Summarisation of portfolio change

By Nigel Simister, January 2016

International NGOs (INGOs) are often expected to summarise results and learning across large portfolios of work, carried out in different regions, countries and sectors. This paper describes some of the different options available to help achieve this. It concludes that it is normally possible to build up a complex picture of change, and to identify the role an INGO plays in contributing to that change. However, it requires a lot of thought and planning, and may require considerable resources.

Introduction

Many organisations based in the North and South are expected to summarise results across an entire portfolio, or broad sections of it. This may be a relatively straightforward exercise, particularly for single-sector organisations working in only one or two locations. But for large, complex organisations, such as International NGOs (INGOs), working across many different sectors and countries, it can be very difficult.

Summarisation is often required to meet the needs of institutional donors providing core funding, who want to see the big picture rather than small snapshots of change in multiple projects. But many INGOs also wish to summarise global performance for their own reasons.

Up until now, little has been written on the subject, and the mechanisms for summarising change across broad portfolios of work have not been well explored. In the past there has often been an assumption that the kind of management frameworks that can work well at project level, such as the logical framework, can be applied at organisational level. But in many cases this is not true, and donors and INGOs need to think more creatively about how to summarise change.

This paper attempts to show different ways in which summarisation across complex portfolios of work can be achieved. It is divided into four main parts:

- Part one addresses some of the theory behind summarisation.
- Part two examines how indicators can be used to help summarise change across an organisation.
- Part three looks at other options that can be used, either as a supplement or an alternative to an indicator-based system.
- Part four discusses how summarisation can be applied across different levels.

It is important at the outset to address three issues regarding the scope of this paper. Firstly, this paper is about summarisation. However, the systems and processes needed to help summarise across a portfolio meet only a small part of the monitoring, evaluation and learning (MEL) needs of an INGO. Secondly, summarisation is by and large an accountability or communications issue. There may be learning associated with summarisation across a portfolio but this is not usually the dominant influence. INGOs will need to explore many other different avenues to identify, share and use learning. Thirdly, a

good MEL system implemented by an INGO allows considerable flexibility for different countries and partners to adopt their own MEL processes to serve their own needs in their own particular contexts. These wider issues are not covered by this paper, which focuses on just one aspect of INGO MEL systems – summarisation of change at portfolio level.

PART ONE: Issues in summarisation

This part of the paper addresses some of the theoretical issues that need to be considered when summarising results at portfolio level. For the remainder of this paper, and for want of a better term, results at portfolio level will be referred to throughout as "portfolio change".

1.1 Different terms and their meanings

Some of the different terms used in this paper are described below. These terms do not have standard definitions within monitoring and evaluation (M&E) circles.

- In this paper, **summarisation** of an INGO portfolio means presenting an overview of what has changed, what an INGO's contribution to that change has been, what has been learned, and/or what has changed as a result of that learning. It is a broad term used to cover a range of different concepts.

- On the other hand, **aggregation** has a more specific meaning. Aggregation means the addition of numeric information from different places within an organisation to reach an overall figure that describes the totality of change across the different locations.

- Within this paper, **measurement** of change is reserved for occasions when an objective is SMART (Specific, Measurable, Achievable, Relevant and Timebound) and can be measured with some degree of accuracy. In some circumstances unexpected changes can also be measured but this is rare.

- In reality, many INGO strategic objectives are not SMART and cannot be measured directly. In these cases INGOs often develop and present evidence that suggests whether, or how far, an objective has been achieved. This evidence is then balanced against any counter-evidence and an **assessment** of change is made. Assessment is considered a weaker term than measurement, but is often more realistic in the context of portfolio change.

- In some cases specific changes are presented in order to **illustrate** wider changes, often in the context of case studies or stories of change. The stories may or may not be used as representative of wider change.

1.2 Trade-offs between quality and coverage

Accurate summarisation of portfolio change requires two things: Firstly, INGOs need to be rigorous in their collection, analysis and presentation of evidence of change. Secondly, that change needs to be assessed in many different locations.

There is, however, an inevitable trade-off between the two. INGOs may spend a great deal of time and resources collecting high quality information on change from a single location. But it is not sensible to combine this information with low quality information from other areas. So if INGOs require information to be summarised or aggregated across many different locations they need to carefully assess how much they intend (or need) to spend.

In some cases, the answer is to sample – in other words to produce high quality information from a small number of locations that are representative of a much larger number of locations. But even then the costs of information collection and analysis may be very high, and there will need to be a trade-off between the desire to carry out high quality monitoring and evaluation (M&E) work and the desire to cover as much of an INGO's work as possible.

1.3 Different types of INGOs

Different INGOs have different types of objectives, and the work of some organisations is more easily summarised than others. On the one hand, some INGOs have clearly defined global objectives, which can be measured with some degree of accuracy, such as the Jubilee 2000 Debt Coalition with its focus on the removal of unsustainable third world debt. Organisations with a sector focus (such as WaterAid or Sightsavers) may also find it more straightforward to articulate and measure global objectives.

Conversely, many INGOs work in multiple sectors with multiple objectives. In these cases it may be much harder to articulate global objectives, especially at beneficiary levels.

INGOs also work through diverse strategies, and this can also influence the ease with which portfolio change can be summarised. For example:

- It is most easy to aggregate results when an INGO is engaged in service delivery. This is true even where work is carried out through partners, whose capacities to carry out M&E work may vary widely in different countries and contexts. Aggregation is simplest in sectors such as health and water and sanitation which have clearly defined, industry-standard methods of measuring change.

- Advocacy or policy influencing is a more complex area to assess. Policy change can take a very long time to achieve, and issues of contribution and attribution may be complex. Summarisation of policy influencing work is more likely to focus on the policy level itself rather than on beneficiaries, with perhaps only illustrations of wider changes.

- Capacity building or partnership work is similarly long-term, and the results of capacity building work can be difficult to assess, especially for programmatic capacity development where eventual results are not narrowly defined (as they are in more technical capacity building). Summarisation of capacity building tends to focus on organisational level – whether that be other INGOs, partner NGOs or small community-based organisations – again with perhaps only illustrations of wider changes.

- Some INGOs focus almost entirely on conducting pilot studies or seeking out innovative solutions to development challenges. In these cases it may be much more appropriate to summarise portfolio change by focusing on a few successful projects rather than trying to cover an entire portfolio.

In summary, INGOs and donors need to recognise that different summarisation challenges are faced by different INGOs depending on the nature of the organisation and the type of work it undertakes. In all cases there are solutions that will enable some degree of summarisation. However, these solutions need to be adopted (and adapted) on a case-by-case basis.

1.4 Frameworks

The type of framework used by an INGO to capture portfolio change may also have an influence over how change is summarised. In recent years, many INGOs have been asked to represent portfolio change in a single logical framework or linear results framework. However, logical frameworks were primarily developed to help plan, monitor, evaluate and summarise projects – timebound interventions with clearly stated purposes – not the work of entire organisations. There are a number of reasons why logical frameworks are not appropriate for the work of entire organisations. But two are most important when assessing the prospects for summarisation.

