
GOLDILOCKS TOOLKIT

Monitoring for Learning and Accountability



Copyright 2016 Innovations for Poverty Action. Monitoring for Learning and Accountability is made available under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/). Users are free to copy and redistribute the material in any medium or format.

FEBRUARY 2016

Acknowledgments: The Goldilocks toolkit was authored by Mary Kay Gugerty, Dean Karlan, and Delia Welsh, with editing and design by the IPA Communications Team (David Batcheck, Laura Burke, Jennifer Cowman, Heidi McAnnally-Linz, Megan McGuire).



ipa
INNOVATIONS FOR
POVERTY ACTION

Monitoring for Learning and Accountability

At the Goldilocks Initiative, we argue that organizations should be doing two things: monitoring what they do and evaluating the impact of what they do. And we've argued that the impact part of the equation is often prioritized over the monitoring part. As a result, we are often evaluating the impact of programs before we know whether they are well implemented.

Consider what happens if we boil down the recipe for organizational impact to a simple formula:

$$A \times B = \text{Impact}$$

In this formula, "A" means doing what you said you would do and doing it efficiently, and "B" means choosing good ideas that actually work.

If only life were that simple.

Although not everything sorts quite so cleanly, the terms monitoring and evaluation roughly align with this formula. Think of monitoring as "A" and evaluation as "B." Much academic work focuses on "B," evaluating the social impact of programs, particularly the work of development economists running randomized evaluations. But organizations should never lose sight of "A," and in this book we argue that good monitoring is seriously undervalued.

Consider what happens if A is forsaken, for example. Good ideas implemented poorly are unlikely to produce impact. We can all agree that bad ideas don't work, but without information on implementation, we can't really distinguish between a bad idea and a good idea poorly carried out. This chapter focuses squarely on A – how to use your theory of change and the CART to better understand and improve what you do.

Why does monitoring matter?

Monitoring has gained a bad reputation, partly for good reason. When monitoring data consists largely of outputs that are not clearly connected to a theory of change, monitoring data can appear trivial. Instead of reassuring donors and the public that funds are well spent, reports appear to be mere "bean counting." Monitoring also doesn't work well when done to appease outside actors, rather than to support the organization internally. When connected to a theory of change and focused on building organizational learning, monitoring systems can provide credible and actionable data, enabling organizations and donors to gain important insight into how to manage and improve programs—information that is far more valuable than the results of a poorly run impact evaluation.

Monitoring data have two main purposes: demonstrating program accountability and helping programs improve.

Accountability and Transparency

Accountability seeks to answer a seemingly simple question: did an organization do what it said it was going to do? Transparency is one reasonably straightforward way of addressing accountability

issues. By showing how they are implementing programs, organizations are living up to their obligation to steward resources in a responsible way on behalf of those they serve.

Organizations typically face a range of accountability demands from different stakeholders. Governments often require nonprofits to report on their financial and legal status. Individual donors, both large and small, want some idea that their donations are making a difference. Charity watchdogs keep an eye on administrative costs and fundraising practices. And finally, institutional donors, such as foundations or development agencies, often require detailed reporting on the use of their funds and implementation success or failure.

The challenge with many of these accountability demands is that they do not always produce the information organizations need to run high quality programs. This gap is another reason why monitoring data are often perceived as unhelpful or unconnected to organizational needs.

We argue that if organizations develop accountable and transparent monitoring systems, they can not only meet external accountability requirements, but can go beyond such requirements to demonstrate performance to stakeholders. At the same time, they will be collecting data that helps them improve their own performance. Stakeholders should hold organizations accountable for developing and reporting on strong monitoring and learning systems, rather than imposing burdensome systems with pre-determined indicators that do not support learning and improvement.

Learning and Improvement

To be of value, monitoring efforts should go beyond simply reporting on program implementation to also seek and use information to support program learning and improvement. We've already discussed how monitoring has gotten a bad reputation – it often serves external stakeholders more than internal organizational needs. And, all too frequently, monitoring data are not connected to actual organizational decisions. Often the data doesn't arrive in a timely fashion, or it isn't appropriate for the decision that needs to be made. And all too often decisions get made on some other criteria, altogether, rather than being based on evidence from the program.

