
Signs to look for:

Criteria for developing and selecting fit for purpose indicators

*Our Land and Water
National Science
Challenge*

October 2017



Private & Confidential

Dr Richard McDowell
Chief Scientist
Our Land and Water National Science Challenge
AgResearch – Lincoln Science Centre
Private Bag 4749
Christchurch 8140.

28 October 2017

Indicator thinkpiece

Dear Richard

As per our contract received 4 July 2017, please find attached our final report on the qualities of fit-for-purpose indicators. This report is a think-piece that presents criteria to guide future work on indicator development. Please note that this report should be read in conjunction with the restrictions at the end of this document.

Yours sincerely

A handwritten signature in black ink, appearing to read 'W. Kaye-Blake', with a long horizontal flourish extending to the right.

Dr Bill Kaye-Blake
Director
Consulting
E: bill.h.kaye-blake@nz.pwc.com
T: 04 462 7033

Table of contents

| | |
|--|----|
| Introduction | 2 |
| Defining indicators | 3 |
| What makes an indicator fit-for-purpose? | 6 |
| Discussion | 14 |
| Conclusion | 17 |
| Appendix A – Full list of qualities of good indicators | 18 |
| References | 19 |
| Restrictions | 21 |

Introduction

The Our Land and Water National Science Challenge (the Challenge) is looking to make indicators currently collected for agricultural productivity and environmental impacts more useful for more stakeholders in the agriculture sector. A large amount of research has gone into developing new metrics and indicators, but very little of this has been focused on defining what indicators are, and there are a wide range of views in the literature on the key criteria for fit-for-purpose indicators. This guide categorises and selects the most important and useful of these criteria and examines the trade-offs between them.

The Challenge has two key goals which are to:

- enhance primary sector production and productivity
- maintain and improve our land and water quality for future generations.

PwC on behalf of the Challenge has convened a group of interested parties from central government, the science sector and agricultural industries to cooperate on the good use of indicators. This indicators working group supports the use and development of agri-environmental indicators by:

- coordinating work programmes on indicators and data
- disseminating results
- driving specific projects on the use of indicators in agricultural practices.

The indicators working group is working alongside the Challenge to ensure indicators chosen and collected support the Challenge's two key goals.

Addressing the Challenge's goals require action across many disciplines, including economics, environmental science, agronomics, ecology and sociology. These multiple disciplines have to share information in ways that are understandable to non-experts. That will mean using indicators to summarise and communicate information. To that end, we have developed a definition of what an indicator is and what qualities make for the best and most useful indicators. This definition aims to cross multiple disciplines and can be applied across the Challenge's work.

Policymakers may find this useful to guide decisions on which indicators they should use to measure policy effectiveness. Researchers can use this definition and associate criteria to ensure that the data they collect is relevant to policymakers and land managers.

Defining indicators

The development of indicators to support policy has gained popularity across many disciplines over recent years, particularly in sustainability and sustainable development (Ramos and Caeiro 2010). While different disciplines use different indicators, much of the literature on good indicators skips over defining what an indicator is. “I shall not attempt to define [it]... but I know it when I see it” is the famous quote from Justice Stewart referring to definitions of pornography.¹ Like Justice Stewart, much of the scientific literature on indicators skips straight to the qualities that make a good indicator in particular fields and implicitly assumes that the definition of an indicator is something that everyone knows. Medical indicators, environmental indicators, economic indicators – these are described in depth by how well they relate to the underlying phenomenon they are trying to reveal.

By putting a definition around what we say an indicator is, we establish a reference point that can be used in discussions for a shared and commonly understood and used definition. That allows us to bring clarity of meaning to what can be an otherwise fuzzily defined conversation and focus discussion on the merits of different indicators.

Here we provide and justify a definition of an indicator that is sufficiently generic it can be applied to multiple fields.

Definitions from the literature

We define an indicator as a relevant variable, measured over time and/or space that provides information on a larger phenomenon of interest and allows comparisons to be made.

In developing this definition, we reviewed several competing definitions. Here we discuss key elements of different definitions that influenced our thinking.

Sustainable Cities International (SCI) (2012), provides a definition: “An indicator provides information on the state or condition of something.” Similarly, Hammond (1995) defines an indicator as “something that provides a clue to a matter of larger significance or makes perceptible a trend or phenomenon that is not immediately detectable (a drop in barometric pressure, for example, may signal a coming storm). Thus an indicator's significance extends beyond what is actually measured to a larger phenomenon of interest.” (Hammond, et al. 1995).

An indicator is a relevant variable, measurable over time and/or space that provides information on a larger phenomenon of interest and allows comparisons to be made.

Hammond goes on to say that indicators have two defining characteristics:

- Indicators **quantify** information so its significance is more readily apparent.
- Indicators **simplify** information about complex phenomena to improve communication.

Astleithner et al (2004) defines an indicator as a “policy-relevant variable defined in such a way as to be measurable over time and/or space.”