Firstly, the work of an organisation is ongoing, and does not begin or end at the same time as the logical framework period. An organisational logical framework can therefore only be seen as a slice of an organisation's ongoing work, with different initiatives beginning and ending at different times that may have no significant overlap with the logical framework period.

Secondly, results achieved in a period covered by a logical framework may be the result of work carried out in that period, but are just as likely to be the result of work carried out in previous planning cycles (or a combination of both). The traditional link between outputs and outcomes in a logical framework may therefore be broken.

There are signs that many donors are drawing back from the imposition of logical frameworks and are increasingly supporting more flexible results frameworks.[1] Sometimes this involves keeping the horizontal logic of a framework (objectives, indicators, sources and assumptions) but relaxing the rules governing the vertical logic. This allows INGOs to link global results frameworks to global impact pathways or theories of change if so desired.

There have also been recent calls for a greater emphasis on learning to be articulated and captured within results frameworks, particularly for INGOs that are concerned with pilot or innovation projects, and are therefore more concerned with learning from the pilots than with reporting on impact across a relatively small number of people. This could be achieved by placing a series of learning questions at the heart of a results framework (see part three of this paper). Organisations could then be held partly accountable for what they have learned.

The more flexibility that a global framework allows, the more INGOs are able to develop their own individual frameworks that suit their own purposes. However, this comes with the added responsibility of developing frameworks that are fit for purpose, and that honestly reflect INGO ambitions at portfolio level.

PART TWO: Indicators

Indicators are by far the most common mechanism used to collect, analyse and present change within INGOs. This section looks at a range of different types of indicators, and explains how they can be used to help summarise portfolio change.

---

[1] See ICAI 2013.

2.1 Types of indicators

It is sometimes possible to develop indicators at portfolio level that allow INGOs to assess portfolio change directly. Where this is not feasible, the most common response is to develop indicators at a global strategic level and then to link these to indicators at lower levels, such as country, sector, programme and project levels. This can be done in a variety of ways, as outlined below.[2]

2.1.1 Direct indicators

Sometimes INGOs set indicators at a global level, and collect these indicators using their own head office resources or secondary sources. This means that reporting against the indicators does not need to rely on information collected through field offices or partners. There are two circumstances where this is likely to be achievable.

Firstly, if an INGO is big enough or important enough in a narrow field of work then it may be able to directly measure change at global level or across a range of different countries. There is an important caveat here. A contribution can only really be claimed if an organisation has sufficient critical mass to justify the argument that it directly influences the change it is measuring. This means it needs to be able to argue that any movement in the change being measured is at least partly down to its efforts. This can only happen when the INGO is a very important and influential player in a wider field, or where it is working in a very narrow area (e.g. livelihoods of ex-miners with disabilities).
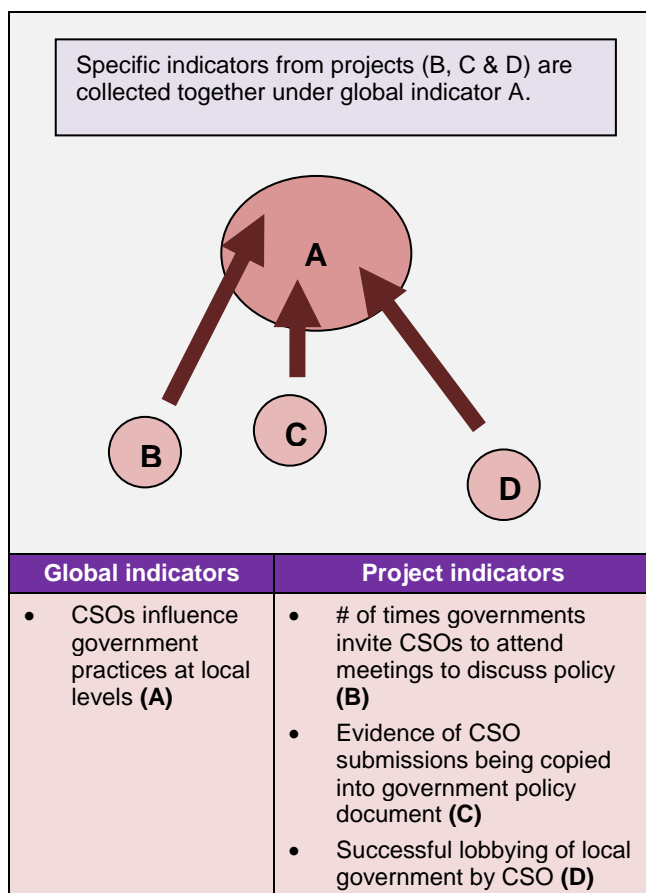
Secondly, indicators may be developed that focus narrowly on what an INGO is trying to achieve with a small group of policies or organisations. For example, an INGO could carry out a survey with all its partners in order to assess the extent to which its partnership (or capacity development approach) is valued, and then develop some simple indicators from that survey. Or it could focus indicators on one or two key policies at global level it is trying to influence.

Direct indicators are often the best kind of indicators for assessing portfolio change, but there are only a few circumstances in which they are appropriate. In most cases, reporting against global indicators relies on information collected in other parts of an organisation.

2.1.2 Framing indicators

Broad framing indicators, otherwise known as basket indicators, are the preferred option for many INGOs. Framing indicators are not specific, and cannot be collected directly at global level. Instead they are used to identify broad areas or domains of change. More specific indicators at programme or project level are then captured and summarised under the framing indicators.

---

[2] Note that indicators can be linked between many different levels of an INGO, but for ease of explanation this section confines itself to links between project level and global level.

Specific indicators from projects (B, C & D) are collected together under global indicator A.

Global indicators	Project indicators
• CSOs influence government practices at local levels **(A)**	• # of times governments invite CSOs to attend meetings to discuss policy **(B)** • Evidence of CSO submissions being copied into government policy document **(C)** • Successful lobbying of local government by CSO **(D)**

Box 1: Framing indicators

Sometimes framing indicators are cascaded downwards; they are very broad at global level, quite broad at regional level, and become successively narrower and more focused at country, programme and project levels. The information collected at lower levels can then be collected together and summarised at higher levels.

In Box 1, "*CSOs influence government practices at local levels*" is an intangible piece of information which cannot be collected directly at global level. Instead it relies on tangible information being collected at lower levels, such as the examples shown. These three tangible pieces of information can be collected under the one indicator at global level to illustrate ways in which civil society is influencing local government across different countries and regions.

