge is information that has been interpreted and made concrete in the light of the individual's understanding of the context (World Bank, 1999)

Communication: the transmission of data, information or knowledge between two or more points.

(Saywell and Cotton, 1999)

Indicators:

- summarise trends at a national level (as well as at provincial and/or local levels where appropriate);

- help integrate information management across resource issues, as well as administrative, policy and scientific boundaries;

- promote more effective sharing of existing approaches, technologies, data and knowledge between relevant agencies;

- improve access to and availability of information for all resource managers, users and the public;

- enable comparison from local to national level and compatibility with regional and global indicators, to the extent possible;

- promote more informed decision-making;

- improve implementation; and

- lead to increased accountability.

Who are measurements for?

"In God we trust; all others must bring data."

W. Edwards Deming (1900 – 1993)

Measurements are useful for all stakeholders involved with a water and sanitation project. Well chosen indicators can ensure the services delivered are efficient and effective. Monitoring

Available from: http://www.unaids. org/ en/media/unaids/contentassets/ documents/document/2010/8_2- Intro-to-IndicatorsFMEF.pdf [accessed 6/12/20012]

UNDP no date. *Handbook on Planning, Monitoring and Evaluating for Development Results.* Available from: http://web.undp. org/evaluation/handbook/ch2-4.html [accessed 6/12/20012]

in inappropriate solutions being imposed.

Guidelines may not met all the time, so some measure of compliance needs to be established – for example the parameter must meet the guideline value 80% of the times it is measured, or perhaps being allowed to miss the guideline value sometimes but never falling beyond some more critical level.

Bibliography and references

ALEGRE, H., BAPTISTA, J.M., E CABRERA, E. Jr, CUBILLO, F., DUARTE, P., HIRNER, W., MERKEL, W. and PARENA, R. 2006 *Performance Indicators for Water Supply Services.* 2nd ed. IWA Publishing London

BROWN, G and EDMUNDS, S. 2011. *Doing pedagogical research in engineering.* Engineering Centre for Excellence in Teaching and Learning: Loughborough University, Loughborough.

CAIN, E. 2003. *Quality Counts. Developing indicators in Children's Education.* Save the Children. London UK Available from: http://toolkit.ineesite. org/toolkit/INEEcms/uploads/1089/Quality_counts_developing_indicators.pdf [accessed 7/12/2012]

CAIRNCROSS, S. no date. *Measuring the health impact of water and sanitation.* WELL Fact Sheet. Available from: http://www.lboro.ac.uk/well/ resources/fact-sheets/fact-sheets-htm/mthiws.htm [accessed 6/12/20012]

DEPARTMENT OF ENVIRONMENTAL AFFAIRS AND TOURISM (DEAT). 2001. *National Core Set of Environmental Indicators for the State of Environment. Reporting in South Africa: Scoping Report.* DEAT, Pretoria. Available from: http://www. environment.gov.za/soer/indicator/background.htm [accessed 9/2/2007]

DFID 2003. *Tools for Development. A handbook for those engaged in development activity.* Available from: http://webarchive.nationalarchives.gov.uk/ + http://www.dfid.gov.uk/Documents/publications/toolsfordevelopment.pdf [accessed 6/12/20012]

SAYWELL, D. and COTTON, A. 1999. *Spreading the Word.* Loughborough, UK. WEDC, Loughborough University. UK. Available from: https://wedc-knowledge.lboro.ac.uk/details.html?id=7500 [accessed 7/12/2012]

THOMSON, M., OKUNI, P.A. AND K SANSOM, K. 2005. *Sector performance reporting in Uganda – from measurement to monitoring and management.* 31st WEDC International Conference, Kampala, Uganda. WEDC, Loughborough University. UK Available from: http://wedc.lboro.ac.uk/resources /conference/31/Thomson.pdf [accessed 7/12/2012]

UNAIDS MONITORING AND EVALUATION DIVISION. 2010. *An Introduction to Indicators.* UNAIDS

Indicators show performance. In logframe analysis and planning, they play a crucial role:

- They specify realistic targets (minimum and otherwise) for measuring or judging if the objectives at each level have been achieved;

- The process of setting indicators contributes to transparency, consensus and ownership of the overall objectives, logframe and plan.

Who sets indicators is fundamental, not only to ownership and transparency but also to the effectiveness of the indicators chosen. Setting objectives and indicators should be a crucial opportunity for participatory management.

A variety of indicator types is more likely to be effective; the demand for objective verification may mean that focus is given to the quantitative or to the simplistic at the expense of indicators that are harder to verify but which may better capture the essence of the change taking place.

The fewer the indicators the better. Measuring change is costly so use as few indicators as possible. But there must be indicators in sufficient number to measure the breadth of changes happening and to provide the triangulation (cross-checking) required.

(DFID, 2003)

progress (is the project on time and on budget?) and quality of products (is it going to do what was planned?) enables users and donors to hold implementers to account.

Particular indicators will be of interest to certain groups of stakeholders:

- project managers may be interested in the use of resources such as project expenditure to date or the number of staff employed;

- both project managers and users will be interested in project progress;

- specialists such as groundwater experts will be concerned about the distance of a pit latrine to a well;

- hygiene promoters will want to know how close a water supply for handwashing is to a latrine; and

- researchers may be interested in how an indicator varies over time, between groups or from place to place.

Who chooses indicators?

Indicators are chosen by a range of people.

- Donors and governments may be addressing international targets or meeting regulatory standards or guidelines.
- Specific surveys, most notably at the beginning and end of a project may be determined by external experts based on standard practice.
- The users of a service may be trained to monitor the facility, empowering them to measure performance.

- Researchers may take a rigorous and independent approach to selection and measurement.

Because of the relevance to all stakeholders, indicators should be selected using a process that involves all interested parties, especially those who can sometimes be excluded from normal decision-making channels. Sharing indicators between interested parties is not only more efficient, but helps align disparate activities towards a shared goal.

Some stakeholders will be very influential and can dictate what indicators are used.