These definitions have largely focused on what indicators *do* rather than what they are. “Indications indicate...” as Simon & Garfunkel sang (1968). Astleithner et al (2004) referred to indicators as being policy-relevant variables, and to us this categorisation usefully pins down indicators to a type of information. Other definitions (“something that provides a clue to a matter of larger significance”) focus on what it does, but are vague as to

¹ *Jacobellis v Ohio*. 378 U.S. 184 (1964). Justice Stewart.

what an indicator actually is. ‘Relevant variable’ describes a type of information that can be an indicator, and we use this to pin down what it is that is providing an indication of something.

Key elements of different definitions

From these definitions there are some common elements and some differences. Each of the definitions has referred to indicators as being something:

1. that is measured over time and/or space
2. that provides information on a larger phenomenon of interest

To these elements we have added two additional elements, that indicators

3. are relevant variables
4. allow comparisons to be made.

The first element, that indicators are measured over time and/or space is not in dispute. However the word ‘measured’ here is used broadly to include other ways of gathering data. Traditionally, measured data refers to continuous variables where an infinite number of values are possible, even within a bounded range. Here our definition includes other kinds of quantitative and qualitative information including count data, rank data, and nominal data (qualitative).

Quantification is one element where experts have mixed views on whether it is a defining requirement of indicators. Hammond says “... indicators provide information in more quantitative form than words or picture alone; they imply a metric against which some aspects of public policy issues, such as policy performance, can be measured.” Hammond goes on to include quantification as one of two defining elements of indicators.

Astleithner on the other hand says that “[Indicators] need not be quantified; measurement can be on the basis of qualitative scales. But the key feature of an indicator is that measurement can take place and this, in turn, allows comparison.”

The extent to which quantification is necessary will vary somewhat by discipline. Some disciplines rely mostly on measured or counted information. For other indicators and disciplines ranked (ordinal) information is used. Other indicators use less quantitative and more qualitative information. For example, the colour of coral on a reef is a measurable, policy relevant variable that provides information about a wider phenomenon (ocean acidification). Colour is usually a qualitative categorisation or nominal variable, but it can be quantified.² As the level of quantification is so variable across different indicators, we do not include quantification in our definition of an indicator.

The importance of the second element – that indicators provide information on a **wider phenomenon of interest** – should not be understated. This goes to one of the key elements of indicators, that they are not collected or measured for their own sake. Indicators can only ever tell part of the story. They inform users about the state and trend of different phenomenon, but they are not the principal matter of interest. Indicators are a tool for providing information on other complex phenomena, and it should be remembered that they are just a tool and not an end in and of themselves. Particularly when indicators are used to set objectives and targets, it should be clear that the indicator is merely a way to measure the ultimate goal and is not a goal itself.

Indicators necessarily simplify complex information in order to provide insight, and to be comprehensible to different audiences. “Indicators are inevitable approximations. They are not the same as the desired change, but only an indicator of that change. They are imperfect and vary in validity and reliability” (Patton 1996). Indicators however are only necessary when the change to be measured is abstract. “Where the desired change

² Colours can be quantified using different levels of primary shades – either additive colour model such as Red/Green/Blue or a subtractive model such as Cyan/Magenta/Yellow/Key.

is concrete, tangible, and measurable, indicators are not needed” (Church and Rogers 2011). This simplification is part of how indicators provide information on a phenomenon of interest.

Most definitions of indicators emphasise the effects and uses of indicators while avoiding types and categories of indicators. We define indicators as **relevant variables** to provide some precision of thinking about what it is that is being used or influencing change.

Accordingly, the first part of our definition is that an indicator is “a relevant variable that can be measured over time and/or space and provides information on a larger phenomenon of interest.”

We argue that **comparison** is also a key element of indicators. Indicators are used to make comparisons either across time, space or to a baseline or benchmark. While this will often involve quantified information, quantification is not necessary to make comparisons. For example, the colour and shade of coral reefs can be compared across space and time without needing to be quantified, and this provides information on a wider phenomenon of interest (ocean acidification). Comparison is an essential feature of the purpose of indicators. Indicators cannot provide useful information without comparison. This is implicit in Astleithner’s definition that an indicator is measurable over time and/or space. For the purposes of making a clear definition we are including the ability to make comparisons as an explicit element of an indicator.

Our definition of an indicator then is:

- a relevant variable
- that is measurable
 - over time and/or space
- that provides information
 - on larger phenomena of interest
- and allows comparisons to be made.

For the Challenge, the larger phenomena of interest relate back to the Challenge’s economic and environmental goals. Accordingly its indicators will be variables that provide information on the economic value and productivity of agriculture in New Zealand, and variables that illustrate the impacts of agricultural activity on the environment.

In the next section we discuss the six criteria that make indicators most useful.

What makes an indicator fit-for-purpose?

We selected six criteria that are the most important for ensuring indicators are useful. Fit-for-purpose indicators are:

- accepted by stakeholders
- valid
- clearly defined and standardised
- based on available or easily accessible data
- easily communicable
- performance-based.