Note that the project indicators may be quantitative or qualitative or a mixture of both. Indeed, they need not necessarily be indicators at all, and information on unexpected or negative changes can be mapped onto the global indicator as well.
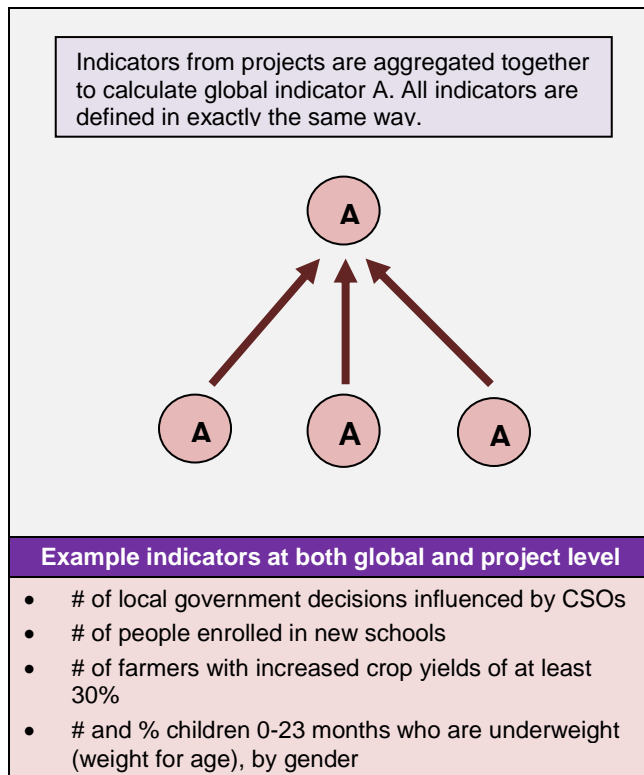
One big advantage of this type of indicator is that programmes and projects can collect indicators and produce reports which are useful to them and reflect their own needs without referring to the global indicators. Indeed there is no technical reason why a project should know about the global indicator at all.

However, it is important to recognise that this kind of indicator cannot be used to measure portfolio change. Instead, it is used to generate a number of examples or illustrations of the type of changes that are occurring. The indicator is basically a mechanism for collecting together and summarising a number of specific changes under a common theme.

2.1.3 Aggregated indicators

Indicators from different projects can also be added together at global level to provide an aggregated indicator. If an INGO intends to aggregate indicators in this way it needs to ensure that all projects define and use the indicator in exactly the same way. In practice this often means using common tools and approaches to collect the indicator, over similar timeframes. Specifically:

- the same indicator definitions need to be used;
- all relevant projects need to collect the information (or a statistically significant sample of projects);
- change needs to be assessed over similar timescales;
- the same (or similar) tools and methods need to be used;
- the quality of information collection needs to be consistent; and
- contributions to change should be broadly similar.

Indicators from projects are aggregated together to calculate global indicator A. All indicators are defined in exactly the same way.

Example indicators at both global and project level

- # of local government decisions influenced by CSOs
- # of people enrolled in new schools
- # of farmers with increased crop yields of at least 30%
- # and % children 0-23 months who are underweight (weight for age), by gender

Box 2: Aggregated indicators

The risk otherwise is that different numbers may be added together that are meaningless. For example, an INGO could develop an aggregated indicator such as *"# of local government decisions influenced by CSOs"*. But it would be meaningless to add together numbers from different projects if there were different understandings of what was meant by "*local government decisions influenced*". For example, one project might only report on cases where local government officials explicitly recognised the contribution of the CSO, whereas another project might include all cases where a CSO representative attended a decision-making forum.

Aggregation is often desired because it is seen as enabling INGOs to measure performance across multiple interventions. And indeed it is quite possible at output level; for example counting the number of schools built or the number of groups supported. However, aggregation is extremely difficult at outcome level, even where the time gap between outputs and outcomes is relatively short, as in the examples provided below, and where industry standard definitions exist. And as outcomes become more complex and intangible, the difficulties in aggregation become greater. This means it is rarely used in complex sectors such as governance, conflict resolution or civil society, where outcomes are very context specific.

The major implication of using aggregated indicators is that all relevant projects/partners need to be told (or persuaded) to use the standard indicators. At best this may involve some level of administrative burden on projects. At worst, it might mean projects have to collect information on indicators that are of no use to them, and maybe inappropriate for their work as well.

Aggregated indicators are most commonly used within INGOs to assess reach – the number of people (or groups or organisations) that an INGO claims to be supporting. Assessments of reach are useful to explain the scale of an INGOs work, and are also valuable where INGOs attempt to sample people or communities to derive more in-depth assessments of change with a view to extrapolating findings over larger populations.

2.1.4 Core indicators:

As a simpler alternative to aggregated indicators, core indicators may be collected in different projects and programmes and then reported across a portfolio, but not necessarily aggregated. Core indicators can be used to show similarities between results in different areas but also differences.

The basic idea of core indicators is that a cohort of standard indicators is developed and then different projects pick and choose from those indicators. But when they use them they use them to the same standards and definitions. This means results can be compared. For example, an INGO could capture information on improvements in child nutrition in 20 projects using the same indicator, but, crucially, not attempting to aggregate them. This allows for summary of different projects at different points to be collected together, as in the table below.
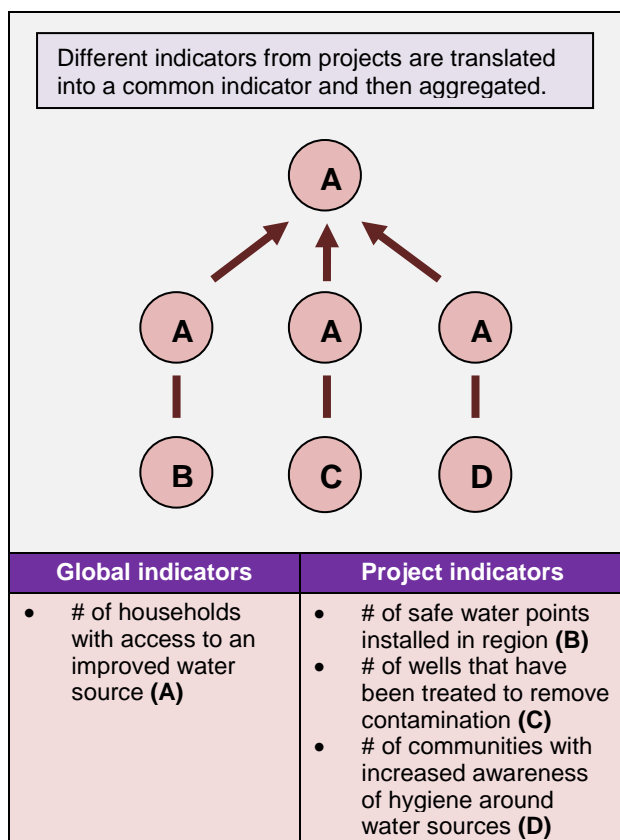
Indicator: % of boys and girls, aged 6 – 59 months, *stunted*			
Project	Baseline	Timeframe	Current
Project 1	50%	2012-	30%
Project 2	30%	2015-	-
Project 3	25%	2009-2014	5%

Whilst it would not make sense to aggregate together these figures from different projects, starting at different times and over different periods, core indicators can allow a measure of summarisation if presented appropriately. Some level of aggregation might still be possible on a sample of projects (those with common timeframes), but the information may actually be richer and more informative when differences can be seen as well as similarities.