Measuring health impact of water and sanitation projects

A review of the published and unpublished results of the best health impact studies of the Water Decade concluded that health impact studies are not an operational tool for project evaluation or 'fine tuning' of interventions. The results are not only unpredictable; they frequently offer no firm interpretation. Even some studies supervised by eminent specialists have produced almost useless or meaningless results, after taking years to complete and costing substantial sums of money.

What we do know from the existing literature on impact studies is that, where health impact was found, the provision of water supply or sanitation was accompanied by improvements in hygiene. Instead of attempting to measure disease rates, studying patterns of hygiene behaviour has far greater diagnostic power, in terms of indicating opportunities for project improvement.

Since it is further back up the causal chain, it is easier to attribute behaviour change to the project intervention. Behaviour can also be assessed at the project design stage. This will not only help to establish a baseline against which to compare evaluation results; it can also be used to improve project design.

adapted from Cairncross, S. (undated).

the data collection is and enables the planner to consider high, low and most probable forecasts.

Several observations may give similar readings but, whilst these are precise, they are not necessarily accurate. Several other observations may show a range of values (so are not precise) but the average of them all may be accurate.

Collating, storing and presenting information

Once collected, the data needs storing and collating. Whilst a report or a database is useful for recording the data, graphical techniques such as graphs, diagrams, maps and tables allow patterns to emerge and provide some sense of order. Analysis and judgement are required to make the data meaningful. Comparisons, for instance with agreed standards or average levels.

Standards and targets

Targets, standards, benchmarks, goals, specifications and guidelines allow the indicators to have some meaning compared with other projects and places. Setting standards can be seen as a way of improving services and ensuring adequate performance, but this can have problems as well as benefits. The standard will be defined by one or more indicators, so the correct choice measurement will be important

Standards are mandatory levels that have to be reached. Guidelines are advisory, showing what is good practice.

Benchmarks are comparisons, seeing how one project or organization compares to other similar entities. Targets are aspirations, setting out where the project, organization or place wants to be. Setting targets that are too low and easy to meet may mean that potential progress may be lost. Too difficult a target leads to inevitable failure and disengagement. Targets should stretch the resources to provide a challenge but not be unrealistic. Some target driven approaches lead to bias as all the focus is on achieving the issue being measured, to the detriment of other factors that may be just as important to the success of the wider project but are not being monitored.

Just as with selecting indicators, setting standards needs consideration. For example regarding a VIP latrine as the minimum standard for sanitation may preclude poor people from having any latrine, as they may not be able to afford what is permitted. Having a VIP latrine standard however can legitimise on-plot sanitation if the prevailing thought is for unsustainable sewerage facilities.

Meeting standards can be a way of promoting quality but can stifle innovation and restrict choice. Selecting a standard hand pump may help with organizing spares, training mechanics and ensuring economies of scale, but it can hamper improvements and reduce competition. A 'one size fits all' approach neglects the local context and may result

Metadata

Each set of data should have some information attached to it, such as:

- When was it collected?
- Who collected it?
- How was it gathered?
- Who "owns" the data?

One useful principle is that of triangulation or crosschecking, where two independent sources of information are used to corroborate and support each other. If there is a discrepancy, then further investigation and seeking additional views and sources of information are required.

A judgement has to be made as to how reliable a search of data is likely to be. A reputable, local independent NGO may be more trustworthy than a remote government official who has political considerations to take into account. This bias may be deliberate or may be due to the observer's cultural, educational and social background.

The information can be validated against known data, for example other similar situations or standards. This can give a rough order of magnitude, i.e. the number of latrines to be expected in one area might be 5% coverage, so a measured figure of anything between 2

and 15% might be reasonable, but 0.5 or 50% coverage is unlikely, though still possible. This would require checking.

Precision (the consistency of repeated observations) can be indicated by the level of detail provided. This can be seen numerically with:

- Between 5 and 20
- 10m ± 5
- About 9m
- 8.5m
- 8.4873m

demonstrating increasingly higher levels of precision. Where there is a high level of uncertainty, such as in the early stages of an emergency response, a range of figures should be used rather than a single number. This shows how confident

Validity, reliability, bias

Validity: The extent to which a measurement or test accurately measures what is intended to be measured.

Reliability: The consistency of the data when collected repeatedly using the same procedures and under the same conditions.

Bias: Any effect during the collection or interpretation of information.

UNAIDS, 2010

Other important stakeholders may be excluded from decision-making and their priorities may be neglected unless a positive effort is made to reflect their interests in the selection of performance indicators.

Data produced by one sector may provide information for another sector – for example, population statistics are important when planning infrastructure but may be collected by health professionals. Working together to collect data can save time and effort overall even if it adds more questions to a particular survey. Population statistics are useful, but it is better to have them broken down into age and gender groups; this will take longer to obtain but will provide more information.

What is an indicator?

"Knowledge is theory. We should be thankful if action of management is based on theory. Knowledge has temporal spread. Information is not knowledge. The world is drowning in information but is slow in acquisition of knowledge. There is no substitute for knowledge."

W. Edwards Deming (1900 – 1993)

Many factors influence a project or programme. Some of these are constant and cannot be changed. (For example, there are only ever 24 hours in a day.) Factors that change or can be changed are called variables. These variables can be measured (perhaps in several

different ways and at various times) and each measurement results in a piece of data (a 'datum') that can be a quantity or a measure of quality. Huge amounts of data can be difficult to manage and absorb, so some refinement is required.

Significant variables may be called parameters, especially if they define or describe important elements of the project. Some variables can give extra information about the wider context and are used as indicators. They indicate something about the status of the project, programme, organization or environment as well as the variable that is being measured. A lion is an indicator species in ecological terms. If they are present then their prey (antelopes, for example) will also be present. If a healthy population of antelopes is present then there is enough vegetation to feed the
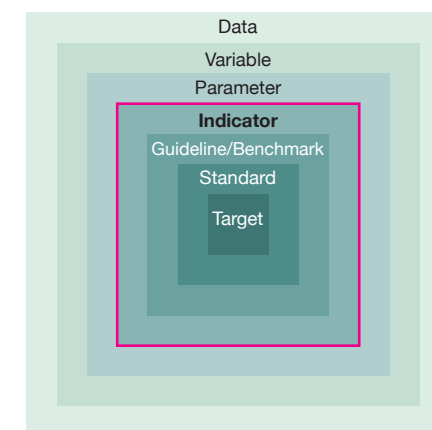
Figure 2. From data to target

herd. Thus surveying the number of lions infers something about the health of the whole ecosystem.