These criteria were selected based on existing research and the aims of the Challenge to facilitate changes in land-use practices to improve both the value derived from agriculture and the environmental effects of agriculture. Most of the criteria we reviewed related to one of two broad themes:

- Technical validity – eg precision, performance based, standardised
- Usability – eg accessibility of data, relevance, easily understood.

Different agencies will draw a different balance between the relative importance of these two broad themes. For the purposes of the Challenge, both technical validity and usability are important. The value of agricultural production can be easily measured at a variety of scales and there are many useful economic indicators that can be used to inform how different aspects of agricultural production are changing over time. The impacts of agriculture on land and water quality however are generally localised, and the trade-offs between the validity of indicators and their practical usefulness are more difficult. Just as indicators that can be easily applied but are not technically valid are not useful, so too indicators that are absolutely technically correct, but impractical to apply are also not useful. Indicators are only useful in as much as they can be used to effect change in the world.

For these reasons we selected a balance of criteria from each theme, three relating to technical validity (valid, clearly defined and standardised, performance based) and three relating to their usability (accepted by stakeholders, based on accessible data, easily communicable).

The six criteria are illustrated in Figure 1 below, and this paper discusses the reasons why each criteria is important, and some of the trade-offs involved between them.

A fit-for-purpose indicator should be widely accepted by a range of major stakeholders

Why is stakeholder acceptance important?

The ultimate purpose of any indicator is to facilitate a change in the world. To do this indicators need to be able to influence decision-makers. If key stakeholders and decision-makers do not accept the validity of an indicator, it won't be useful for influencing people to make change. For this reason, the strength of an indicator depends on its having the approval of stakeholders (Moller & MacLeod, 2013). Indicators that are accepted or endorsed by a range of interested parties are likely to be seen as more legitimate and be used in practice (UNECE/OECD/Eurostat Working Group, 2008). By seeking input from stakeholders early and often, it is much more likely that the resulting indicators will not be disputed later once they are being collected. It ultimately ensures that indicators are useful and used for a purpose, and not simply collected for the sake of measurement.

An additional benefit of consulting stakeholders is to get input and opinions from a wide range of groups with different interests and backgrounds. This will ensure indicators reflect a diverse set of values, and are less likely to omit significant aspects or facets of the issue at hand (United Nations, 2007).

Stakeholder consultation can help to ensure that indicators are practical and sensibly defined. In particular, involving early those who will actually collect data and use the indicator can provide a useful reality check for whether it will be feasible to get data and how the indicator will be applied. It also helps to ensure that the indicator is relevant to its intended purpose.

The ultimate purpose of any indicator is to facilitate a change in the world. To do this indicators need to be able to influence decision-makers.

Involving stakeholders early also helps to ensure their accountability. If stakeholders such as policy-makers and businesses are involved in the development process they are more likely to be held accountable for implementing and using the indicator. Especially for the purposes of public reporting, stakeholder involvement helps to ensure trust and legitimacy. The public is much more likely to view an indicator as legitimate if it has been accepted and approved by experts and other stakeholders (Brown, 2009).

Who should be consulted?

Which stakeholders should be involved depends greatly on the purpose and audience for which the indicator is intended. There are some groups who are likely to be involved more often than others. Indicators are typically used to aid in policy decisions. Therefore it is sensible that policy-makers, from both local and central government, should have a role in the selection or creation of indicators.

Indicators will impact a range of different organisations throughout society. Businesses and non-government organisations are likely to use or be affected by the flow-on effect from the use of indicators. It is crucial to consider the impacts on businesses, and the economy more generally, when making any kind of policy decision. Accordingly it is important that industry groups and business leaders are involved in the process of selecting and developing indicators (Brown, 2009).

Indicators are always used in a specific context. Whether they are economic indicators, health indicators or sustainability indicators, indicators are used to convey the state of certain phenomena of interest. Many indicators are used only by those in their chosen fields, while others are used to inform a wider audience or group of stakeholders. Regardless of the purpose or audience, it is important that indicators are developed with consultation from experts in the specified field (United Nations, 2007). Relevant experts could be academics, scientists, or any person working in the field of interest.

Another important group of stakeholders is the general public. The ultimate purpose of many indicators is to communicate information to the public (United Nations, 2007). The primary reason for consulting the public is to ensure that the indicators selected relate to things that the public care about and that they can be understood by the public (de Olde, et al., 2016; Moller & MacLeod, 2013). If indicators intended to inform the public are not relevant or comprehensible, then resources invested in them may be wasted.

When and how should stakeholders be consulted?

It is best to involve stakeholders throughout the entire process of selecting and developing indicators right through to implementing and using them in practice. It is much more likely that an indicator is accepted by stakeholders if they are involved in its selection or creation (United Nations, 2007).

Participation of major stakeholders in the development and design of indicator sets and frameworks for informing policy will allow indicators to be more trusted and practical from the outset. A diverse set of perspectives throughout the process will ensure that indicators remain practical and useful. It is also more likely to ensure that indicators present an impartial view (Brown, 2009).