High quality monitoring that follows the CART principles is essentially the same as good management. CART-based monitoring means figuring out what credible data can be collected, making the commitment to use these data to manage programs, and to do so in a responsible (cost-effective) way.

Five Types of Monitoring Data Everyone Should Collect

We highlight five types of monitoring data that are critical to learning and accountability. Two of these – financial data and activity tracking – focus on tracking program implementation and its costs. The other three -- targeting, take-up and engagement, and feedback – are less commonly collected but are critical for program improvement.

Financial data: Financial information helps decision-makers understand how resources are allocated along the theory of change. Organizations should track spending on operations and any expected revenues, not just for annual financial reports but especially for understanding the true costs of program provision. Data such as costs per service provided or revenues generated per site allow an organization to make critical performance improvements. Once the organization is ready to

measure impact, such data can also support analysis of cost effectiveness. Information on direct provision costs for each activity or output helps inform future project costs and budgeting. Data on indirect costs helps an organization see true costs of provision and adjust management or programs structure.

Activity tracking: Activity tracking collects data on program implementation, including data on key activities and outputs from the theory of change. Activity tracking could include information about everything from how many chlorine dispensers an organization distributed to the number of financial products offered to (and taken up by) the unbanked.

Activity tracking helps improve performance by allowing an organization to understand where programs are working well and where performance can be improved. Organizations can learn from high performing locations and use this information to support other locations in doing better.

To be useful, tracking data also require systems for getting data to decision-makers in real time. If management is in the dark about how the program is proceeding, they will not be able to correct glitches or deeper flaws in implementation. Ideally, these activity data will be connected to financial data to get a better understanding of the costs of different activities.

Targeting: Targeting data consist of information on the people in a program, serving two purposes: identifying who enters a program, and for those in the program, determining what type of services they should be provided. Basic program data might include information on an individual's age, gender, marital status, and socioeconomic status. But additional data on health status, educational achievement, and level of financial inclusion, among other indicators allow an organization to direct programs or benefits to certain groups.

Targeting data help organizations understand if they are reaching their target populations and undertake changes (e.g., to outreach efforts or to the program design) if they are not. To be useful, targeting information must be collected and reviewed regularly so that corrective changes can be made in a timely manner.¹

Take-up and engagement: Good targeting data make analysis of take-up possible, since take-up data compare the percentage of people actually using a product or service against the total number of people who were offered that good or service. Take-up data support program learning and improvement because they help an organization understand whether a program is meeting client demand. If services are being provided, but take-up numbers are low, organizations may need to go back to the drawing board, as this could indicate a critical flaw in design. Low take-up could suggest that the program is not well advertised, too expensive (in time or money costs), or doesn't match the needs of people adopting it.

Engagement data—data on how people interact with the product or service—can also provide important information to support learning and improvement. In particular, engagement data can

¹ Note that targeting data is very important for efficient program implementation, but it cannot and should not double as outcome data for reporting impact. A finding that 75% of a program's 1,000 clients are the rural, poor women that they are trying to reach is important information, but claiming that this is equivalent to program impact ("750 rural women's lives were improved by our program") is just plain wrong.

help programs test assumptions behind the theory of change. If initial take-up is high but actual use is low, for example, this may signal that program is not easy to use or is a poor fit for the context.

Collecting these data is an important first learning step before thinking about whether impact evaluation is a right-fit for your organization. After all, if people do not adopt and then use something, how can it possibly make a difference?

Feedback: Feedback data give information about strengths and weaknesses of the program from the perspective of those it seeks to help. While businesses often receive immediate feedback from customers in the form of sales as customers “vote with their feet,” social organizations are not always in the business of selling goods and services and may need to be more intentional about seeking out feedback. Low take-up and engagement may be signs that more feedback is needed. Feedback data can come from a number of sources: benchmarking exercises, conversations or focus groups with clients, or brief quantitative surveys that help identify adopters and non-adopters.² Program staff can be a valuable source of feedback as well.