Some variables can be difficult or expensive to measure in any meaningful way, so an indicator should be verifiable. Indicators may need some level of analysis (extrapolated to become an average or maximum value, for example) and so the indicator becomes a statistic.

Thus, indicators represent a huge amount of data in a form that is easier to comprehend. The process can be taken further by comparing a few indicators against guidelines or benchmarks, to judge performance. One or two indicators can be used to determine the status of all the other indicators and beyond them, the variables. These can then be used as a target or a goal.

Using indicators

Indicators are used to communicate information about progress towards a specific goal in a simplified manner, condensing information about complex issues for decision-making, management, monitoring and reporting purposes. Indicators will provide a signal to an issue of greater importance or make more evident a trend or phenomenon that is not immediately detectable.

Like any form of information, there are limitations to their use. The acceptability of any indicator depends on the availability and confidence of

the data, as well as the interpretation of the indicator, as indicators tend to provide only the essence of a situation rather than the whole picture. Indicators reduce the number of measures required to understand changes in society, the economy or the environment and simplify the communication process of transmitting information to the user.

The WASH sector uses a wide variety of indicators:

- environmental conditions such as temperature or rainfall;

- human characteristics such as literacy and life expectancy;

- physical status such as leakage rates in piped water systems or cracking in concrete tanks;

- economic state of affairs such as energy costs or wage levels; and

- social circumstances such as the number of women on a committee.

Classifying indicators

There are several ways of describing indicators that demonstrate some of the issues to consider when selecting parameters to measure.

Data can be divided into quantitative (i.e. it can be measured objectively, often numerically) and qualitative data (i.e. it is measured subjectively; it is the views, opinions and perspectives of an individual or group). The number of latrines built (quantity) is different from whether the

Data quality

Data need to be of an appropriate and known quality. This can be expressed as validity/ accuracy, reliability/ precision, and bias. Just looking at the output data will not tell you about its quality. Looking at how and when it was collected, who carried out the survey, how they were trained and what experience they have will give an indication of its reliability.

This 'data about the data' is called metadata. Original records of data (e.g. survey books, notes from meetings, questionnaire forms) should be kept even after the data has been transferred to sketch maps, databases, spreadsheets or reports, as a weak point in information management is the transfer of data from one form to another. Having the raw data enables checking at a later date.
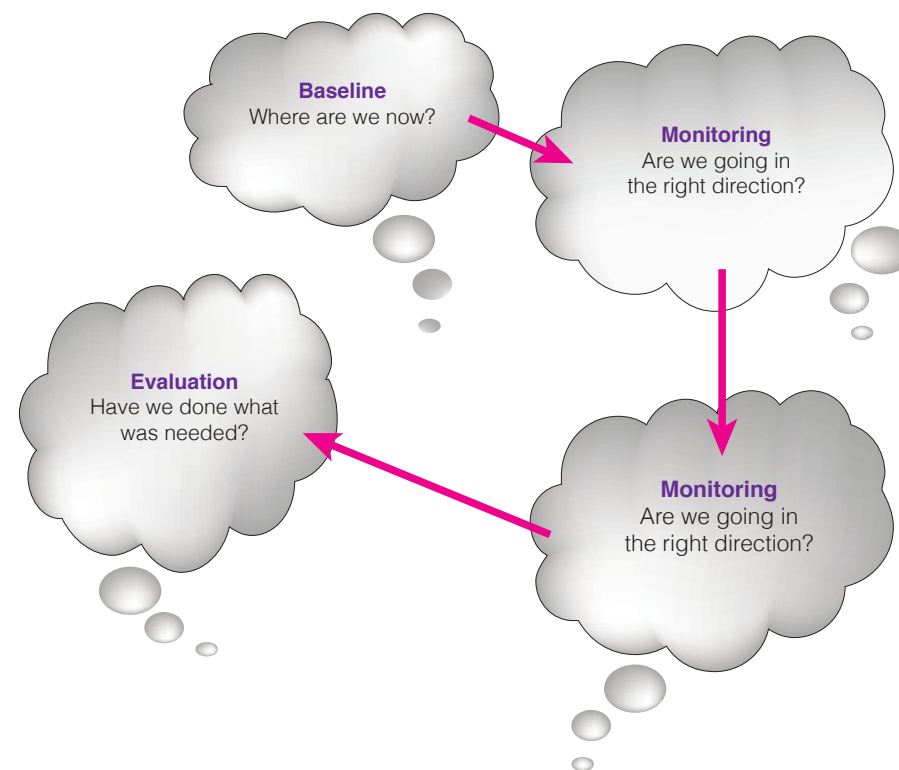
Figure 8. Assessments occur throughout the project cycle
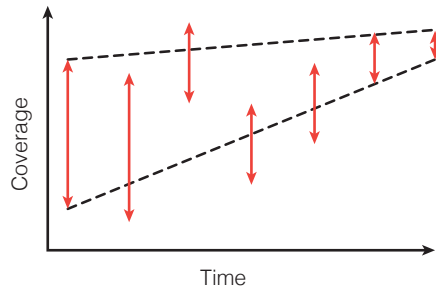
Figure 7. Changes in readings over time

The pattern may be irregular. If a factory washes out chemical containers at the end of every batch, then sampling time should coincide with this cycle, rather than being fixed. Similar patterns can occur socially with daily and annual patterns. Trying to interview people when they are preparing food or harvesting crops may not be the best time.

Understanding the rate of change may be important for the frequency of readings. Rainfall can vary from minute to minute, river flows vary from hour to hour but groundwater levels may only vary discernibly from week to week or even month to month. If the frequency of observations is too frequent (e.g. daily ground water readings), then effort is being wasted. If the frequency is too infrequent, then important details can be lost.