Figure 1: The six criteria for good indicators



There are two main approaches to the collaborative development of indicator sets and frameworks: top-down and bottom-up (Sustainable Cities International, 2012). Top-down is less inclusive, and relies on experts and policy-makers to define goals and collect data for indicators. The top-down approach can create highly technical indicators that require experts to correctly interpret and communicate. The bottom-up approach involves extensive collaboration and consultation with the wider community. More viewpoints are taken into account in the bottom-up approach, but indicators may end up being basic and broad.

There are advantages and disadvantages to both of these approaches. The top-down approach does not include as wide a range of viewpoints and may lead to indicators that are less meaningful to the general public. In contrast, the bottom-up approach may be inefficient in its use of resources, and ultimately may create indicators that are overly simplistic and ill-suited to measuring the underlying phenomenon. The purpose and intended audience of the indicators will determine which of these approaches is best.

Consensus is not necessary

While it is important to get input from a range of stakeholders when developing indicators, it can also make the process run less smoothly. Involving stakeholders is not always easy, especially when it involves public consultation. The more groups that are involved, the more resources are necessary to get their input.

Another trade-off to involving stakeholders is that not all viewpoints will be compatible with each other. Some groups may never agree with otherwise suitable indicators that paint them in a bad light. Ultimately, it will be difficult to satisfy all stakeholders with a single indicator or set of indicators. It is not necessary to get full consensus from different stakeholders on the indicator or set of indicators. As long as different stakeholders have been involved and feel heard, compromises can often be reached. The approach of using a portfolio of indicators is one solution to this issue. Weaknesses in one indicator can be made up for by strengths in other indicators. Where stakeholders cannot agree on one indicator, their objections may be reduced if the portfolio includes other indicators they do support.

A fit-for-purpose indicator must be valid

An indicator is valid where it adequately reflects the phenomenon it is intended to measure and is appropriate to the needs of the user (Brown, 2009). Indicators should be as relevant and practical as possible. Where indicators are used for policy-making, it is also important that they provide information at an appropriate level for those decisions to be made.

It is important that indicators are valid as they need to provide useful and trusted assessments of the phenomena they relate to. It ensures that they are suitable for use for their intended purpose.

For an indicator to be useful, it must have a meaningful relationship to the underlying phenomenon it is purporting to reveal. Informative relationships are often not clear-cut and it can be difficult when there is dispute over the direction of causation, or over the strength of a relationship between the proposed indicator and the underlying phenomenon to be measured. There needs to be a clear link between the indicator and the phenomenon it measures that is based on robust theory or scientific evidence (Brown, 2009). A valid and meaningful indicator should be analytically sound, using appropriate data that has been collected using acceptable methods. They will also be coherent over time, and sensitive to changes in the underlying phenomenon (OECD, 2003).

A fit-for-purpose indicator must be clearly defined and standardised

Indicators simplify complex phenomena to inform us about how they are changing. Clear definition and standardisation allows indicators to be used to observe trends and make comparisons.

Why do indicators need to be clearly defined and standardised?

Clear definition and standardisation is important as it helps to ensure that data is valid and usable for robust analysis. If indicators are not well defined or data is not collected appropriately, then it is likely that the

indicator will not be useful for its intended purpose (OECD, 2003). However, perfect definition and standardisation is rarely, if ever, possible. Usually some compromise between standardisation and cost will need to be made to ensure data can be collected and indicators can be used.

One of the primary purposes of indicators is to use data in a way that enables us to observe trends over time or to compare between different regions or countries. If different definitions or data collection methods are used, then indicators may not be able to be used for making comparisons. It is therefore important that wherever possible, standard definition and methods of data collection are used to ensure coherence and consistency (Statistics NZ, 2017). Consistency across time and jurisdictions allows for meaningful comparison that accurately show how the underlying phenomenon is or is not changing.

Indicators, and the phenomena they are intended to measure, should be clearly defined before being used for reporting or decision-making. Where possible given other constraints, it is important that they are defined as accurately and narrowly as possible so as to be unambiguous. It is important to make it clear exactly what the indicator is measuring. For example, the percentage of people living in poverty is often used as an indicator of economic health of a city or region. This is not meaningful unless there is a clear definition of what living in poverty is. Different countries use different thresholds for considering poverty, so an indicator is only able to be compared if it states the definition of poverty it uses.

In order for an indicator to meet this criteria, it must be based on verifiable and scientifically acceptable data that is collected using accurate and validated methods. However the level of precision necessary will depend on the phenomenon of interest and the underlying aims of goals. It is not always possible for indicators to be defined and standardised as well as we would like them to be. Limitations associated with data availability and collection technology mean that it is not realistic or practical for these methods to be used at all times (OECD, 2003). Budgetary constraints mean that gold-standard data collection is often implausible and compromises must be made.