2.1.5 Translated indicators

Translated indicators are a halfway house between framing indicators and aggregated indicators. Projects can report on different indicators, as in the case of framing indicators. However, those indicators are then 'translated' into a common indicator later on in the process. The common indicator can then be treated as an aggregated or a core indicator. In other words, it can be aggregated or summarised as required. Theoretically, this can be done early on in the process by project staff, but in practice the translation is more commonly the task of global programme staff as this allows for greater consistency.

In the example shown in box 3, three different project indicators are all converted into a common global indicator before being aggregated. For instance, it should be possible to measure or estimate the number of households with access to newly installed water points (indicator B). It should also be possible to measure or estimate the number of households being served by wells treated for contamination (indicator C). It may be much harder to estimate the number of households benefitting from increased awareness of hygiene around water sources (indicator D), but it would still be possible. After translation, the three numbers could then be aggregated under the common programmatic indicator "*# of households with access to an improved water source*" (A).

<table>
<tr><td colspan="2">

Different indicators from projects are translated into a common indicator and then aggregated.

(A)

(A) ← (A) → (A)

(B) (C) (D)
</td></tr>
<tr><td>**Global indicators**</td><td>**Project indicators**</td></tr>
<tr><td>

- # of households with access to an improved water source **(A)**
</td><td>

- # of safe water points installed in region **(B)**
- # of wells that have been treated to remove contamination **(C)**
- # of communities with increased awareness of hygiene around water sources **(D)**
</td></tr>
</table>

Box 3: Translated indicators

There are many examples of where these kinds of indicators have been used within planning and M&E work. For example:

- In cost-benefit analysis, different kinds of social benefits are routinely translated into monetary values to enable comparison of costs and benefits.

- Health programmes often convert different indicators collected within projects into DALYs (disability-adjusted life years) – understood as the number of years of healthy life.

- An indicator such as '# additional jobs created' can be developed to capture project indicators on employment where some businesses are increasing jobs, others are saving jobs and others are shedding jobs in order to save further losses in employment.

Using translated indicators is seldom an easy exercise, and – as with aggregation – should not be undertaken lightly. There is a significant risk that the value of some project indicators may be lost in translation; for example through placing arbitrary economic values on social benefits.

However, the main benefit of using translated indicators is that some level of aggregation can be achieved even if project indicators have not been defined or collected in the same way from the start of a project. This can help avoid the necessity of imposing indicators on project staff, which may be of little or no use to them.

2.1.6 Ranking and rating indicators

Many tools and methodologies of data collection and analysis result in ranking or rating information being produced, often from qualitative information. Examples include the kind of rankings and ratings used to assess capacity development within Organisational Capacity Assessment Tools. These rankings and ratings can then be summarised or manipulated using standard numeric procedures.

Ranking and rating tools are widely used within areas of work such as policy influencing, capacity development and partnership. They are a form of translated indicator, but are more likely to be coded at an earlier stage in the process. For instance a supported organisation may rate or rank its own capacity over time, rather than passing data upwards to be ranked or rated at a higher level.

Some INGOs also produce rankings and ratings directly. For example, projects may be ranked (or rank themselves) according to whether or how far they achieve their objectives.

RAG (red, amber, green) ratings are another common example of the type of rating used to assess project performance. One example of a tool designed to allow effective rating and ranking is the CAFOD Voice and Accountability Tool (see case study below).

Case study: CAFOD Voice and Accountability Tool

CAFOD's Voice and Accountability Tool: is an example of a tool that uses rankings and ratings to look at the ability of supported CSOs to carry out advocacy. It is basically a self-rating tool that asks CSOs to score themselves in four key areas – involvement in government processes, advocacy strategy development, community and constituency building, and involvement in structures. A rating scheme with different, pre-defined levels is used to help CSOs assess their level in the key areas. The rating scheme is specific enough to enable some consistency in ranking and rating, yet broad enough to allow flexibility across different contexts and locations. Discussions with CSOs are meant to help them plan and develop within their own contexts, and the categorisation of change on a common scale enables CAFOD to perform some level of aggregation at portfolio level.

Source: (CAFOD 2011)

The great advantage of ranking and rating indicators for INGOs is that they allow quantification of aggregate change, even though the changes are different and inherently qualitative.
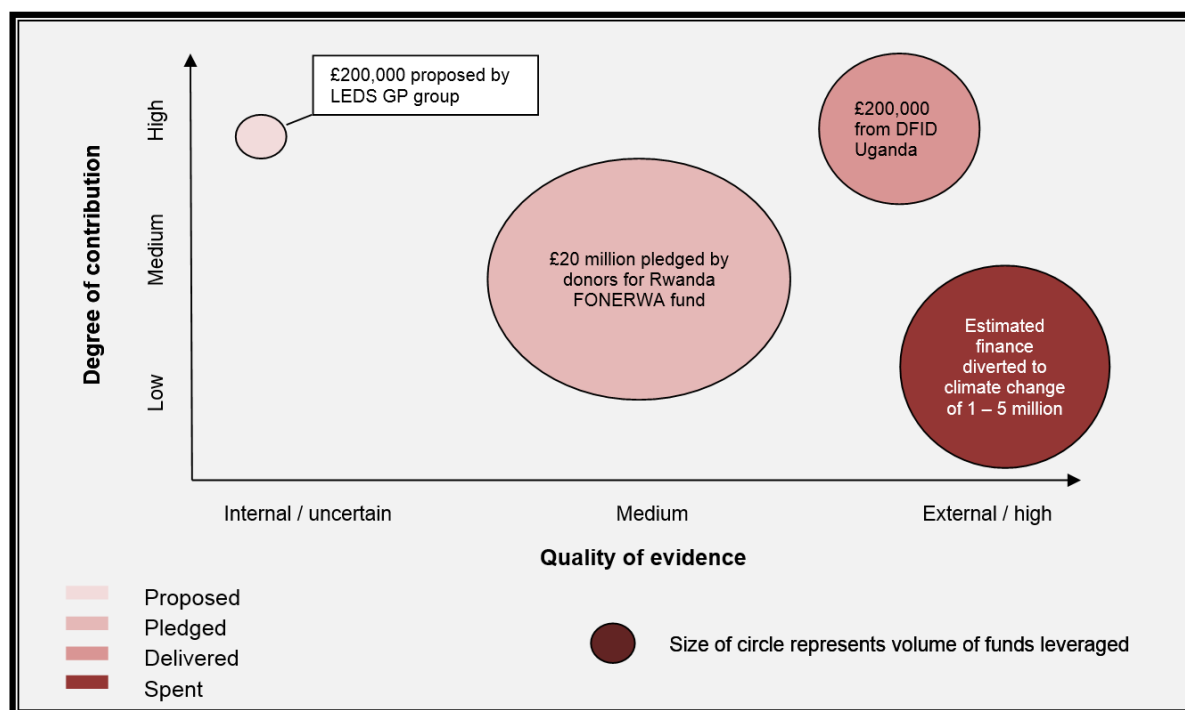
2.1.7 Mixed indicators

An indicator of lowered maternal mortality rates or increased education enrolment may be clearly understood if used in isolation. But other kinds of numeric indicators need to be supplemented by qualitative information if they are to make any sense. For example, indicators such as "*numbers of policy changes", "uptake of pilot studies"*, or "*number of organisations with increased capacity"*, make no sense on their own. They may give some idea of scale, but deeper information is required. For example there is a world of difference between being invited to add something to the wording of a policy that is about to be implemented by a sympathetic government, and achieving deep and real support for a policy that was never before on a government's agenda.