Projects
Projects have a particular cycle. Different indicators may be needed to measure:

- inputs
- outputs
- outcomes and
- impacts.

Within the project cycle, the same set of indicators may need to be measured several times to provide:

- baseline data before the project starts (where are we now?);
- if time is limited in emergencies, the baseline study is focused only on critical needs;
- monitoring data at regular intervals or specific milestones (are we going in the right direction at the right speed?); and
- evaluation data at the end of the implementation stage (did we get where we wanted to go?).

There is also a longer-term project view, where capital works implemented in a project are then operated, with Key Performance Indicators demonstrating operational effectiveness.

Longitudinal studies
Some issues have a strong time element and so observations are made at intervals, perhaps over many years, to produce a time series of data. This may focus on a cohort, such as children starting school or hand pumps installed in a single year, observing how they change over time.

users like them or the latrines are clean (qualities).

Qualities of data
'Quantitative' and 'qualitative' is a simple classification; this distinction can be expanded. Data can be:

Nominal i.e. a name with no 'value' e.g. blue/ green/ yellow, yes/ no, male/ female.

Ordinal e.g. 'first choice, second choice' or 'very bad, bad, good, very good' so the data is ranked but there is no measure of the strength of the choice/ opinion. May is after January but May cannot equal January x five.

Interval values provide an order (one value being more than another) and in addition provide an indication of the difference between the values (e.g. 50°C is hotter than 25°C and slightly cooler than 55°C). People may be asked to rate preferences on a scale of 1 to 5.

Ratio – most natural measurements (e.g. height, weight, age). These have a zero value.

Assigning 1 = 'yes' (or 'very satisfied' or 'prefers blue') enables qualities to be expressed as numbers but they are still qualities.

Data can be discrete where the measurement or description of a variable that can only take specific values (e.g. days are Monday or Tuesday – you cannot have Wednesday and a half). A road is

Quantities or qualities?
Whilst both types of data indicate the reality of the situation, numbers have an illusion of being more valid but a precise number may be less useful than the views of an informed local person. For example, consider how long it takes to travel 50 km on a road in a vehicle that can move at an average of 25 km/hour. The quantitative answer may be two hours, but a qualitative response may be longer, as an experienced person allows time for checkpoints, punctures, rest stops and the unexpected.

concrete, tarmac or earth, numbers of people attending a meeting must be whole numbers, not, for example, 43.73. In contrast, continuous data can be any value within limits (e.g. the well can be 4.58 or 4.61 m deep).

One particular type of discrete variable is when continuous data is grouped into classes or categories. Children are grouped into classes at school even though their ages vary by a year (or more). Having a block of data makes handling the figures easier. Examples of classes of data include:

- annual utility profit values;
- national sanitation coverage;
- death rates of children under five; and

- the total (cumulative) flow in a river in June (the sum of the continuous values throughout the whole month).

Exclusive values are when the measurement of a variable can only take certain unique values – a pump may be broken or working, a committee chairperson can be a man or a woman, a person is dead or alive.

Disaggregated data ensures that a class can be broken down into sub-classes. The term is often used when more detail is required than the 'headline' indicator can provide. An example is rather than stating how many people were trained on hand pump maintenance, the number of men and women trained is given.

Some measurements are direct, with the variable that is of interest being examined first hand (such as observing how many people wash their hands after defecation). Indirect measures attempt to evaluate the variable where direct measures may not be viable (e.g. asking people if they wash their hands after defecation). Proxy indicators attempt to measure the variable by looking at a related issue (e.g. is there soap in the latrine and is it being used?). Indirect and proxy measures are used when the cost, timing or other limitations preclude the use of direct measures.

Primary data is gathered by the assessor whilst secondary data is derived from the work of others (e.g. using figures published in a research paper or textbook).

Data may be grouped by sector or theme. For example:

- economic data;
- social data;
- physical data;
- environmental data; and
- human data.

Indicators may be categorized by level, with globally relevant figures, national statistics, regional performance monitoring and local measurements all related but focusing of different aspects. Different levels can be nested, with each lower level providing more detail and supporting the higher levels.

Relationships

Variables are not easy to understand in isolation so need to be viewed in relationship with other variables.

Variables can be classified as independent and dependent. A hygiene promotion campaign may result in increased sales of soap but an increase in sales of soap will not create a hygiene promotion campaign, so the sales are dependent on the campaign and the campaign is independent of the sales figures.

Confounding factors may appear to show a level of dependency where no real relationship exists. Malaria may decrease as the price of maize falls, but the two unrelated issues are dependent on the

Case studies recognise that the whole population of interest is very diverse. Rather than studying across the whole range, it may be better to study one or two specific examples in detail to provide depth. It is recognised that this will not be representative of the total population, but may give insights that would not be possible by looking more generally.

When to measure?

Another aspect of choosing what observations to make, besides the number of them and the location, is when to sample. Temporal variation (changes over time) need to be considered and the time and frequency of readings needs to be planned. Figure 6 shows how readings at the wrong time can give 'false' information, shown by the dashed line compared to the 'real' solid line. Some critical data is needed immediately, for specific needs, some will be needed in the future and can be gathered later to spread out the assessment process. Some may be needed in the future, but it should start to be collected as soon as possible. Rainfall and river flows for example cannot be characterized by a single 'snapshot' but need a long data record before they can be analysed. Single observations of a changing variable are called instantaneous readings, recognising that they are a part of a longer continuous pattern.