Comparability vs locale richness

Where there are internationally accepted data and methods of collection it is important that they are adhered to (Moller & MacLeod, 2013). However different jurisdictions may use different benchmarks and have different priorities for the collection and measurement of data. As a result, international standards may be less important than being locally relevant (Statistics NZ, 2017). There may be some indicators that are only used in New Zealand, or that are focused on in New Zealand more than in other countries. Where that is the case, it may be important to rely on local knowledge and expertise rather than aligning with international data.

There is often a trade-off between indicators being comparable between regions or jurisdictions and indicators providing a richness of locally specific information. Using a standardised approach may conflict with a desire to collect information on a locally-specific aspect of the phenomenon of interest that isn't relevant outside a local region.

One example of this has been seen in water quality monitoring. Regional councils are charged with monitoring water quality within their region to promote the environmental, economic, cultural, and social wellbeing of people in their region. Many local authorities measured water quality at sites where there were known to be problems. By focussing on problem areas, the councils could use up-to-date information to advise people about when they were safe for swimming, fishing and other recreation uses. Sites that were known to have excellent water quality were monitored less often, or not at all as they did not pose a risk to human health. This meant that when the Ministry for the Environment attempted to describe a national picture of New Zealand's water quality, the available data was not suitable as it was collected for a different purpose. The data suggested that more sites were of poor quality than actually was the case because mostly poor quality sites had been selected for monitoring. The councils had selected sites for local concerns and the as a result the data was not sufficiently representative at a national level. This highlights the tensions involved in collecting indicator data for different purposes.

Indicators should make use of data that is either readily available or easy and affordable to collect

Indicators that are able to be measured easily and affordably are more likely to be used by a wider range of participants and to be monitored more regularly. The purpose of using indicators is to measure progress and make comparisons. Increased participation and measurement means indicators are more effective at keeping track of changes, which ultimately makes them more useful.

When developing frameworks and indicator sets, selection of indicators requires extensive consideration of data availability. Appropriate data is not always available for some indicators. Where it is not readily available, it is important that the time and costs of measuring or collecting the data are investigated before choosing indicators (Moller & MacLeod, 2013).

Keeping the accessibility of data in mind when designing and choosing indicators ensures that later monitoring can be done cost-effectively.

Cost and practicality are the key drivers of this feature. Designing indicators that are perfectly accurate, with excellent data will often be too costly to implement and actually collect data for. Where data is unavailable or particularly costly, it may be possible to replace some data with more accessible proxies. However in many of these cases, costs will ensure that only data for the most essential indicators will be collected. Keeping the availability and accessibility of data in mind when designing and choosing indicators means that many of these issues can be pre-empted and indicators can be chosen that can be cost-effectively monitored and relied upon.

Indicators should be meaningful, easily communicated and comprehensible

Indicators should be easy to communicate to those not directly involved and should be clear, simple and easy to interpret (de Olde, et al., 2016). Indicators that are easily and widely understood are likely to have increased interest from stakeholders and the public.

People are unlikely to be interested in an indicator that they do not understand. Most indicators benefit from high levels of interest from stakeholders and the public. For indicators to be relevant or meaningful to the public or to decision-makers, they have to be easily communicated.

Know your audience

For an indicator to be meaningful, it needs to be appropriate to the needs of the user or audience. It is important to understand who the audience and end-users of an indicator are in order to determine how simple it should be. If the audience is made up of experts then it may need greater technical complexity to retain its usefulness. If the audience includes the public then it should be simplified to the extent necessary to ensure it is understood and holds public interest. If it is for making policy decisions, then it must provide enough information at an appropriate level for sound decisions to be made (Moller & MacLeod, 2013).

One key function of indicators is to simplify communication of complex data and inform the public (OECD, 2003; United Nations, 2007). Where this is one of an indicator's purposes it is important that it presents this information in a way that is easy to understand and use. Decision-makers need to be able to use indicators to inform their decisions. Further, not all policy-makers will be experts in the technical aspects of indicators, so it is important that indicators convey information clearly and simply.

Simplifying indicators allows them to be more easily understood by more people. That may not be a good thing, depending on the intended use and audience of the indicator. Some indicators are specifically used by experts, and require the use of highly technical or scientific data. Oversimplifying the data may lose some of the accuracy and validity necessary to make that data useful (Sustainable Cities International, 2012).

Context is key

Contextual information is key to making indicators easily understood. Putting indicator values in the context of benchmarks or quality ranges enables them to be understood by a wider audience without losing any of their validity and meaning (Brown, 2009). For example, a measurement of *E. coli* concentration of 37 cfu per 100 ml of river water is not that useful absent the ranges of how that value should be interpreted. When that value is put in the context that the target value for an A grade is a level below 130 cfu per 100 ml, the value may be interpreted as showing good water quality.

Ease of access to supplementary information and metadata is important for ensuring the interpretability of data (Statistics NZ, 2017). Quantitative indicators give highly precise information, but it is not always clear what the number is referring to or whether that number is considered good or bad. Qualitative indicators help to interpret the data as instead of giving a number they may provide a rating from low to high, or good to bad. While this loses some precision, it makes the indicator clear and simple to understand for audiences without needing further information.