The use of mixed indicators can be seen as a response to this. Mixed indicators allow some level of quantification, whilst recognising that many aspects of change are wildly different. An example is *"# and description of changes to policy"*. This can be reported via a number that shows the scale of changes (e.g. *"twenty policies changed in child rights across Asia"*) with descriptions of the twenty changes that show the depth of change. This is another way of producing aggregated numbers without pretending that all changes have the same value or importance. Mixed indicators are essentially a way of grouping together changes around a common theme, whilst recognising that there may be significant differences between them. As with framing and translated indicators, they do not rely on standardisation at project level.

A significant advantage of mixed indicators is that they can be used when an INGO is expected to define baselines, milestones and targets as well as indicators. If an INGO were to attempt to produce a qualitative baseline on twenty different policy influencing projects it would quickly run out of space in a logical framework. But mixed indicators allow milestones and targets to be defined in quantitative terms whilst still providing qualitative reporting.

Mixed indicators also allow endless opportunities for change to be summarised graphically. For example, the Climate and Development Knowledge Network (CDKN) is expected to report to its donors against a standard indicator, *"volume of public finance mobilised for climate change purposes"*. Rather than attempting to aggregate numbers from many different examples, CDKN has developed its own indicator *"number and description of incidents where public finance has been leveraged"*. It then charts different types of leverage on a graph, showing not just the volume of resources but also the contribution of CDKN to the leverage, its confidence in the findings, and the state of the leverage (i.e. proposed, pledged, delivered, spent). A portion of this graph is shown in the diagram below.
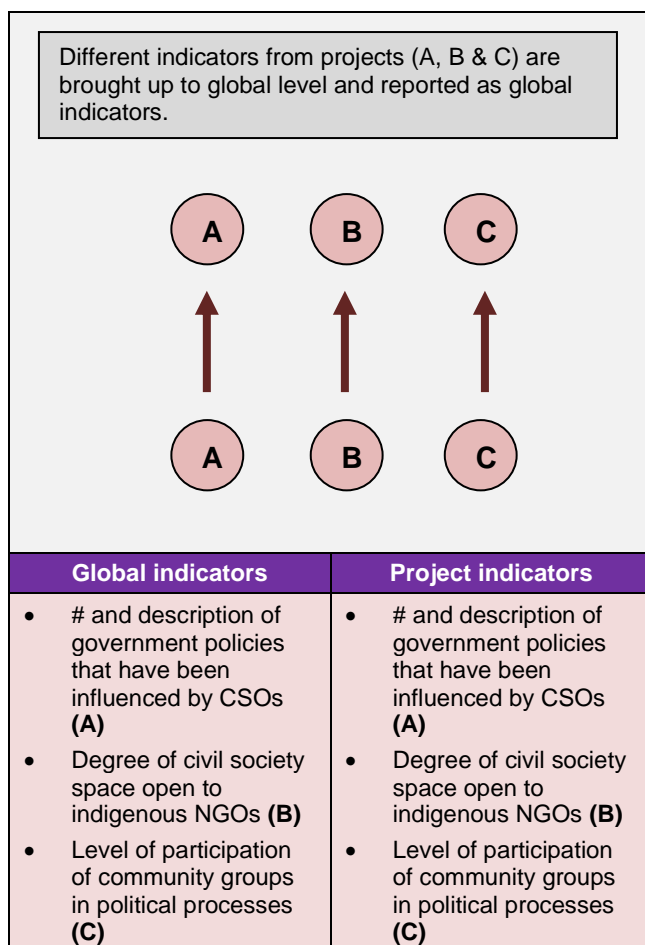

With the assistance of technology the graph could also be made interactive to show longer, qualitative descriptions when a cursor is hovered over different cases. Again, this provides opportunities to see the big picture in numbers whilst investigating the more qualitative, in-depth changes at the same time.

Mixed indicators can easily be used in conjunction with framing indicators (see 2.1.2), simply by inserting *"# and description of cases where..."* at the front.

2.1.8 Cluster indicators:

Cluster indicators are used when results are reported at global level from different lower levels as distinct entities with no attempt to pretend they are anything more than disparate results from different projects. Cluster indicators are most commonly used in programmes where work at lower levels has little or no interaction, and the total programme change can therefore be viewed as the sum of the individual parts.

Box 4: Cluster indicators

In the box:

> Different indicators from projects (A, B & C) are brought up to global level and reported as global indicators.

Global indicators	Project indicators
• # and description of government policies that have been influenced by CSOs **(A)** • Degree of civil society space open to indigenous NGOs **(B)** • Level of participation of community groups in political processes **(C)**	• # and description of government policies that have been influenced by CSOs **(A)** • Degree of civil society space open to indigenous NGOs **(B)** • Level of participation of community groups in political processes **(C)**

In the example in Box 4, three different indicators collected by projects have been brought up to the global level to show a range of changes taking place within individual projects. This enables some level of aggregation by showing different indicators from different projects together. But it does not attempt to show how the whole is greater than the sum of its parts.

Cluster indicators are most commonly used by INGOs to bring results up from country to global level. Thus a global results framework may contain a series of different indicators from different country programmes. It is possible to do this when an INGO works in just a few countries. But for INGOs that work in many different countries the number of indicators concerned quickly becomes unmanageable in a single framework. For example, if an INGO works in 15 countries and brings up three indicators from each it would end up with 45 indicators in its global results framework.

One example from the most recent round of Programme Partnerships Agreements (PPAs) (DFID's core funding to INGOs) concerned an organisation working on humanitarian assistance that made no attempt to develop consolidated indicators into its organisational logical framework, but instead included a different set of output indicators for each country in which it was working. This was possible because it was only working in five countries. For INGOs working in 20 or 30 countries this would be an impractical solution, as the logical frameworks would quickly become very large.

2.2 The relationship between tools and indicators

Some indicators can be collected through many different tools or methodologies of information collection and analysis. In most cases, framing indicators, mixed indicators and cluster indicators are not dependent on any particular tool or methodology. But for direct indicators, aggregated indicators, core indicators and ranking/rating indicators there is normally a closer relationship between an indicator and the tool used to collect (and/or analyse) it.

Indeed, in many cases an indicator may have no meaning without an understanding of the tool being used to collect it. For example, if an INGO administers a partnership survey to a partner asking it to rate the partnership as 'equal', 'uneven' or 'highly uneven' it might make sense to develop an indicator such as *"# and % of partners that regard the partnership as*

'uneven' or 'highly uneven'". But without the tool the indicator would be meaningless. Some other examples of indicators that only have meaning when associated with a particular tool or methodology are contained in the table below.