Seasons and cycles

Where the parameter to be measured varies in some sort of pattern, it is

important to understand this in order to obtain valid observations. For example, if daily temperature readings are being made, they could be the maximum (or minimum) or at a fixed time (e.g. 9 am). Taking one reading at noon and the next day's reading at 8 pm would not be valid as there is a diurnal (daily) pattern.
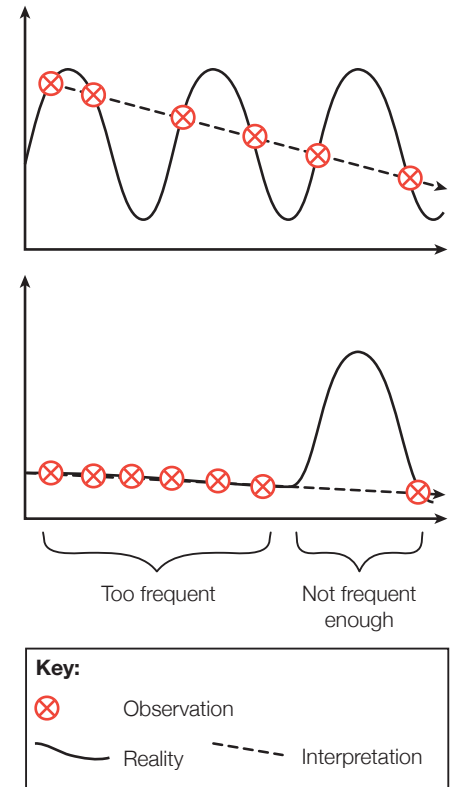
Figure 6. Choosing when and how often to observe can be important
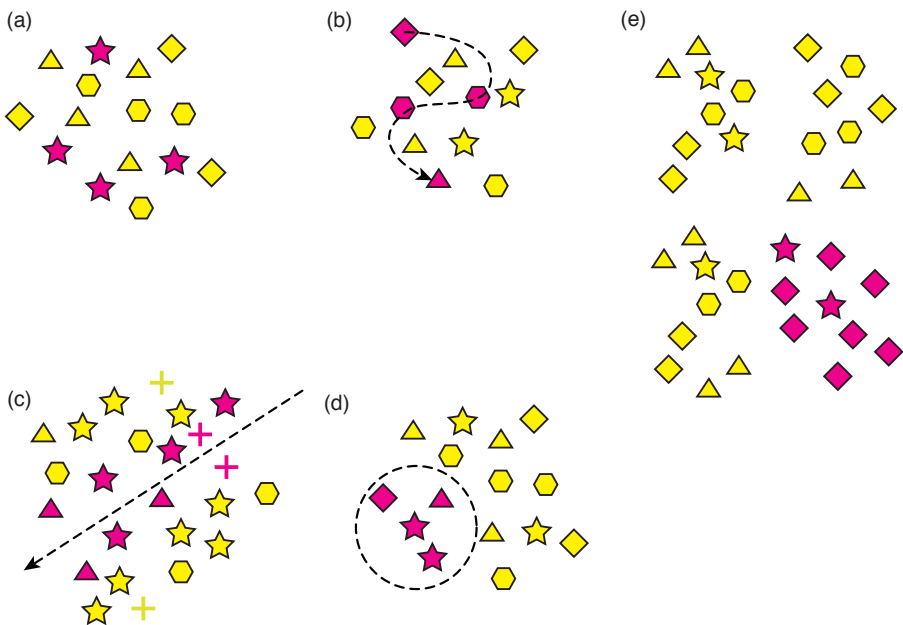
Figure 5. Non probability sampling techniques
a) purposive b) snowball c) transect walk d) opportunistic e) case study

Opportunistic or convenience sampling. An element of convenience is likely to enter into all sampling decisions. With limited time and resources, it makes sense to opt for informants who are available today rather than wait a week for others to return from leave. Given two (apparently) equivalent villages as potential case studies, it makes sense to choose the village that is only half an hour rather than three days' walk.

Complete reliance on convenience is not advised, although it can usefully

supplement other sampling methods. **Purposely opportunistic** sampling, e.g. interviewing women and children queuing at the hand-pump, can valuably supplement observational data.

Transect walks are another technique of providing convenience. A route is selected through an area of interest, ideally going through different neighbourhoods. Talking to bystanders and householders encountered is opportunistic but some attempt is made to ensure variety.

dry season and the resultant harvest and are independent of each other.

Over a longer period, going beyond a specific project, this cause and effect relationship can be expanded by categorizing the indicators according to DPSIR – see Table 1.

Change
One simple relationship is comparing the same variable over time, seeing if it has changed. Change could be presented as an absolute value or a percentage/ proportion. An absolute value could be, for example, 'the price of water has risen from $5.00/m^3$ to $6.00/m^3$ in a month' whilst a proportional comparison would be that 'the price of water has risen by 20% in a month'. No change can also be significant.

Projects and programmes
A baseline study looks at the conditions before a project. Summative indicators are evaluated at the end of a project or process, whilst formative indicators are measured at stages or milestones to monitor progress. Dividing the indicators into before, during and after a project can be expanded into further detail.

- A project has a goal and the impact can be measured.

- The impact may take time to become apparent and may depend on other factors, so the success of the project purpose can be reviewed by measuring the outcomes.

Table 1. DPSIR indicators

Driving force	The human influences and activities that combine with environmental conditions that bring about change
Pressure	The action of driving forces on society, environment or infrastructure
State	The condition of a community, environment or infrastructure
Impacts	The results of the pressures on the state, which may occur in a certain sequence.
Response	These are the actions taken by society to address the impacts.

Adapted from DEAT, 2001

- The aim or purpose of the project is addressed by carrying out activities, which result in outputs.

- The activities can be measured as they are carried out by process indicators or can be monitored by the provision of inputs or resources.

Aggregation and indices
Indicators focus and condense information about complex issues for decision-making, management, monitoring and reporting purposes. Indicators will provide a signal to an issue of greater importance or make more evident a trend or phenomenon that is not immediately detectable. In short, indicators quantify and simplify

phenomena to help us understand complex situations. Although indicators may be aggregates of raw and processed data, they can also be further aggregated themselves to form complex indices. High-level decision-makers dealing with sustainable development issues, routinely call for a manageable number of indices that are easy to understand and use in decision-making.

Globally, a number of indexes exist (e.g. the Human Development Index). These aggregate indicators from a core set, into meaningful indices that will support high-level decision-making.

How to select indicators

The process of identifying selecting indicators can be a research project in itself, with several stages before the most suitable measures can be chosen. The following is a case study of the selection of indicators for the South African State of the Environment Report (DEAT 2001) and illustrates the stages that need to be considered.