Where a concept is particularly abstract, the concept can sometimes be defined in reference to the indicators used to measure it. For example, 'Swimmability' may be defined in reference to a certain threshold of bacteria in waterways. In these cases, particular care must be taken in choosing the indicators as they go beyond measurement and indication to defining potentially contentious terms.

Most indicators need some form of comparison to make the values useful and give them context. That comparison can be to another time period, another place, or to a benchmark, baseline or target. This makes the indicator more meaningful and relevant to most people than simply its value (de Olde, et al., 2016).

For indicators used in policy-making, it may be necessary to provide information at multiple scales of detail (Sustainable Cities International, 2012). Detailed quantitative data may be used by experts, such as scientists or academics, to precisely measure states or changes. The same indicators at a more contextual scale using qualitative range (such as satisfactory to unsatisfactory) may be used to report the results to the public, and can be simplified to communicate to a wider audience.

Indicators should be performance-based rather than practice-based where possible

A performance-based indicator measures the performance towards target outcomes (OECD, 2003). Performance-based indicators give relevant information about any actual changes and progress that may be occurring. Practice-based indicators, on the other hand, inform how well certain practices are being adopted or adhered to. For example, kilometres of waterways fenced is a practice-based indicator for looking at water quality. Water clarity and total nitrogen levels are performance-based indicators that measure the quality of a waterway more directly.

Both practice-based and performance-based indicators are one step removed from the phenomenon of interest. Depending on the situation one or the other may be appropriate. Both practice-based and performance-based require a suitable link with outcomes to be shown or assumed. For practice-based indicators the link to be shown is the causative effect. If these practices are done to a certain level, this will cause a certain outcome. If it is the practices that are measured, the causative link to the outcomes of interest must be well-established.

For performance-based indicators, the link is that they show or illustrate a larger phenomenon. Here the causative link runs in the opposite direction. The outcomes of interest cause the indicator to vary in predictable ways. Where the outcome cannot be easily measured directly, performance-based indicators can provide good information on the underlying outcomes of interest. However, the causative link between the outcome and the indicator must also be well-established and sufficiently direct that it does not vary due to other influences.

Where an issue is affected by multiple causes practice-based indicators are less-useful as they may only include one cause of many. For environmental issues, performance-based indicators will often be more useful, as it will often be easier to establish a direct link between the phenomenon of interest and indicators that show the state or trend under multiple different pressures.

It is important that indicators measure outcomes as directly as possible as it helps to inform whether current practices do actually promote the outcome for which they are intended. While adhering to good-practice is desirable, it can lead to a false sense of progress if the link between activity and outcome is not well established. This is a particular risk where practice-based indicators are used because performance-based indicators are unavailable or not practical. In these cases authorities measuring and reporting indicators need to be especially clear on the limits of what is being measured.

Practice-based indicators can sometimes be more cost-effective for looking at system-wide performance. While performance-based indicators are better measures of outcomes, it is also important to have an awareness of the factors that influence outcomes (Brown, 2009). As long as practice-based indicators are not over-interpreted, they can be useful alongside performance indicators for evaluating the effectiveness of different interventions and activities.

Discussion

The key criteria we have selected for choosing fit-for-purpose indicators are not always easily compatible, and show that there are trade-offs involved in indicator selection. The criteria we have developed are not the only set of criteria for indicators. In particular Statistics New Zealand have their own criteria for their environmental reporting programme. The two sets of criteria are broadly similar but not entirely the same. Here we discuss the trade-offs between different criteria, and the comparison with Statistics NZ's criteria set for environmental indicators.

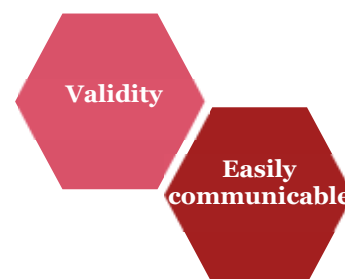
Trade-offs between criteria

We see that there are four key trade-offs within our key criteria:

- Validity vs Easily Communicable
- Validity vs Accessible data
- Validity vs Performance-based

Validity vs Ease of communication

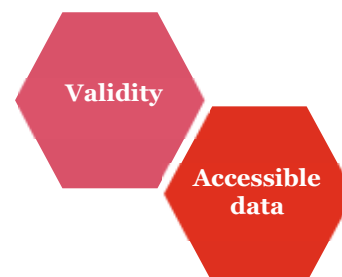
That an indicator must be valid is not in dispute, but validity is not a binary property but a matter of degree. Because indicators are necessarily one step removed from the underlying property they inform about, they are never perfectly precise and valid. There is an inherent trade-off between the validity and the simplicity of indicators. Indicators necessarily simplify complex or abstract phenomena in order to provide useful information. However, environmental systems are inherently complex and interrelated and some information is inevitably lost when an indicator simplifies those systems.



The exact balance to be struck between simplicity and validity will vary from case to case, but is important to be aware of. Greater simplicity can make indicators more practical and understandable, but at the expense of richer information about the phenomena of interest. However, with greater meaning comes greater complexity, and for particularly abstract concepts, greater levels of information are not always more useful.