Method	Possible indicators
Surveys	• % of beneficiaries that have increased knowledge of HIV transmission in at least one area • # and % of partners that are 'very satisfied' or 'satisfied' with support provided by INGO
Organisational assessment tools	• average capacity score in areas of M&E, human resources and leadership • # of organisations showing an increase in capacity score in at least one area of support over a 2-year period
Outcome mapping	• # and % of organisations where at least 60% of expect to see markers and 30% of like to see markers are realised • % of outcomes (represented by expect to see, like to see and love to see markers) realised
Process tracing	• Average policy influencing contribution scores, as ranked by evaluators conducting process tracing
Voice and accountability tool	• percentage of partners demonstrating increased engagement and influence in policy and advocacy work in at least one area of support • # and % of CSOs showing at least a one level rise in 'involvement in corporate structures'
Tracer study	• # of partner staff trained by INGO that are still employed in NGO sector two years after training

Essentially, if INGOs adopt common tools they find it much easier to develop indicators that are capable of showing change across a broad portfolio. This reverses the commonly accepted protocol which is derived from logical framework thinking, and suggests that indicators should first be derived and then tools or methodologies selected to provide the information. When working at global level it is often the other way round: the methodology for data collection is selected first and the indicators developed afterwards.

2.3 Early and late coding

A key difference in how indicators are treated within an organisation concerns early and late coding. Early coding occurs when INGOs develop categorisations and definitions for indicators and then ask lower levels to collect the information accordingly. This has to be the case for aggregated and core indicators. Early coding means partners and projects have to comply with the requirements of the INGO, which may mean the information collected is of little use to them, but certainly makes it easier for the INGO to process data at a head office level.

Late coding, on the other hand, means information is presented from project level in many different ways, and is then sorted, categorised and translated at a later stage. This is easier on the project or partner, but requires greater capacity for data management at head office level. Late coding is a better approach for those INGOs that are concerned about the effect their information requirements may have on their partners.

Again, there are trade-offs here and tensions that need to be managed on a case-by-case basis. There may be tensions between INGOs and their partners, or within different levels of an INGO. But these tensions are a necessary transaction cost of core funding, and cannot just be ignored.

PART THREE: Further options

Summarisation can be achieved through other means than indicators. This part of the paper deals with three other options:

- narratives;
- evaluation or learning questions; and
- summarisation through strategic assessment.

Any of these options can be used as an alternative to indicators. For INGOs, though, they are more often used as a supplement to indicators contained in broad results frameworks. None of the three options listed here are mutually exclusive.

3.1 Narratives

Narratives (e.g. qualitative reports, stories of change, case studies, testimonials) are valid ways of reporting activities and change that do not rely on indicators. Most INGOs invest heavily in stories of change and use them extensively in both M&E reports and communications materials.

There are two major issues that need to be addressed when using stories for summarisation purposes. Firstly, the stories themselves need to be properly researched and developed if they are to be useful. Secondly, if a spread of stories is used for summarisation purposes then it needs to be clear what the stories represent and how they have been developed. If the method of production of stories is transparent and systematic then stories are better able to be used for summarisation, and the risk of narratives being dismissed as anecdotal or misleading is reduced.

There are many different methods of summarising change through narratives. Some examples are shown below.

- Most Significant Change methodology seeks to produce information-rich stories representing the most significant changes brought about by an organisation within pre-specified domains. Within an INGO, the domains can be defined to represent broad areas of change such as strategic objectives or framing indicators. The method of story generation and selection is transparent and replicable, and, if the methodology is followed correctly, stories are verified by a range of different stakeholders before being used. Stories are deliberately acknowledged to be generated through purposeful sampling.[3] A set of stories produced at global level can help summarise the work of an INGO because the rationale for the stories is clear, and the methodology transparent.

---

[3] See Davies and Dart 2005.

- A set of narratives based on a random selection of cases can be treated as representative. However, there may be accusations of bias if the sample is not large enough. This can be offset either by selecting a large enough sample in the first instance, or by gradually building up a large number of cases over time. Narratives can range from short descriptions of individual outcomes to much longer pieces of work, such as evaluations or effectiveness reviews (see case study below).

- Simple story telling can be used to support summarisation, provided the cases are acknowledged. For example, some organisations divide stories into different categories such as best cases stories, worst case stories, learning cases, representative cases and comparative cases. If enough stories are told, readers should be able to establish the best and worst of an organisation, along with key lessons that are being learned along the way. But this is only the case where INGOs are honest about what the stories represent. Stories used in this way can provide a window into the richness and complexity of an organisations' work. But it can be hard for outsiders to tell the difference between genuine cases and those developed for marketing or communications purposes. A transparent categorisation of the narratives certainly helps.

- Qualitative research methods can also be used to draw trends or generate learning from a wide variety of narratives. For example, a global report could seek to draw trends across many different project reports. This helps support summarisation by describing trends, similarities or differences across a wide variety of cases. One possible methodology that can be adapted is the Sense Maker technique, developed by Cognitive Edge (2015), which claims to be a way of assessing trends and patterns across a large database of micro narratives. The author has never seen the technique in use, but a large number of UK-based INGOs are currently interested in trialling it.

- In some cases numbers can be produced from qualitative methodologies. For example, the findings of qualitative case studies carried out on random or representative supported organisations or beneficiaries can be extrapolated to identify trends or estimate outcomes/impact in numeric terms.

Case study: Oxfam Effectiveness Reviews

Oxfam operates in 52 countries, across 227 different programmes, and with over 1000 projects, with an annual budget of over £200 million. Oxfam uses effectiveness reviews that aim to evaluate the impact of its work using standard methodologies across different sectors of its work. As Oxfam states on its website, *"[T]o be fully accountable, we want to determine whether our work is resulting in positive change and why. The undertaking of rigorous evaluations is the only credible way of doing this. With over 400 projects completing each year, doing full impact evaluations on each of these would be too resource intensive. Instead, projects are randomly selected and assessed under the thematic areas."* This does not necessarily result in a statistically significant sample of projects. But it does result in a series of reviews that can be used summarise change across different projects. The fact the projects are randomly selected means that Oxfam cannot be accused of 'cherry picking'. Note that Oxfam also aggregates quantitative outputs across all its projects.

Source: (Oxfam 2015)

A combination of these different methods can be used to summarise the work of an organisation through a rich tapestry of narratives. Too often, however, INGO annual reports simply use stories as illustrations of positive change, and make no attempt to produce stories in a methodological way (or if they do it is not communicated clearly to the reader). It is not surprising, therefore, that many people treat INGO narratives with a fair degree of scepticism and suspicion.

One of the reasons that many INGOs feel unable to use qualitative evidence to the same extent as quantitative evidence is that qualitative evidence is more difficult to fit into a logical framework or similar results framework, especially when donors request baselines, milestones and targets as well. Whilst producing a short narrative baseline and target produces no real problems for an individual project, producing baselines and milestones for fifty or a hundred stories does. This is where the use of mixed indicators can help (see previous section). Mixed indicators can be a vehicle that allows for reporting on complex changes across distinct themes through rich narratives. But it also allows organisations to set simple numeric baselines and targets identifying the number of cases they will cover.