Case study: example of the process of selecting environmental indicators
The indicators should address [current] priority issues and emerging issues. The framework for reporting on indicators

Figure 3. A results chain
http://web.undp.org/evaluation/handbook/ch2-4.html
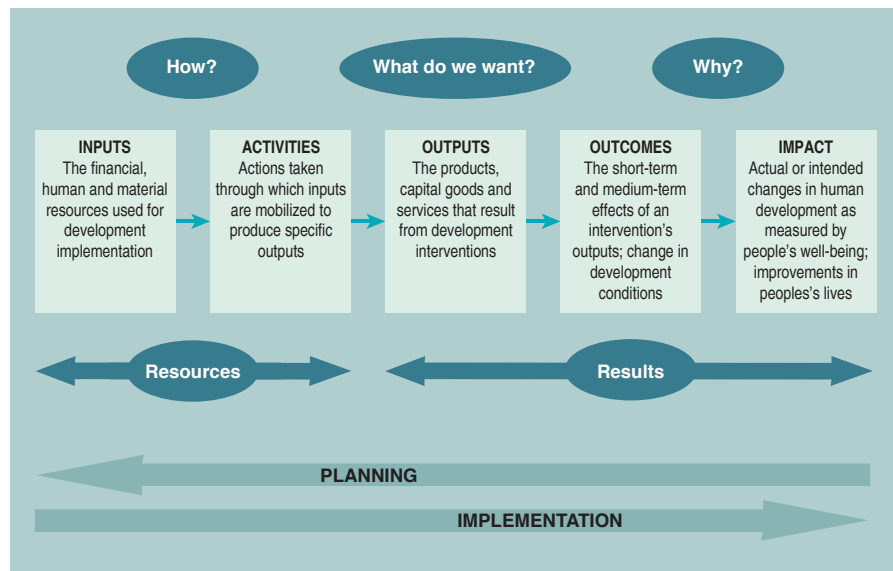
representative than just selecting people at random – which may have led to more men or women being selected, depending on the circumstances. For example, if every fourth householder was interviewed during the day and men were out working somewhere, there may be more women consulted. The same survey later in the day, when men are back home and women are cooking, may lead to more men being interviewed.

Cluster sampling. Rather than trying to survey a whole area, cluster sampling involves stratification of the population, but only taking a sample of the strata (or clusters), e.g. blocks of housing. Within clusters, data can be collected on all units, or a sample taken using one of the above methods. This is then used to predict the findings from all areas. Care has to be taken to select areas that are representative, to avoid bias.

For example, a District is divided into six sub-Districts, with six parishes in each sub-District. Cluster samples: two out six sub-Districts are selected, three parishes in each of the two selected sub-Districts (total six parishes), then random sampling in each parish.

Types of non-probability sampling
These are illustrated in Figure 5.

Quota sampling may appear to be similar to stratified random techniques, but differs in that there is an element of choice rather than random selection. The

surveyor has a quota, such as 15 women over the age of 40, but they can select who they want with these criteria.

Purposive sampling. In the context of qualitative research, this is the most common form of sampling. Informants are selected based on particular characteristics that are likely to provide the richest and most informative responses to the research questions. These could include people:

- with a particular expertise or position, e.g. NGO staff, chair of water committee;

- belonging to a minority group – ethnic or religious, disabled, living with HIV/AIDS;

- with a particular experience, e.g. trained as a pump attendant, participated in a project evaluation; and

- who provide a politically important perspective.

A few informants may need to be chosen out of courtesy or political necessity.

Snowball sampling is often used when the range of possible informants is not known in advance. The starting point may be only one or two informants who then provide further contact details of informants who they think will be able to provide further information, and so on, so the number of informants gradually gets larger – like a snowball that grows as it rolls downhill.

members of the population. This is a powerful tool, but often difficult to employ outside of very small studies.

Systematic random sampling. Here every fifth, tenth, twentieth etc. person on a list, in a queue or passing by, or every sixth house on a street is sampled.

Stratified random sampling. Here the whole population is divided up into groups (e.g. by age, gender, location) and then within that group, individuals are selected by simple or systematic random sampling. Thus if 35% of the population was male and 65% female, interviewing 35 men and 65 women may be more

(a)

(b)

(c)

(d)
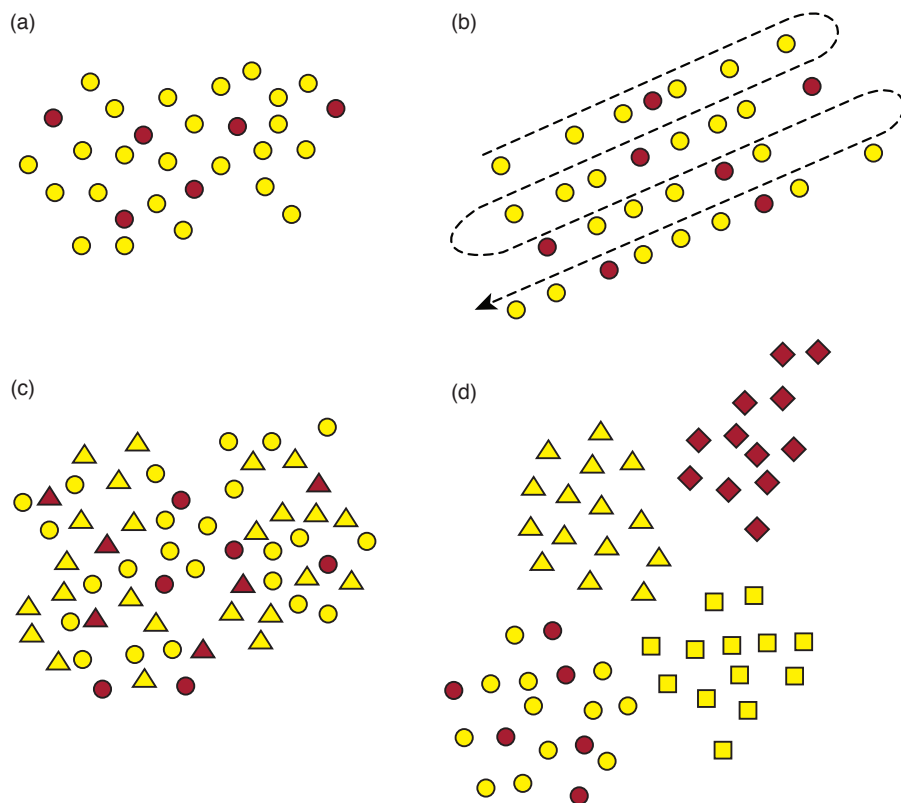
Figure 4. Probability sampling techniques
a) simple random, b) systematic random, investigating every 4th household, c) stratified random (with two groups ● and ▲), d) cluster sampling of two groups out of four

will be flexible to allow continuity between different levels of reporting (local to global). Selection took place through four phases:

Phase 1. The scoping phase included a review of the strategic context of the programme, a review of existing indicator programmes, and design of the project process and stakeholder consultation.

Phase 2: The selection of issues and criteria included the formulation of criteria for selecting indicators, the preliminary identification of issues through a review of policy and legislation, and discussions with experts on emerging issues. In addition, the public was consulted through a survey to determine key issues facing them. A stakeholder workshop was held to contribute to the refinement of the issues for which indicators were to be selected.

Phase 3: Selection of indicators was the longest phase in the project, divided into specialist studies to select existing and/ or develop new indicators.

The indicators were also categorised into three different levels based on the availability of data for each particular indicator. The indicator levels are defined as follows:

- **Level 1:** adequate data are available now for all components of the indicator and can be used to support the indicator without significant additional costs;

- **Level 2:** the indicator is presently feasible, but cannot be provided without additional investment in the data collection process; and

- **Level 3:** no data currently exists for the indicator, and there is no immediate intention to collect the data.

The indicator selection process should always be a participatory process that involves various different stakeholders. Stakeholders are often able to provide excellent insight into what needs to be measured through indicators, and have a diverse range of views and expertise that contributes to selecting the best set of indicators.

Phase 4: Implementation was concerned with the development of an implementation plan for the indicators. A database was developed to enable the indicators to be displayed in a dynamic and customizable way. The implementation plan provides details on:

- Data collection procedures (including the legal requirements for providing access to information, instructions on how to go about accessing existing data and how to collect new data);

- Conceptual information system design (including information on data collection, system architecture, user interface, archiving and security);

- An implementation plan (including a plan for establishing data collection

mechanisms, determining data standards and procedures, determining data conversion routines, developing a user interface and developing the system;

- Costs of implementation (including hardware and software);

- Funding mechanisms (including novel ideas on sponsorship and contributions in kind to ensure that the programme can continue to exit independently of donor funding);

- Use and interpretation of the indicator set (including information on how the indicators can be used to fulfil international and national reporting obligations, the relevance of the indicators at provincial and local level, aggregation and use of the indicators in reporting).

The set of indicators is divided into themes, designed to address the priority issues. Each indicator has an information sheet associated with it, including:

- description of the indicator;

- reasons for selecting the indicator;

- linkages to other indicator initiatives;

- information on data sources, data acquisition and data limitations;

- example(s) of reporting on the indicator, including methodologies for calculation where appropriate; and

- outline of any needs for further developing the indicator.

The indicators are arranged according to the Driving Forces-Pressure-State-Impact-Response (DPSIR) framework.

What to measure?

"The most important things cannot be measured."

W. Edwards Deming (1900 – 1993)

Deciding what information will be useful can be difficult. As a rule of thumb, if you cannot think of a use for a set of data, there is probably no purpose in collecting it. There are many factors to consider. This section introduces some of the theory behind the selection of indicators.

Assessments need to be planned to ensure that the measurements are relevant and good enough for the purpose they are being collected for. Indicators need to be selected and measured in in a meaningful way – even if the assessment is only going to be rapid. Basing plans on the wrong data will mean that actions are also wasted.

It may be that the full set of indicators may not be obvious at the beginning of a project. This is particularly the case in 'process' work, where iterative processes are employed. At the outset of a process-led task, it may be very difficult, and undesirable, to state precise outputs. Instead, outputs and activities may be

various reasons, e.g. records may be incomplete. In some cases, sampling may destroy sampling units (e.g. removing water samples for water quality analysis).

In spite of careful sampling, it is very unlikely all the sample will be observed, depending on the method used. For instance, carrying out a survey with a 'captive' audience such as a class of schoolchildren, everyone there can be reached. A door-to-door survey will depend on whether people are at home then. For surveys that require an active response such as those sent by mail or email, rates of reply are likely to be lower than those carried out face to face. A response rate as low as 10% is not unusual in some cases. There are ways to improve response rates, however, expect the worst and allow for it in the total sample size.

Sampling methods

The sample may or may not represent the whole population, so there are various methods of trying to ensure that the response from the sample selected is as near as possible to the response from the whole population (if it was physically possible to collect data from everybody). There are two basic approaches to sampling – 'probability' sampling and 'non-probability' sampling.

Probability sampling is based on the idea that the chosen sample of respondents or study targets are expected to be a representative cross-section of the whole

population or range of possible study targets. As a result, the research findings are more likely to be generalizable to the wider population.

Non-probability sampling has no expectation that the selected respondents or study targets will be in any way representative of the whole, and findings are therefore unlikely to be applicable generally.

Where or who to measure?

When collecting data, a compromise has to be made between time and accuracy; digging a trial pit at the site of every pit latrine will tell you what ground conditions to expect, but will take a long time. Digging a few pits at certain locations will however tell you what could roughly be expected over a wider area. Asking one or two people what sort of latrine they would like hopefully will give a rough indication of what the whole community would like; asking more people should give a clearer, more representative idea of the whole population's views.

The population may vary spatially (from place to place) or socially (from person to person). The sampling strategy needs to take this variation into account.

Types of probability sampling

These are illustrated in Figure 4.

Simple random sampling. Here any member of the population could be selected, independently of other

Consider some surveys of a human population:

- A sample size of 'one' is unrepresentative, because the views of only one person are obtained and other people are likely to have different opinions. The larger the sample size, the more likely it is that opinions will be representative of the whole population.