Building a Monitoring System According to the CART Principles

The starting point for any monitoring system is the theory of change. This visual map of a given program outlines each stage in program implementation (activities, outputs, and outcomes) as well as the assumptions that have to be in place for the program to work as intended. A strong theory of change is absolutely essential to monitoring. Without a clear program logic, organizations cannot identify the data needed for learning and improvement. And clearly articulating the assumptions underlying program elements helps to identify the reasons why programs are or aren't working according to plan.

But how do you get from a potentially long list of data that could be collected to a right-fit monitoring system? By using the CART principles. These four principles can help guide your decisions on routine monitoring – deciding which data to collect as well as how to collect it.

Credible

Collect high quality data and accurately analyze the data.

Collecting credible data means two things: first, only collect data that can be measured with high quality. What does this mean in practice? First, it means ensuring that for each component of the theory of change you want to measure, you are able to collect data that are a good measure of that component.

Second, applying the credible principle means that there are very few situations in which organizations should be collecting outcome data in the absence of a counterfactual. This is an important (and potentially controversial) element of the C in the CART system: monitoring efforts should focus on activities and outputs – not on outcomes.

Actionable

Commit to act on the data you collect.

² A number of organizations and initiatives seek to increase feedback loops and constituent voice in development. These include Feedback Labs and Keystone Accountability.

Once an organization has developed a detailed theory of change, collecting information on every element from activities through outputs could easily create a bloated system of data with too many indicators to be useful. We need some principles that can help us identify which data are most important. The actionable principle helps clarify essential data: only collect data your organization will use. Ask yourself if the information you collect can be (and will be!) used to change the course of action at your organization. If the answer is no, don't collect it. If you will continue the same course of action no matter what the data say, you are wasting money by collecting the information.

The argument that a program's data collection should connect back to management or operational decisions within organizations is not new.³ But the actionable principle goes even further. The actionable principle pushes organizations not just to commit to using the data, but to specify how they will use it.

There are two parts to creating actionable data. First, articulate a clear organizational response for each piece of data collected. Second, build efficient data storage systems that get information to those who need it, when they need it.

Articulating a response to data requires three things: each indicator is clearly linked to an action that someone will take based on the findings, resources are available to implement the decision, and decision-makers commit to using the data to inform decisions.

Of course, implementing the actionable principle is easier said than done. Many organizations work in rapidly changing environments that make it hard to plan activities for the next week, much less commit to action six months or a year in the future. But using the actionable principle as a filter for data collection decisions should greatly reduce the amount of data that needs to be collected, while increasing its usefulness.

Responsible

Ensure the benefits of data collection outweigh the costs.

Although organizations may have many credible and actionable questions they want to answer through data collection, all data have costs, and there are cognitive and resource limits to what can be collected and used.

First, collecting data carries direct costs to your organization. Staff need to be paid for their time designing forms and monitoring program events. They need money to get out into the field. It even costs money to print forms. Analyzing data also carries staff costs in making sense of the data.

However, the cost of data collection goes beyond these direct costs. Time and money spent collecting data can't be used to do something else, from expanding the program to improving equipment at the office or one of the many other uses. In economics, this other cost is called the opportunity cost.

In monitoring, the Responsible principle helps organizations think through the costs of data collection by weighing the total amount of data collected against the opportunity cost of alternative

³ There have been various efforts to help solve the problem that organizations collect too much unhelpful information. More and more practitioners have called for collecting real-time data that informs decision-making through "feedback loops" -- a pathway for data to travel from those collecting it to those who need it to make decisions and back again.

activities. To minimize costs and fit a monitoring system against the responsibility principle, organizations should find the right balance and investigate cheaper ways to collect the data.

Transportable

Collect data that will generate knowledge for other programs.

Although the main focus in data collection should be to constantly inform and improve the program at hand, data collected through monitoring should also be relevant to other programs. Organizations should be sharing what they have learned along the way so that others do not have to reinvent the wheel. While we commonly think about sharing the results of impact evaluations, monitoring data can be highly transportable as well: organizations can share their implementation successes and failures so that others can build on success and avoid making the same mistakes. They should also share their experiences with indicators and data systems so that others can learn from successes and innovation in data collection.

Organizations should share successful methods of program implementation and information about which indicators are good measures and which are not. In addition, they should share their failures. Sharing what hasn't worked can help others avoid the same mistakes.