3.2 Evaluation or learning questions

Some level of consistency in reporting can potentially be achieved through the development of a common set of questions, which can be asked across different projects. This is routinely done in evaluations, which focus around a core set of evaluation questions. However, there is no reason why this cannot be done at organisational level as well. Questions may focus on change, but might also focus on process and learning as well. Some examples are as follows:

- What significant changes have occurred in the lives of poor women, men and children?
- What changes in policies, practices, ideas, and beliefs have happened?
- To what degree has [the INGO] contributed to the changes?
- How cost-effective have the interventions been?
- What particular lessons have been learned?
- How might these be applied in other locations?

The questions could and should be much more focused when applied to the work of an individual INGO, and much more aligned with its core mission and ways of working.

Broad evaluation questions or learning questions such as these can easily be included in INGO results frameworks. Up until now this has not happened because the logical framework approach, with its focus on specific objectives and indicators, has dominated donor thinking, and it does not leave room for wider evaluation questions or specific learning questions. There are signs that this is changing, and it is possible that in the future the idea of including evaluation questions or learning questions in a results framework may become mainstream.

3.3 Summarisation through strategic assessment

INGOs have often been put under pressure to aggregate results, on the assumption that overall effectiveness is largely the sum of work carried out at different levels. When engaged in service delivery, capacity development or partnership work this may to some extent be true. But there are areas of some INGOs' work that should not be summarised through

assessment of the entire portfolio, but more through in-depth assessment of change at strategic points.

Some examples of work that could be summarised through more strategic assessment are:

- key lessons learned and disseminated that are being used by other agencies to bring about change;
- pilot projects that are being mainstreamed or scaled-up by other agencies;
- examples of transformational changes that are changing the way other agencies do business; or
- examples of innovative projects or projects that diverge from business-as-usual approaches, with rigorous assessments of strengths as well as weaknesses.

All these examples could lead to impacts that go far beyond those that would traditionally be captured in an INGO results framework. Reporting in this way recognises the fact that – for some INGOs at least – the bulk of their impact actually comes through a small portion of their work.

PART FOUR: Different levels of change

The previous two sections have described mechanisms for summarising change across different levels of an INGO – global, region, sector, country, programme and project. This part of the paper examines the different levels of change that can be summarised.

4.1 Goal level

Many INGOs work towards explicit or implicit global-level changes that they hope to influence, even if only in a small way. Indicators representing these global-level changes may already be captured by other agencies, e.g. the Millennium Development Goals or the Human Development Index. However, as explained in section 2.1.1, an INGO can only real claim a contribution to these kinds of high-level changes if it can show that it has sufficient critical mass to directly influence the change.

If an organisation does not have this kind of critical mass – and arguably most INGOs do not – then it is still possible to measure the overall change to which an INGO is trying to contribute. But the challenge then becomes how to construct a narrative describing how the work of the INGO has supported any positive trends (or helped to minimise downward trends) in the area being measured.

In essence, this is more of a communications exercise that attempts to link INGO work at different levels with the overall goal it hopes to influence. It is valid in the sense that it provides a structure for explaining what the organisation is attempting to achieve and where it believes it is contributing to change. But the overall change cannot be taken as a measure of organisational performance.

Case study: Transparency International

For a long time, Transparency International (TI) has implemented different indices and barometers that seek to measure national changes in areas such as transparency, accountability and corruption. Yet TI was reluctant to use these as part of its M&E systems as it was concerned that it was a relatively small player in a large market and thus, on its own, had no measurable effect on these changes. More recently, TI has begun to use the results coming out of the indices and barometers to contribute to a narrative of change. This means reporting changes in transparency, accountability and corruption at national level, and then providing examples of changes brought about through TI work within and across those countries to demonstrate how TI is contributing to the national changes it measures. TI does not seek to claim any contribution to overall changes in transparency, accountability or corruption within different countries, nor judge its own performance based on these changes. Instead it uses the goal-level changes to construct a narrative to help explain how its efforts fit into the wider picture.

4.2 Outcome level

Some agencies seek to aggregate change at beneficiary level. This is always difficult to do well, but is relatively easier when INGOs are engaged in supporting service delivery in areas with industry-standard indicators, such as health and livelihoods.

INGOs that focus more on areas such as capacity development and policy influencing are more likely to attempt to summarise change at organisational or policy level, representing the areas over which they may have a direct influence. On occasion it may be possible to go beyond this and illustrate how changes are filtering down to beneficiaries. But these cases remain illustrations, and it is rarely possible to perform any sensible degree of aggregation at beneficiary level.

4.3 Output and activity level

In theory it is easier to aggregate change at output or activity level, and many INGOs have developed systems to do this. Outputs may include beneficiaries, groups or organisations supported; policy influencing work undertaken; trainings carried out; money disbursed; or any kind of deliverable.

Most INGOs are reasonably adept at producing output level summarisation, although the author's experience is that most also find it very difficult. This is largely down to practical, rather than theoretical issues; in particular the difficulty of getting consistent, high-quality, timely information from many different projects and partners. Tolerance, perseverance and endless patience are often the keys here, rather than theoretical understanding on technical issues.

4.4 Added-value

The ultimate justification for the existence of many INGOs lies in the value they add to the work of their partners in the South. This is a difficult area, and INGOs in the past have been reluctant to monitor and evaluate their added-value work. But there are plenty of prospects for INGOs to summarise added-value work – either through developing indicators, building up narratives, or addressing key evaluation/learning questions as outlined in previous sections.

Some added-value work, such as capacity development of partner organisations, is routinely captured and summarised by INGOs. But there are many other areas which are either not covered in INGO reports and results frameworks, or are reported anecdotally with little attempt to provide rigorous analysis. These include areas such as:

- facilitating synergies and cross-learning within and between projects and programmes;
- provision of technical assistance to partner projects;
- networking between partners and between local, national, regional and international levels;
- linking advocacy on different levels;
- producing and disseminating research;
- developing tools and methodologies, or promoting new methodologies and practices;
- resource mobilisation;
- designing and overseeing programmes; and
- piloting or testing new methodologies.

In the current economic climate it would probably be unwise of INGOs to abandon all attempts to summarise at the level of outcomes and outputs, as these are most easily understood by external audiences. But there is an argument that this should be supplemented by much better M&E at the added-value level, and a better and more rigorous assessment of what INGOs bring to the table.

4.5 Organisational change

The final level of summarisation is probably the least interesting for external audiences, although very important for adaptive management. Most INGOs have a set of internal, organisational indicators that management uses to assess how they are performing as organisations. Many INGO strategic plans contain indicators at this level. Sometimes changes can be assessed at the level of the organisation itself (e.g. fundraising, staffing, finance). But for larger INGOs information may also be aggregated up from regional or country office levels.

Measures of organisational performance can also include rankings and ratings that are used to assess performance within projects and programmes. Whilst only relative, these can perform a useful quality assurance role that can help an INGO assess whether or how performance at project or programme level is improving or getting worse over time.