Validity vs Accessible data

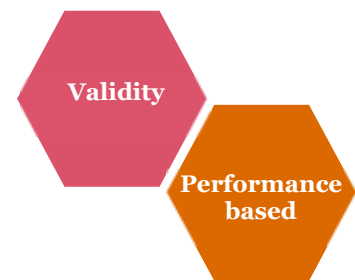
Another trade-off that must be made with an indicator's validity is with the data that is available and accessible. An excellent indicator may exist that perfectly responds with changes in the underlying phenomena and meets the other criteria. However, very often data does not exist, or it may be too expensive to collect. In these cases a compromise must be made and a less suitable indicator must be used instead.



For most purposes a less than perfect indicator about which data is collected and available is better than a perfect indicator that is never more than an idea to be discussed. To make a difference, indicators must be used and when defining and setting indicators the real-world constraints must be acknowledged. Very often, the next best alternative is sufficiently suitable and the level of precision and accuracy gained from moving from sufficient to perfect is not as great as might be first thought.

Validity vs Performance-based

For some environmental and agricultural issues there is a trade-off between indicators being performance based and being valid and useful. This can be because environmental systems can take a long time to show changes. Where environmental effects operate over decades or longer, the effect of policy changes will take a long-time to manifest in performance-based indicators. For example, world average temperatures can be used as an indicator for progress on climate change. However, greenhouse emissions of carbon dioxide from the last 60 years will continue to warm the planet for many more decades to come, and the effects of policy changes today will be difficult to perceive using temperature as an indicator. Instead, a practice-based indicator of the amount of greenhouse emissions per year is used to see if governments and businesses are making progress. For cases such as these performance-based indicators will not be suitable and practice-based indicators will be more appropriate.



Portfolio approach

One way of managing trade-offs between indicators is to use a portfolio of indicators, rather than rely on a single indicator. By using multiple indicators to look at a phenomenon from different angles, weaknesses in one indicator can be made up for in the strengths of another. It will be difficult to find a perfect indicator that meets all six criteria, and by using multiple indicators, this perfection becomes unnecessary.

Other criteria sets

Statistics New Zealand

The criteria we have chosen are not the only criteria available. Statistics New Zealand has a list of six criteria for data used in environmental reporting. Those criteria are:

- relevance
- accuracy
- timeliness
- accessibility
- coherence
- interpretability.

These are largely compatible with our list as shown in Table 1. Two of the major differences are our inclusion of stakeholder acceptance, and Statistics New Zealand's inclusion of timeliness. Neither list is intended to be exhaustive, and of course there are many other characteristics that can make an indicator useful. The full list of useful characteristics for indicators found in the literature is included in Appendix A.

Table 1: Criteria sets for good indicators

| Our criteria | Statistics New Zealand Criteria |
|----------------------------------|---------------------------------|
| Clearly defined and standardised | Coherence |
| Available or accessible data | Accessible |
| Performance based | Relevance |
| Easily communicated | Interpretability |
| Valid | Accuracy |
| Accepted by stakeholders | Timeliness |

Timeliness refers to the length of time between data collection and publication. Timeliness is a sensible criteria to apply in the national framework where different gradients need to be distinguished between types of datasets

and indicators based on the quality and rigour of the underlying data. Statistics New Zealand, as the country's official statistics agency rightly emphasises rigour and validity highly.

For the purposes of Challenge however, timeliness is an important criteria, but not one of the most important. This is because the Challenge places a greater emphasis on usefulness and application than Statistics New Zealand does. Rigour and accuracy are still very important (as reflected in the criteria Clearly defined and standardised, and Valid), but as the purpose and audiences of the Challenge's work are more narrowly defined, they place a greater emphasis on usefulness and usability.

In addition, we view Timeliness as part of a broader criteria already included – accessible data. One facet of data's accessibility is its timeliness, and data that is not timely is not accessible for the purpose of an indicator.

SMART

For objectives and goal setting, SMART is an acronym that represents five key criteria:

- Specific
- Measurable
- Achievable
- Relevant
- Time-bound.

SMART is widely used in management theory as criteria for good objectives. While these criteria are not specifically directed at indicators, there are broad parallels with the criteria we have chosen. Specific and Measurable are important elements of indicators being well-defined. Achievable relates more closely to objective setting than indicator development, but a good analogue for indicator development is the accessibility of data, which we have included as one of our criteria. Relevance is a part of an indicator's validity and its acceptance by stakeholders.

Time-bound is specific to objective setting. Goals need to have a specified target date for achieving them while indicators can inform at a point in time or over a period of time. The similar criteria for time-bound relating to indicators would be specificity. Just as goals need to be specific about when they are to be achieved, indicators need to be specific about what they measure, where they relate to and what period of time they cover.

Conclusion

We have developed here a general definition of what an indicator is. An indicator is a relevant variable that is measurable over time and/or space that provides information on larger phenomenon of interest and allows comparisons to be made.