- A sample of people in the streets if a town may be unrepresentative, because it will only include people who are able and who choose to be in town (and compare the sample collected during a quiet time during the working week with a peak shopping time, late at night or on a national holiday).

- A telephone sample may be unrepresentative because people who have a telephone many not be representative of the whole population.

- A sample of people from within one age group (e.g. students) may be unrepresentative because people of different ages may hold different opinions.

Sample selection has to consider:

- the size of the sample;

- the location of the sample; and

- the timing and frequency of observations

Samples may be split into groups, for example comparing people or places that have had a certain intervention (such as training or provided with water) with a similar sample that has not experienced this. This second group is called a control group.

How many to measure

In theory, the larger the sample, the more representative of the whole population the data will be. Statistics such as the average and standard deviation of a sample should approach that of the whole population as the sample size increases. Good sampling means that this convergence between the data from the sample and the actual data from the population becomes similar with only a small sample size. Poor sampling requires larger sample sizes to obtain good data.

It is difficult to give guidance on numbers of units of assessment (be they people, water samples etc.) required. As an indication, Brown and Edmunds (2011), looking at impacts of teaching methods on students, suggest that 30 per subgroup is the minimum number for a small-scale survey that allows statistical analysis of similarities and differences. For more complex forms of analysis (e.g. cluster analysis or multiple regression) a larger sample is needed and should be four times the number of items on the questionnaire.

Response rates

Availability may be an issue; it may not be feasible to obtain all possible data for

devised for the first stage or year; then later outputs and activities are defined based on lesson-learning.

Standard indicators

Before selecting indicators, review what is already being used. Besides saving time and money, these indicators will have a track record, enabling different projects and places to be compared and developments over time to be assessed using a common measure. Alegre et al (2006) looks specifically at Water Utility Performance Indicators.

Criteria for a good indicator

"It is much easier to make measurements than to know exactly what you are measuring"

J.W.N. Sullivan (1886 – 1937)

There are many factors that could be taken into consideration. At it is simplest an indicator should reflect Quantity, Quality and Time (QQT). Other guidance is summarized by the mnemonics SMART (Table 2) and SPICED (Table 3). SMART indicators tend to be quantitative and SPICED, qualitative. The letters can be used for more than one issue, for example, 'R' in SMART can be 'Realistic', but that is covered by 'Achievable'. The 'I' in SPICED can be 'Indirect' rather than 'Interpretable'.

The SMART and SPICED gives some qualitative guidance, but there are also other factors that are needed to inform the choice.

Table 2. SMART indicators

Specific	What things does the project intend to change?
Measureable	Can the indicator be measured independently and objectively?
Attainable/ achievable	Is it possible for the project to attain the indicator?
Relevant	Is the indicator relevant to the project and practical/ cost-effective to use?
Time bound	When should the indictor be achieved by?

Adapted from Cain, 2003

- 'Does the indicator have a clearly stated title and definition?

- Does the indicator have a clearly stated purpose and rationale?

- Is the method of measurement for the indicator clearly defined, including the description of the numerator, denominator and calculation, where applicable?

- Are the data collection methodology and data collection tools for the indicator data clearly stated?

- Is the data collection frequency clearly defined?

- Is any relevant data disaggregation clearly defined?

- Are there guidelines to interpret and use data from this indicator?

- What are the strengths and weaknesses of the indicator and the challenges in its use?

- Are relevant sources of additional information on the indicator cited?

- Do indicators have a proven track record – i.e. demonstrated performance in field-testing or operational use – before they are broadly deployed.'

(UNAIDS, 2010)

How many to measure?

Generally, it is best to avoid collecting huge amounts of information as it may not be collected accurately, people may resent having to provide it and it may take too long to analyse. A small assessment system that works is much better than one that is large and unwieldy.

Measureable indicators

It is generally easier to measure behaviour than feelings; behaviour can be observed. So if an objective is 'to increase people's confidence in meetings', it may be appropriate to measure this by observing how often they speak and whether they speak clearly.

DFID, 2003

Table 3. SPICED indicators

Subjective	Informants may have unique insights which give reliable information which is valuable but anecdotal
Participatory	Indicators should be developed together with those best placed to asses them
Interpreted and communicable	Indicators defined by local groups may need to be explained to external audiences
Cross-checked	Check information by comparing different indicators or progress and using different informants and methods
Empowering	The process of setting and using indicators should be empowering by helping groups and individuals reflect on their changing situation
Diverse	Using indicators set by different groups (e.g. men and women); information gathered should reflect these different perspectives.

Adapted from Cain, 2003

Population and samples

One of the important statistical concepts is the idea of the population, which is the total number of observations that could be made. Sometimes observations are made for the whole population – for example in a national census or elections when the statistical 'population' is … the population. However this is expensive and often there is only the time and resources to look at a proportion of the population. This is called a sample and is made up of a certain number of observations – the sample size.

Care needs to be taken when selecting a sample to try to ensure that it is representative of the whole population. The types and numbers of observations included in the sample must be large enough to make it likely that the range of results will be representative of the whole population.

Selection criteria – South Africa State of the Environment Report

1. The indicator must be based on good quality data that are available at a reasonable cost.

2. The indicator should provide information that measures something that is important to decision makers.

3. The information can be presented in a way that is easily understood and appealing to the target audience.

4. The indicator must relate to goals, targets or objectives.

5. The indicator must provide timely information (to allow for response).

6. The indicator must be able to detect small changes in the system.

7. The indicator must be relevant to policy and management needs within the [relevant] context. The indicator must therefore be associated with one or several … policy issues.

8. The indicator must be based on data that are accurate, reliable, statistically sound and scientifically valid. Metadata should define the quality of the data in the data set and include information on sensitivity, uncertainty, variability, precision, accuracy and error.

9. The data must be available and accessible, particularly in the long term.

10. The indicator must be based on data of the correct spatial and temporal extent. Sufficient historical data must be available to identify trends over time.

11. The data collection process should have minimal environmental impact.

DEAT, 2001