In recent years there has been more interest focused on this area. As an example, one of the recommendations of the recent UK PPA (Partnership Programme Agreement) evaluation was that DFID should pay more attention to organisational performance and governance monitoring in both selecting INGOs and in assessing value for money. Organisational change is arguably the easiest level of all to summarise, as it is straightforward to develop and collect indicators centrally.

4.6 Putting it all together

Some of the larger, international agencies are now routinely producing results frameworks that capture change at many different levels (see case study below).

<div style="border:1px solid #000; background:#f5d5d5; padding:10px;">

Case study: International Fund for Agricultural Development (IFAD)

Based on the results frameworks of institutional donors, IFAD has developed a hierarchy of indicators at different levels that it uses to summarise performance (IFAD 2011).

- Level 1 indicators represent development outcomes to which IFAD contributes, but that are not attributable to IFAD alone. These indicators are based on World Bank indicators such as global 'crop production index' and 'level of official development assistance to agriculture'.
- Level 2 indicators measure country-level development outcomes, and are based on IFAD projects and programmes. They include a number of indicators based on ratings of IFAD projects in areas such as relevance, effectiveness, efficiency and gender equality. They also include aggregated indicators at beneficiary level based on common approaches to data collection – such as 'household asset ownership index' and 'level of child malnutrition'
- Level 3 indicators track outputs delivered by IFAD-supported programmes. These include indicators such as 'people trained in crop production practices', 'value of loans and savings mobilised' and 'roads constructed (km)'.
- Level 4 indicators monitor IFAD's operational effectiveness, focusing on the quality of country programme and project design, and implementation support. Indicators here are mostly based on ratings of projects in areas such as quality assurance, portfolio management and risk assessment. A client survey is also used to assess partnership.
- Finally, level 5 indicators aim to capture progress on organisational performance, covering areas such as gender balance of IFAD staff, percentage overhead costs and internal budget management.

Source: (IFAD 2011)

</div>

Summary and conclusions

Some INGOs find it relatively easy to summarise portfolio change, particularly if they are focused on a single sector or are mostly concerned with service delivery. These INGOs may be able to find indicators that allow them to directly measure change across complex portfolios of work, or aggregate changes up from project level.

But in INTRAC's experience this is rare. More often, INGOs will need to use a range of different processes and methodologies at different levels to summarise change across their portfolios. Even then, this usually stops short of measurement of portfolio change, and consists more of assessments of change based on balance of evidence (see case study on next page).

Pursuing efforts to summarise performance across a portfolio does involve trade-offs. There are trade-offs between the resources required to carry out M&E work and the value this brings to the INGO. There are trade-offs between getting field offices or partners to present information in a way that allows for easy aggregation or summarisation, and allowing them more freedom to collect and present information that serves their own purposes. And there are trade-offs between pursuing M&E processes that suit the internal purposes of an INGO and meeting the needs of external stakeholders such as donors.

Decisions need to be made on a case-by-case basis, and these trade-offs need to be constantly managed and reassessed. This paper has outlined some of the common methods through which INGOs summarise portfolio change. But each INGO is different and each is likely to need a different mix of approaches. There are no panaceas.

Case study: CDKN

The Climate and Development Knowledge Network (CDKN) works at the intersection of climate change and development, working in areas such as mitigation and adaptation. It works in over 60 countries. There are many actors in this field and CDKN is not large enough to claim an influence over global changes. CDKN's M&E system relies on a range of different methods to summarise overall performance.

- CDKN organises plans and reports around seven key dimensions of change, including policy and capacity change.
- Output indicators are routinely collected and aggregated in areas such as people trained and publications produced.
- CDKN uses mixed indicators extensively, such as *"# and description of countries where CCD policies have been changed with CDKN contribution"* or *"# and description of cases where public finance has been leveraged"*.
- A system of ranking and rating is used to assess the proportion of CDKN's projects that achieve their longer-term objectives – measured through a consistent review process that occurs one to two years after projects have finished. CDKN also ranks projects according to stakeholder assessments of satisfaction.
- A consistent Outcome Map is used to assess changes in the capacity of developing country negotiating groups to influence climate change talks.
- A series of stories of change are produced each year, divided into best, worst, comparative and learning cases. These provide a window into CDKN's work, and are reflected in the organisational logical framework.
- A partner survey is used in an attempt to measure CDKN's role as a trusted partner directly.

Together, these processes help CDKN provide an overall assessment of portfolio change by building up a complex picture of change. However, focused as it is on climate change, CDKN works in a complex and changing environment, and recognises it is unable to provide absolute measures of portfolio change. In the end, any assessment of the performance of CDKN is down to the judgement of different stakeholders. The role of the M&E team is to help inform that judgement.

None of the measures used by CDKN are particularly costly or require extensive resources to be used. And none individually is capable of measuring performance across CDKN's portfolio. Yet together they allow the organisation (and its donors) to summarise change at different levels and arrive at a critical assessment of portfolio change.

History suggests that wiser INGOs develop their own systems and procedures to assess their portfolio change, rather than focusing too much on the difficulties. Many organisations have developed effective systems for summarisation and their experiences should always be used as a starting point. The danger of doing nothing is arguably much greater than the costs of trying, even if it does not work at first.

For organisations that impose results frameworks on others – whether they are governments, institutional donors, intermediary organisations from the private sector, philanthropic organisations or INGOs themselves – the key is to be realistic about what can actually be achieved. It is perfectly reasonable to expect those receiving your funds to do the best they can to summarise their portfolio change, within the boundaries of what is possible. It is unreasonable to expect absolute measurement of performance at portfolio level where that is theoretically and practically impossible. Indeed organisations that use funding relationships to insist that others do so are, in the opinion of the author, stepping dangerously close to abuse of power.

Further reading and resources

A review of results frameworks of different institutional donors and multilateral agencies can be found in the OECD publication (2014) listed below. This also includes details of the IFAD approach to measuring results. A companion paper on aggregating results from capacity building programmes is listed below (see Simister 2015).

References

- Cognitive Edge (2015). http://cognitive-edge.com/sensemaker/#sensemaker-about
- Davies, R. and Dart, J. (2005). The 'Most Significant Change' (MSC) Technique: A guide to its use.
- Earl, S.; Carden, F. and Smutylo, T. (2001). Outcome Mapping: Building learning and reflection into development programmes. Evaluation Unit, International Development Research Centre (IDRC). Ottawa, Canada.
- ICAI (2013). DFID's Support for Civil Society Organisations through Programme Partnership Arrangements, Independent Commission for Aid Impact.
- IFAD (2011). Results Measurement Framework, 2013-2015.
- OECD (2014). Measuring and Managing Results in Development Co-operation: A review of challenges and practices among DAC members and observers.
- Oxfam (2015). http://policy-practice.oxfam.org.uk/our-work/methods-approaches/effectiveness-reviews#contentprimary_0_ctl00_FirstTab
- Simister, N. (2015). M&E of capacity building (forthcoming). INTRAC.