There are a wide range of qualities that make for good indicators. For the purposes of the Challenge, fit-for-purpose indicators are:

- accepted by stakeholders
- valid
- clearly defined and standardised
- based on available or easily accessible data
- easily communicable
- performance-based.

There are inevitably trade-offs and compromises to be made among these criteria when selecting indicators. A portfolio approach using multiple indicators with strengths in different criteria is one approach to managing these compromises and trade-offs.

Appendix A – Full list of qualities of good indicators

The following list includes all qualities of good indicators found in the literature.

- Transparent
- Broadly accepted
- Simple
- Accurate
- Easy to interpret
- Quantifiable
- Meaningful
- Timely
- Standardised
- Specific
- Affordable measurement
- Accessible data
- Sensitive
- Performance-based
- Easily understood
- User-friendly
- Valid
- Relevant
- Precision
- Capacity to upscale
- Clearly defined.

References

- Astleithner, F., Hamedinger, A., Holman, N., & Rydin, Y. (2004). Institutions and indicators - the discourse about indicators in the context of sustainability. *Journal of Housing and the Built Environment*, 19, 7-24.
- Brown, D. (2009). *Good Practice Guidelines for Indicator Development and Reporting*. Statistics New Zealand.
- Church, C., & Rogers, M. (2011). *Designing for Results: Integrating Monitoring and Evaluation in Conflict Transformation Activities*. DME for peace. Retrieved from <http://dmeforpeace.org/learn/designing-results-integrating-monitoring-and-evaluation-conflict-transformation-activities>
- de Olde, E., Moller, H., Marchand, F., McDowell, R., MacLeod, C., Sautier, M., . . . Manhire, J. (2016). When experts disagree: the need to rethink indicator selection for assessing sustainability of agriculture. *Environmental Development Sustainability*.
- Hammond, A., Adriaanse, A., Rodenburg, E., Bryant, D., & Woodward, R. (1995). *Environmental Indicators: A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development*. World Resources Institute.
- Korzybski, A. (1933). *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*.
- Moller, H., & MacLeod, C. (2013). *Design Criteria for effective Assessment of Sustainability in New Zealand's Production Landscapes*. New Zealand Sustainability Dashboard.
- OECD. (2003). *OECD Environmental Indicators: Development, Measurement and Use*.
- Patton, M. (1996). *Utilization-Focused Evaluation* (Third edition ed.). SAGE Publications.
- Ramos, T. B., & Caeiro, S. (2010). Meta-performance evaluation of sustainability indicators. *Ecological Indicators*, 10, 157-166.
- Simon, P., & Garfunkel, A. (1968). You Don't Know Where Your Interest Lies. *Bookends*. Columbia Records.
- Statistics NZ. (2017, January). *Environmental Indicators Te taiao Aotearoa*. Retrieved from Statistics NZ: http://www.stats.govt.nz/browse_for_stats/environment/environmental-reporting-series/environmental-indicators/Home/About.aspx#topics
- Sustainable Cities International. (2012). *Indicators for Sustainability: How Cities are Monitoring and Evaluating their Success*.

Sustainable Cities International. (2012). *Indicators for Sustainability: How cities are monitoring and evaluating their success.*

UNECE/Eurostat/OECD Task Force. (2013). *Framework and Suggested INdicators to Measure Sustainable Development.*

UNECE/OECD/Eurostat Working Group. (2008). *Measuring Sustainable Development.* New York: United Nations.

United Nations. (2007). Indicators of Sustainable Development: Guidelines and Methodologies. *Economic and Social Affairs.*

Restrictions

This report has been prepared solely for the purposes stated herein and should not be relied upon for any other purpose. We accept no liability to any party should it be used for any purpose other than that for which it was prepared.

This report is strictly confidential and (save to the extent required by applicable law and/or regulation) must not be released to any third party without our express written consent which is at our sole discretion.

To the fullest extent permitted by law, PwC accepts no duty of care to any third party in connection with the provision of this report and/or any related information or explanation (together, the “Information”). Accordingly, regardless of the form of action, whether in contract, tort (including without limitation, negligence) or otherwise, and to the extent permitted by applicable law, PwC accepts no liability of any kind to any third party and disclaims all responsibility for the consequences of any third party acting or refraining to act in reliance on the Information.

We have not independently verified the accuracy of information provided to us, and have not conducted any form of audit in respect of the organisation for which work is completed. Accordingly, we express no opinion on the reliability, accuracy, or completeness of the information provided to us and upon which we have relied.

The statements and opinions expressed herein have been made in good faith, and on the basis that all information relied upon is true and accurate in all material respects, and not misleading by reason of omission or otherwise.

The statements and opinions expressed in this report are based on information available as at the date of the report.

We reserve the right, but will be under no obligation, to review or amend our report, if any additional information, which was in existence on the date of this report was not brought to our attention, or subsequently comes to light.

This report is issued pursuant to the terms and conditions set out in our contract with the Our Land and Water National Science Challenge (via AgResearch) received 4 July 2017.

